Sensing Dynamic Retail Environments

Thesis submitted in accordance with the requirements of the *University of Liverpool* for the *Degree of Doctorate in Philosophy* by *Susannah Philp*

August 2022

Acknowledgements

I would like to express my deepest gratitude to those who have supported me throughout this journey and who without, this thesis wouldn't have been possible.

Firstly, I'd like to thank my primary supervisor Dr Les Dolega for his expert advice, consistent support and steady encouragement for the past four years. I have been incredibly fortunate to have worked with you on this project. I'd also like to thank my secondary supervisors – Dr Mark Green and Prof Alex Singleton. Mark – for your excellent advice and brilliant use of GIFs, and Alex – for always finding the best solution to the (several!) roadblocks I've faced. I'll never forget how you seemed to pull together this entire PhD topic, complete with data partner, in less than 24 hours. I am grateful to have had the support of such a wonderful supervisory team.

I could not have undertaken this project without the CDT for Data Analytics and Society who funded and provided training for this PhD. I am proud to be a part of the first cohort of students of this scheme and have had the pleasure to see it grow each year. I'd like to thank Mark Birkin, Eleri Pound, Hayley Irving and everyone involved in creating and coordinating the CDT for the wonderful opportunity and for the invaluable skills it has taught me.

I would also like to express my gratitude to those at the Local Data Company for providing data and funding without which this project wouldn't be possible. Thanks should also go to the Consumer Data Research Centre for their part in data provision.

My thanks extend to the department of Geography and Planning, particularly the GDSL and the PhD students I have had the privilege to share an office with. Thank you for the professional and personal support, and for your wonderful company.

My final thanks go to my family and friends. This journey hasn't been smooth was only made more challenging by the impacts of the COVID-19 pandemic. My office has moved from the Roxby Building to my childhood bedroom in my parents' house, then to my flat in Liverpool, back to my childhood bedroom and now to London. My parents and family (and our Westie, Albie) have been with me each step of the way, providing unwavering support and love. I would like to dedicate this thesis to them, and to the loved ones who are no longer with us. To everyone who has supported me through this journey, I express my deepest thanks.

Abstract

High streets play a vital role as retail locations, economic hubs and physical cores of culture and community. Their preservation is crucial to many. How and where people shop has evolved significantly over the last few decades. Consumers are attracted to many convenient and budget-friendly options, often to the detriment of UK high streets. There is a consensus that a more in-depth understanding of high street vitality and viability is needed to adapt to these changes effectively. Recent advancements in sensing technology have made the high street more measurable than ever before, offering novel data-driven insights into high street performance and the factors that influence it. These insights can be valuable in helping high streets achieve resilience and become sustainable for the future.

This study applies data analysis and machine learning methods to investigate Local Data Company's pedestrian count data from over 1,000 retail locations across the UK. It examines how retail footfall is influenced by the world around us, quantifying the impact of characteristics such as the proximity to transport hubs and anchor stores on footfall. The high temporal resolution of the data is harnessed to give novel insights into the impact of temporary events such as local festivals, extreme weather, and the festive season on different retail contexts. An accessible classification of footfall context is presented, which can be applied to high streets across the UK to give quick insight into what factors impact footfall. This research is assembled to provide the groundwork for a prediction model to provide a footfall prediction for any British retail address at any time.

This research has applications for high street revitalisation strategies and valuations used for determining rents and business rates. It can help retailers make effective and efficient decisions in the location and running of their stores and inform high street revitalisation policy as we enter a post-pandemic world.

Table of Contents

1 Introduction

On an international scale, shopping and leisure are engrained into everyday life and imbedded into our geography and society. Physical retail centres such as high streets, retail parks and shopping centres can be defined as spatial concentrations of retail stores and services that aim to supply this ubiquitous demand for goods. Retail connects people to each other and to the wider economy on a uniquely regular basis. As such, it is active and reactionary, always evolving in tandem with consumer demand and with wider forces. From the development of new ways to shop, such as out-of-town retail centres or online and mobile shopping, to major political and economic shifts, the retail industry continuously faces new challenges. These can be particularly strongly felt by high streets and other physical retail centres, as they are constrained by the location, socioeconomic and demographic characteristics of the area or their current retail mix and often slower to adapt and therefore overcome these challenges.

Every physical retail centre is unique from its location to its retail mix to its identity. As such, they each experience and respond to these shifts differently. It is important to be able to quantify these changes and understand the mechanisms behind them in order to identify opportunities for growth and adaption. Retail centre performance indicators, such as vacancy rates or footfall – the number of people passing by a location in a given amount of time – can be used to measure these changes, however comparatively little is known about the relationship between these different indicators and retail performance in different contexts. How are retail composition and micro-site characteristics linked to the patterns and magnitude in visitor numbers? To what extent do temporary events impact the performance of different retail contexts? What potential is there to harness data on pedestrian counts to understand retail performance across different locations? This PhD aims to answer these questions by using footfall data to explore patterns and relationships between pedestrian behaviour and the form and function of retail centres.

1.1 The current context of UK retail

High streets are varied, complex and multifunctional places. A high street, often synonymous with town centre or main street, designates the central retail area of a town or city. Although there is no formal definition, recent research has delineated 6,400 different retail centres across the UK (MacDonald et al., 2022). However, British high streets are struggling to remain sustainable and adapt to the changing context of retail. They play a key role in the country's economy, with 14% of all employment coming from businesses on the British high street, but for more than three-quarters of local authorities this amount has decreased since 2015 (ONS, 2020a). The consequences are more than economic. As Mary Portas writes in her 2011 review of UK high streets:

> "High streets are the heart of towns and communities. They have been for centuries. People are passionate about high streets. They may have different views on what's wrong and what's right, but I don't believe anyone can put their hand on their heart and say they don't care."
> (Portas, 2011)

High streets also have social and community value to the people who use them, helping to form an identity around a place (Yuill, 2009; Portas, 2011). Therefore, the impacts of high street decline are not limited to business closures and loss of household names, but also a loss of social contact and self-identity, leading to possible social exclusion and a correlation with a rise of crime and poor health (Yuill, 2009). UK high streets are essential for cultural and economic sustainability; therefore, they must be resilient and adaptable to current and future changes in the retail industry.

The retail landscape in the UK is constantly evolving. In 2019, 19.2% of retail sales were made online, compared to 6.2% in 2009 (ONS, 2021). The COVID-19 pandemic has further catalysed this trend, with the share of online sales reaching

27.9% in 2020 due to lockdowns causing temporary closures of non-essential businesses (ONS, 2021). Share of online sales peaked during February 2021, when the lockdown caused share of online sales to reach 36.1% (Dalgeish, 2021). The growing dependence of consumers on the internet is only one force in the retail industry change. The 2008-9 economic crisis, the development of out-of-town retail provision and shifting consumer behaviours are all also reshaping the UK retail industry, and high street retail has suffered disproportionately as a result (Burt, 2010; C. Parker et al., 2016; Portas, 2011; Wrigley et al., 2015).

1.1.1 Forces of high street change

Wrigley *et al.* (2015) identified three critical structural forces of high street change: the progressive rise of online shopping, the competition from out-of-town retail developments, and convenience culture.

Online shopping

With the improved accessibility to the internet through computers and smartphones, online shopping has become increasingly prevalent. Instead of visiting the local high street to compare prices and acquire a product, consumers can do the same from the comfort of their home or wherever is convenient for them. The product is delivered to them promptly, with many retailers such as Amazon and Argos offering same-day service. Online shopping offers a level of convenience that the physical high street suffers to compete with. Certain deterrents to using online retailers, such as payment security and trust has become less of a concern as more and more people use the service (Wrigley et al., 2015). The COVID-19 pandemic acted as a catalyst for this, as for some non-essential goods such as clothing or technology, online purchasing became the only option. This may push those who might have never purchased online before to do so, primarily the older population.

The internet has also modified how consumers use the high street; for example, they might browse offers or compare prices online before visiting, impacting the frequency and duration of their trip (Weltevreden, 2007). Some products such as music, books and movies have become digitised themselves, giving online shoppers instant access to their purchased goods resulting in the closure or administration of many well-loved brands, such as Blockbuster in 2013 and HMV in 2013 (administration) and 2018 (closure) (Centre for Retail Research, 2020a). As a result, physical retailers who offer these products have seen the most significant change (Wrigley et al., 2015). Secondary and medium-sized centres may be the most vulnerable to the rise of online shopping, whereas larger retail centres and smaller convenience centres may be more resilient (Singleton et al., 2016). A possible explanation for this could be a polarisation effect, where large centres have a very clear comparison retail[1] purpose and smaller centres with a clear convenience purpose, whereas medium sized or secondary centres have a less clear function (Dolega and Lord, 2020). While smaller centres can often provide essential goods to a small but loyal catchment within a convenient distance, and larger retail centres may have the attraction of being a destination and the ability to offer an experience of retail with leisure, entertainment, events and tourist attractions, medium sized centres may find themselves with no purpose in between, and therefore struggling to attract customers.

Out-of-town retail developments

The second force of high street change identified by Wrigley *et al.* (2015) is the increased competition from out-of-town retail developments. Out-of-town retail developments were encouraged by government policy of the 1980s and early 90s and allowed businesses and supermarkets to build expansive stores located on the outskirts of the city centre (Guy, 2006). Their location and ability to provide plentiful and cheap parking made them more accessible to the car-owning suburban population in comparison to the congested town centre, and they quickly grew in popularity. Food retailers such as Tesco and Sainsbury's also expanded to large out-of-town locations, and began to carry more comparison goods such as clothing, homeware, and electronics, taking advantage of the essential nature of grocery

[1] Comparison Retail: Retailers that sell comparison goods. These are goods that are purchased infrequently and tend to be of a higher value. This includes a wide range of items such as clothing, household goods and technology. They are named as such because a consumer will likely compare different options before making their purchase. This is opposed to convenience goods such as food or services such as salons, banks and restaurants.

shopping. As most people could get many of the items they needed without entering the city centre, many high streets had to adapt to this change in demand. Successful high streets catered to a leisure or entertainment demand, providing a more comprehensive experience in addition to comparison retail (Wrigley et al., 2015).

Convenience culture

The third and final identified force of high street change is convenience culture. The shift in consumer attitudes towards a convenience culture is a less conspicuous force than out-of-town centres or the rise of online shopping; however, it links to both. Convenience is favoured by a consumer because it saves time, making it quicker and more efficient for them to find a desired product. Customers have always factored time as a cost; however, demographic changes such as an ageing population, longer working hours and smaller or dual-income households have acted as drivers to convenience culture in recent decades (Hood et al., 2016; Wrigley et al., 2015). Examples of consumers prioritising convenience could be through visits to 'one stop shop' supermarkets and retail developments, the desire to shop or browse online or on mobile, and through convenience grocery retailing that fits around the consumer's day-to-day life (Wrigley and Lambiri, 2014a). The growth of the convenience grocery market is further evidence of this. It has nearly doubled in value between 2000 and 2015 (Hood et al., 2016). Major grocery retailers such as Sainsbury's and Tesco invested heavily in this market, opening their Local and Express branches respectively, and have become prominent players in a former independent and small chain dominated sector (Hood et al., 2016).

Impact of the 2008-9 economic crisis and recession

These three forces of medium to long-term retail change have been compounded by the economic crisis and recession of 2008-9 and the following period of austerity. Unemployment rose by up to 4.7%, with the manufacturing, wholesale and retail trade industries accounting for the most significant loss in jobs (Campos et al., 2011). Consumer confidence crashed, and emerging budget retailers, such as Aldi and Lidl, and eastern European and Asian markets (Burt, 2010) increased competition and pressured retailers to cut their already tightened profit margins to compete (Birkin et al., 2017). The reduction in disposable income meant consumers increasingly shopped around, looking to online retailers to get the most out of their money. Value and budget retailers such as Poundworld grew in popularity, increasing sales by close to 50% in both 2011 and 2012 (Stevens, 2018).

Smaller secondary and tertiary centres such as market towns and district centres were particularly hard hit (Wrigley and Dolega, 2011). 72% of retailers saw a drop in footfall post-recession (Srinivasan and Sivakumar, 2011) and many companies needed to rebrand and invest in their online presence to survive (Genecon LLP, 2011; Sharma, 2011; Srinivasan and Sivakumar, 2011). Recognisable high street names, such as Woolworths, entered administration in 2009. Ten years later, retailers continue to struggle. Forty-one chains, including Clintons and Forever 21, went into administration in 2019 (Centre for Retail Research, 2020b). The loss of recognisable high street names knocked consumer confidence (Portas, 2011; Wrigley et al., 2015) and vacancy rates for UK high streets rose by 9.3% from 2008 and peaked at 16.3% in 2012 (Wrigley and Lambiri, 2014b).

Other factors of high street change

It should be noted that Wrigley et al. (2015) does not contain an exhaustive list of factors of high street change. Although it concisely groups and summarises the major drivers, there are many other determinants of high street change, many of these interrelating with those mentioned previously. These other factors of high street change could include:

1. Increasing business rates. Business rates are the tax which non-domestic properties pay to their local authority and are based on the rateable value of a property and a national multiplier. For many retailers, business rates make up a large proportion of the outgoing costs and can be a major factor when deciding whether to retain or close a store (British Retail Consortium, 2021b). A previous business rates revaluation in 2017 had an inconsistent impact across the country, however, on average rates did increase (Ministry of Housing, Communities & Local Government, 2019). It is yet to be seen what will be decided when properties are revalued again in 2023, with many stakeholders calling for reform of the system.

2. Technological advancements. In addition to the significant impact of online shopping, technological advancements have changed how people shop in other ways, for example, the innovation of contactless payments systems, self-service and mobile checkouts and the use of augmented reality. Companies such as Nike, Ikea and Apple have implemented technology which allow consumers to visualise using the product before purchase (Marr, 2021).
3. Demographic changes. These were previously discussed in relation to convenience culture however, demographic changes in general are changing the consumer base of the UK, and therefore the retail demand. The UK is becoming increasingly diverse, and different age groups, ethnicities and cultures can interact with retail in different ways. The UK population is also steadily increasing through natural increase and immigration, meaning that there are more consumers (ONS, 2020b).
4. Urbanisation and suburbanisation. In addition to a higher population of people, there are also a steady increase of people living in urban and suburban areas. This fundamental underlying change can be a catalyst for other changes such as the increase in out-of-town retail centres and convenince shops.
5. Changes in consumer behaviour. In addition to the trend towards more convenient retail, there has been other shifts in consumer demand and behaviour which have shaped the UK retail market. For example, consumers are generally becoming more concerned with the sustainability and ethical impacts of their shopping habits, consuming only what they need, reducing meat consumption and opting for low carbon emission transport (Deloitte, 2022). Consumers might be more drawn to the high street as it avoids the environmental cost of online shopping.

1.1.2 COVID-19 pandemic

In 2020 and 2021, UK high streets faced the unprecedented challenge of the COVID-19 pandemic and lockdowns. The government has implemented three national lockdowns, making the opening of all non-essential retail shops illegal for over 16 weeks in 2020 and 15 weeks in 2021 (Institute for Government, 2021).

Restrictions were stricter again for the leisure and entertainment sector. Music venues and nightclubs were closed for over a year, and restaurants, pubs, gyms, and cinemas allowed to reopen had to adhere to strict social distancing, capacity limits and curfews.

In 2020, retail sales fell by 1.9%, the most significant annual decrease on record, and sectors such as fuel, clothing and hospitality were particularly hard hit (ONS, 2020c). Footfall decreased 43.4% when comparing year on year with 2019 (British Retail Consortium, 2021a). With most workers adhering to the stay-at-home policy and 2.2 million vulnerable people advised to shield and not leave their house (ONS, 2020d), online sales increased dramatically for non-essential retail and essential goods, such as food delivery services. The share of online sales decreased when non-essential retail reopened; however, as of May 2021, it remains 10% higher than pre-pandemic levels (ONS, 2021). Available research into the impact of COVID-19 on retail is limited to specific sectors or locations at time of writing (examples include Brewin, 2020; Goddard, 2021; Naseri, 2021), and the lasting impact on UK high streets is for now unknown. However, it is feasible there might be permanent shifts as workplace strategies adopted rapidly due to the pandemic, such as working remotely or from home, could be here to stay (Kniffin et al., 2021).

1.2 Adapting to change

The changes and structural trends outlined in the previous section show that retail centres are under constant and consistent pressure to adapt and evolve. Understanding these changes, the impact they have had, are having and will have on a retail centre is vital to understand when looking to the future of a high street and its vitality and viability. In UK policy, vitality is defined as how busy a retail centre is, both at different times and in different spaces, and viability is the ability of the retail centre to attract sustainable and continuing investment. Key performance indicators help to achieve this, and when they are consistently monitored, they can give insight to short-, medium- and long-term trends that a retail centre experiences. These include quantitative indicators, such as footfall or vacancy rate, as well as qualitative indicators such as consumer perception, purpose for visit and attractiveness. A closer look at key performance indicators will be given in the next chapter (Section 2.1.3).

Key performance indicators became a vital piece of the puzzle of high street revitalisation in the wake of the 2008-09 financial crisis. In 2011, the government identified a great need for more information about the high street and research became a political priority. They commissioned retail expert Mary Portas to conduct an independent review and give her recommendations for high street resilience and recovery (Portas, 2011). In the following years, there was an emergence of reports and reviews from academia and private sector that aimed to provide references and tools to revitalise the high street (Coca-Stefaniak, 2013; Genecon LLP, 2011; Grimsey et al., 2018, 2013; Hart et al., 2014; Millington et al., 2018, 2015; C. Parker et al., 2016; Parker et al., 2014; Portas, 2011; Wrigley and Dolega, 2011; Wrigley and Lambiri, 2014b). Although the challenges of the COVID-19 pandemic differ from those of the 2008-9 financial crisis, the knowledge base going into recovery and revitalisation is significantly larger.

Among their recommendations was the need for more data such as vacancy rates, Goad maps, visitor numbers, footfall, and sales. In combination with local knowledge, this information can help measure performance and better understand the underlying forces and processes in the high street. When communicated to stakeholders, it facilitates a shared vision of the current state of a high street and can be used to design achievable and measurable goals for the future.

Technological advancements have enabled the automated collection of vast quantities of data, or 'big data'. Loyalty cards, mobile tracking and social media all provide a wealth of insights into customer behaviour, urban flows, and marketing strategies. Increasingly measurable through advancements in sensing technology, footfall provides a commonly used heuristic of retail centre vitality.

1.2.1 Footfall – the 'lifeblood' of the high street

Often cited as the 'lifeblood' of high street vitality and viability (Birkin et al., 2017), footfall - defined as the count of people travelling through a shopping area at a given point in time (Lugomer et al., 2017) - is a key measure of high street sustainability and a widely used proxy for economic performance (Coca-Stefaniak, 2013; Millington et al., 2018).

Footfall is widely considered to be one of the most influential factors in high street performance (C. Parker et al., 2016), and it is used in many town management strategies (Hogg et al., 2004). The importance of footfall as a key performance indicator will be discussed in more depth in the following chapter, however it is unique in it is ability to capture both economic and cultural strength of a retail centre. In addition, it is very responsive to external events. For example, when the UK entered an economic recession in 2008, footfall dropped by 10.4% during 2008-2011 (Genecon LLP, 2011). Footfall is a measure of activity, which is the heart of high street vitality and viability, and this is what makes it such a popular and powerful indicator.

Despite this, relatively little is known about footfall and its connection to other characteristics, particularly on a micro-scale. The work of Reilly (1931) and Huff (1963) was seminal in underpinning the relationships between pedestrian counts and meso- and macro-scale factors. In addition, connections have been drawn between footfall and national consumer attitudes, as well as factors such as weather (Makkar, 2020). However, attempts to investigate, quantify or generalise the relationship between footfall and micro-scale characteristics remained unknown until the relatively recent advent of sensing technologies, which allowed footfall to be continuously measured and monitored with minimal human effort.

This thesis harnesses data generated by one of these sensing technologies – Wi-Fi sensors that were deployed and used as part of the SmartStreetSensor project between University College London and retail data provider Local Data Company.

These sensors collect probe requests that are sent out by Wi-Fi enabled devices in close proximity to the store front the sensor is installed in. These counts are cleaned and filtered by Local Data Company to ensure to the best of their ability that only smartphones from passing pedestrians remain. These counts are aggregated to 5 minutes and are measured continuously over hundreds of locations across the UK. This methodology is explained in more depth in Chapter 3, in addition to the accuracy and sources of error that need to be considered.

This wealth of data, novel in its temporal and spatial resolution, has implications for research into footfall, urban flows and retail geography. Along with other data sources, it allows the micro-scale relationships between footfall and other characteristics to become more quantifiable over both space and time. These insights can allow for a much greater understanding of the variation and similarities between retail centres, how they are impacted by different changes and why.

1.3 Outline and Objectives

This thesis aims to use CDRC/LDC SmartStreetSensor footfall counts to explore the capabilities and potential of footfall as a measure of high street performance. The novel spatial and temporal resolution of the data provides new opportunities to gather a greater understanding of footfall and its relationship to the characteristics and evolutions of the surrounding environment. Through the application of machine learning and statistical analysis, it will illuminate trends in high street footfall and quantify how footfall is impacted by change.

When commencing a research project, it is important to establish clear research aims, objectives and rationale. This can help to provide clarity and direction to the work undertaken and establish its wider value and novelty. There are several key points to consider in this:

- ◊ What question will the research try to answer?
- ◊ What similar research exists and how will this research approach differ?
- ◊ What data and resources will be necessary to achieve this aim?[2]
- ◊ What potential applications or value could result from this research?

There will be three analysis chapters in this thesis, and each will aim to answer a research question about footfall, rooted in an identified literature gap, and provide insights that are applicable and valuable for future research.

[2] This will be defined and established in Chapter 3

1.3.1 The world around us

This first analytical chapter, which will be Chapter 4 is entitled 'The world around us - Quantifying temporal variations in footfall in relation to micro-locational characteristics'. It primarily focuses on the relationship between footfall and the immediate context surrounding it. There are many characteristics that are posited to impact footfall based on empirical evidence – vacancy rate, the retail mix, the proximity to transport hubs are some examples – however, there is little quantitative evidence that explores this. This chapter will explore this across two dimensions – space and time – through the following research aims:

The world around us	Investigate how different characteristics and contexts of the immediate environment impact footfall magnitude and signature
	Using characteristics of retail and footfall context, develop a classification that captures these main differences
	Identify how trends in footfall magnitude and signature differ between these different retail contexts

This will contribute to the literature in this field by exploring to what extent factors such as form and function impact footfall, quantifying the strength of this relationship and exploring how it differs over time.

An activity-based classification system is also outlined and proposed. The system is based on the environmental context, as opposed to solely being based on footfall data which is the case for previous footfall classifications. This could allow new insights in how footfall varies across different retail contexts and provides a system that could be applied to areas where there is no footfall data collected.

1.3.2 What happens there

'What happens there - Exploring event-related temporary fluctuations in footfall magnitude and their relation to micro- and meso-scale characteristics' is the title for Chapter 5, and this analysis will specifically focus on temporary events and fluctuations in footfall. This includes shopping events such as Black Friday or Christmas, weather events such as heatwaves or storms and local events such as festivals or sports events. It is well-documented that these occurrences can have beneficial or detrimental impacts on footfall and the vitality and viability of a retail centre. Often footfall figures are given to quote the success or impact of an event (e.g. Edinburgh festivals in Naylor *et al.* (2016), Storm Deirdre in BBC News (2018a)) however there is little to no publicly avaliable research that compares how these can differ between retail centres. Through comparison of four key case study locations, this chapter will provide new insights into how the impact of temporary events can differ in different retail contexts through achieveing the following objectives:

What happens there	Identify events which significantly impact footfall.
	Investigate how factors of both the immediate environment and in the wider context could influence this impact
	Explore the trends and similarites between footfall of different events in different locations and what they could imply about retail footfall

This analysis will contribute to the literature by building a greater understanding of how temporary events impact footfall, the size of their impact and investigating whether there are any factors that might mediate or control any impacts. As of writing, no research has been found that focuses specifically on footfall and events across multiple retail environments and contexts. Therefore, this analysis will be the first of its kind and facilitate more potential research on this topic in the future.

1.3.3 What remains unknown

The final analysis chapter, Chapter 6, will be called 'What remains unknown - Investigating the potential for a spatio-temporal prediction model for footfall data' as it will explore footfall modelling and prediction, and the capability of this data to be used to estimate footfall numbers in places where data is not collected.

Much of the research into footfall modelling and prediction is limited, both temporally and spatially. Limited time spans are used, or temporal considerations are omitted altogether, and research aims tend to limit the sample to an area or neighbourhood of interest. Datasets like the SmartStreetSensor project footfall dataset collect data with unprecedented spatial and temporal coverage, allowing the opportunity to develop a footfall prediction model that is not limited temporally or spatially. This would have a wealth of applications for decision makers, retailers and high street stakeholders allowing locations which may not have the resources for consistent footfall measurement to benefit from the insight which footfall measurements can give. Chapter 6 will explore different methodologies that could be used to achieve this, through the objectives that can be found on the next page.

What remains unknown	Define the criteria and use case for a footfall prediction model and identify appropriate methdologies to achieve it
	Create a preliminary model that predicts footfall that is location and time dependent
	Critique the performance of this model, identifying opportunities for improvement

1.3.4 Thesis Structure

The previous sections have established the research aims of the three analytical chapters, which will sit in the centre of the thesis. The next chapter, Chapter 2, presents a literature review, where the literature gaps expressed previously are explored in more depth. It will give an overview of the the breath of literature and research into high street revitalisation and footfall, exploring the methods of footfall collection, applications and how it is impacted by macro-, meso- and micro-scale characteristics.

Chapter 3 will explore the data which will be used in this thesis. This includes the aforementioned SmartStreetSensor footfall data, in addition to other supplementary datasets. It introduces and explores the data source, its collection and its sources of error and considerations and how it will be used to answer the research questions in the analytical chapters.

Chapter 4, Chapter 5 and Chapter 6 form the analytical section of this thesis, with each chapter presenting a unique and novel approach to a research question established in the previous chapters. They are entitled 'The world around us', 'What happens there' and 'What remains unknown'.

Chapter 4 – the world around us – links footfall with micro-locational characteristics and aims to quantify their impact on footfall and how this differs across space and across time.

Chapter 5 – what happens there – focuses specifically on temporary fluctuations in footfall and uses a case study analysis to compare the impact of events such as Christmas, Black Friday, local festivals and weather events.

Chapter 6 – what remains unknown – provides a novel exploration of footfall prediction methods and how these could be applied to the SmartStreetSensor footfall dataset for the purpose of spatial and temporal prediction.

The final chapter Chapter 7 revisits the results and conclusions drawn from the previous three chapters and discusses the wider context of footfall and retail research. It sets out the contributions of this research and recommendations for high street stakeholders and for future researchers working with footfall data.

This thesis will be an investigation of footfall data, exploring what quantitative data analysis can reveal to us about the interrelationships that exist in high street environments during the study period[3]. This will not only expand the knowledge base on high streets, but also on footfall data, its strengths and limitations and what considerations may need to be made when analysing and utilising this data to make real decisions.

[3] The majority of the data used is from 2015-2019. It should be noted that, due to the unprecedented impacts of the COVID-19 pandemic, the applicability of these insights to the current high street may be limited.

2 Literature Review

The retail industry has evolved significantly over recent decades. Demographic and social changes such as longer working hours and an increase of dual-income as opposed to single-income households have put more pressure on consumer's time, further emphasising the value on convenience in retail. Out-of-town retail developments began to flourish in the 1980s and 1990s, encouraged by government policy, and they offered an accessible alterative to the congested city centre for car-owning suburban consumers. The priority of convenience has also been a driving factor in the increase of online and mobile retail since the 2000s, as consumers do not have to leave their homes to compare and purchase the items they require.

However, these evolutions have reduced the demand on the high street for retail purchase, causing many UK high streets and town centres to suffer without appropriate adaption (Wrigley et al., 2015). In particular, the growth of online retail has been rapid and it has the capability to evolve quickly to meet any changes in consumer demand, a speed that high street retail is now forced to contend with (Reynolds, 2000). Events such as the 2008-9 economic recession and the COVID-19 pandemic further catalysed these impacts. High streets and town centres are economically, culturally, and socially important places; therefore, it is critical they adapt and overcome these changes and become sustainable and resilient for the future.

There is a consensus that more data driven evidence is needed to inform and support high street sustainability (Portas, 2011; Wrigley and Dolega, 2011). Measures of performance can provide key information on the health, vitality and viability of a high street. Footfall, cited as the 'lifeblood of high street vitality and viability' is a commonly used metric to measure economic and social performance of a high street (Birkin et al., 2017). Despite its popularity, relatively little is known regarding the behaviour of footfall in relation to characteristics of the surrounding environment and wider events. How does footfall vary in different kinds of high street environment? How does it respond to wider events? This literature review will explore three key topics: high street revitalisation, footfall and how they intersect.

The first section, Section 2.1, will explore how high streets have attempted to adapt to the forces of change outlined in the previous chapter – online shopping, out-of-town retail developments and convenience culture. It will discuss the different strategies used, their successfulness and common barriers to implementation. The next section, Section 2.2, focuses on footfall, its definition, how it is collected and how the data can be utilised in a variety of different research fields. The final section, Section 2.3, looks at footfall as an indicator of high street performance and what is known about its interrelationship with factors on a macro-, meso- and micro-scale. Section 2.4 will summarise the chapter, concisely presenting the strengths and weaknesses in this research field currently and defining how this research will contribute.

2.1 High street revitalisation and performance

Evolutions in the retail industry over recent decades have presented many challenges for high street sustainability. Many high streets have struggled as a result (Wrigley et al., 2015). In response to this, there have been many attempts to revitalise the high street from government, industry and academia.

2.1.1 Attempts to revitalise the high street

Since 1993, government retail planning policy has recognised the importance of maintaining the vitality and viability of UK high streets (Findlay and Sparks, 2014). The Town Centre First policy aimed to curtail new out-of-town retail developments by implementing a 'sequential test' in 1996, which required conclusive proof that there were no viable locations for the proposed development in any established town centres, and the policy has been somewhat successful. More new developments since the mid-1990s have been in town centres (Department for Communities and Local

Government, 2009) and, although Town Centre First has not stopped the decline of retail sales and space in town centres, it has decelerated it and mitigated the more detrimental impacts of out-of-town developments exhibited by the downtown areas in some US cities (Wrigley et al., 2015). Town Centre First policy was also a driving force for the convenience grocery market, as it became more difficult for supermarket retailers to construct new sites on green-field land (Hood et al., 2016).

Despite the efforts of the Town Centre First policy, the financial crisis of 2008-9 had a marked impact on high streets. Commissioned in 2011, the Portas Review (Portas, 2011) aimed to provide an independent insight on the state of UK high streets and recommendations for resilience and recovery. Through secondary research, fieldwork and interviews, Portas gave 28 recommendations to help strengthen high streets across the UK that focused on:

- "*Getting our town centres running like businesses*" highlighted the importance of cooperation between stakeholders and a coherent vision for the future.
- "*Getting the basics right to allow business to flourish*" identified the need to update business rates, parking charges and other regulations and 'red tape' that are barriers for business.
- "*Levelling the playing field*" focused on empowering small businesses and dismantling the monopoly of large chains and out-of-town retail.
- "*Defining landlords' roles and responsibilities*" encouraged community engagement in landlords to reduce the amount of vacant or unattractive storefronts.
- "*Giving communities a greater say*" highlights the need for communication with local people to understand the evolving demand and function of a high street.

Thirteen 'Portas Pilot' towns were selected to try these recommendations; however, they found limited success. A headline from the Daily Mail in 2013 claimed, "Towns have spent just 7% of Mary Portas' £10 million fund to rescue High Streets... and most of it went on reindeer and Peppa Pig costumes" (in Parker *et al.*, 2016) and Mary Portas herself, in hindsight, is critical of the scheme. In 2017, of the government action to her recommendations, she said, "It was a weighted PR campaign which looked like 'Hey, we're doing something' and I hoped it might kick start something - but it didn't" (Fenwick, 2017).

The Grimsey Review, a report published in 2013 by private-sector retail experts, critiqued the Portas Review. It did not dispute her recommendations — also highlighting the need to address business rates, landlord responsibilities and stakeholder cooperation for future high street vitality and viability — however, Grimsey *et al.* (2013) felt that retail should not take centre stage in the future of the high street. Instead, the Grimsey Review stated:

> "Town centre/high street plans must encompass a complete community hub solution incorporating; health, housing, education, arts, entertainment, business/office space, manufacturing and leisure, whilst developing day time, evening time and night time cultures where shops are just a part of the total plan."
> (Grimsey et al., 2013)

The report focused on the multifunctionality of high streets and town centres and how technology could facilitate that through customised deals, flexible workspaces, and high street monitoring. Although the scenarios discussed in the Grimsey Review (Grimsey et al., 2013) are hypothetical visions for the future, the later Grimsey Review 2 (Grimsey et al., 2018) can provide some successful examples, such as the implementation of data portals and the collaboration of the Scottish Towns Policy Group. However, there is no critical or independent evaluation of the recommendations of the Grimsey Review.

The authors of the Grimsey Review were not alone in their critique of the Portas Review. Swinney and Sivaev (2013) in their report 'Beyond the High Street' for Centre for Cities, also emphasised how the Portas Review and consequent government policy put too heavy focus on high street shopping, and instead needed to consider job locations and the city centre economy as a whole. They identified vacant shops and failing businesses not as the problem itself, but as a symptom of a demand reduction, either from a lack of residents, jobs or leisure amenities. In some UK cities, businesses and offices have relocated to out-of-town locations with cheaper rents, more parking spaces, and better connections to the road network; therefore, workers have less need to visit the high street. Swinney and Sivaev (2013) suggest that policy should focus on the centralisation of private sector businesses. For some cities, that involves maintaining city centre attractiveness through congestion control or caps on cost. Other cities should improve the city's attractiveness to business through flexibility in development or conversion of unused city centre space, balancing office space and housing according to need and improving city centre transport links.

Reviews by Portas (2011), Grimsey *et al.* (2013, 2018), and Centre for Cities (Swinney and Sivaev, 2013) are among many bodies of work that present recommendations to improve high street vitality and viability. Others include Genecon LLP, 2011; Wrigley and Dolega, 2011; Coca-Stefaniak, 2013; Hart *et al.*, 2014; Parker *et al.*, 2014; Wrigley and Lambiri, 2014b; Millington *et al.*, 2015, 2018 and C. Parker *et al.*, 2016. Many of these reports highlight how a 'one approach fits all' mentality is not appropriate for high street planning, relying on the judgement of town planners and local government to understand the processes behind these recommendations and discern which would be most suitable for their location. Government guidance is similarly broad, with general recommendations such as providing a wide range of complementary uses and fostering evening and night-time economies (Ministry of Housing, Communities & Local Government, 2014).

However, there is limited evidence of these policy recommendations being implemented and implemented effectively, even on recommendations that garner heavy consensus, such as the concept of a community hub. Grimsey *et al.* (2018) found that, of the 276 local authorities surveyed, 21% had established a long-term ten-year plan for their town centres, and 55% carried out "regular" health checks to assess town centre performance. Fortunately, the wealth of knowledge on UK high streets and town centres is much more significant as we tentatively enter post-pandemic recovery than after the financial crisis; however, there are still barriers to implementation. From the research, there are no 'wrong' recommendations for high street vitality and viability. The consensus is that different high streets need different views and approaches and in locations where one recommendation might be beneficial, another may harm. Therefore, a challenge that many town planners and stakeholders face is discovering which course of action would be most appropriate for their retail centre, and then convincing investors and interested parties of its benefits, garnering consensus and cooperation (Millington and Ntounis, 2017; C. Parker et al., 2016). Two key focuses could help stakeholders achieve this:

1. More efficient methods of transferring knowledge from academia and the private sector to high street stakeholders and decision-makers
2. More measuring and recording key indicators of high street performance.

   Both are explored in Sections 2.1.2 and 2.1.3 respectively.

2.1.2 Understanding high street performance

There is an identified need for more efficient methods of transferring knowledge from academia and the private sector to high street stakeholders and decision-makers (C. Parker et al., 2016). Accessibility is an issue, with many high street datasets and reports behind paywalls, and silos in knowledge from private businesses who may be unwilling to share findings and lose a competitive advantage (Wood and Reynolds, 2012). However, there is plentiful open access research available to all. The universal challenge is time. It takes time to read, synthesise and understand the results and recommendations of reports, and this is time which every stakeholder or member of a town team may not be able to dedicate. Therefore, developing more precise and succinct methods of understanding high street performance can greatly assist in filling any knowledge gaps, allowing for more efficient understanding of key issues and therefore a better construction of a shared vision for the future of a town centre or high street.

In response, several reports and projects have attempted to synthesise the data and information on high street performance to provide decision-makers with succinct key material.

The High Street UK 2020 [HSUK2020] project was one such project that aimed to collect the existing knowledge of retail and high street change and to deliver it to individual locations to assist them in developing a sustainable high street by 2020 (C. Parker et al., 2016). Their method relied on engaged scholarship, where they approached a diverse range of experts and stakeholders to understand what range of factors impact high street vitality and viability (C. Parker et al., 2016). They identified over two hundred factors which were influential to high street vitality and viability, 165 of which local stakeholders could influence. These 165 became 25 top priorities which encompassed many of the recommendations given by other projects, such as community leadership, management, barriers to entry, vision & strategy, retail diversity, liveability, and accessibility (C. Parker et al., 2016). These priorities could be more tangible than the narrower recommendations given by previous reviews, inspiring creative and critical thought from stakeholders. However, their non-specificity could prove intimidating and leave room for debate, constructive or otherwise. Nonetheless, these priorities were communicated to local stakeholders from a wide range of sectors in half-day workshops in ten identified partner locations, and feedback showed that, on average, those who attended strongly understood the HSUK2020 framework and saw how its application to their town centres.

With the positive feedback from their workshop and the success of the HSUK2020 project in partner towns such as Altringham and Ballymena (C. Parker et al., 2016; Theodoridis et al., 2017), the HSUK2020 project has demonstrated that communicating key material and practical strategies to stakeholders through workshops can be a successful method of knowledge transfer. However, there is an immense amount of preparation, logistics and organisation that needs to occur to organise this for the many town centres and high streets in the UK that could benefit.

Classifications can also be tools used to inform stakeholders of the characteristics present in their town centre and compare them to neighbouring towns and cities. An appropriate and consistent classification system can aid both discussions into consumer behaviour and purpose and compare empirical research between different high streets or town centres (Guy, 1998). For example, a decision makers can identify the class their retail centre fits within and analyse other retail centres of the same class to see the impact different intervention strategies had on their high streets. A reliable classification system can be quickly understood and conceptualised, which aids the efficient transfer of knowledge without extensive reading or in-person workshops.

In the past, retail centres have been classified in largely hierarchical structures, mainly concerned with their retail function. Understanding and documenting the retail mix present in different centres helps to inform land use and retailer location planning, and the hierarchies and rankings built from such information had commercial value (Dolega et al., 2021) and the simplification they offered is favoured for decision-making in business and policy (Brown, 1991). Rankings today such as the HDH Vitality Index (Harper Dennis Hobbs, 2021) are powerful marketing tools for retail centres.

However, with recent changes in the UK retail sector diversifying the purpose and demand of high streets, there is a need for more holistic classifications (Dolega et al., 2021). These classifications may concern socio-demographic and cultural aspects, in addition to economic to provide a more rounded classification. Many of these being post-hierarchical in these classifications could emphasise how retail centres may serve different functions or purposes but are equally important to the community and catchment they serve. It also removes the constraints and limitations of a pre-determined structure, such as Central Place Theory. On the other hand, removing the hierarchical aspect could limit the application of these classifications for marketing and comparison purposes. However, by using a holistic mix of indicators and a post-hierarchical structure, it could be argued that these classifications are more accurate reflections of modern retail centres.

ATCM's town centre classification matrix is an example of one of these classifications (Coca-Stefaniak, 2013). Their four classes of town centre reflected both the size of the catchment, the type of visitors they wanted to attract and the economic vs socio-cultural focus of the retail centre. It is a flexible classification, with a centre potentially shifting from 'specialist' or 'community-focused entrepreneurs' to 'global celebrity' dependent on the season (Coca-Stefaniak, 2013). Although there are exemplar case studies for each of the four classes, the system relies on town centres to classify themselves using a 'town personality test' or survey, acknowledging that the vision for the town centre may not align with the users' perceptions.

Other classifications have bypassed this by using a quantitative, data-based methodology. Mumford *et al.* (2021) applied k-means clustering to annual footfall signatures to determine four classes of centre based on what months of the year they were busiest. This produced a useful and informative classification; however it was only limited to 155 towns and cities with footfall data. Dolega *et al.* (2021) based their town centre classification primarily on retail occupancy and vacancy data, enhanced with related demographic factors such as employment and income. Five Supergroups, and within them, fifteen nested Groups were created, each with unique and prominent characteristics. Notably, the retail centre typology developed by Dolega, Singleton and Pavlis has near-complete coverage for Great Britain. It is openly available to download and map form on the Consumer Data Research Centre website, making it highly accessible and applicable.

Classification systems do have limitations. They are assumptive and make generalisations instead of treating each high street as unique and tailoring strategies to their individual context. In addition, as a result of data limitations and the relatively recent need to develop them, few retail centre classifications encompass the multifunctionality of high streets. However, classifications facilitate a quick conceptualisation of vast amounts of data and information, applying illustrative examples and characteristics that may benefit time-restricted stakeholders. Achieving a shared understanding of their town centre or high street, its context and its relationship with surrounding locations is a crucial foundation for developing a future strategy with a clear, cohesive vision.

2.1.3 Indicators of high street performance

One of the challenges local authorities and planners face is difficulty identifying, understanding, and communicating their market position to other stakeholders (Millington and Ntounis, 2017). Key indicators of high street performance can allow stakeholders to achieve and relay an interpretable insight into the health of their high street. In addition, having a current and historical database can allow the quick identification of trends – positive or negative – and give communicable evidence to support current and future investment.

An indicator is a measure used to evaluate performance. A robust and efficient system of performance measurement must employ several interlocking indicators (Mikušová and Janečková, 2010). For example, a combination of 'hard' and 'soft' indicators, such as sales with customer satisfaction, and indicators that capture all stages of the process from input, efficiency, output, and satisfaction with output. Most companies and businesses use key performance indicators [KPIs] to gauge their short and long-term performance and how effective and efficient any investment is. Hristov and Reynolds (2015) investigated different methods of measuring innovation in retailing business and used their findings to develop a framework which classified these into broad categories on financial and non-financial, with the former including measures such as sales and profitability and the latter encompassing store image, conversion rate and customer satisfaction.

Although, as Mary Portas advised, the town centres could run as a business, the vitality and viability of a high street are typically more complex and comprehensive than that of a business. Culture, identity, and place play a significant role, and town planners have generally less control over their town centre environment than a business has over its operation. Therefore, it is even more challenging to develop a reliable system of KPIs.

In 1997, Urbanism Environment Design [URBED] outlined several KPIs for town centre managers, with the intent that they can be used both to monitor high street performance and to be presented to stakeholders and investors to display the potential or impact of their investment (URBED, 1997). Despite acknowledging that town centre performance is social and cultural as well as economic, URBED recommended measuring retail focused KPIs. These included: footfall, sales trends, business surveys, vacancy rates, business formation/churn, property value, investment and retail diversity. However, with omnichannel[4] retailing weakening the retail function of a high street, recent recommendations have shifted to favour a more holistic selection of KPIs.

[4] Omnichannel: Retail from multiple sources (online, mobile and physical)

In 2013, ATCM published their indicator-based performance toolkit, which compromised twenty KPIs under four headings: people and footfall, diversity and vitality of place, consumer and business perceptions, and economic characteristics (Coca-Stefaniak, 2013). In addition to the more traditional KPIs such as footfall, vacancy, and retail offer, ATCM's toolkit also included community spirit, crime - both perception of crime and number of reports - and cultural and leisure offer. From the guidelines in Mikušová and Janečková (2010), this is a significantly more robust selection of KPIs. It includes a mixture of 'hard' and 'soft' indicators and KPIs that monitor actual output alongside consumers perception of output. Current government recommendations cover a similar range of KPIs, albeit less succinctly organised, including consumer experience and behaviour, accessibility, and environmental quality in addition to vacancy, rents, and footfall (Ministry of Housing, Communities & Local Government, 2014).

The prominence and robustness of KPI usage within local authorities varies across the country. Hogg, Medway and Warnaby (2004) found that all 84 town centre managers the surveyed used KPIs to measure the successfulness of their schemes, the most widely adopted being car park usage, footfall figures, vacancy rate, retail mix and crime. Grimsey *et al.* (2018) found that only 55% of local authorities undertook "regular" health checks to assess town centre performance, indicating that the monitoring of KPIs may not be as thorough in practice. It should be noted that Grimsey does not define the frequency of "regular". In addition, although all the town centre managers surveyed by Hogg, Medway and Warnaby (2004) used KPIs, the quantity and combination of KPIs were not discussed. The findings considered the problematic use and potential manipulation of KPIs from a marketing standpoint: their primary collection and usage was to "sell" a town centre as opposed to accurately representing or monitoring progress. It is clear that a representative and robust system of KPIs may not be the priority for some town centre managers.

Hart *et al.* (2014) and Parker *et al.* (2016) also discuss the limited or non-existent monitoring of KPIs. Hart *et al.*, (2014) touch on how evidence of town-centre regeneration is mainly anecdotal, with any measurement focusing heavily on economic indicators and not social. Whereas Parker *et al.* (2016) highlight how many of the partner towns in the HSUK2020 project were not in the practice of collecting primary data, such as footfall and vacancy rates, prior to taking part in the project.

The appropriate measurement and application of a system of key performance indicators can assist in a clearer understanding of the problems a town centre faces or may face. Even though there are robust and modern frameworks for measuring town centre performance (e.g. Coca-Stefaniak, 2013), the capacity for town centres to collect this amount of data can be limited. Time, money, and perceived lack of value can all be barriers to measuring KPIs. In addition, the collection of data may not be straightforward. Town centre KPIs are a mix of qualitative and quantitative data: KPIs such as consumer perception, purpose for visit and attractiveness may have to be collected by methods such as interviews with high street consumers whereas quantitative KPIs such as vacancy, retail mix and footfall may be collected by in-person counts or surveys. There is little information regarding the limitations and assumptions of KPIs, which may be needed for decision makers to appropriately critique and analyse the measurements.

Information that does exist includes 'Healthy High Street? A healthcheck for high streets and town centres': a guide released by the Department for Business Innovation & Skills in 2010. Among the guidance was a section entitled 'Not necessarily signs of decline' that explained how changes such as a decrease in footfall or rents or an increase in vacancy rate may not always indicate decline, as would be the first assumption. Instead, they may indicate change: a necessary step for high streets to evolve and adapt. Similarly, in Wrigley and Dolega (2011) it is discussed how a rising or high vacancy rate is not uniformly negative for town centres or high streets. A healthy high street is constantly in a state of readjustment or churn as businesses move in, leave, and relocate to more effectively meet the market demand. Structural vacancy – a measure of how many units have been vacant for a prolonged period of time – may be more indicative of fragility or weakness in a high street. However, it should be noted this relies on consistent and referable historical measurements.

Footfall is one of the most influential factors of high street performance (C. Parker et al., 2016) and is one of the most measured KPIs (Hogg et al., 2004). Despite this, there is limited guidance for town management regarding its measurement, influences and impact. The following sections will focus on this, discussing the different collection methods and applications of footfall data and how it can fit within the broader picture of the high street.

2.2 Footfall data – collection and applications

Footfall is the number of people that travel through a specific area at a given point in time (Lugomer et al., 2017). Also referred to as pedestrian counts, people counts or foot traffic, footfall is relevant in retail (e.g. Harding and Powell, 2011; Lugomer and Longley, 2018; Mumford *et al.*, 2021), marketing (e.g. Kirkup and Rafiq, 1999; Denison, 2005; Yiu and Ng, 2010), and urban planning literature (e.g. Islam, Jones and Dye, 2014; Graham and Peleg, 2017).

2.2.1 Footfall Applications

Applications of footfall tend to fall under four headings: footfall as an indicator, footfall classifications, footfall as potential and modelling footfall.

Footfall as an indicator

Footfall is used as an indicator in primarily retail literature and town centre management to assess high street health and performance. Widely recommended by government, academia, and private industry (Coca-Stefaniak, 2013; Graham, 2016; Ministry of Housing, Communities & Local Government, 2014), footfall is a key performance indicator in 87% of town management schemes (Hogg et al., 2004). Warnaby and Yip (2005) found that increasing footfall was a key focus for promotional planning and as a performance indicator for out-of-town regional shopping centres, and the HSUK2020 project highlighted the importance of available footfall data to inform and monitor high street revitalisation projects and to act as evidence of their successfulness (Millington and Ntounis, 2017; Ntounis and Parker, 2017). Morley, a town in Yorkshire and one of the HSUK2020 test towns, installed a footfall sensor to generate data to inform their revitalisation strategy and Ballymena, in Northern Ireland, measured an 8.1% increase in footfall comparing December 2016 to December 2015 partly as a result of their improved Christmas marketing campaign (Parker et al., 2017).

Footfall can also be a marketing indicator for private companies. Typically, sales are the sole key performance indicator for a company promotional strategy (Denison, 2005) as the data is generally automatically collected by the Point of Service checkout system and is easy to analyse. However, when sales data is combined with footfall it can give a more powerful image of the success of a store or a particular advertisement campaign. Denison (2005) demonstrated how footfall could give new insight into what particular aspects of a campaign (e.g. TV advertisements, radio advertisement or store window displays) were most effective at bringing customers in to the store, even if they did not spend money on that visit.

In addition, footfall can be an indicator of the impact and recovery of a place after a major disaster. Harding and Powell (2011) successfully used footfall to measure the impact of the 2011 earthquake in Christchurch, New Zealand on businesses, and monitor their subsequent recovery. Likewise, Bras *et al.* (2021) analysed footfall data from Dutch shopping streets to determine the impact of lockdowns, face masks and social distancing measures.

However, the prominence of footfall as an indicator can also be problematic. Most studies or methodologies do not consider other influences on footfall besides the promotion, scheme or event being measured. Factors such as weather, school holidays or the purpose of visit (whether due to a promotion or otherwise) are not taken into account. Some do not clarify the day of the week or the time of day of measurements, which could also have a significant effect. There is an assumption that the only significant influence on footfall is the phenomenon being measured. This assumption is, to an extent, unavoidable when using footfall as a measure. However, it is rare to see it acknowledged or discussed and therefore difficult to discern if the author assumes this inherent understanding of the reader or if they have considered these factors in their interpretation at all.

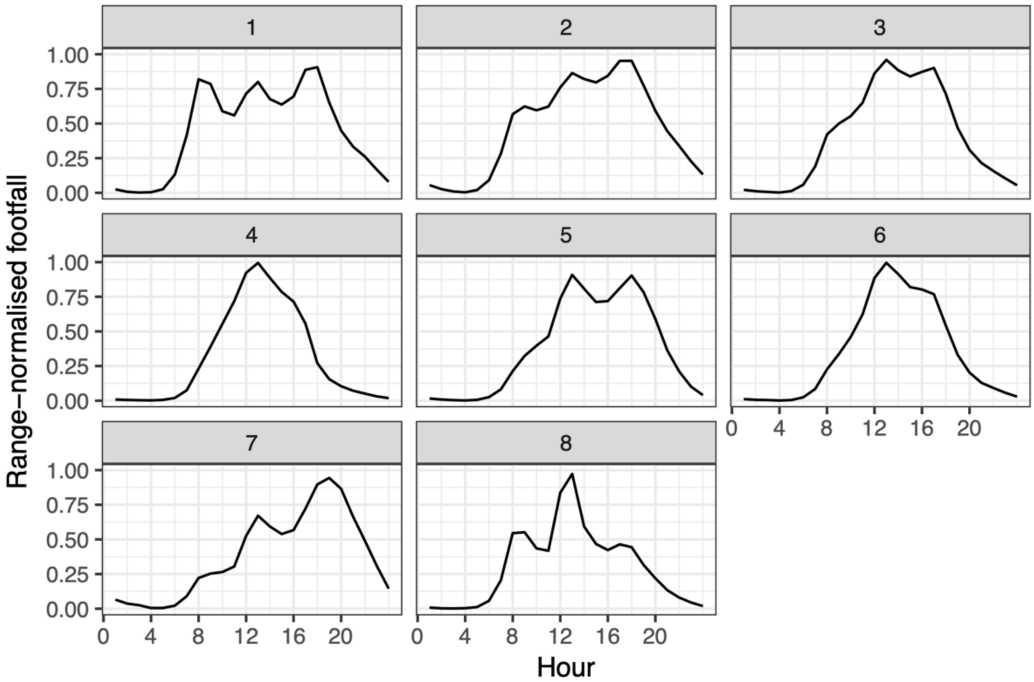
Footfall classifications

In recent years, footfall data has been used to develop retail centre classifications based on temporal footfall signatures. Classifications can be efficient and accessible tools to transfer knowledge and to better understand retail centres (this is discussed at further length in Section 2.1.2). As high streets and town centres shift from a retail focus to a multi-functional focus, classifications that are based on solely economic and retail indicators may not be fit for purpose. Classifications that are based on footfall or activity may be a better measure in an omnichannel retail context, as they capture aspects of community as well as economic strength, which are both needed for high street vitality and viability.

The implementation of automated footfall collection methods (discussed in further detail in Section 2.2.3) allows for the generation of continuous footfall datasets, such as Local Data Company's [LDC] SmartStreetSensor data and Springboard's footfall data. These datasets contain powerful information regarding the patterns of footfall across the country that can be used to define new activity-based classifications. These can be applied to any retail context where footfall is monitored, and data is collected.

As the technological developments that have made footfall more monitorable are relatively modern, analysis of this footfall data remains an evolving field. Currently, there are two main footfall based classifications based on UK footfall data: Lugomer and Longley (2018) and Mumford *et al.* (2021).

Lugomer and Longley (2018), used representative, weekly footfall signatures to develop a micro-scale classification based on Local Data Company's footfall data. By using a clustering algorithm to group similar locations together, they created eight distinct classes, the daily footfall patterns for each are given in Figure 2-1. These were named according to the shape and/or an assumed usage, such as 'Quiet mornings, busy evenings' and 'Busy lunchtimes with both commuting peaks'. The classification Lugomer and Longley (2018) made is based on micro-site data, as opposed to data that covers an entire retail centre; therefore, their classification is applicable for other micro-sites, such as streets or one shopping outlet.



Cluster	Proposed name	Cases	Percentage (%)
1	Commute and lunch	84	13.88
2	Gradual rise	80	13.22
3	Consistent afternoons	169	27.93
4	Midday top	119	19.67
5	One-directional commute	29	4.79
6	Lunch time with minor afternoon commuter inflow	90	14.88
7	Quiet mornings, busy evenings	19	3.14
8	Busy lunchtimes with both commuting peaks	15	2.48
	Total	605	100.00

Figure 2-1 Footfall classes from Lugomer and Longley (2018)

(Adapted from Figure 1 and Table 1 in Lugomer, K. and Longley, P. (2018) 'Towards a comprehensive temporal classification of footfall patterns in the cities of Great Britain', in Leibniz International Proceedings in Informatics, LIPIcs. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, p. 43.)

Mumford *et al.* (2021) also built a classification from weekly footfall signatures, applying footfall data from provider Springboard. In contrast to Lugomer and Longley (2018), their classification covered entire retail centres and for retail centres instead of micro-scale or street level. They produced four distinct clusters, named according to the distinctive function of the town centre: holiday (tourism), comparison, speciality and multi-functional. The average monthly footfall signature for each is given in Figure 2-2.

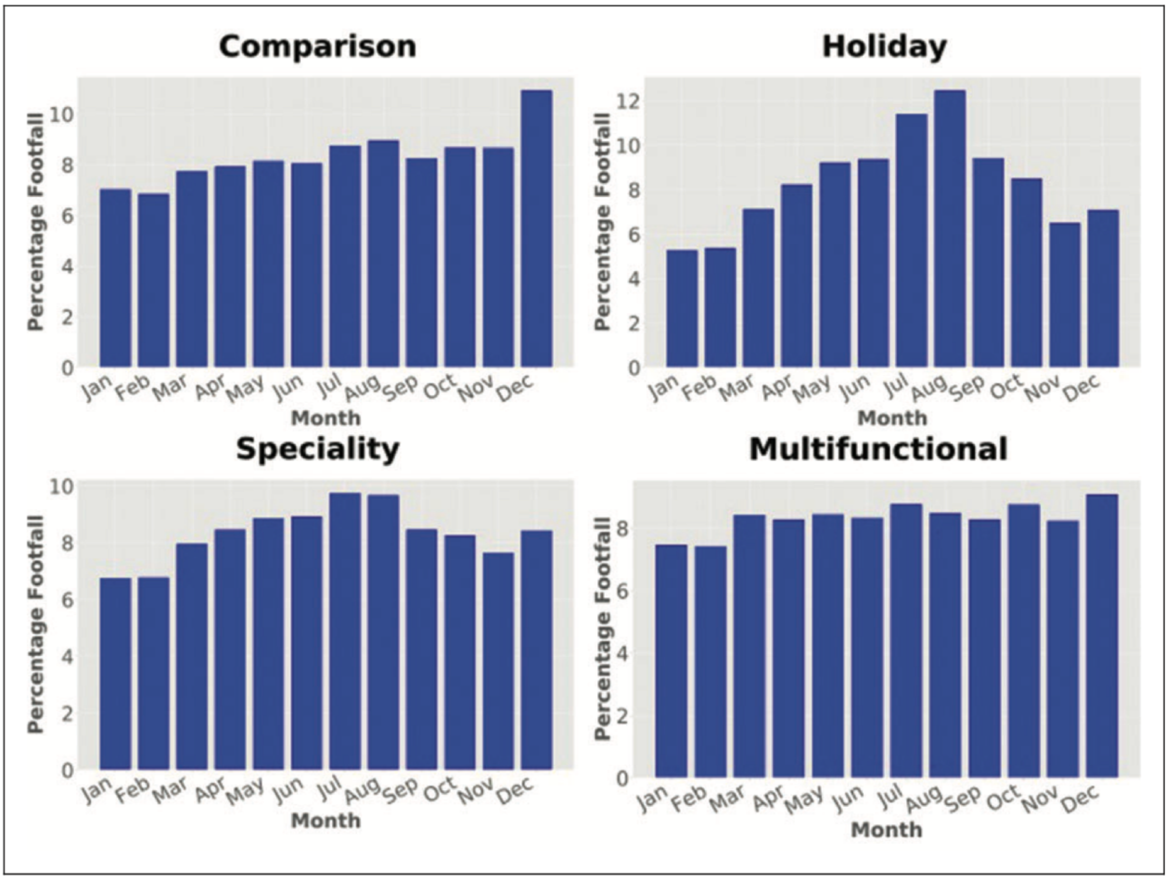


Figure 2-2 Classification from Mumford et al. (2021)

(Adapted from Figure 1 in Mumford, C. et al. (2021) 'Footfall signatures and volumes: Towards a classification of UK centres', Environment and Planning B: Urban Analytics and City Science, 48(6), pp. 1495–1510.)

The two classifications differ in their data sources and in their applicability, with the classification by Lugomer and Longley (2018) concerning to micro-site locations, and the one by Mumford *et al.* (2021) concerning entire retail centres. To be useful and applied to a new location, both rely on some prior knowledge or footfall monitoring of the respective area. This could be intuitive, as many people concerned with a retail centre and it's performance will likely have a perception of when it is busier and when it is quieter. However, relying on these perceptions could introduce bias. Both classifications contain qualitative descriptions of their different classes (although these are more developed in Mumford *et al.* (2021)) which could be used to bridge this gap.

However, neither integrates data sources other than footfall data and activity levels into their classification. This information does have value alone, particularly as footfall captures the economic and social uses of a retail centre or location. However, more insights could potentially be gained from integrating other quantitative data sources, such as retail mix, retail vacancy, transport connectivity and function, into these footfall classifications. This has the additional benefit of potentially allowing a classification to be applied in places where footfall data isn't monitored or collected.

A further limitation of these initial attempts at footfall classifications is that they are strongly linked to the function of a location; however, a function may change over time or cyclically throughout a year or even a week or day. This is particularly the case for Mumford *et al.* (2021), whereas Lugomer and Longley (2018) is more descriptive than functional. The inclusion of retail centre function (e.g. Holiday destination, Comparison Retail centre) is valuable extra information and insights that are a valuable quality of the Mumford *et al.* (2021) classification, yet any retail centre that does not fit a clear function is termed as 'Multi-functional', instead of reflecting how that centre's function switches from Holiday to Comparison and back over time. Flexible classifications that could be adapted and changeable over time could be an interesting avenue for further research of footfall classifications, and result in classifications that are more reflective of their area of interest. However, this is still quite a novel research area, emerging from the availability of continuous footfall data. It is currently limited to these two examples, but there are many interesting developments that could be made.

Footfall as potential

Footfall can also be applied as a measure of the potential customers a store could have. A buyers-to-shoppers ratio, or conversion rate, can be defined as the proportion of people that visit a shop to the number that make a purchase (Yiu and Ng, 2010). Conversion rate has been measured at 32-48% (Underhill, 2009; Yiu and

Ng, 2010). Underhill (2009), used surveys and questionnaires to determine the conversion rate in New York department stores at 48%, while Yiu and Ng (2010) found direct observations of shoppers was a more reliable methodology and reduced over-reporting, measuring the conversion rate in Hong Kong malls at 32%.

Attraction rate, as opposed to conversion rate, is the number of pedestrians who pass the shop and then decide to enter (Graham, 2016). However, as Graham (2016) notes, research on attraction rate is more minimal, and there is little to no research investigating a combination of footfall density, attraction rate and conversion rate.

In his work, Graham (2016) proposes an initial empirical generalisation, or law, that states that no matter the footfall, the attraction rate will be about 4%, and the conversion rate, 42%. Therefore, of the people walking past a store at any given time, 4% will enter. Of that 4%, 42% will make a purchase. If proven, this would improve the usefulness of measuring footfall for businesses and retailers, as footfall could essentially predict the success and profitability of their store. However, no further work has tested the findings, and the original study relies on multiple assumptions: it only considered stores that sell a skincare brand and measurements only took place between 14:00 and 15:00 and on Wednesdays and Saturdays. The study did consider outlets in a variety of retail settings including stations and shopping centres; however, the geographic scope was limited to just London. The findings of Denison (2005) also contradict this law, as they measured a higher in-store footfall, or attraction rate, during their promotional campaign. Graham (2016) would suggest this increase would be because of higher footfall in general.

Modelling footfall

There exists a wide range of different analyses that analyse and model the predictability and patterns of footfall data in order to learn more about pedestrian behaviour. Reviewing or comparing the wealth of research on modelling and analysing footfall can be challenging due to the heterogeneousness of the evidence base. Studies can be unique in the goal of their research, what factors are measured, the spatial scale, the wealth or limitations of the data, the methodology used, and many other considerations that can call for different approaches. There are certain patterns and commonly used methodologies, but no established framework for the analysis and modelling of footfall data.

Micro-simulation models can be a common approach when analysing footfall at an indoor or small-scale urban areas, such as intersections, shopping centres or stations. Typically, an agent-based model is applied, which simulates the decisions of pedestrians moving through space. These can consider utility and efficiency, short-term prediction, barriers and obstacles and the behaviour of other pedestrians (agents) in the model (Hoogendoorn, 2004; Kitazawa and Batty, 2004). These micro-simulation models can be useful for understanding crowd flows and best practise for evacuation (Batty et al., 2003), in addition to helping understand human behaviour and how pedestrians interact with space (Kitazawa and Batty, 2004; Turner and Penn, 2002).

Network analysis models of pedestrian volumes tend to encompass larger areas, such as cities or neighbourhoods. They estimate pedestrian counts or footfall for street segments and intersections, and can also be combined with variables of demand, such as tourist attractions, retail mix or size. In some transport planning literature, these are called 'direct demand models' (Cooper et al., 2021; Munira and Sener, 2017), however this naming is not consistent. Works such as Stavroulaki *et al.* (2019), Bolin *et al.* (2021) and Sevtsuk (2021) apply a direct demand approach without using this term.

In many cases, a space syntax approach is used. Space syntax is a technique that breaks down the components of a street network and represents them as maps and graphs that reflect the accessibility and integration of those places (Hillier and Hanson, 1984; Omer and Kaplan, 2017). Axial maps describe streets by the minimal set of visual sightlines needed to cover the open space entirely. Evidence shows that topological analysis of the connectivity and integration of these axial lines is a reasonable predictor of pedestrian flows, which allows for the generation of reasonably accurate maps that predict pedestrian flows over a spatial area.

Lerman, Rofè and Omer (2014) predicted pedestrian flows in Bat Yam in Israel, reporting an R2 value of 0.62. In Amsterdam, Read (1999) applied a similar method which was able to predict 60-70% of pedestrian flows, and studies in London have produced similar levels of accuracy (Hillier et al., 1993; Jiang, 2009). When combined with information on transport accessibility, land use, and street capacity in a multi-level model or 'direct demand model', Desyllas *et al.* (2003) could model pedestrian flows in London, giving an R2 value of 0.82. Using a space syntax approach, roughly 60-70% of pedestrian flows or footfall can be predictable (Jiang, 2009).

However, the process of generating axial maps can be time-consuming and subjective, even with the assistance of graphical software. Analyses are often limited to one location or neighbourhood, as creating larger or multiple axial maps is laborious. In addition, axial maps use the longest line of sight, which relies on the assumption that all space is usable by the pedestrian, and this is not always the case, particularly for non-pedestrianised roads. Liu and Jiang (2012) developed a methodology that instead used street centre lines to generate axial lines, which they successfully applied to six street networks. More studies have found success applying this method in other locations (e.g. Sun, 2012; Shen and Karimi, 2016). Although it is less time-consuming than the alternative, Liu & Jiang's method still remains a somewhat extensive task, reliant on an appropriate data source.

A significant challenge when reviewing works that have applied a network analysis or direct demand approach is that the model is often designed for a certain context and research goal. The model may be reasonably successful at predicting footfall for that one city or location; however, it is unclear how transferable that model is to other locations, and how generalisable any observations and conclusions are. Stavroulaki *et al.* (2019) aimed to investigate this, building a footfall prediction model that was applied to London, Amsterdam and Stockholm. They found that there were certain predictors of footfall that were significant in all locations, namely the connectivity of the network through space syntax measures and the proximity of public transport hubs. The other factors tested (proximity to schools and proximity to retail markets) were found to be insignificant. Their findings show that there may be some element of transferability in footfall modelling, and that a model that works for one location might also work for another. However, it should be noted that their study was limited to three European capital cities and did not include regional centres or smaller retail centres. It also applied data from a variety of neighbourhoods within those cities, which were not necessarily retail centres. There remains a clear research gap here in terms of investigating if the factors that determine footfall are consistent across retail centres, regardless of their differences or similarities in terms of size and function.

The temporal dimension is something that can be overlooked or disregarded in these studies. Often, this is due to data limitations. For example, Stavroulaki *et al.* (2019) surveyed their 53 neighbourhoods during a three-week period in October, only taking measurements from 6:00 am until 10:00 pm on weekdays. Therefore, they are unable to explore whether their study might yield different results on weekends or other months of the year. Aggregation of footfall data by location is also common. Lee, Yoo and Seo (2020) collected data from 7:30 am until 8:30 pm on weekdays for their locations in Seoul; however, they aggregated this to improve ease of analysis and removing any temporal aspect to focus on the spatial relationships.

This is not to say the temporal aspect of footfall is left unexplored. There are many studies that employ time series analysis methods to study the patterns within previously collected footfall data. These patterns are used to gain a greater understanding of an area, such as through the classifications of Lugomer and Longley (2018) and Mumford *et al.* (2021), mentioned in the previous sub-section. They can also be used to forecast future footfall and predict how many people might pass a location at a certain date or time, which is valuable information for retailers and many other stakeholders. To our knowledge, there is yet to be a study that combines time series methods and network analysis or direct demand models to acquire a footfall prediction over space and time. This is likely a result of data limitations - both in terms of number of locations within a sample and the amount of historic data - the complexity of the challenge and the relatively modernity of footfall modelling. Most studies focus on a singular city, retail centre or store (e.g. Klapka *et al.*, 2013; Omer and Kaplan, 2017; Wang *et al.*, 2017; Chen and Zhou, 2020; Cooper *et al.*, 2021; Liu *et al.*, 2021) or have a specific research aim, for example increasing pedestrian safety (e.g. Schneider *et al.*, 2012; Munira and Sener, 2017; Liu *et al.*, 2021), or the specific impact of a certain factor on footfall (e.g. Makkar, 2020;

Cooper *et al.*, 2021; Martínez-de-Albéniz and Belkaid, 2021). Replicability and generalisability are not necessarily priorities in their investigations.

The multidisciplinary nature of the field can also lead to difficulties in collation and comparison as different fields use different terms and frameworks for essentially identical or similar phenomena. For example, many studies incorporate consumer demand or factors of attraction into multi-level models to predict footfall. In transport planning literature this can have the specific term 'direct demand models' whereas in retail literature it will often just be described as a multi-level or regression model.

In summary, the relative modernity, multidisciplinary nature and data limitations within studies which model footfall has resulted in two major research gaps in the spatial modelling of footfall data: the transferability and generalisability of methods over different environments and the consideration of temporal patterns and changes over time.

2.2.2 Footfall data collection methods

There are many methods of collecting footfall data, but the cheapest and historically used technique is manual counts, where a person stands in a location for an amount of time and counts the number of people who pass using a tally. Many studies apply this method (e.g. Desyllas *et al.*, 2003; Harding and Powell, 2011; Graham, 2016; Omer and Kaplan, 2017). Despite its wide usage, it does have significant limitations.

Firstly, manual counts are limited in duration. Due to the nature of in-person measurements, counts tend to last for an allotted period, consisting of a few hours of a day, or a few days of the week. For example, in Graham (2016) measurements were taken on Wednesdays and Saturdays between 2:00 and 3:00 pm. Often the justification for the times selected is based on expertise and local knowledge, not quantitative observations, and this could result in important fluctuations being overlooked. Secondly, manual counts are also limited spatially, with the number of locations with simultaneous measurements being controlled by the number of researchers counting. Physical safety may also be another consideration researchers have to make while conducting manual counts, particularly in areas that aren't well lit or have high crime rates. Finally, manual counts are reliant on human perception and subject to error. Denison (2005) described manual footfall counting as "very difficult to do accurately, because human fatigue and boredom soon set in and that leads to huge inaccuracy".

Recent technological advances have resulted in the adoption of automated methods of footfall data collection, which present a solution to the challenges manual counts face. An automated method relies on a permanently fixed sensor or device that measures the number of people passing by, eliminating the need for a human counter. There is a wide selection of electronic footfall counters using different technologies, including infrared, radar and thermal imaging, with the most modern using Wi-Fi counting or object detection from video recordings (Javare et al., 2020; Lugomer et al., 2017; Soundararaj et al., 2020; Sruthi, 2019).

The key benefit of automated data collection over manual counts is the potential wealth of data. Automated methods do not have the same time limitations as manual methods, and can continuously collect data throughout the day, week, month, and year. This sufficient data allows for more powerful analysis, including understanding variation in hourly and weekly fluctuations of footfall, identifying and contextualising periods of extreme fluctuations in footfall and developing an understanding of long-term trends. In addition, automated methods avoid human measurement errors and bias that hinder manual counting methods.

Despite their benefits, the adoption of automated data collection methods of collecting footfall data has been slow. Manual counting is still common, with many studies not exploring the option of automatic or electronic methods in their methodologies (e.g. Desyllas *et al.*, 2003; Graham, 2016; Omer and Kaplan, 2017). Those that do, such as Harding and Powell (2011) quote price, maintenance, superfluous data analysis and an inability to investigate anomalous results as drawbacks to automated methods, in their case, infrared counters. Kirkup and Rafiq (1999) investigated why businesses, in particular fashion retailers, may hesitate to adopt automated methods and found that the accuracy of the data was a concern. In particular, the inability to distinguish between individuals or groups and needing regular maintenance and verification to ensure that they are working optimally. This potential for inaccuracy could lead to concerns over the validity of the data, which limits its applicability as a key performance indicator, and there was scepticism over the realistic 'value-added' of installing a system.

Despite the significant developments in technology since Kirkup's research, achieving accurate footfall counts is still a challenge faced in research today (Javare et al., 2020; Lugomer et al., 2017). Lugomer *et al.* (2017) investigated the accuracy of Wi-Fi sensors, which count smartphones as a proxy for passers-by, and found the measurement error compared to manual counts varied from 68.5% to 25.0% depending on the sensor location. Javare *et al.* (2020) reported an accuracy of 71.4% from their automated CCTV object detection algorithm. Some of the limitations these methods face, such as counting in crowds or poor lighting, might also be challenges for accurate manual counts; however, many are problems unique to the sensor type and automated methods[5]. As such, there is an application for manual counts to be used in combination with automated methods in order to validate these measures.

As technological advances change the way footfall can be measured, awareness of tracking and monitoring presents new concerns. Current law protects individuals, with strict regulations over who can view and use personal data. However, there is distrust and scepticism from consumers. There are privacy concerns, with consumers unnerved by footfall counting, feeling as though they are "being followed", and of exploitation, with companies making more money from their data without compensation (Kobsa, 2014). Even when their information is aggregated and anonymised, as per current law, consumers are still hesitant to disclose data to retailers (Kobsa, 2014). Increased anonymisation and security of smartphones by manufacturers is also a growing source of inaccuracy for Wi-Fi footfall sensors.

Automated methods of footfall data collection generate a wealth of continuous data that can give valuable new insights into flows and patterns of use; however, they can

[5] These issues of privacy and technological limitations will be explored in more depth in Chapter 3.

produce more error than manual counts and take time and skill to collect, clean and analyse. Sensors can be expensive and are an investment that companies are sceptical about making when manual counts are more cost-effective. Consequently, manual footfall counts are still an attractive option for many stakeholders, despite their limitations.

2.3 Footfall as an indicator of high street performance

High street performance in the UK is discussed alongside two terms: vitality and viability. Vitality is the busyness of a retail centre across different temporal and spatial scales. Although this can encompass many modes of transport and inter and intra urban travel, footfall is an intuitive measure of high street vitality. In combination with its potentially cheap and easy measurement, the relevance of footfall could explain why it is one of the most popular KPIs for town management to measure (Hogg et al., 2004).

Footfall is also related to viability – the ability to attract sustainable and continuing investment – as there is a positive correlation between footfall and spend (Bras et al., 2021; Graham, 2016; Warnaby and Yip, 2005). In addition, consistent or rising footfall indicates that the retail centre meets the current demand of consumers, which increases the likelihood of a return on investment (Graham et al., 2019). The value of footfall as a KPI is its capacity to capture both retail vitality and viability.

There is a clear link and understanding between footfall and the performance measures it indicates; however, relatively little comprehensive research covers the factors that influence footfall. Evidence tends to be anecdotal, focusing on one city and the impact of one event or influence, and there is no standardised guidance for what influences and factors to consider when analysing footfall measurements and how significant an impact they could have. For example, empirical evidence shows that footfall is negatively affected by poor weather, but this often is not quantified, and the impact of other factors such as retail mix or centre size are not considered. Understanding and quantifying the impact of different influences on footfall is of particular importance to town centres that measure footfall using manual counts, as they are tasked with building a complete picture from a snapshot of data.

Footfall can be determined by many factors on different spatial and temporal scales, from national influences that impact long-term trends such as economic health or political climate to micro-locational factors that affect footfall on a minute-by-minute basis, such as proximity to a station. The factors that influence footfall are complex, interrelated, and challenging to quantify, and this comprehensiveness can present problems when attempting to analyse the causes and processes behind fluctuations and patterns in footfall.

In order to gain a better understanding of the breadth of factors that influence footfall, the following section compiles the relevant literature and research on the topic, from large-scale national and regional factors to smaller-scale local and micro-locational factors. The micro, meso, macro framework is used, where micro-scale considers factors of the retail centre itself, meso-scale relates to the relationship between different retail centres and macro-scale contains national or large-scale factors, such as politics or the national environment.

2.3.1 Factors that influence footfall

Macro scale factors

A factor that influences footfall on a macro scale already discussed in this chapter is the political context. The 2008-09 financial crisis and the coronavirus pandemic have had a long-term negative impact on UK footfall through a drop in consumer confidence and disposable income and policies that discouraged or prevented visits to local high streets and town centres. However, government policy can also have a positive impact on footfall. For example, the 'eat out to help out' scheme offered a 50% discount for dining establishments during August 2020, increasing evening footfall by 18.9% and lunchtime footfall by 9.6% (Kollewe, 2020). The impact of political context is not uniform across the UK and is challenging to quantify independent of local influences; however, it is essential to acknowledge when comparing footfall measurements, particularly those taken over several years.

A similar impact on footfall to be considered for England and Wales is the 1994 Sunday Trading Act (GOV.UK, 2021), which reduces the permitted opening times for stores larger than 280 square metres to six hours on Sundays. Larger stores also act as crucial footfall anchors for smaller stores around them (Williamson et al., 2006). As a result, retail centres in England and Wales consistently experience lower footfall on Sundays.

Another factor that can influence footfall on a macro scale is national holidays and events. Black Friday, Bank Holidays, January Sales and the Christmas shopping season are examples of events that might affect footfall. Annual events such as these are widely considered drivers of footfall; however, there is little openly available data that quantifies this. Reports often focus on when footfall is below the norm; for example, Storm Deirdre hit the UK on a key pre-Christmas shopping weekend in 2018 and footfall dropped by 9% (BBC News, 2018a), and Black Friday footfall was down by 58% in 2020 compared to pre-pandemic (Partington, 2020). It is relatively intuitive to consider the impact of annual events on footfall, as they often occur at a similar time each year. However, there are some exceptions. Sporting events such as footfall and rugby competitions that are culturally celebrated in pubs and bars in the UK could also drive footfall, yet they are less intuitive considerations.

Weather is another macro factor that has a temporary impact on footfall. Bad weather such as snow or storms can negatively impact footfall on an hourly, daily or weekly temporal scale, as consumers may not want to travel outside in poor weather for non-essential purchases. Parsons (2001) found that temperature and rainfall had a negative impact on the number of visitors to a shopping store, whereas humidity and sunshine were insignificant. Makkar (2020) successfully predicted the impact of weather forecasts for future footfall; however, the magnitude of the influence is not stated.

Although political and economic climate, governmental policy, national holidays and events and weather are macro-scale factors that impact a large regional or national area, the impact they have locally may vary due to a range of different factors including the function of the centre, the provision of services and the local demographics, culture and identity.

Meso-scale factors

At the meso-scale, a factor that influences footfall is the proximity and offer of competing retail centres.

One of the earliest attempts to model the number of visitors to a retail centre was Reilly's Retail Gravitation Theory (Reilly, 1931). Building on Newton's Law of gravitation, Reilly's Law states that two retail centres will attract consumers in proportion to the population of the retail centre and inversely in proportion to the distance between the consumer and the retail centre. Therefore, at the midpoint between two retail centres, the consumer will be attracted to the one with a larger population.

Reilly's Law proved a robust theoretical basis that many researchers have developed since, most notably Huff. Huff's Model of Trade Area Attraction (Huff, 1963) built on the same principles as Reilly's work; however, it applies shopping centre square footage instead of population and travel time instead of distance. It also incorporates a parameter that reflects the willingness of a consumer to increase travel time depending on the type of trip. Huff's work was the first to recognise the importance of the consumer's perspective in retail gravitation modelling. Further work (e.g. Bucklin, 1971; Nakanishi and Cooper, 1974) has generalised Huff's store size attribute to measure attractiveness based on several micro-scale to retail centre scale attributes, including a mix of retailers, store image, price, ambience and service quality.

Although there are criticisms of applying retail gravitation models in today's omnichannel retail context, they have been valuable tools for determining retail centre catchments and footfall for decades. The 'attractiveness' (often defined by the researcher) of a retail centre to those nearby and the connectivity or travel time for a consumer are crucial determinants of footfall.

Connectivity can be split into two key definitions: interconnectivity and intraconnectivity. Interconnectivity refers to how connected a retail centre is to other retail centres, and intraconnectivity considers the connectivity inside a retail centre, at a micro-locational scale. Factors of interconnectivity such as access to road, rail and air can make a retail centre more attractive as it is widely accessible to a larger population. Berry *et al.*, (2016) found that proximity to major transports hubs can increase footfall and sales, particularly at commuting times. Having good access to car parking is also important as many consumers will avoid using public transport in favour of the convenience of their personal vehicle. Therefore, the proximity of a retail area to a public car park can influence the number of visitors and footfall for the entire retail centre (Coca-Stefaniak, 2013).

Micro-locational factors

There are also factors that influence footfall that vary within a retail centre on a micro-locational scale, for example, intraconnectivity. A main aspect of intraconnectivity for footfall is walkability. There are many contesting definitions of walkability. Therefore, here it is defined as the attractiveness of a particular street to a pedestrian. Walkability can encompass physical characteristics, safety and security and network connectivity (Lo, 2009). Indeed, specific physical properties of streets have been shown to increase their walkability. Wide streets with gentle slopes that are well lit have been shown to be the most attractive (Erath et al., 2017; Unwin et al., 2017).

Additionally, how the street is situated within the wider network has proven to be a reliable indicator of pedestrian counts (Hillier et al., 1993; Raford and Ragland, 2006). Well-connected streets tend to have higher footfall as it is often the shortest route from their origin to their destination. This can be determined by various centrality measures, including closeness and betweenness, which respectfully capture the closeness of a node to other nodes and the prominence of a node as a bridge between other nodes (Freeman, 1977; Porta et al., 2009). As such, they can be used to predict busy intersections or nodes.

Retail centre function can also impact footfall on a daily, weekly and seasonal temporal scale. The function is the purpose to which a high street or town centre serves its users, and most retail centres are multi-functional – they simultaneously act as workplaces, residential areas, shopping and leisure destinations, throughways or tourist attractions (Millington et al., 2015). Characteristics such as the presence of anchor stores or the tendency towards premium or value goods can all indicate the retail centre identity, who it may appeal to, and consequently, when they may visit (Guy, 1998). For example, locations with a high concentration of employers and businesses typically have higher weekday footfall (Berry et al., 2016; Swinney and Sivaev, 2013) and streets with more convenience outlets may experience shorter, more frequent trips compared to those with a more comparison retail focus (Guy, 1998). Areas with a high concentration of leisure outlets such as restaurants, bars and clubs may experience more footfall in the evenings, from 5pm onwards and tourist destinations such as Cornwall can see grocery retail demand double during on-season (Newing et al., 2018).

In general, larger retail centres offer a more varied retail mix and are more multi-functional. A comprehensive and diverse retail mix has been proven to be a key footfall driver on a retail centre scale and a micro-locational street-level scale. The better the ability of the micro-location retail offer to match consumer demand of the consumers, the busier it can become, increasing the magnitude of footfall . Similarly, a retail centre with a high structural vacancy rate would have a poorer retail offer, therefore decreasing long-term footfall. There are certain popular destinations that act as attractors, concentrating footfall to particular micro-locations (Scheurer and Porta, 2006). Anchor stores, restaurants, entertainment venues and tourist attractions have all demonstrated this ability (Bras et al., 2021; Hart et al., 2014; Teller and Alexander, 2014; Üsküplü et al., 2020; Yuo et al., 2003).

Local festivals and events can also impact footfall at a retail centre and micro-locational scale. When successful, food and cultural festivals, sporting events and other local celebrations can boost footfall temporally for the duration of the event and potentially contribute to a sustainable level of future footfall. For example, the Giant Spectacular Liverpool's Dream event drew in 1.3 million people over 4 days in October 2018 (Giantspectacular.com, 2019), and the annual festivals in Edinburgh festivals attract more than 4.5 million visitors a year with 75% of residents believing that the festivals made Edinburgh a better place to live (Naylor et al., 2016).

2.3.2 Conclusion

Academic literature points to many macro, meso and micro-scale factors which influence footfall, however the evidence base is limited. Most research is empirical or contextual and no literature exists which quantifies the impact of a combination of these influences. In the past, this could be attributed to a lack of data. Manual footfall counting limits the amount of historic and current data available and the temporal resolution. However, advances in sensor technology have made it possible to gather frequent and consistent measurements of footfall through permanently installed sensors. This provides a valuable opportunity now to build on this previous research and establish a quantitative evidence base regarding the different factors that influence footfall.

2.4 Chapter Summary

This chapter has explored the literature and research around the wider topic of high street vitality, viability and revitalisation, evaluating which approaches have appeared more effective than others and identifying the barriers to effective intervention. After the 2008-2009 financial crisis, the future sustainability of the high street became a key priority for government. As such, many reports and research studies were undertaken to learn more about the challenges faced by the high street and what could be done to make them more adaptable for the future. However, there is a disconnect between the recommendations given and those which are successfully applied to retail centres. Two priorities were identified to help achieve this: efficiency in transfer or knowledge and improvement in monitoring of key performance indicators in high street environments. Workshops were shown to be an effective way of achieving the former goal, however they require significant time and investment. Classification systems were identified as an alternative, allowing recommendations and conclusions from recent research to be given in a succinct way that is tailored to the needs of a particular retail environment.

Focus was put on footfall as a widely measured and monitored key performance indicator and one of the most influential factors of high street vitality and viability. The measure has many strengths – it accounts for economic and community uses, it is responsive to changes, measuring by manual counts is very accessible – however, comparatively little is known about the patterns behind footfall itself. For example, there are qualities that are often assumed to impact footfall such as vacancy rates, weather or retail function and mix, however, there is little qualitative or quantitative evidence to support these claims. The question could be asked how much is footfall impacted by each of these factors? Is this something that is predictable, or is there an element of chance or randomness within any impact on footfall? If footfall is so changeable and there is not clear understanding of the reasons behind these fluctuations, should footfall be considered such a strong KPI?

There is seemingly a lack of research that specifically focuses on footfall, given its prominence and importance. Historically, this could be justified through difficulties in collecting enough footfall data to be able to research these questions; however, with technological advances in automated sensing technologies making pedestrian flows more monitorable than ever before, now there is ample opportunity to fill these research gaps and learn more about processes in retail centres through footfall. By reviewing the existing literature on footfall, the research gaps identified include,

- ◊ Footfall classifications which integrate supplementary contextual data, and which could be adaptable to fluctuations in function.
- ◊ Better understanding of the relationship between footfall and spend, including testing and developing the proposed empirical generalisation by Graham (2016) which draws a connection between footfall, attraction rate and conversion rate.
- ◊ In footfall modelling, testing if network analysis or direct demand models are transferable between retail centres of differing size and function and integrating temporal context of the footfall measurements into any analyses
- ◊ Development of a more complete view of the factors that impact footfall, with evidence that demonstrates the potential size and duration of their effects.

The analyses in this thesis will aim to contribute to these identified gaps in the following ways.
Chapter 4 addresses the first gap by creating a new classification of retail centres which is informed by footfall and contextual data. It also addresses the last gap by investigating how different variables of locational context impact footfall, quantifying this amount and exploring how it varies with time. Chapter 5 also addresses the final gap by investigating the specific impact of events and festivals on footfall. On the

other hand, Chapter 6 will aim to address the third aim by creating a model that predicts footfall based on both location and time, and testing how generalisable this model is to different cities.

The one gap that will not be addressed is the relationship between footfall and spend. This is due to data restrictions – it can be challenging to source sales data or revenue from retailers, as this data is very valuable to them and therefore protected. In addition, it is difficult to obtain sales data for a location where there is also nearby, reliable and complete footfall data. Hopefully, further research will be able to address this gap one day in the future.

3 Data and study design

The following chapter will introduce the datasets that will be used in the analytical part of the thesis (Chapters 4, 5 & 6). All three chapters will largely use the same datasets in different ways to achieve their individual research aims. Therefore, the datasets will be given a broad overview in this chapter and issues of data quality and representation will be explored and the assumptions that underlie these datasets will be made apparent. This chapter will also briefly outline the study design and rationale for the analysis completed in the rest of the thesis, introducing the aims, objectives and methodological approach that will be applied.

This chapter will be structured as follows. Section 3.1 will introduce the key supplementary datasets, including Local Data Company's [LDC] Retail Unit Address data and the Consumer Data Research Centre [CDRC] Retail Centre Boundaries and Typology (Pavlis et al., 2018). Then Section 3.2 will introduce the SmartStreetSensor project, a collaboration between LDC, University College London [UCL] and the CDRC and the foundational dataset for this analysis. It will detail how the project employs Wi-Fi-based methods to produce footfall data, the assumptions inherent in the data, the benefits and drawbacks of the technique and summarise the other options available for continuous automated footfall data. Section 3.3 will then be dedicated to detailing the sources of error within the SmartStreetSensor dataset and the potential impact on accuracy. Section 3.4 gives an overall summary of all the datasets. The final section, Section 3.5, will then detail the study design of the following three analytical chapters. It will introduce the research questions, the methodological approach that will be used and the data that will be employed in each analysis.

Sources for all datasets discussed in this chapter will be given in Appendix 3.1.

3.1 Supplementary Datasets

Supplementary datasets will be used in this analysis to give context to the SmartStreetSensor project footfall data and provide insight into any retail, demographic or morphological trends. These data sources contain information such as the distribution and location of different retail units and transport services in a retail centre, delineations of a retail centre itself, the street network and different measures of population. This section will evaluate the data collection, completeness and limitations of each dataset, with particular focus given to LDC's Retail Unit Address Data and the CDRC Retail Centre Boundaries and Typology. For ease of understanding, a list of the supplementary dataset that will be used in this thesis can be found below:

- ◊ Retail Unit Address Data
- ◊ Retail Centre Boundaries and Typology
- ◊ UK Census
- ◊ National Public Transport Access Nodes [NaPTAN] dataset
- ◊ OpenStreetMap
- ◊ Car Parks Dataset.

The first section will focus on evaluation and exploration of the LDC Retail Unit Address Data, a dataset of retail units across the county. Then Section 3.1.2 will discuss the CDRC Retail Centre Boundaries and Typology, which includes delineation and categorisation of retail areas. Section 3.1.3 will evaluate the remainder of the datasets in the list above, exploring their applicability and limitations.

3.1.1 Retail Unit Address data

Collected and maintained by LDC's surveyors, the Retail Unit Address Data contains information about retail units across Britain, including the address and coordinates, whether it is vacant or occupied, and, if it is occupied, the shop name and category in LDC's unique typology. This information is invaluable to understand the retail context which surrounds any footfall sensor. The dataset is updated frequently to ensure it captures the dynamicity of retail centres accurately. For the analyses that will be completed in this thesis, three snapshots were used: July 2017, July 2018 and July 2019. Similar data sources include the openly available OpenStreetMap Buildings, the Valuation Office Agency data which is safeguarded and the Ordnance Survey Address Data which are both licensed dataset.

To preface, it should be noted that the data is not exhaustive – it does not contain every retail unit in Britain – however, it includes a large selection of retail centres from across the country[6]. The most recent data available (from July 2019) contained 618,634 retail units across 1,098 locations. Data from the Valuation Office Agency and Ordnance Survey will be more complete than the LDC data, the former as it is maintained by council tax records and the latter as it is informed by Royal Mail records. It is difficult validate the completeness of the data as these data sources are not openly available. A comparison with data from OpenStreetMap buildings does prove the LDC data to be accurate in the areas where it does have coverage. For example, the same area in central Liverpool included 1,605 individual buildings in the OpenStreetMap data and 1,593 units in the LDC data. Although the LDC dataset may not cover as many unique locations as the other datasets, it has good completeness in the areas which are surveyed and the most up to date snapshot from July 2019 covers almost every town or city with a SmartStreetSensor footfall sensor installed[7]. The benefit which the LDC data does offer that compared to these other sources do not is detailed retail categories and subcategories, which will be discussed later in this section.

The analyses in this thesis will use three snapshots of the Retail Unit Address Data (July 2017, July 2018 and July 2019), but these are not directly comparable. The July 2017 snapshot, which was sent earlier and used in the Chapter 4 analysis, has some

[6] Excluding Northern Ireland

[7] The two sensors installed in Belfast and Bangor in Northern Ireland in May 2018 are not represented in this dataset.

differences compared to the July 2018 and 2019 snapshots that are important to note.

Firstly, as the complete dataset will not be necessary for the analyses in this thesis, the July 2017 snapshot only includes, as requested, a sample of locations of interest, given in more detail in Section 3.4. However, the July 2018 and 2019 snapshots that were received subsequently included the entire Retail Unit Address dataset available at that time. That consisted of 592,656 units for 2018 and 618,634 for 2019.

Secondly, the structure of the July 2017 version differs from the more recent snapshots, as shown in the metadata tables in Appendix 3.2. Some variables in the 2017 dataset were not present in the 2018/2019 dataset, such as 'PremiseStatus', which recorded whether a unit was vacant or not, and 'MultipleName', which identified whether the business was part of a chain. Fortunately, these variables were simple to recreate from the Category information, which identified whether a unit was vacant or not, and from checking for duplicate values for ShopName to see whether or not a business was part of a chain.

Thirdly, the 2018/2019 snapshots introduced a variable that identified multiple retailers within a store, for example, concessions inside department stores. These were not counted in the 2017 version of the dataset. This methodological change inflates the raw count of units in the more recent data. As an example, Harrods Department Store in London had one observation in 2017, but in the 2019 data, it had 150. For a more typical department store, for example John Lewis in Liverpool, the count was inflated from 1 to 23. Therefore, in terms of raw count and proportion of types of units, the 2017 & 2018/9 data are not directly comparable.

Finally, the structure of LDC's categories and subcategories typology differs between the two versions. The following subsection will discuss typology itself in more depth; however, it is clear from data exploration that locations categorised as 'Non-Retail' within the 'Miscellaneous' class are present in the 2017 version of the dataset but not in more recent versions. These include police stations, universities, colleges and some government buildings. As retail units were the primary focus of this analysis, this did not present a problem.

Retail categories

Within the Retail Unit Address data, LDC categorises different units by their use or retail offer. Their typology has three levels: Classification[8], Category and Subcategory. There are seven classifications, 44 categories and 422 subcategories, as summarised in Table 3-1. The most prominent classes in the dataset are comparison stores, leisure units and services, which make up roughly a quarter of all units each. The remaining quarter is predominantly convenience stores and vacant stores, with some non-retail and miscellaneous units. The vacancy rate fluctuates between the three versions of the dataset, but averages at 8.7%, which is slightly lower than comparable figures of the time (9.4 – 12.8%) (British Retail Consortium, 2017; Knight Frank, 2019; Wood, 2019).

[8] 'Classification' is not present in the more recent versions of the dataset. However, it can be inferred from the category and sub-category variables.

Table 3-1 Classes, categories and subcategories in LDC's Retail Unit Address Data and their prevalence in 2017, 2018 and 2019

Classification	Description	Prevalence in the datasets		
		2017	2018	2019
Comparison	16 categories & 140 subcategories, including clothes shops, book shops, charity shops and electrical goods.	24.4%	27.4%	26.8%
Convenience	6 categories & 23 subcategories, including supermarkets, convenience stores, grocers and petrol filling stations.	11.6%	11.3%	11.0%
Leisure	5 categories & 125 subcategories (129 in 2018/9), including restaurants, bars and pubs, hotels and entertainment venues.	25.5%	23.1%	23.1%
Service	12 categories & 82 subcategories (83 in 2018/9), including hairdressers, banks, garages and post offices.	23.5%	25.6%	25.7%
Miscellaneous	3 categories (2 in 2018/9) & 46 subcategories (19 in 2018/9), including medical centres, hospitals, stations and government buildings.	6.6%	3.8%	3.7%
Non-Retail	2 categories & 5 subcategories, including Royal Mail delivery offices and shopping centres and markets.	0.1%	0.2%	0.4%
Vacancy	Vacant units.	8.3%	8.6%	9.3%

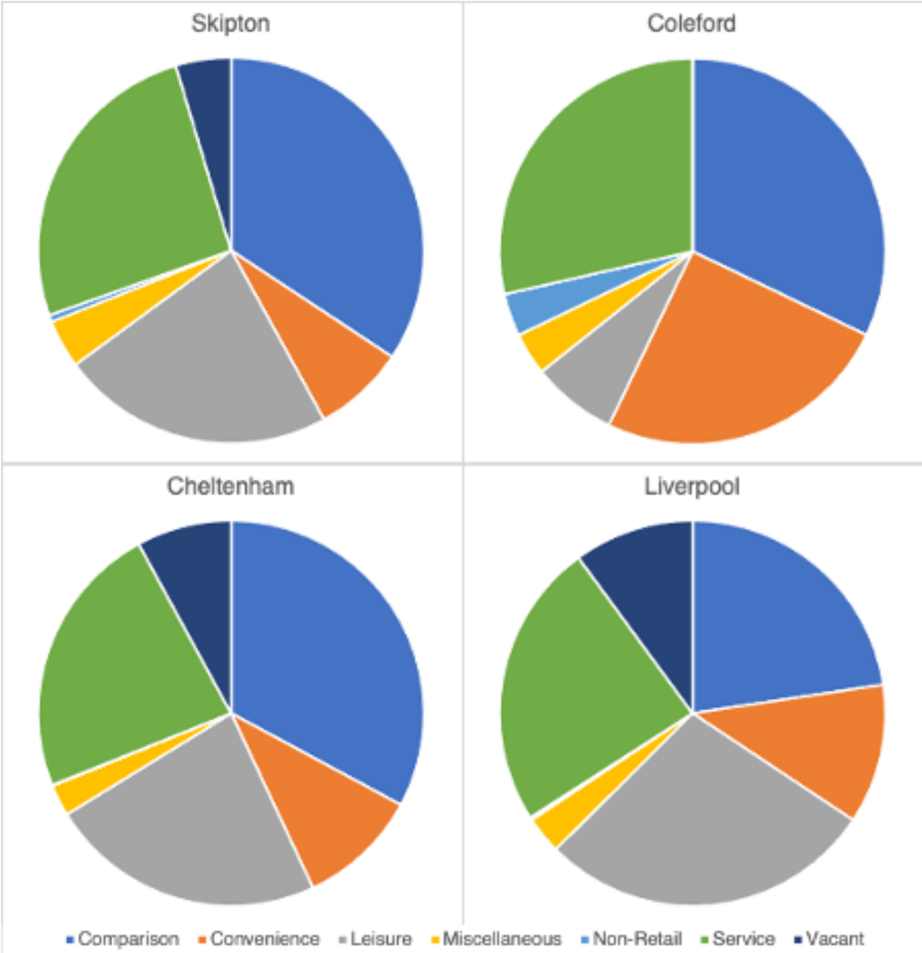


Figure 3-1 Classification of units in four different towns and cities in England[9]

Figure 3-1 shows the how the proportion of units of each class can vary between different cities and towns. These examples are taken from the classification in Coca-Stefaniak (2013). Comparison, leisure and service tend to encompass most of the units, with proportionally more leisure units in a larger city such as Liverpool. Liverpool also has less Comparison units, and more Vacant units. A smaller, local community, such as Coleford in Gloucestershire, has more Convenience units and fewer Vacant units. Skipton is a small town in Yorkshire with local and regional attraction and a specialist retail focus, whereas, Cheltenham in Gloucestershire is classified as a 'sustainable destination' with an established national and international

[9] These four locations were chosen as they are exemplar of the four types of centres identified in (Coca-Stefaniak, 2013). Therefore, it was presumed that they would represent a variation of retail mixes. The Retail Unit Address data represented is from July 2019.

reputation (Coca-Stefaniak, 2013). Despite this, they appear similar in terms of their proportion of different classifications of units.

The classification level provides a useful overview of the dataset, but it can mask complexity within the classes. For example, plumbers and beauty salons are both services; however, they are very different in practice. The categorisation allows for some delineation between these; however, within categories, there remain nuances. For instance, 'Hairdressers, Health & Beauty', the most prevalent category, comprises ~12.5% of all units. Within 'Hairdressers, Health & Beauty' there are the expected salons and barbers, but also gyms and dentists, which fall under the 'Health' label, although not being intuitively linked. The subcategory level aptly captures this nuance but is limited by its size. With over 400 different subcategories, there might not be enough generalisation. This thesis will use a combination of subcategories, categories and classifications which are tailored to what is being investigated.

3.1.2 Retail centre boundaries and typology

An accurate and standardised taxonomy of retail centres in Great Britain is invaluable for national or regional retail analysis. Knowing where a retail centre is located – delineation – and its role and significance in the broader context of retail – typology – assists understanding of trends in the data, draw comparisons between retail centres, and determine how results could be applied on a national scale.

However, as discussed in Chapter 2, it is challenging to create a standard of retail centres as they are constantly evolving in both use and extent. Thurstain-Goodwin and Unwin (2000) established a set of retail centre boundaries are an accurate and verified standard and were adopted by the Department of Communities and Local Government [DCLG] in 2004. However, the dataset is spatially and temporally limited, encompassing just England and Wales and only providing a snapshot of retail geographies from the early 2000s, which is outdated today.

Retail Centre Boundaries

The boundaries defined by Pavlis, Dolega and Singleton (2018) are a more recent and comprehensive alternative. They benefit from the vast amounts of data that have become available in the last decade to create a higher granularity retail delineation than was previously possible. A modified spatial clustering algorithm (DBSCAN) was applied to the LDC Retail Unit Address data explored in Section 3.1.1 to identify dense, spatially collated groups of units indicative of retail centres. Over 3,000 retail centres were defined in their boundaries published by the CDRC, more than double the amount previously identified by the DCLG in 2004.

However, it does have some limitations. Although the boundaries were calculated in 2018, the LDC Retail Unit Address data used was from 2015. A 2021 version was published after completing the analysis using Valuation Office Agency and OpenStreetMap data, which gives a clearer picture of this difference. The most recent version contains over 6,000 retail centres, including 79 retail centres in Northern Ireland. A breakdown and comparison of the DCLG retail boundaries and the two versions of the CDRC boundaries can be found in Table 3-2. The increase in identified retail centres from 2004 to the 2021 boundaries exemplifies how the quality and completeness of data has increased in the last two decades.

Table 3-2 Different retail centre boundary datasets and the number of centres defined in the constituent countries of the UK

Countries	DCLG (2004)	CDRC (2018)	CDRC (2021)
England	1,233	2,873	5,610
Wales	79	128	342
Scotland	0	252	392
Northern Ireland	0	0	79
Total Dataset	1,312	3,253	6,423

For the analyses in Chapters 4, 5 & 6, the 2018 version of the boundaries will be used. Although it is not as complete as the 2021 version, it was available at the time and, has the benefit of the Retail Centre Typology.

Retail Centre Typology

Developed by Dolega *et al.* (2021), the typology[10] incorporates elements of composition, diversity, economic health, size and function to group similar retail centres into coherent classes. The groups and supergroups developed by Dolega *et al.* (2021) are multi-dimensional and based on retail offer, size, and affluence regardless of hierarchical structure. Although this provides a more rounded classification, it does not have an inherent order or hierarchy (e.g. village, town, regional centre, national centre) which could limit its application.

There are five coarse 'supergroups' and 15 nested 'groups', which can be used to draw similarities between patterns observed in different retail centres and infer possible trends in unmeasured retail centres. The classes are summarised in Table 3-3.

[10] The typology is designed to be updatable and potentially can be applied to the 2021 retail boundaries. At time of writing, this is not yet available.

Table 3-3 Retail Centre Typology Supergroups and Groups

Supergroup	Group	Key characteristics
1. Local retail and service centres	1.1 Diverse urban service centres	Upmarket, minor urban centres, densely populated catchments, London-dominated, higher diversity, hospitality service
	1.2 Local urban convenience centres	Urban centre, independent retail and food service, convenience goods and some comparison
	1.3 Inner urban service centres	Inner urban small shopping parades, low diversity, highly independent, focussed on consumer service and non-retail
2. Retail, shopping and leisure parks	2.1 Primary retail, shopping and leisure parks	Large retail parks, extensive catchment, broad offer including mass brand fashion and department stores, very low vacancy
	2.2 Less diverse retail, shopping and leisure parks	Smaller and less diverse retail parks, non-leisure, predominantly comparison goods
3. Leading comparison and leisure destinations	3.1 Premium shopping and leisure destinations	Top regional and sub-regional destinations, affluent market towns, diverse and comprehensive offer, retail, services, leisure, home to top national chains
	3.2 Mass market and value retail large centres	Semi-regional, less affluent destinations, smaller catchments, broad mass and value retail/service, fewer anchors
	3.3 Affluent and premium retail destinations	Small number of centres, affluent catchments, upmarket fashion and multiple retailers, premium brands, non-value, low vacancy
4. Primary food and secondary comparison destinations	4.1 Vibrant secondary urban destinations	Smaller urban district centres, densely populated and less affluent catchments, vibrant, service hubs
	4.2 More affluent district destinations	Town/major district centres, more affluent, high diversity, mass and value retail, local leisure hubs
	4.3 Urban value destinations	Less affluent, higher crime and unemployment, less diverse, value oriented, non-premium, fast food hubs
5. Traditional high streets and market towns	5.1 Traditional high streets of rural Britain	Small market towns, rural Britain, independent, diverse, convenience and comparison retail and leisure offer
	5.2 Suburban and market town high streets	Small suburban centres, commuter belt, less diverse, convenience retail and consumer and business services
	5.3 Diverse and affluent urban leisure destinations	Inner-urban traditional high streets, affluent, independent and speciality, e.g. boutiques, tea-rooms
	5.4 Indie and value oriented high streets	Small high streets, less affluent, deprived, value oriented independent retail and consumer services

(Adapted from Table 2 in Dolega, L. *et al.* (2021) 'Beyond retail: New ways of classifying UK shopping and consumption spaces', *Environment and Planning B: Urban Analytics and City Science*, 48(1), pp. 132–150. doi:10.1177/2399808319840666.)

The typology is also heavily physical retail based, with many of the variables used to group the retail centres pertaining to the mix of leisure, services, and independent stores, or alike measures. In a time when consumers are increasingly reliant on online retail, it could benefit from incorporating a measure of online shopping permeation. However, as is, it facilitates comparison and understanding of similarities between retail centres which will be invaluable for this analysis.

3.1.3 Other datasets

The LDC Retail Unit Address data and the CDRC Retail Centre Boundaries and Typology give an accurate and up-to-date depiction of the micro and macro-scale retail context of locations across Great Britain. However, as discussed in Chapter 2, many non-retail characteristics influencing footfall are not captured in these datasets. Therefore, other supplementary datasets are used throughout the analyses to add demographic and morphological context to the footfall data. These are summarised in Table 3-4.

Table 3-4 Supplementary datasets – provider, year, and the purpose they serve in this analysis

Name	Provider	Year	Purpose
UK Census	UK Data Service	2011	Population and Workplace Population
NaPTAN	Department for Transport	2014	Locations of Transport hubs
OpenStreetMap	OpenStreetMap	Constantly Updated	Road and path network
Car Parks Dataset	Department for Transport	2015	Locations of car parks

These datasets have some limitations that should be considered. Firstly, the UK Census is temporally limited. Although it will provide accurate demographic information regarding daytime population and the prominence of locations as hubs of employment, it is almost a decade out of date, which will introduce significant uncertainty as demographics and population are dynamic. The UK Census covers the entire country, however some measures, such as workplace population, are not available for all areas.

The two datasets provided by the Department for Transport (NaPTAN and Car Parks dataset) are also slightly outdated. However, transport hubs, stations, bus stops, and car parks tend to be permanent, so the uncertainty introduced is negligible. They cover Great Britain but exclude Northern Ireland.

Finally, OpenStreetMap is an open-source database of roads, paths, and locations across the world. It is reliant on crowdsourced data from volunteers which is collected through surveys, GPS devices, local knowledge, and aerial photography. Due to this, the coverage, quality and completeness of OpenStreetMap data can vary, however the accuracy is found to be high in locations of high participation and population, for example, retail centres (Barron et al., 2014; Haklay, 2010; Zhou et al., 2022).

One of the benefits of crowdsourced or volunteered geographic information (VGI) is that it can be very quick to update when there are changes in the environment, for example the construction of a new street or road. In comparison, other data sources are less responsive, for example the OS Open Roads data is updated every six months (Ordnance Survey, 2023). However, a downside to this is that the OpenStreetMap data can be difficult to backdate. It would be very resource consumptive for OpenStreetMap to save historic versions of their dataset as it changes so frequently. Therefore, the data used from OpenStreetMap is a snapshot at the time of analysis, which can lead to inconsistencies if other data sources are older than that.

This section has introduced the key supplementary datasets that will be applied to give context and provide further insights into the patterns within the footfall data. The next section will introduce the SmartStreetSensor project footfall data – the primary dataset that will be used in this analysis.

3.2 SmartStreetSensor Project Footfall Data

The SmartStreetSensor project is a collaboration between retail location data company LDC, University College London [UCL] and the CDRC, a centre that encourages and facilitates academic engagement with industry through consumer data research. The footfall data provided by the SmartStreetSensor project is collected through automated Wi-Fi sensors and will form the cornerstone of these analyses. The following section describes the dataset, giving a brief overview of how the data is collected, exploring the dataset and its distribution and comparing the LDC dataset to other footfall providers.

3.2.1 Data collection

Footfall data is collected from across the UK through Wi-Fi sensors developed by LDC. Wi-Fi enables the exchange of data by radio waves, allowing devices such as smartphones, laptops, tablets, printers and computers to connect to each other and to the internet wirelessly. Probe requests are a standardised mechanism with which these devices deploy to seek out nearby available Wi-Fi access points before connection, to make the process more efficient. Wi-Fi-capable devices deploy probe requests continuously, ubiquitously and passively, making them an effective measure for how many devices are in a location at any given time. LDCs sensors are placed in retail shop windows to record these probe requests. The assumption is that the count of mobile devices would be indicative of users, making it a functional proxy for footfall (Soundararaj et al., 2020).

Passive counting of consumers through Wi-Fi probe requests has many potential uses such as crowd control and real-time crowd management (Bonné et al., 2013; Weppner et al., 2016), monitoring of shopping behaviours and potential user profiling (Ebeling et al., 2018; Hwang and Jang, 2017) and public transport occupancy detection (Kalikova and Krcal, 2017). A reliable count of pedestrians walking past a storefront at any given time has applications for store retailers, town stakeholders and researchers. For retailers, it allows them to calculate store performance metrics like conversion rate, gauge success of advertisement strategies and modify staffing or store opening hours to be more efficient. By measuring footfall trends, town stakeholders can monitor their high street health, ensure efficient deployment of day-to-day management such as policing or cleaning and measure the successfulness of any revitalisation strategies. Footfall data from across the country has allowed LDC, journalists, and researchers to understand how events such as Christmas or the COVID-19 pandemic have impacted physical retail on a national scale.

3.2.2 Data exploration

The SmartStreetSensor footfall dataset (after this point referred to as the dataset) contains five-minute counts from sensors across the UK. The raw data has 11 variables, as shown in Table 3-5. This version does not contain personal data such as individual anonymised MAC addresses. However, it is commercially valuable, and under license conditions, is classed as safeguarded data (CDRC, n.d.). The non-aggregated dataset is available as secure data from the CDRC; however, the extra detail it provides will be unnecessary for the analyses in this thesis. As safeguarded data, registration and project approval was needed.

Table 3-5 Description of variables in the dataset

Variable Name	Class	Description
timestamp	Date time	The date and five-minute interval in Coordinated Universal Time [UTC] (does not consider daylight savings time)
location	Integer	The unique location identifier of the sensor
device	Integer	The unique identifier of the sensor device. This is not synonymous with location.
probes_global	Integer	Count of probe requests that are non-randomised (global)
probes_local	Integer	Count of probe requests that are randomised (local)
macs_global	Integer	Count of MAC addresses that are non-randomised
macs_local	Integer	Count of MAC addresses that are randomised
counts_global	Integer	Filtered non-randomised count
counts_local	Integer	Filtered randomised count
adjusted_local	Integer	Filtered randomised count adjusted according to the ratio of the non-randomised count
imputed	Logical	Shows whether the values are real or estimated to avoid gaps. Any gap of less than 30 minutes is imputed (CDRC 2019).

The five-minute footfall estimates are the sum of “counts_global” and “adjusted_local”. The total dataset ranges almost five years, from July 2015 – May 2020. The data is in CSV format, and it is separated into months to avoid large file sizes. There is also a ‘locations’ file that has information about each sensor's location and positioning, including address, coordinates, the sensor device identifier, and notes about the environment around the sensor. In total, sensors have been placed in 1,037 locations across the UK.

3.2.3 Distribution of sensors

This subsection will explore the distribution of those 1,037 locations in the following ways:

- Across time
- Geographically across countries and regions of the UK
- Across retail catchment size and type
- Across type of store or unit the sensor is located in

It should be noted that the analysis in Chapters 4, 5 & 6 will not use the entire sample of 1,037 sensors. Different samples of the dataset are derived depending on what that chapter will seek to explore and when the analysis will be completed. For ease of understanding, the details of sample derivation, pre-processing and distribution will be discussed in their respective chapters and Section 3.4.3 will give an overview and justification for these three samples.

Distribution across time

The first sensor was installed in Hammersmith, London, in July 2015, and throughout that year, a further 53 sensors were installed. Figure 3-2 is a Sankey diagram that visualises how the distribution of sensors grew and shrunk from 2015 to 2020. The installation of sensors did depend on uptake from retailers; therefore, it was sporadic. Most were installed during 2016 and 2017, with 495 and 358 sensors being installed each year, respectively. Fewer were installed in 2018 and 2019, and none in 2020. The largest sample of active sensors was in 2018, with 877 sensors.

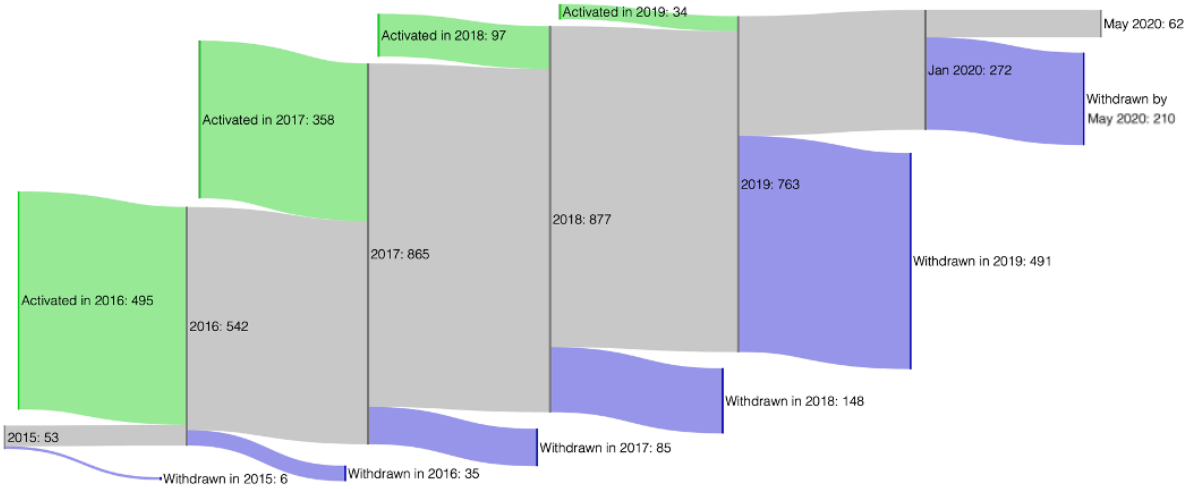


Figure 3-2 Sankey diagram showing the distribution of sensors over time[11]

Until 2018, sensor retention was strong, with 90-94% of sensors for 2015, 2016 and 2017 still active the following year. The sensors that were withdrawn during the time were either in locations that were found to be unviable or in locations where the retailer withdrew interest in the project.

However, sensor retention began to decrease in 2018 and 2019. Only 31% of the sensors active in 2018 were still active by January 2020. This correlates to the introduction and growing prevalence of MAC address randomisation by iOS and Android devices, a topic which will be discussed in the next section. As the reliability and accuracy of Wi-Fi-based methodologies decreased, many of the sensors were withdrawn, and stakeholders in the SmartStreetSensor project began to search for other alternatives to capture footfall data.

This continued throughout 2019 and into 2020, where the number of active sensors decreased dramatically over the first five months of the year. Only 23% of the sensors that were collecting data at the start of the year were still doing so by May.

[11] A sensor 'Activated in 2017' would be present in 2017 but not in 2016. A sensor 'Withdrawn in 2017' would be present in 2017 but not in 2018. The duration the sensor was active in any year was not taken into account for this diagram. For 2020, the difference in the number of active sensors between January and May was so significant that the decision was made to include it here.

The main reason for this drop in sensor numbers was the COVID-19 pandemic. By law, non-essential retail and services such as clothing stores, department stores and restaurants had to remain closed from the 26th March 2020 until the 15th June. As a result, many retailers switched off power to the sensor, either for economic reasons or because the footfall data would not be used if they were closed.

On average, each location has 25 months of data, with the maximum being 50 months and the minimum just one month. However, this is often not consecutive. Figure 3-3 displays the temporal distribution of a sample of sensors in the dataset from 2016 until 2019. It is apparent that most sensors have some gaps. For example, sensor 476 in Figure 3-3 has two gaps: a week in September 2016 and 80 days in the latter half of 2017. In addition to the sensor-specific gaps, there are gaps in the entire dataset. Data is missing for periods of a few days in September, October, and December 2019.

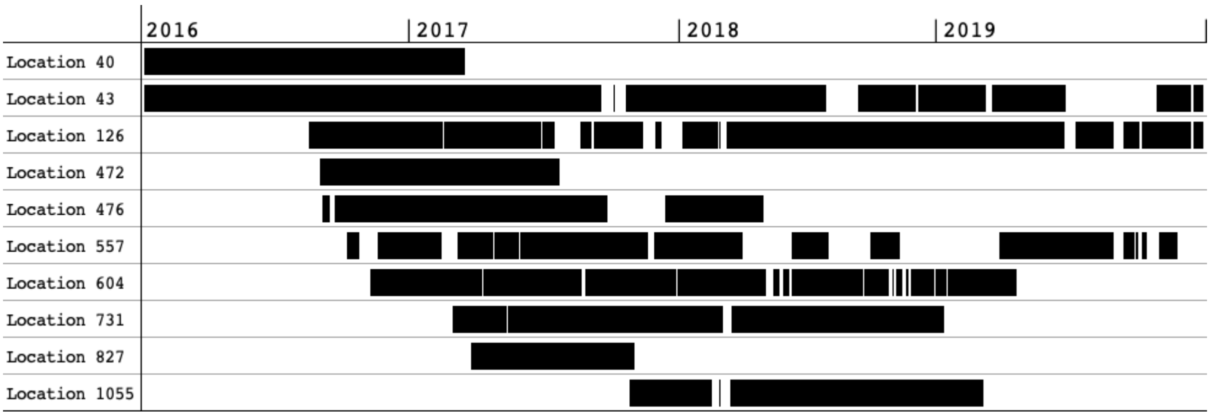


Figure 3-3 A random sample of ten sensors and the completeness of their data

The temporal quality of the data is sensor dependent. 131 of the sensors were missing <1% of data; however, 71 had >30% missing, and 4 had >90%. On average, sensors were missing 11.6% of data, limiting the dataset's applicability for temporal or time series analysis.

A time series can be defined as a series of data points in time order. Time series data can be used to understand seasonal patterns and underlying trends and forecast into the future. A typical application of time series data is in stocks and shares. Time series modelling based on the past value of a company's value could be used to forecast how they might rise or drop in price in the future.

Gaining the ability to apply forecasting methods to footfall data and accurately forecast footfall over space and time is one of the major benefits of investing in automated sensor technologies. To model annual seasonal fluctuations, at least two complete years of data are needed. However, the limitations in the temporal distribution of this data, both in the range of time captured and from gaps within the data, restricts the capability to use more complex analyses. Some researchers navigate this by using a representative average week of footfall (Lugomer and Longley, 2018; Mumford et al., 2021), which allows meaningful research to be undertaken while can miss out on annual or monthly fluctuations.

Geographic distribution

The sensors are distributed across all regions of the UK, as shown in Figure 3-4. The majority are installed in England (n = 910, 88%), predominantly in Greater London, the South East and Yorkshire and the Humber. Out of the remainder, most are in Scotland (n = 107, 10%), leaving Wales and Northern Ireland with 16 and 2 sensors, respectively.

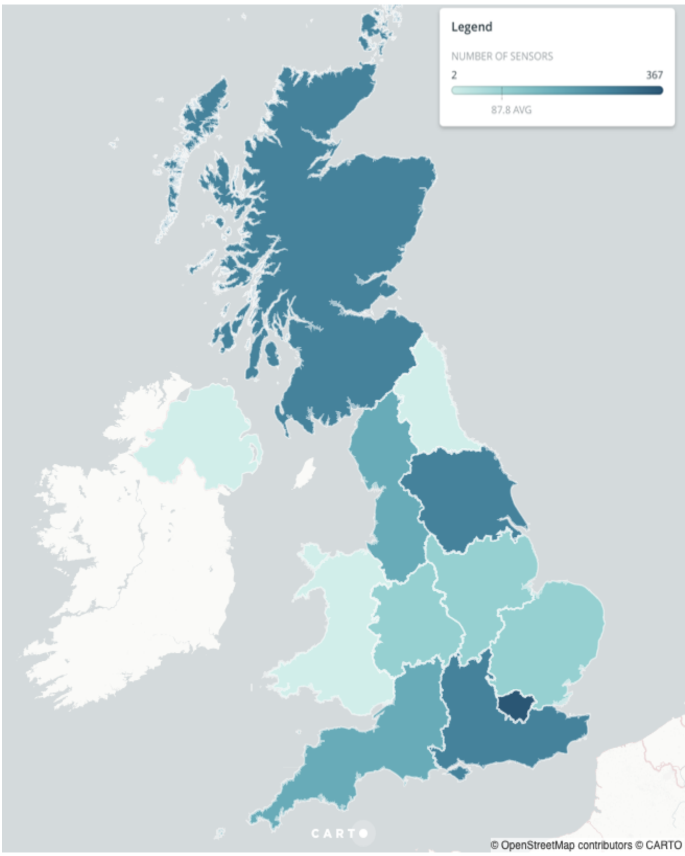


Figure 3-4 Distribution of sensors across regions of the UK, coloured by quantile

Although the distribution of sensors appears skewed, Figure 3-5 shows that the number of sensors in Wales, Scotland and Northern Ireland is roughly proportional to the population, as is the case for regions in England. The exception is Greater London, where there are far more sensors per person than any other region. In general, the dataset has good coverage of all regions in the UK, with a bias towards Greater London.

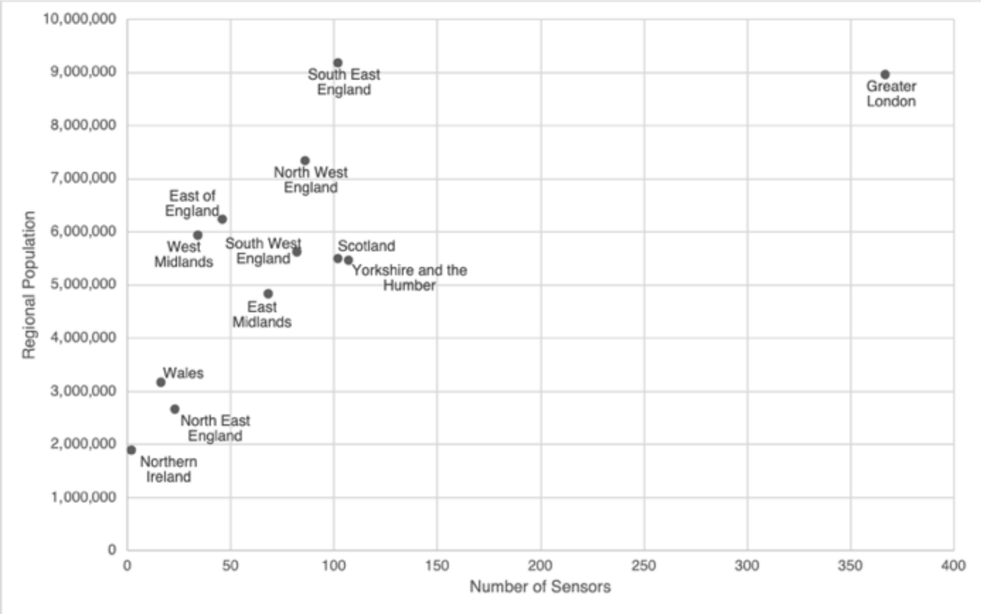


Figure 3-5 Number of sensors in each region by population[12]

However, intra-regionally the sensor distribution is more sporadic. Table 3-6 gives the frequency of sensors in each city and town, and although large cities such as London and Manchester are represented with more sensors, as a while this distribution is not proportional to population. Smaller cities such as Wakefield, Gloucester and Kingston upon Thames[13] also have a significant number of sensors, whereas cities such as Birmingham, Cardiff and Newcastle are comparatively underrepresented. This is due to the distribution of sensors being largely dependent on uptake from clients. For example, The Ridings Shopping Centre in Wakefield installed 29 sensors, which resulted in the city having more sensors than any other in the Yorkshire and the Humber region. 95 towns and cities had fewer than 15 sensors, and 49 of which were only represented with 1 or 2 sensors. 108 towns and cities are represented in the dataset in total.

[12] Population from (ONS, 2020e)

[13] Included in the Greater London region

Table 3-6 Number of sensors in each city or town

City or Town	Region	Sensor Count
London	Greater London	302
Edinburgh	Scotland	45
Manchester	North West	31
Wakefield	Yorkshire and the Humber	30
Nottingham	East Midlands	28
Glasgow	Scotland	25
Gloucester	South West	25
Leeds	Yorkshire and the Humber	23
Kingston Upon Thames	Greater London	23
Brighton	South East	22
Aberdeen	Scotland	20
Reading	South East	19
Liverpool	North West	19
Bristol	South West	18
13 towns and cities		11-15
15 towns and cities		6-10
18 towns and cities		3-5
49 towns and cities		1-2

Retail catchment & typology distribution

In the dataset, sensors are labelled according to their town or city; however, towns or cities can encompass many different retail centres, such as retail parks, suburbs and inner-city retail, each with a unique retail offer. Therefore, to gather a greater understanding of how the sensors are distributed across different retail types, the CDRC Retail Boundaries and Typology will be used. 942 of the sensor locations were able to be attributed to a retail centre (92%). The 108 towns and cities that were labelled in the footfall dataset accounted for 179 retail centres according to the typology.

The biggest contributor to that increase was London. The 302 sensors are distributed across 49 different retail centres. Although the majority are found in 'Central

London' (78) and 'Edgware Road Paddington London' (29), sensors are distributed across the boroughs. Figure 3-6shows that there are sensors present in 22 of the 32 London boroughs, which gives a good representation of the different range of retail centres within the capital. Other cities which were split into multiple retail centres according to the typology included Glasgow, Aberdeen, Bristol and Edinburgh.

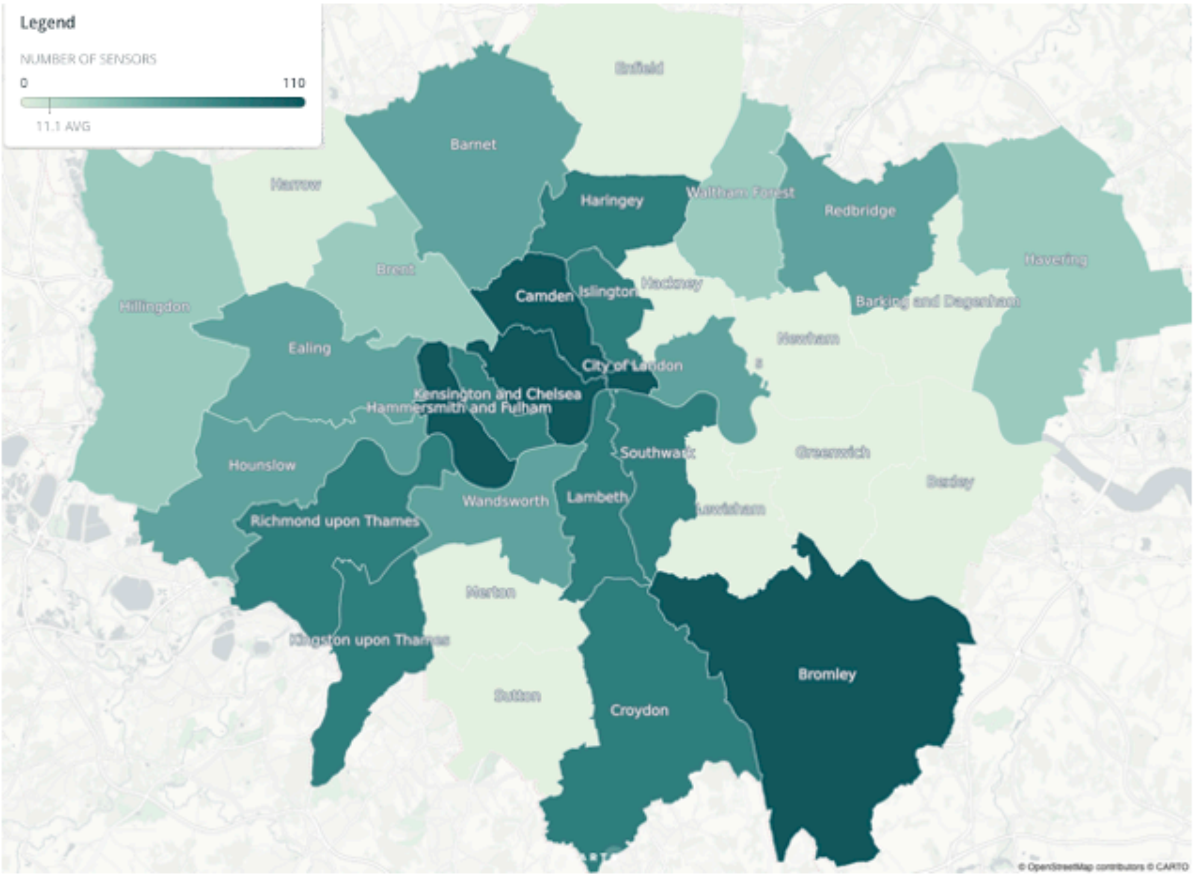


Figure 3-6 Distribution of sensors across London boroughs

Table 3-7 shows that the sensors are predominantly located in Supergroup 3, leading comparison and leisure destinations. Within that supergroup, most (85%) are in group 3.1, which are premium shopping and leisure destinations. Supergroup 3 represents 14% of retail centres in Great Britain, therefore 'Leading comparison and leisure destinations' (for example, city centres such as Central London) are significantly over-represented within the dataset, both in terms of individual retail centre, and by number of sensors. 'Local retail and service centres' and 'Traditional high streets and market towns' are under-represented.

This could be due to client-bias, as the retailers who invest in the project may be more inclined to place footfall sensors in stores in 'Leading comparison and leisure destination' stores. Although only eight sensors were identified to be place in Supergroup 2 'Retail, shopping and leisure parks'; however, this does not include shopping centres that are within a wider retail centre. For example, sensors in The Ridings Shopping Centre in Wakefield are inside the 'Wakefield' retail centre, which is group 3.1 'Premium shopping and leisure destinations'.

Table 3-7 Frequency of each supergroup of the CDRC Retail Centre Typology

Supergroup	Sensor Frequency	Frequency of retail centres with a sensor present	Distribution of supergroup in entire dataset
3- Leading comparison and leisure destinations	775 (82%)	113 (63%)	446 (14%)
4- Primary food and secondary comparison destinations	86 (9%)	30 (17%)	559 (18%)
5- Traditional high streets and market towns	22 (2%)	13 (7%)	642 (21%)
1- Local retail and service centres	18 (2%)	16 (9%)	1189 (38%)
2- Retail, shopping and leisure parks	8 (1%)	6 (3%)	274 (9%)

Distribution across store type

The LDC Retail Unit Address Data introduced in Section 3.1.1 was joined to the sensor location data to determine the type of store in which the sensor was located. This was done by matching the sensor's address to the store's address in the most applicable version of the Retail Unit Address Data (2017, 2018 or 2019). 788 (76%) of the sensors could be linked to a viable store. Due to errors and incompleteness within the data, 231 sensors did not have matching addresses, and 18 matched with vacant properties and were removed.

Out of the 788 locations where perfect joins were possible, the most popular categories were Fashion & General Clothing with 108 sensors (13.7%) and Charity & Secondhand Shops with 103 sensors (13%). Over half (452) of the sensors were in comparison retail stores, whereas 216 (27%) were in restaurants, bars, cafes, and the hospitality sector. 62 (7.9%) were in convenience grocery stores.

The distribution of LDC's sensors is dependent on uptake from clients – if a restaurant chain invests in footfall monitoring for thirty of their outlets, the dataset may be skewed to footfall trends displayed in food courts or areas with a collection of restaurants. Although no names of clients will be shared, 460 of the sensors are distributed across just 24 retailers making this a significant consideration when conducting analysis with this dataset.

In conclusion, primarily as a result of the sensor location and duration being determined by clients, the dataset has inconsistent distribution over time, space and store type. There are several considerations and bias that could impact or limit analyses undertaken using this dataset.

3.2.4 Alternative providers of footfall data

There are a range of data providers in the UK who, like LDC, aim to provide retailers, Business Improvement Districts, local planners and researchers with accurate footfall data. Companies such as Blix and Proximity Futures use a Wi-Fi-based methodology reminiscent of LDC's; however, image-based methodologies are a popular alternative. A non-exhaustive summary of these companies can be found in Appendix 3.3.[14]

Image-based methodologies use video recordings to count pedestrians. Typically, cameras are installed overhead in areas of interest, such as shop or station entrances (Sidla et al., 2006). The recordings are analysed using an AI algorithm trained to distinguish the shapes of people in images. The output is a count of pedestrians captured passing by in the recording. More complex algorithms can detect the trajectory of pedestrians or their demographic characteristics (Antipov et al., 2017; Sidla et al., 2006; Sun et al., 2018).

One company that uses a primarily image-based methodology to record footfall data is Springboard. Springboard has been tracking footfall across the UK since 2006, and their current footfall system can provide hourly and real-time footfall counts, visitor movement and dwell time and customer demographics such as age or gender (Springboard, n.d.).

LDC collect their data through a Wi-Fi-based methodology, which has been shown to result in some data quality and completeness issues, future discussed in Section 3.3. Springboard's image-based methodology is prone to some of the same logistical errors which LDC's sensors suffer from such as power outages, in addition to obstacles obscuring the camera's field of vision, that LDC's sensors do not experience. Without comparison of Springboard's footfall data to manual counts, it is

[14] The exact methodology that a company applies is often unclear from the information that is publicly available. Many companies stated using footfall 'cameras' without specifying if they applied the image-based methodology discussed here or instead referred to time-of-flight cameras – an alternative method for pedestrian counting which utilizes LIDAR.

difficult to know if their sensors might also undercount footfall, and if they are more or less accurate than LDC's.

Like LDC, Springboard's data is used in retail reports, news articles and academic research (e.g. BBC News, 2021; Komninos, Dunlop and Wilson, 2021; Wright, 2021). In particular, their footfall counts have been used by researchers at the Institute of Place Management [IPM] at Manchester Metropolitan University and Cardiff University to create activity-based classifications of footfall, as described in Chapter 2 (reference: Mumford *et al.*, 2021).

It is challenging to find a complete overview of the spatial and temporal distribution of Springboard's database, as neither the company itself nor data provider Urban Big Data Centre publicly provides this information. However, the supplementary material for the IPM footfall classifications (Mumford et al., 2020) gives some insight into the size and quality of the dataset.

Operating since 2006, Springboard has been collecting footfall data for significantly longer than LDC, who began in 2015. Mumford *et al.* (2020) do note that, like LDC, more sensors become installed over time, therefore not every location has data dating back to the beginning of the tenure, but from this information alone it seems likely that Springboard's data has better temporal coverage.

Significant weekly or monthly gaps are also mentioned in Mumford *et al.* (2020); however, there is also enough collected data in Springboard's dataset to be able to calculate monthly averages for locations with at least two years of data, which is not possible for many locations in the current LDC SmartStreetSensor dataset as the tenure of the project is shorter. This is especially valuable information as it could allow for annual and seasonal trends could be drawn from the data. Mumford *et al.* (2020) quoted the completeness of the Springboard dataset at 97%, which is higher than the LDC SmartStreetSensor dataset's completeness of 88.4%.

Springboard data represents at least 155 cities and towns with 600 sensors (Mumford et al., 2020). In comparison, LDC's has data representing 108 towns and cities, using over 1,000 sensors. This means that some cities and towns in the dataset, such as

London, Edinburgh and Manchester are represented to a higher spatial resolution, although fewer unique locations are covered by the LDC SmartStreetSensor dataset.

Both companies publish reports and quote footfall figures to media and news outlets frequently; therefore, future research which could directly compare the two datasets would be useful and relevant. To give a general conclusion based on the limited available information, LDC's data is more suited to for studying micro-locational or street-level variations in footfall within a city. Springboard's data is better at capturing long-term trends and investigating seasonal fluctuations; however, may be limited on the number of sensors per location.

3.3 Sources of error and issues of data quality

The previous section introduced the dataset, explored its qualities and distribution and compared it to other similar data sources. This following section - Section 3.3.1 - will explore the different sources of error within the SmartStreetSensor dataset. Unfortunately, the dataset is prone to errors from many sources including technical and logistical issues in addition to assumptions made about the people being measured. It will also explain measures that LDC have taken to mitigate the impact of these sources of error and evaluate the consumer privacy concerns that arise from Wi-Fi based methods such as the ones employed by LDC.

In Section 3.3.2, the measured counts will be compared and validated against manual counts conducted by LDC's in-person researchers. This will quantify the impact of these sources of error on the measured data. Then in Section 3.3.3, data exploration highlights several key examples of erroneous results.

3.3.1 Sources of error and issues of data quality

The primary sources of uncertainty and error in the SmartStreetSensor dataset are,

- Overcounting of devices
- Consumer privacy and MAC address randomisation
- MAC address collisions
- Undercounting of devices due to mobile ownership
- Probe request frequency and signal propagation
- Human and electrical error.

Many of these are related to the Wi-Fi based methodology implemented in data collection, which was described in Section 3.2.1. As a summary, the sensors use the Wi-Fi probes from nearby smart phone devices as proxy for footfall, counting how many unique devices pass by within 5-minute intervals. This can leave the data prone to error from technological and logistical barriers, in addition to assumptions made about the pedestrians themselves. The following section defines and explores these sources of error, the attempts by LDC to mitigate their effects and the issue of consumer privacy related to Wi-Fi-based methods.

Overcounting of Devices

The sensors are prone to overcounting in several cases. For example, when one consumer has multiple devices, such as a smartphone and a laptop, or there are Wi-Fi enabled devices in the vicinity that are not indicative of a consumer, such as routers and printers. In particular, this impacts sensors close to mobile phone or electrical shops, offices and residential buildings. These may have many Wi-Fi-capable devices that are not carried by passing pedestrians, therefore not representative of footfall. Sensors could also overcount as they cannot automatically distinguish between the devices of staff within a store or nearby residents and the pedestrians passing by.

When a device sends out a probe request, it includes a unique device fingerprint called a Media Access Control address [MAC address]. A MAC address includes an

Organisational Unique Identifier, or OUI, which is unique to the manufacturer (Apple, Microsoft, Samsung), and a device-specific identifier, which can be used to track a single device. It is possible to filter these probe requests and distinguish the devices of interest through the OUI in the MAC address. For example, a device with EPSON's OUI will likely be a printer or scanner. However, many popular manufacturers offer a wide range of Wi-Fi-enabled products such as smartphones, tablets, laptops, computers, and smartwatches; therefore, it is impossible to discern between them. Using the OUI in the MAC address is not viable to identify which probe requests belong to devices of interest and which do not (Redondi et al., 2016; Soundararaj et al., 2020).

Therefore, for the SmartStreetSensor footfall data, Local Data Company uses a different technique to filter out devices such as printers or scanners and the devices of residents or workers in buildings nearby. Probe requests are aggregated to five-minute intervals for each sensor, and within the five-minute interval, repeating MAC addresses are removed. The result is a count of unique devices measured by the sensor every five minutes (CDRC, 2019).
Then, MAC addresses that appear during consecutive intervals within a half-hour period are removed. Therefore, printers or mobile devices owned by residents or store employees are only counted once every half hour instead of every five minutes, reducing the risk of overcounting from 'long dwellers'.

However, this method has some limitations.

Firstly, by removing any consecutive instances of a unique MAC address within a half-hour, valid counts could potentially be removed. For example, if a person walked one way past a sensor at 12:15 and back at 12:19, they would only be counted once, leading to a potential undercount.

Secondly, devices such as printers or staff-owned smartphones will still be counted once every half hour. Depending on the cleaning process, this could all occur within the same five-minute interval resulting in boosts at, for example, ten minutes and forty minutes past the hour, resulting in overcounting and clear discrepancies in five-minute counts. In future analyses in the thesis, this issue was handled by removing the inflated count and calculating the average of the 5-minutes before and after to interpolate the resulting gap.

Finally, the introduction of MAC address randomisation has made this method inapplicable to an increasing number of devices in recent years. This will be explored in the following paragraphs.

Consumer privacy and MAC address randomisation

Wi-Fi probe requests make devices into wireless beacons that periodically and unprompted will advertise the location of that unique device. Whereas this makes them appealing for non-invasive automated continuous pedestrian counting, it presents a huge privacy concern. Unique MAC address identifiers can allow the tracking of an individual device across locations, and Wi-Fi-based methods introduce a serious risk of infringing on consumer privacy through the collection of MAC addresses. Individual MAC addresses could leave a consumer vulnerable to some very serious, malicious actions, including stalking, targeted attacks or collecting sensitive information (Cunche, 2014).

To LDC and many companies which utilise Wi-Fi-based methods, the privacy and security of consumers is a priority. Since 2018, MAC addresses have been considered personal data under the UK and EU General Data Protection Regulation or GDPR; therefore, their collection and application are closely monitored and protected (GOV.UK, 2016). Therefore MAC addresses are hashed and anonymised by LDC to protect consumer privacy (CDRC, 2019; Demir et al., 2014), although the extent to which this protects the consumer from de-anonymisation attacks is contested (Demir et al., 2014).

The GDPR should protect consumers from mobile tracking misuse from UK and EU controllers and companies; however, there is still trepidation in the public sphere. In 2013 leaked classified documents made the wider public aware of surveillance methods the US National Security Agency and the Five Eyes Intelligence alliance used that were generally considered invasive and a breach of personal privacy. Since consumers have become increasingly aware of the risks posed by mobile device tracking, particularly from sources that might be exempt from this regulation, such as foreign intelligence agencies from countries such as China or Russia, and private companies (Kerr, 2013; Martin et al., 2017; Tucker, 2019). Demonstrated in 2020 and 2021 by the trepidation to download the NHS Test and Trace App during the COVID-19 pandemic (Kleinman, 2020), consumers are becoming more protective of their data and who has access to it.

Today's consumer demands reassurance and a level of transparency and security from device providers and app developers. As a result, device manufacturers and mobile operating systems have progressively adopted MAC address randomisation strategies that periodically scramble the MAC address so that the device cannot be tracked. Android and iOS—the two largest smartphone operating systems globally—have implemented MAC address randomisation in their software that ensures users privacy until they willingly connect their device to a Wi-Fi Access Point (Martin et al., 2017). Although uptake has been slower for Android devices than iOS, with the introduction of stricter regulation around MAC addresses and with consumer privacy concern increasing, MAC address randomisation is likely here to stay.

The implementation and adoption of MAC address randomisation poses a significant challenge to Wi-Fi-based pedestrian counters, such as those employed in the SmartStreetSensor project. An essential step in their cleaning and validation procedure involves filtering out MAC addresses that appear in consecutive five-minute intervals.

The frequency at which scrambling occurs, and the severity of this issue, depends on the device and manufacturer. As an example, iOS 10.1.1 devices generate a new randomised MAC address every time the device is locked or unlocked, Wi-Fi is activated or deactivated, or the device attempts a connection to a Wi-Fi Access Point (Oliveira et al., 2019). When the MAC address is being randomised, it is difficult to identify when the device may appear consecutively. There are methods to de-anonymise these MAC addresses, such as timing attacks, scrambler attacks and linking through sequence numbers (Cunche and Matte, 2016; Freudiger, 2015; Martin et al., 2017; Matte et al., 2016). However, these methods risk exposing the consumer's privacy and are unreliable in the long term as operators seek to strengthen security and remove these potential vulnerabilities.

In the current SmartStreetSensor dataset available via the CDRC, LDC mitigates this issue by adjusting the randomised probe counts using the non-randomised probe counts. As repeating devices and long-dwellers can still be identified and removed for non-randomised MAC addresses, the ratio of the before count and the filtered count is used as an adjustment ratio for the randomised probe count (CDRC, 2019). Although the widespread adoption of MAC address randomisation that began in 2017 has reduced the effectiveness of the cleaning and validation procedure LDC employ, their solution protects the consumers' privacy and identity while producing continuous footfall estimates. The uncertainties introduced are unavoidable and well-justified.

Unfortunately, this solution may only be temporary, as more operators adopt MAC address randomisation and the proportion of non-randomised counts to randomised decreases. Wi-Fi-based methods have many benefits and positive applications; however, they are limited by a trade-off between data quality and consumer privacy, of which the latter is becoming increasingly relevant ethically and legally. Alternative technologies such as radar, infrared and image analysis could provide the same continuous pedestrian counts without the risk to consumer privacy.

MAC address collisions

Another source of error in the SmartStreetSensor dataset is a results of MAC address collisions. MAC address collisions occur when two separate devices appear to share the same MAC address. MAC addresses are intended to be unique, so this should not happen; however, MAC address collisions have been observed by LDC in the data. Unique hashed MAC addresses have appeared at two separate locations within a short amount of time, less than the time taken for the user to travel from one location to another feasibly. These MAC address collisions are thought to result from limitations in the hashing algorithm used to anonymise the data and aggressive MAC address randomisation (CDRC, 2019; Soundararaj et al., 2020).

MAC address collisions do not present an obstacle for data cleaning when the matching identifiers are across two different locations, as each location is parsed and cleaned separately. However, it indicates that some of the duplicate counts registered by a single sensor may not originate from the same device but two separate devices, which would be impossible to distinguish. This error could lead to undercounting in the data, as these false, duplicate counts are removed. However, this is not a prevalent issue (<0.01% of instances), and no adjustment has been made to the data to account for it (CDRC, 2019).

Undercounting of devices due to mobile ownership

A sensor is liable to undercount footfall when the consumer does not own or carry a Wi-Fi-enabled device. A survey by data platform Statista found that, in 2021, 92% of people aged 16 and over in the UK own a smartphone (O'Dea, 2021); therefore, there is a potential 8% of consumers that may not be counted using probe requests. Smartphone ownership decreases with age, and the majority of these potential missed consumers would fall in the 65+yr range (Boyle, 2019), which may disproportionately impact locations that serve an older population and under-represent the 'grey market'.

On the other end of the scale, although surveys have shown children 8-16yrs have a similar smartphone ownership level to adults (O'Dea, 2020), children younger than 8yrs are unlikely to carry their own device. This may be a minor issue when counting footfall from an economic standpoint, as young children are unlikely to have significant purchasing power. However, considering footfall as an indicator of a retail centre's vitality and social and cultural strength, all ages are equally relevant. The relationship between smartphone ownership and age could be combined with demographic data such as the census to estimate the pedestrian undercount; however, LDC does not adjust for this in their SmartStreetSensor dataset. With overcounting from pedestrians carrying multiple mobile devices also not adjusted for, this could mitigate the undercounting error to an unknown extent (discussed further in the next section).

Probe request frequency and signal propagation

Another source of undercounting can come from probe request frequency. Freudiger (2015) demonstrated that the average probe request frequency was 55 times an hour, just less than once a minute. Therefore, there is an element of chance as to whether a mobile device will send out a probe request while the consumer is in the proximity of a sensor. Probe request frequency also varies significantly by the device manufacturer and the number of known networks (Freudiger, 2015). Android devices tend to send more probe requests than iOS and Blackberry, and continue to send probe requests even when the Wi-Fi functionality is switched off (Freudiger, 2015; Soundararaj et al., 2020). As a result, Android users would be the least likely to be undercounted by Wi-Fi sensors, followed by iOS whose users may be undercounted due to lower probe request frequency and disabling of the Wi-Fi functionality. Blackberry devices do not broadcast probe requests, and they will consistently be undercounted by sensors.

Adjusting for undercounting can be challenging as it is easier to determine how much extra data is collected than how much is missed. To estimate the significance of undercounting due to probe request frequency, the amount of time a consumer spends in the range of a sensor would have to be known. However, due to variations in signal strength and propagation, there is not a uniform or standardised sensor range. From device to sensor, multiple variables determine the strength of the signal sent and how far it might travel, such as battery level of the device, device manufacturer, if the screen is on or off, humidity, temperature, and environmental obstacles such as doors, walls, or other people (Freudiger, 2015; Jiang et al., 2015; Soundararaj et al., 2020; Su and Jin, 2007). The orientation and positioning of a sensor within a store window can also determine the count of probe requests. As demonstrated through the examples in Figure 3-7, the surrounding environment of different sensors is unique; therefore, there can be no standard range that determines the devices measured. Some sensors are installed closer to the street within a store than others. Obstacles on the pavement such as bus stops, seating or other pedestrians could also act as barriers to a probe request. Even for an individual sensor, the range might vary during different times of the day with higher signal attenuation when it is busier and during different weather conditions. It is difficult to adjust for undercounting from probe request frequency because sensor range would be time and resource-intensive to calculate.

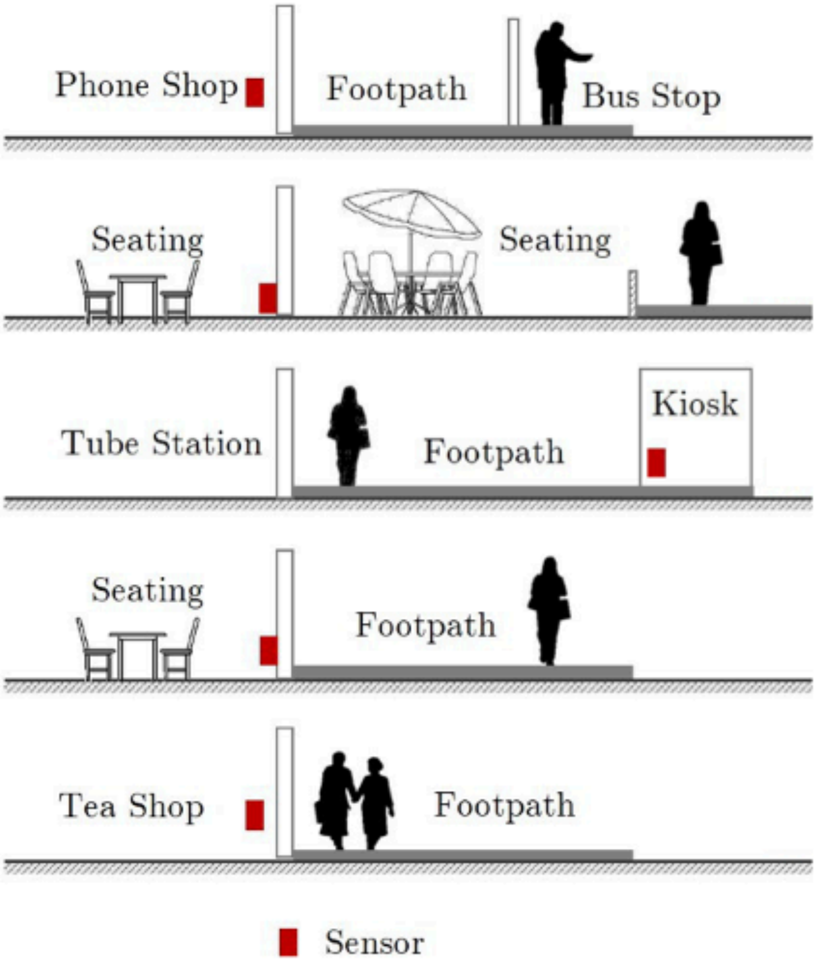


Figure 3-7 Variation of settings and environments surrounding a sensor

(Adapted from Figure 6(a) in Soundararaj, B., Cheshire, J. and Longley, P. (2020) 'Estimating real-time high-street footfall from Wi-Fi probe requests', *International Journal of Geographical Information Science*, 34(2), pp. 325–343.)

Logistical and human error

Finally, because the sensors are installed in real-world stores and locations, they are subject to disruption and malfunction from various sources. These include, but are not limited to,

- Tampering, damaging, or unplugging the sensor
- Power cuts
- Software updates
- Sensor hardware or software malfunction
- Loss of internet connection
- Store closure or power loss

These errors can cause significant gaps in the dataset, of varying lengths. For gaps that are shorter than thirty minutes, LDC imputes the missing data. To impute, or fill in, these gaps, they apply Kalman smoothing (CDRC, 2019). This involves using the observed data points to predict the most likely value to fill the gap. Unlike using the last observed value, or a mean value, it considers the underlying trend and error in the known data points to predict the unknown ones (Bishop and Welch, 2001). LDC uses the R package 'imputeTS' (Moritz and Bartz-Beielstein, 2017), which includes a Kalman smoothing algorithm specifically for time series data. This helps to standardise their imputations and makes the process more efficient. Figure 3-8 provides an example of a sensor in Coventry on 3rd November 2016. Logistical or human error has resulted in six gaps in the data, five to ten minutes long. The dotted line shows the values that LDC has imputed.

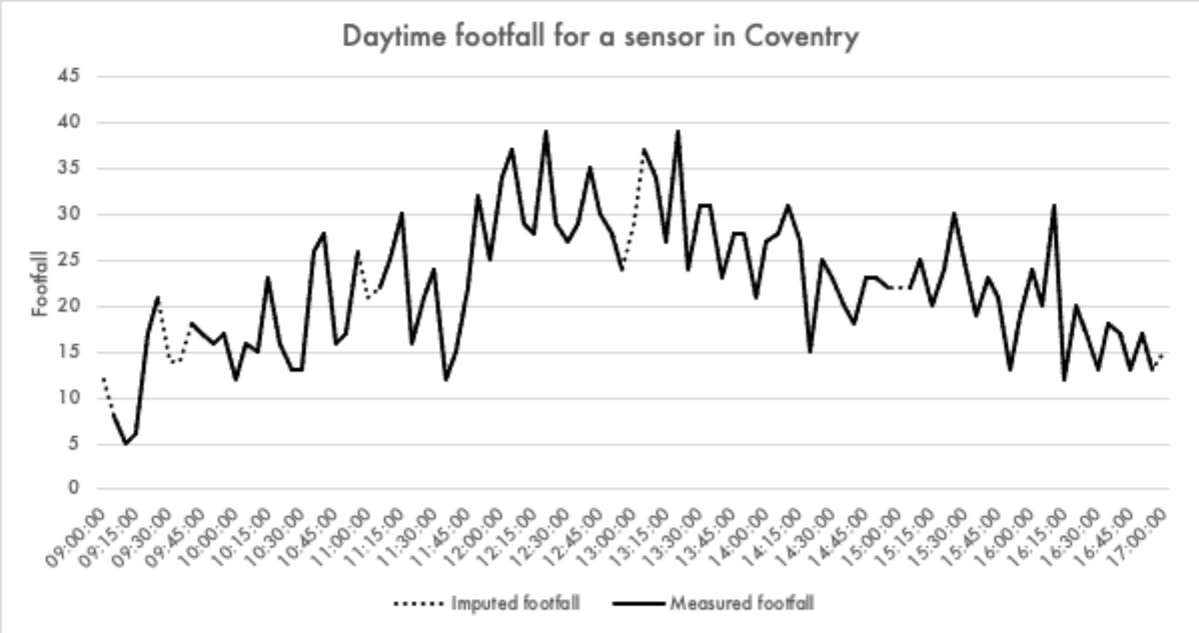


Figure 3-8 Example of missing data for a sensor being imputed by LDC

6% of values are imputed in the dataset, which is reasonably low; however, this can vary depending on location. Some locations have no imputed data, while others have up to 25%. Gaps in the data that are longer than thirty minutes are not imputed by LDC and left in the data. The extent of this issue was explored previously, in Section 3.2.3.

3.3.2 Data Validation through manual counts

In order to ground-truth the data and quantify any error, LDC conduct manual counts at sensor locations. These involve counting every pedestrian that passes by within a timeframe. This is usually ten minutes, however, there are some fifteen- and twenty-minute counts. The manual count can then be compared to the sensor count to determine whether the sensor is over-counting or under-counting and by how much. Although LDC records manual counts, these are not adjusted for in this dataset[15].

[15] In the analyses in Chapter 4 & 5, it was wrongly assumed that LDC had adjusted for these manual counts within their dataset, therefore the analyses were completed on the unadjusted data. The

Comparison of manual counts to sensor counts

The manual counts data from LDC contains over 4,000 observations for 862 locations from 17th May 2016 until 3rd October 2019. They took place at different times of day (from 5 am – 11 pm) and days of the week (Monday – Friday, with limited weekend measurements). These manual counts were compared to the sensor counts data for that measurement period, and a ratio was calculated by dividing the sensor counts by the manual counts, as in the equation,

$$Adjustment\ ratio = \frac{Manual\ count}{Sensor\ count}$$

Therefore, if the ratio is less than 1, this means that the footfall sensor is undercounting compared to the manual counts, and if the ratio is more than 1, it indicates that the sensor is overcounting. Counts that were 0 were removed to avoid undefined or biased ratios. Imputed sensor counts were found to make little difference to the overall ratios. Therefore, they were kept in.

Figure 3-9 shows the distribution of the ratio between the sensor counts and manual counts. Out of the 4,000 observations, only 337 of the sensor counts were within 10% of the manual counts. For the majority of observations (72%), the sensors significantly undercounted footfall. 1,803 of the observations taken, the sensor measured footfall at less than half of the count the surveyors reported. This occurred over a range of dates and footfall magnitudes, from quieter areas where the manual count was less than 10, to busier areas where it was over 1,000. There were also instances of significant over-counting from the sensors, however, these were less prevalent. 20% of the observations were deemed as overcounting (>10% of the reported manual count). The median ratio for all observations was 0.56, and the average absolute error was 66.7%.

---

manual counts themselves were included in a later update received in 2019 and the analysis in Chapter 6 does adjust the footfall data to in order to externally validate the data.

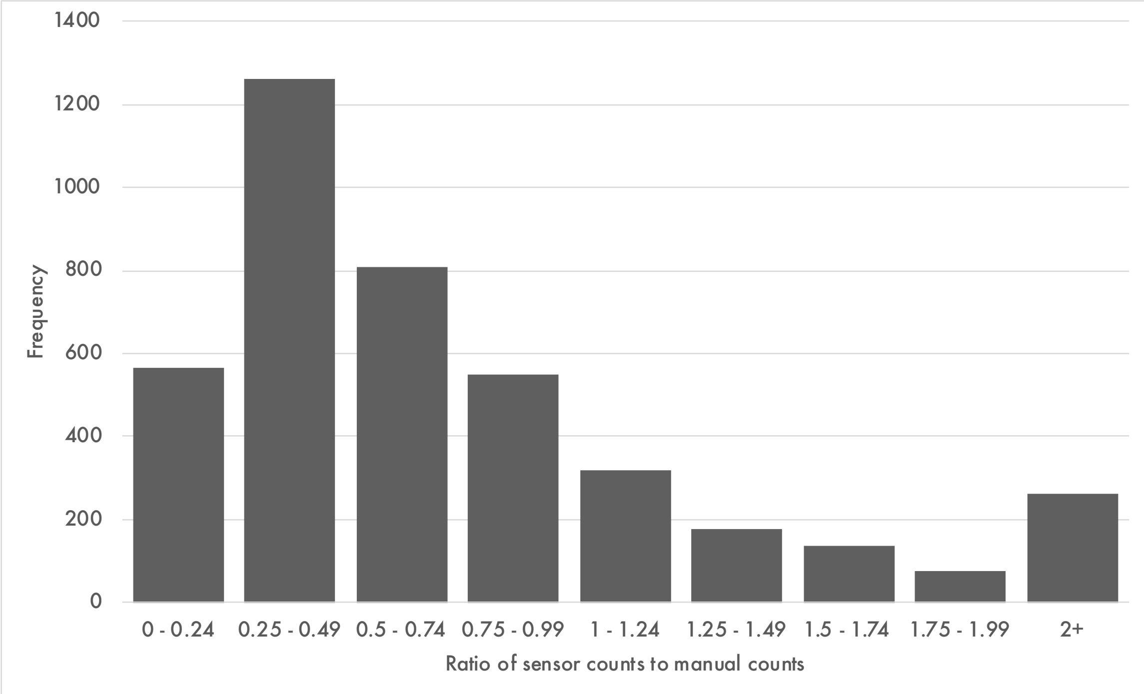


Figure 3-9 Distribution of the ratio of sensor counts to manual counts

This shows that, despite LDC's pre-processing and adjustment, there is still a significant difference between the sensor counts and the manual counts completed in person. Although some of this difference may result from inconsistent counting or human error in the data collection for the manual counts, a substantial number of sensors report a count different from the ground truth, implying that the error is introduced by the data collection methodology for the sensors.

Error in the dataset

Potential sources for error have been discussed at length in the previous subsection, however, they are reiterated in Figure 3-10. Infrequent probe requests, MAC address collisions, device ownership, signal propagation and MAC address randomisation could all be a factor in the significant undercounting observed in the data. MAC address collisions are unlikely to have a significant impact, as they occur in an estimated 0.01% of instances (CDRC, 2019). Device ownership might cause the sensors to undercount by ~8%, but this could be mitigated by the overcounting from pedestrians with multiple devices. Out of those identified influences, infrequent probe requests, signal propagation and MAC address randomisation have an unknown impact.

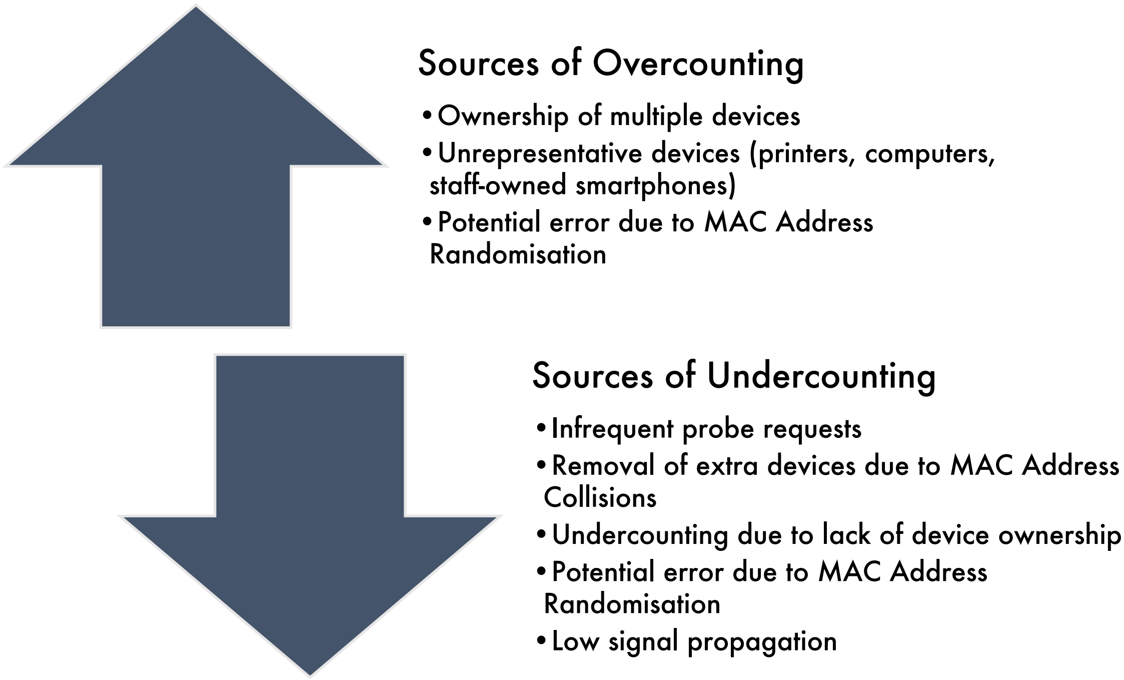


Figure 3-10 Sources of Overcounting and Undercounting in the dataset

As discussed in Section 3.3.1, MAC address randomisation was mitigated by applying the same ratio of pre-cleaned to post-cleaned counts from non-randomised devices onto randomised devices. However, if this process is too severe, it could cause undercounting. MAC address randomisation was introduced in late 2017 and adopted throughout the following year. If the undercounting resulted from an over-removal of devices in LDC's pre-processing, there might be a steady increase of undercounting sensors over time. However, Figure 3-11 shows the opposite. As time passes, the average ratio begins to move closer to one (accurate counting). However, four times as many observations took place in 2016 and 2017 (2,994) in comparison to 2018 and 2019 (743); therefore, they may not be directly comparable.

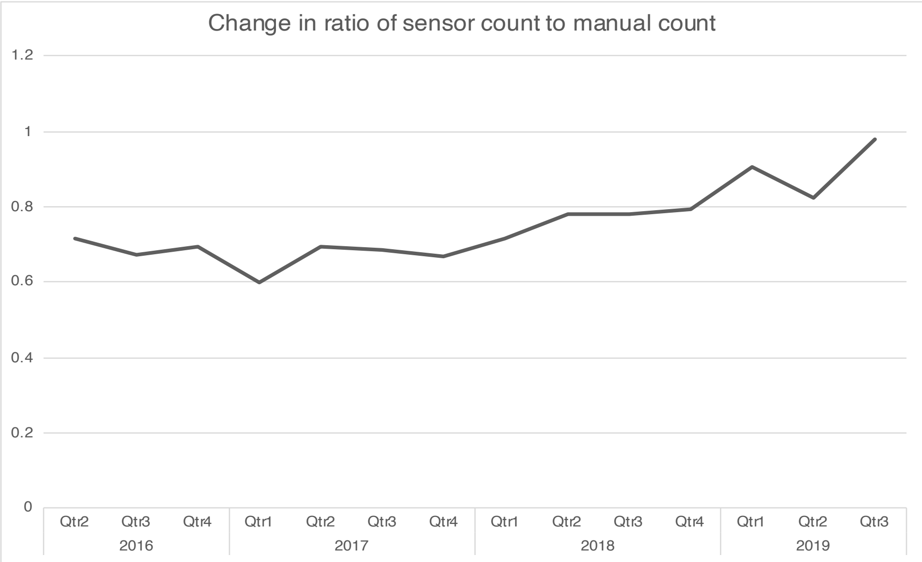


Figure 3-11 Change in the ratio of sensor counts to manual counts over time[16]

The significant undercounting of the sensors could also be the result of low probe request frequency. Work by Freudiger (2015) finds the frequency which a device emits a probe request is 55 times an hour while on standby. Typical walking speed ranges from 0.9 – 1.35 m/s (Schimpl et al., 2011). In the 1.09 minutes in between probe requests, an average pedestrian might travel between 59 and 88 metres. There is an element of chance as to whether this occurs within the range of a sensor.

Additionally, the range of each sensor is unknown and difficult to calculate as it is dependent on many factors about the surrounding environment and the mobile device in question. As this is the case, there is potential that the manual in-person counters might tend towards a larger range than the sensor measures. For example, if they were counting pedestrians on both sides of the pavement, yet the sensor only counted the side nearest, this could cause one count to be double the other. Inconsistent sensor range and infrequent probe requests are both feasible sources for this chronic undercounting.

[16] Outliers were removed (±5% of the data)

There is considerable error in the dataset, which could be the result of assumptions in the data collection methodology, errors in the manual count or another unknown source. Chapter 6 explores this in more depth including how error varies according to time and characteristics of the micro-location, and the dataset is adjusted as much as possible to account for this.

3.3.3 Data exploration and erroneous results

This section will explore the qualities of the SmartStreetSensor dataset and identify and evaluate erroneous results. The dataset has more than 65 million observations of footfall, to five-minute accuracy. The mean footfall per 5 minutes is 46 (standard deviation = 51.3), the minimum is 0 and the maximum is 4,377, which was recorded in Cambridge at 13:00 on 3rd July 2017. However, after some research it appears that it is likely this is an erroneous result. This will be discussed later.

Figure 3-12 gives the average daily signature for footfall across the entire dataset. The data is given in UTC, and it has been adjusted to GMT/BST, as the hours of day are relevant. It presents a pattern with three peaks, roughly at 9:00, 12:00 and 17:00. During the early hours of the morning, footfall is close to 0, and, on average, it peaks at 80 people per five minutes at midday. The distribution appears jagged, and some data exploration has found usually regular peaks at ten past the hour, which were assumed to be the result of LDC's cleaning procedure, as mentioned in Section 3.3.1.

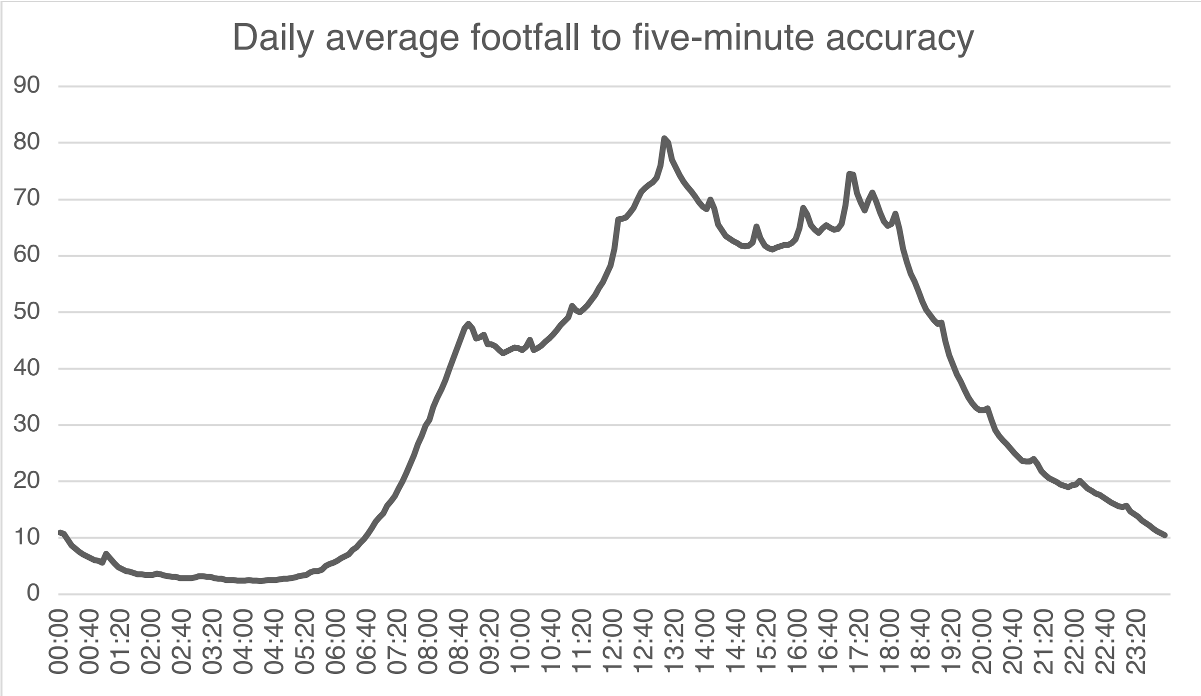


Figure 3-12 Daily average footfall to five-minute accuracy

Outlying patterns and results can also be observed at the individual sensor level. Figure 3-13 gives examples of four sensors which show unusual patterns of footfall. Sensor 165 (top left) shifts from a mean footfall of 200-400 people per hour to 1,000+ before returning to 200-400 people again in mid-2018. Sensor 260 (top right) appears to be one of the busiest locations at first look at the dataset; however, in context, the measurements appear to be outliers. Sensor 517 (bottom left) appears to be consistent through most of its tenure, however, has a month of faulty low results in August 2018, and Sensor 461 (bottom right) is the opposite, with a few months of faulty high results in 2019.

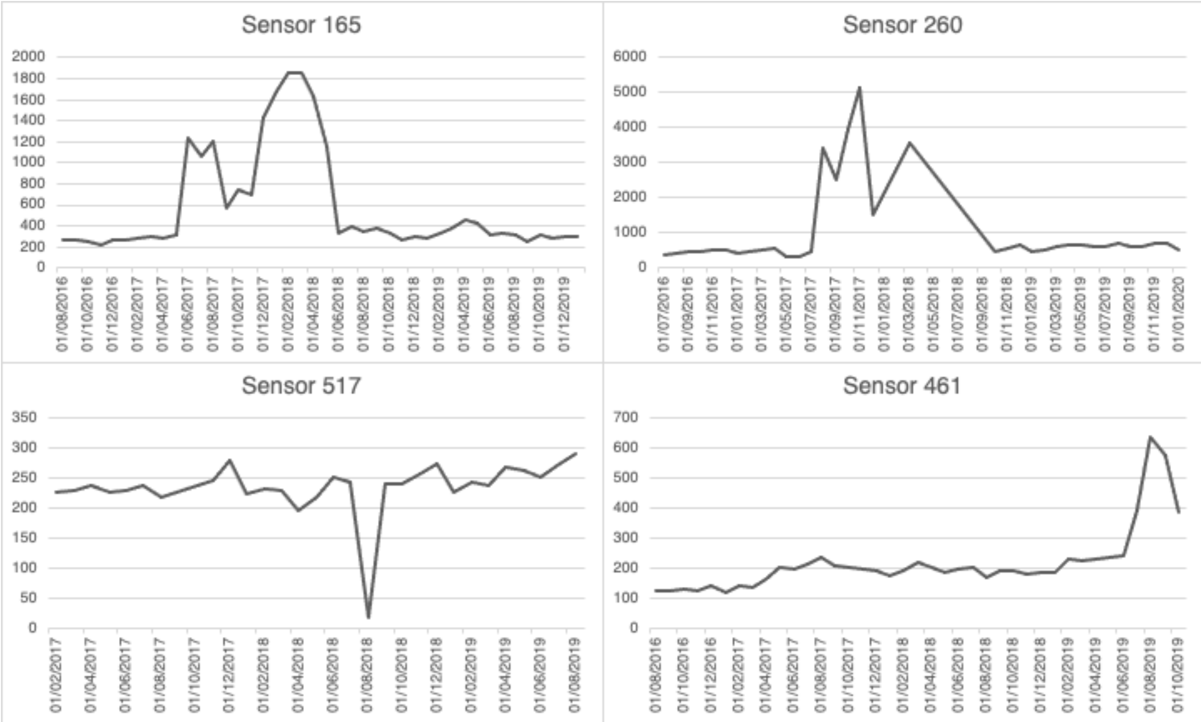


Figure 3-13 Examples of sensors with outlying results

If not correctly dealt with, erroneous results could skew any models or aggregate values calculated. When viewing erroneous results, not all values are equal. Sensor location, month, weekday and hour of day can all determine whether a value is erroneous or not. For example, a count of 0 at 12:30 pm on a Saturday on a busy high street sensor would likely be an outlier, but the same value at the same location but at 3:00 am would not be. Ideally, every sensor would be verified through individual visual inspection, but with over 1,000 different sensors, this would an incredibly time-consuming job to complete. In addition, local knowledge of every sensor would be needed, as events like road closures, festivals or shopping events could be legitimate causes for results that may appear to be erroneous. It is also subjective as what appears an uncharacteristic footfall distribution to one researcher might appear as a normal fluctuation to another. Automated outlier removal, such as removal by z-score or by applying DBSCAN could be alternatives; however, these should be applied appropriately, accounting for normal temporal and spatial fluctuations.

This section has identified some of the sources of error within the dataset. Comparisons with manual counts have shown that they have a significant impact on the quality of the data, with the majority of sensors found to undercount footfall, and the average absolute error measuring 66%. In addition, there are many inconsistencies and measurements that appear to be erroneous within the data, and these must be handled and processed before any meaningful analysis can take place.

The approach and methodology that will be employed to achieve this differs for each of the analyses within this thesis, as they each use separate subsets of the data and approach a different research aim. These approaches are summarised in Section 3.4.

3.4 Study Design

This chapter has explored the different datasets which will form the foundations of the analyses within this thesis. The key dataset for these analyses will be the SmartStreetSensor project footfall data, provided by LDC and the CDRC, however, supplementary datasets are also used to give insight into the environment and context the measurements have been taken in. These include population data, transport data, retail data and street network data.
It has explained how the data is collected, what it measures and the limitations and assumptions that are made regarding its application.

This following section will give an overview of the analytical chapters which will follow: Chapter 4, Chapter 5 and Chapter 6. Each analytical chapter will investigate a specific research aim established in Chapter 1 and reiterated here.

Chapter 4 is called 'The world around us - Quantifying temporal variations in footfall in relation to micro-locational characteristics' and explores the relationship between footfall and the immediate context surrounding it through the following research aims:

The world around us	Investigate how different characteristics and contexts of the immediate environment impact footfall magnitude and signature
	Using characteristics of retail and footfall context, develop a classification that captures these main differences
	Identify how trends in footfall magnitude and signature differ between these different retail contexts

Chapter 5 is entitled 'What happens there - Exploring event-related temporary fluctuations in footfall magnitude and their relation to micro- and meso-scale characteristics' and this analysis will specifically focus on temporary events and fluctuations in footfall, including shopping events such as Black Friday or Christmas, weather events such as heatwaves or storms and local events such as festivals or sports events. Through comparison of four key case study locations, it will explore the following aims:

What happens there	Identify events which significantly impact footfall.
	Investigate how factors of both the immediate environment and in the wider context could influence this impact
	Explore the trends and similarites between footfall of different events in different locations and what they could imply about retail footfall

The final analysis chapter, Chapter 6, is called 'What remains unknown - Investigating the potential for a spatio-temporal prediction model for footfall data' as it will explore footfall modelling and prediction, and the capability of this data to be used to estimate footfall numbers in places where data is not collected. It will achieve this through the following aims:

What remains unknown	Define the criteria and use case for a footfall prediction model and identify appropriate methodologies to achieve it
	Create a preliminary model that predicts footfall that is location and time dependent
	Critique the performance of this model, identifying opportunities for improvement

They will employ a sufficient subset of the data and an appropriate methodology. Altogether they contribute to the wider research aim of the thesis – to explore the relationship between footfall and its context.

All three analytical chapters will employ quantitative methods and data analysis on the datasets introduced in this chapter in order to achieve their specific research aim.

3.4.1 Methodological approach

All three of the analysis chapters

Chapter 4 will use statistical methods and machine learning through cluster analysis to analyse footfall data and how it relates to the surrounding context and environment.

Chapter 5 will uses data exploration of case study locations, employing a bottom-up methodology to analyse how temporary events can impact footfall in different cities. The final analytical chapter, Chapter 6, uses machine learning through regression and classification to create a spatio-temporal prediction model of footfall. More detail on the justification and derivation of each approach will be found at the appropriate point within their respective chapters.

All analysis will be done in R and Python with QGIS, CARTO and Microsoft Excel used for data visualisation.

3.4.2 Dataset Subsets

As previously mentioned, the analyses will not use the full dataset of 1,037 footfall sensors. Samples of the footfall data were drawn and cleaned individually depending on the specific research aims and when the research was conducted.

Chapter 4 investigates the relationship between different micro-locational factors and footfall on UK high streets. It used a sample of 640 sensors out of the 840 that were available at the time of analysis. This sample excluded sensors that were not placed in high street locations (e.g. shopping centres, stations) and sensors that had less than nine months of footfall data. In addition, Scottish sensors had to be excluded from this sample as the analysis used census daytime population, which is not available for Scotland. The data used was from January 2017 until August 2018.

Chapter 5 explores event-related temporary fluctuations in footfall and how these vary due to micro- and meso-scale factors It employs a case study analysis methodology and, as such, uses a sample of 20 sensors from across the three case study cities: Liverpool, Manchester and Edinburgh. This analysis used two calendar years of data: 2017 and 2018.

Chapter 6 examines the potential of the dataset for spatiotemporal extrapolation – predicting footfall at a high temporal resolution for unmeasured locations. It employs the largest sample of the dataset at 668 locations. The locations excluded had less than 90 days of data, displayed abnormal patterns or fell outside the areas of interest.

3.5 Chapter Summary

This chapter introduced the datasets used in the subsequent analytical chapters, their data collection and limitations. Six supplementary datasets are used to bring context to the SmartStreetSensor footfall data, with a particular focus on LDC's Retail Unit Address Data and the CDRC's Retail Boundaries and Typology.

LDC's Retail Unit Address Data is an invaluable resource which details the retail, leisure, service and vacant units in over 1,000 towns and cities across Great Britain. The three-classification system allows for different levels of analysis, and the dataset paints a picture of the retail environment that directly surrounds a footfall sensor. The CDRC Retail Boundaries by Pavlis, Dolega and Singleton (2018) identify more retail centres across Britain than the previous DCLG boundaries, and the 2021 version identifies more again. The matching typology by Dolega *et al.* (2021) groups the retail centres based on elements of their composition, function, affluence and size to give a granular overview of retail centres in Britain. It can be used to draw similarities between patterns observed in different retail centres and infer possible trends in unmeasured retail centres. Other datasets include the UK Census, OpenStreetMap and the Department for Transport's NaPTAN and Car Parks datasets, which give information on the transport, demographic and morphological context of the footfall sensors, yet some temporal and spatial limitations.

The SmartStreetSensor footfall dataset forms the cornerstone of the analyses in this thesis and is a footfall dataset that is near-unique in its temporal resolution and its quantity and spread of sensors. The data is collected through Wi-Fi-based methods that endeavour to count the number of smartphone devices passing on a street as a proxy for footfall. However, it is a methodology prone to considerable sources of error including, logistical and electrical error, assumptions over smartphone ownership, and the increasing efforts of smartphone manufacturers to safeguard their customers from malicious use of their information. When compared to manual counts of footfall, the majority of sensors were found to undercount footfall, with the average absolute error measuring 66%.

The data is also limited temporally. Although LDC's footfall data has a high temporal resolution at five minutes, to complete meaningful timeseries analysis considering annual seasonal trends, at least two years of data is needed. The average sensor has data for 25 months, which could make this possible. However, on average 11.6% of data for each location is also missing before any outlier removal, which limits the applicability of forms of timeseries analysis on individual sensors.

The footfall dataset covers 108 towns and cities with over 1,000 sensors, however, has an inherent bias due to sensors being placed in locations favoured by clients. Leading comparison and leisure destinations have proportionally more sensors than other types of retail centres, and with almost half of the sensors distributed between 24 retailers, where the store is within the retail centre is also an important consideration.

In conclusion, the SmartStreetSensor footfall data is a powerful and vast resource that captures continuous pedestrian counts for an unprecedented number of sensors across the UK, which can offer new insight into pedestrian behaviour in retail areas. Although the short tenure and data quality issues can limit applications, a thorough and systematic process of data cleaning and validation is needed to produce valuable insight. Without the necessary pre-processing, analysis could be flawed, and erroneous conclusions could be drawn.

4 The world around us

Quantifying temporal variations in footfall in relation to micro-locational characteristics

The world around us	
	Investigate how different characteristics and contexts of the immediate environment impact footfall magnitude and signature
	Using characteristics of retail and footfall context, develop a classification that captures these main differences
	Identify how trends in footfall magnitude and signature differ between these different retail contexts

The world around us

This first analytical chapter, 'The world around us', investigates the relationship between footfall and the immediate environment around it. Footfall can be influenced by a range of factors at macro-, meso- and micro-scale contexts, as discussed in Section 2.3.1. It is well-established through the work of Reilly (1931), Huff (1963) and others that factors such as the proximity of retail centres, the attractiveness of nearby retail centres and population all impact the number of visitors to a retail centre and, therefore, footfall magnitude. The coronavirus pandemic and subsequent lockdowns have been a stark demonstration of the impact which macro-scale political factors can have on footfall. However, this chapter focuses primarily on the micro-scale context because datasets such as the SmartStreetSensor footfall dataset allow footfall to be explored in this context more closely than ever before.

This immediate environment is referred to as a micro-location or micro-scale and encompasses an area larger than the sensor range but smaller than the retail centre itself. Micro-locational characteristics could include the retail mix of a street, the proximity of a train station or bus stop, or an anchor store or entertainment venue nearby. In reports and academic literature, the impact of these characteristics is often treated as inherently 'known' through observation and experience of the authors themselves. Anyone who writes about retail centres has their own experience of high street environments and how pedestrians ebb and flow, when and where it might be busy, and why. An example would be the commuter rush. A micro-location characteristic, such as a bus stop or train station, may result in higher footfall, mainly after 5 pm on workdays. Another example might be an area with a high concentration of bars and clubs correlating with higher footfall on Friday and Saturday evenings. However, the internal validity of this evidence – the potential risk of bias – is high, as it is based solely on experts' opinions (CEMBa, n.d.). It cannot be quoted as fact by other researchers, and it cannot be robustly applied to any other context outside the one it is discussed.

With advancements in sensing technology and the availability of higher resolution data for cities, it is now possible to validate these observed patterns, quantify the impact different micro-locational characteristics can have on footfall and, potentially, identify some previously unobserved trends.

Recent research has helped to fill this gap. Lugomer and Longely (2018) identified different temporal footfall signatures across Great Britain, some of which were linked to the function of their micro-location. This research was built on by Lugomer (2019), who demonstrated that the magnitude and pattern of footfall are statistically significant to the hierarchy of the street and the workplace classification of the location, in addition to investigating how different retail stores can result in different footfall patterns, for example, fast-food outlets and restaurants showing a more pronounced lunch-time peak and shopping centres showing to be busier at midday.

Berry *et al.* (2016) investigated how workplace population and proximity to stations and transport hubs impacted footfall and sales for a selection of Co-Op grocery stores in Inner London, showing how both impact weekday footfall. They showed how the different positioning of the three stores determined their footfall, with stores near to stations experiencing higher footfall at commuting times, and a store on a major throughway showing a more pronounced lunchtime peak, but no commuting peaks.

Another example of research that has explored the impact of different influence on footfall or visitors number is the Hart *et al.* (2014) report, 'The customer experience of town centres'. They collected a wealth of quantitative and qualitative data from focus groups in six different UK towns and cities to explore how consumers use and experience the high street. Their findings showed hot spots in retail locations near major fashion and chain retailers and gave an insight into factors that data-driven approaches can oversee. For example, participants were quoted as avoiding retail areas which are known to be busy. Therefore, locations with high footfall can equally act as footfall deterrents.

Even so, there continues to be a clear gap in research that quantifies the impact of a micro-scale factor, and studies which explore how that impact varies at different times of day and days of the week as well as across different towns and cities. Examining micro-scale relationships has particular value as they account for the variation in footfall within retail centres, providing insight into why two locations in the same meso- and macro-scale contexts might experience different footfall.

Identifying and quantifying relationships between micro-locational characteristics and footfall can have many applications. For retailers, it can inform potential locations for new stores based on their surroundings' influence on footfall. It can also help with logistical planning and ensuring that a store is appropriately stocked and staffed at peak times. There are also applications for determining rents and business rates. Retail rents and business rates are both high in the UK, and one of the biggest determining factors as to whether a store is viable or not (British Retail Consortium, 2021b). Therefore, retailers and store managers want to ensure the location they pay for receives good footfall and at the times which would best suit the store type (Sanderson and Edwards, 2014). By relating business rates and retail rents to the footfall, this could be a fairer way of determining how much a retailer has to pay. Also, if a specific characteristic is proven to increase footfall significantly, a premium price could be charged to open a store nearby. In addition, it contributes to the knowledge of micro-locational patterns that can exist within retail centres, which could be used to inform spatio-temporal predictions or forecasts of footfall. This could hold value for retailers, town planners and stakeholders. There is a consensus that data driven empirical evidence is needed to support high street performance and revitalisation strategies (Portas, 2011; Wrigley and Dolega, 2011) and, footfall, often cited as the 'lifeblood' of a high street vitality and viability (Birkin et al., 2017), is a key measure for the successfulness of these strategies and a widely used proxy for their economic performance (Coca-Stefaniak, 2013; Millington et al., 2018) . Therefore, the more that is known about footfall and the factors that influence it, the better any insights from footfall data can be applied.

This chapter investigates the impact of a range of functional and morphological micro-locational characteristics on footfall magnitude and signature. Footfall magnitude and signature are referenced considerably in the following chapters. Footfall magnitude is defined as the number of people measured within a set time period (for example, five minutes, one hour or a day). Footfall signature refers to the variation of footfall magnitude according to the time of day, day of week or month of the year. It is important to investigate both, as they might show a different relationship to micro-locational characteristics.

This chapter has three objectives, as outlined on the chapter title page. The first objective will be to investigate how different characteristics and contexts of the immediate environment impact footfall magnitude and signature. First, Section 4.1 will identify the micro-locational characteristics that influence footfall and derive quantitative representations using available data sources. Section 4.2 will then focus on this first research objective by investigating the correlation between footfall signature and magnitude of the different micro-locational characteristics derived for each location in the previous section. This will quantify how these different factors can impact footfall, and how much that can change over time.

The second objective will be to use these characteristics of retail and footfall context to develop a classification. In Section 4.3, the second research objective will be fulfilled, applying culture analysis to define three classes of footfall context, where the sensors within a class have similar micro-locational characteristics.

The third objective will be to identify how trends in footfall magnitude and signature may differ between the classes. This will be explored in Section 4.4, which looks more in depth at the characteristics of each of the defined classes. Finally, Section 4.5 provides a chapter summary.

This chapter has been adapted from the article 'Archetypes of Footfall Context: Quantifying Temporal Variations in Retail Footfall to Micro-locational Characteristics' published in Applied Spatial Analysis and Policy (Philp et al., 2021). The full paper can be found in Appendix 4.1.

4.1 Deriving micro-locational characteristics that impact footfall

To investigate the quantitative impact of micro-locational characteristics on footfall, the attributes of interest must first be identified and derived. The process for this is outlined in this section. First, in Section 4.1.1, thirteen micro-locational characteristics which may impact footfall are determined from the literature. Section 4.1.2 introduces the methodology and datasets used to quantify these characteristics, and in Section 4.1.3 they are described and summarised.

4.1.1 Identifying micro-locational characteristics

There is a broad range of factors that might be influential on footfall. In Chapter 2, a number of these were identified through literature and grouped under three headings: Macro-scale factors, meso-scale factors and micro-locational factors. These are restated in Figure 4-1. The latter, micro-locational factors, will provide the focus of this chapter. These are characteristics of the immediate surrounding environment that might impact footfall, such as proximity to specific stores or how that street is situated within the broader network of the retail centre. This emphasis on micro-locational characteristics should not imply that macro- and meso- factors do not impact footfall. However, there is a significant research gap when considering the effect of micro-locational factors on footfall magnitude and signature. Thus, this chapter will focus on these. Meso- and macro-scale factors will be explored in more depth in the analyses in Chapter 5 and Chapter 6.

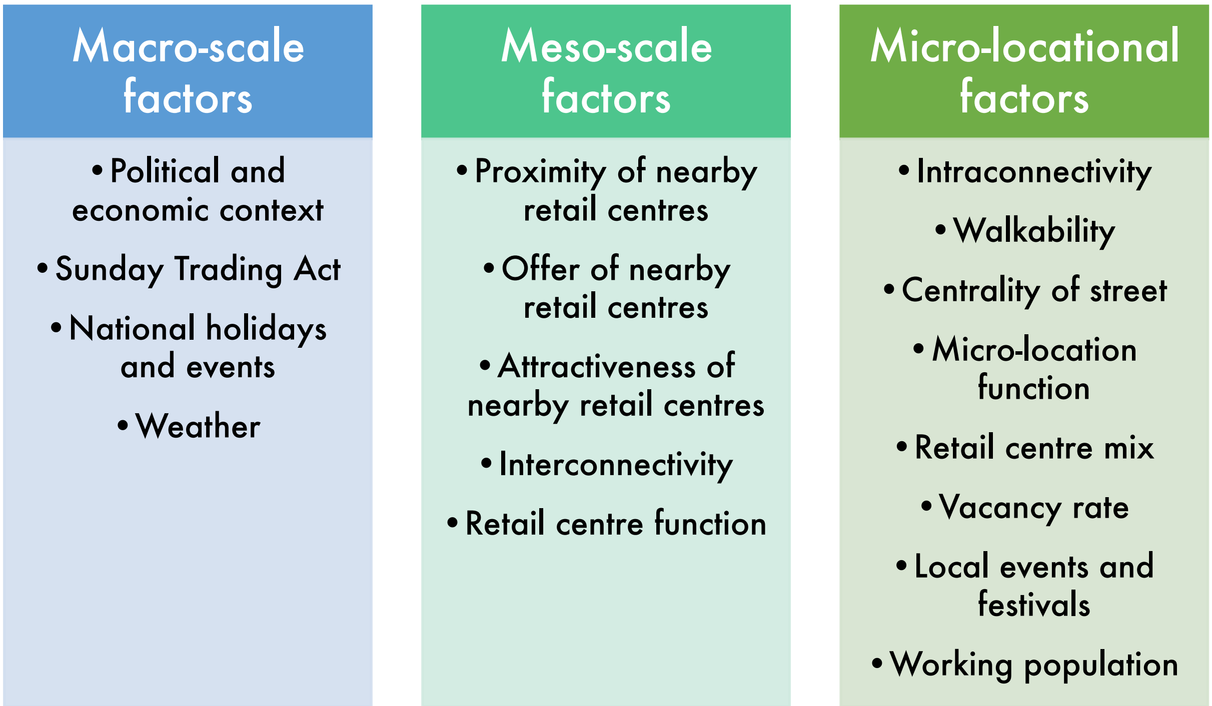


Figure 4-1 Footfall influences identified from the literature

There has been limited research on the quantitative impact of these micro-locational characteristics on footfall, however there is evidence of their effects. These micro-locational effects can be summarised under two main headings: ‘functional’ and ‘morphology’. The ‘functional’ category captures aspects of context that may attract people to a retail area. The purpose for any patronage of a retail site is logically linked to a temporal factor. For example, food outlets will attract more people during mealtimes, and an area rich with bars and restaurants would draw people in the evenings aligned to opening hours. The ‘morphology’ category encompasses features of walkability and attractiveness, such as transport accessibility, the density of units, and the street’s centrality within the retail centre network.

Functional

The function of a retail centre is the purpose it serves to users. Most retail centres are multi-functional, simultaneously performing several purposes (Millington et al., 2015). The function of a retail centre or micro-location impacts footfall magnitude and signature.

Firstly, having a varied and cohesive retail mix has been shown to boost retail centre vitality and attractiveness (Millington et al., 2015; Tyler et al., 2012). If the offer of retail centre or micro-location matches the demand of the consumers, the magnitude of footfall will increase (C. R. Parker et al., 2016; Portas, 2011).

Secondly, research shows that the function of a retail centre is closely aligned to both diurnal and other periodic patterns of use. For example, retail centres in locations with a high concentration of employers and businesses typically have higher daytime footfall (Berry et al., 2016; Swinney and Sivaev, 2013). Such relationships have been shown to drive footfall and sales during weekdays, especially in the early morning, midday, and early evening (Berry et al., 2016). Characteristics such as the presence of anchor stores or the tendency towards premium or value goods can all indicate the retail centre identity, who it may appeal to, and consequently, when they may visit (Guy, 1998).

Morphology

How the street is situated within the wider network has also proven to be a reliable indicator of pedestrian counts (Hillier et al., 1993; Raford and Ragland, 2006). Well-connected streets tend to have higher footfall as it is often the shortest route from their origin to their destination. The connectivity of a road can be determined by various centrality measures, including betweenness, which can capture the prominence of a node as a bridge between other nodes (Freeman, 1977; Porta et al., 2009).

Streets can also have high footfall if they are close to access points for other forms of transport, such as train stations, car parks or bus stops (Mazumdar et al., 2020). As popular origins and destinations, these features can concentrate footfall to particular micro-locations (Scheurer and Porta, 2006). Anchor stores, restaurants and entertainment venues have similarly demonstrated footfall attraction (Bras et al., 2021; Hart et al., 2014; Teller and Alexander, 2014; Üsküplü et al., 2020; Yuo et al., 2003). The proximity of stores to significant transport hubs has increased their footfall and sales, particularly at commuting times (Berry et al., 2016). Having good access to car parking is a demand of retail areas and many consumers will avoid using public transport in favour of the convenience of their own vehicle. Therefore, the proximity of retail space to a public car park can influence the number of visitors and impact footfall for the entire retail centre (Coca-Stefaniak, 2013; Tyler et al., 2012).

Academic literature points to many functional and morphological influences on footfall. However, no literature exists to our knowledge that quantifies the impact of a combination of these influences. Therefore, a data-driven exploration of footfall spatial and temporal patterns will add quantifiable evidence to the existing evidence base in this research area, particularly to observed relationships between footfall and the characteristics of the surrounding micro-location.

4.1.2 Representing micro-locational characteristics

To perform a robust quantitative analysis on the relationship between micro-locational characteristics and footfall, variables need to be derived that represent these characteristics. These variables must match the footfall data both spatially and temporally, covering the same areas over roughly the same time periods. The LDC Retail Unit Address data (2017), the UK Census (2011), OpenStreetMap (2018), the CDRC Retail boundaries and the Department for Transport NaPTAN (2014) and Car Parks (2015) datasets were all utilised to calculate different micro-locational characteristics. More information on the limitations and coverage of these datasets can be found in Chapter 3, with sources to the datasets in Appendix 3.1.

A range of potential variables theorised to significantly impact footfall magnitude and signature were assembled using these datasets. Thirteen variables were selected, and these are gathered in Figure 4-2. Details of how each variable was derived can be found in Table 4-1.

Functionality

- Distance to the nearest anchor store
- Distance to the nearest premium store
- Distance to the nearest entertainment activity
- Proportion of vacant stores
- Proportion of independent stores
- Proportion of value stores
- Proportion of night-time economy locations
- Workplace population
- Ratio of service to retail stores

Morphology and Connectivity

- Distance to the nearest transport hub
- Distance to the nearest car park
- Density of stores
- Centrality of the street

Figure 4-2 Micro-locational characteristics that influence footfall

Table 4-1 Key features of the variables used as micro-location footfall descriptors

Variable	**Derivation**
Distance to the nearest anchor store	Euclidean distance (metres) to nearest anchor store, identified by their brand name (e.g. John Lewis, Primark, full list in Appendix 4.2)
Distance to the nearest premium store	Euclidean distance (metres) to the nearest premium store, identified by their brand names (e.g. The White Company, Burberry, full list in Appendix 4.2)
Distance to the nearest entertainment activity	Euclidean distance (metres) to the nearest venue for entertainment (e.g. Cinemas, Arcades, Museums). These were identified using the LDC's typology (full specification in Appendix 4.2)
Proportion of vacant stores (vacancy rate)	The proportion of vacant stores identified within a 100m straight-line buffer of the sensor
Proportion of value stores	The proportion of stores identified as value stores by their brand name (e.g. Aldi, Home Bargains, full list in Appendix 4.2) within a 100m straight-line buffer of the sensor
Proportion of independent stores	The proportion of stores identified as independent by the singular instance of their store name in the dataset within a 100m straight-line buffer of the sensor
Proportion of night-time economy locations	The proportion of locations within a 100m straight-line buffer of the sensor that offer evening appeal (e.g. bars, clubs, restaurants, fast food) identified using LDC's typology (full specification in Appendix 4.2)
Workplace population	The average of the daytime population densities of the workplace zone the sensor falls into and those which border it.
Ratio of service to retail	The ratio of number of service locations over number of retail locations within a 100m straight-line buffer of the sensor (further specifics in Appendix 4.2)
Distance to the nearest transport hub	Euclidean distance (metres) to the nearest group of bus stops or train stations as identified in the NaPTAN dataset.
Distance to the nearest car park	Euclidean distance (metres) to the nearest car park as identified by the Department for Transport
Density of stores	The number of store units within a 100m straight-line buffer of the sensor
Centrality of the street	The edge betweenness of the street the sensor is located on.

For several descriptors, such as the proportion of vacant stores or the density of stores, a 100m circular buffer around the sensor was calculated to select the stores close enough to be considered within the immediate retail environment of the sensor. 100m was chosen as it encompasses a reasonable sample of stores to derive a complete picture of the retail environment but is not so large as to remove the micro-locational variation of interest. This method assumes that the sensor is located in an area with a dense concentration of retail units and that the circular shape can appropriately capture this. Sensors with fewer than five units within the buffer area (a total of 5 sensors) were removed from the sample as there are not enough stores to get a representative understanding of the proportions within the retail environment. The number of stores in the buffer ranged from 7 to 189. This was used to define the density of stores variable. This number was also used as the denominator to calculate the proportion variables, including the proportion of independent, vacant and value stores.

Representing density of stores through number of stores makes the assumption that all stores are a similar size when this is not the case. A limitation of this measure is that it might discount the impact of stores with large floor space and overcompensate for locations which have lots of small stores or concessions. An indicator based on floor space might be more representative of the density of retail area, however no such data was available.

For some features, a Euclidean distance was used instead of a proportion. These included anchor stores, premium stores, transport hubs, and features that appear in most retail centres, though not in multitude. When a proportion was calculated for these features, they returned measures with a significantly constrained variation as there would be so few instances within a buffer. As such, distance was deemed to be a more appropriate measure.

It should be acknowledged that network distance may have been a more appropriate for this task. Due to difficulty accessing certain software and technology at the time of analysis, it was not feasible to complete this analysis using network distance and Euclidean distance was chosen instead. Repeating this analysis and instead applying network distance to calculate buffers and distance may produce different results.

The street centrality measure was calculated from networks generated by the OSMnx python library (Boeing, 2017). OSMnx uses data from OpenStreetMap to create a network graph of a road structure within a boundary. The CDRC retail centre boundaries were used to generate the pedestrian network around a sensor. The edge betweenness centrality of the sensor street was then calculated to give the street centrality measure. Edge betweenness was chosen as the centrality measure because it can be applied to streets instead of intersections, where most footfall measurements are taken, aiming to capture the prominence of that edge as a pass-through route.

4.1.3 Variable descriptions

After outlier removal, the micro-locational characteristics dataset used in this chapter consisted of 13 variables for 640 sensor locations (640 x 13). As shown in Appendix 4.3, no variables were strongly correlated with each other. The strongest correlation was 0.44 between 'Distance to Anchor Store' and 'Proportion of Independent Stores', both of which capture different aspects of a retail environment, despite their high correlation. The absolute mean correlation value was 0.16.

Out of the distance variables (anchor store, premium store, entertainment venue, transport hubs and car parks), anchor stores and transport hubs were the closest to sensors being 139m away, on average. The furthest away were premium stores, which were 284m. On average, 46% of the stores within 100m of the sensor were independent, and just 3% were value stores. 22% were night-time economy outlets, and 8% were vacant. The mean number of stores within 100m of the sensor was 65. However, all thirteen variables have a high variation, with different variables being more prominent in some locations than others, which is useful as it can investigate how the presence or absence of that micro-locational characteristic can impact footfall. The complete summaries of the micro-locational characteristics, including mean, median, maximum, minimum and interquartile range are given in Appendix 4.3.

4.2 Micro-locational characteristics and footfall

The last section identified and represented thirteen functional and morphological micro-locational characteristics, including retail mix, proximity to transport hubs, working population and network centrality. Observational and anecdotal evidence to suggests that these characteristics may have a significant impact on footfall, however, this impact is not quantified. This section will employ the SmartStreetSensor footfall dataset to achieve the first research objective: Investigate how different characteristics and contexts of the immediate environment impact footfall magnitude and signature

First, the footfall data is cleaned and pre-processed to remove noise and reduce error and subsection 4.2.1 will briefly outline this process. Then, in subsection 4.2.2, the correlation between footfall magnitude and micro-locational characteristics is investigated, and the potential impact is quantified. Finally, in subsection 4.2.3, the relationship between the micro-locational characteristics and footfall signature is explored. Micro-locational characteristics are tested, exploring if their effect on footfall differs depending on the time of day or the day of the week.

4.2.1 Pre-processing the footfall data

Before any analysis was completed, the footfall data was cleaned and pre-processed. Footfall measurements from January 2017 until August 2018 were used to reflect the same temporal coverage as the Retail Unit Address Data. Only sensors with over nine months of data were used to remove any bias from new or temporary sensors, which may only have footfall data for busier or quieter times of the year. The data was adjusted from UTC to GMT/BST to reflect local time and date.

As of August 2018, LDC has sensors in 840 locations in 88 towns and cities across the UK. Due to limitations in coverage of the supplementary datasets, these variables could only be derived for 640 sensors, covering 40 high street retail sites in England and Wales. Locations with internal sensors (footfall sensors inside the store, as opposed to the storefront) were excluded, as were sensors inside larger buildings such as shopping centres or stations. In addition, some outliers were removed in the analysis process (in Section 4.3.2).

The distribution of sensors is particularly biased towards London (n=291), with 45% of the sensors, as well as larger cities such as Manchester (n=18), Liverpool (n=16) and Cardiff (n=8). Excluding London, sensors per location ranges from n=20 in Kingston-upon-Thames to n=1 in Gateshead and Windsor. Although most sensors in the sample are in larger cities, some regional centres and market towns are also represented, such as Taunton (n=6) and Market Harborough (n=13). The complete geographical distribution of the sample can be found in Appendix 4.4.

As discussed in Chapter 3, there is significant noise and error within the dataset. To avoid extreme values skewing the results, the top and bottom 5% of values for every sensor were removed[17]. In addition, footfall was aggregated for each sensor.

For this analysis, footfall was aggregated to a daily mean for each sensor. This daily average will be compared to each of the micro-locational characteristics via a Pearson's product-moment correlation to achieve an impression of the relationship independent of the time of day or the day of the week. Mean daily footfall for the locations was 9,435 people with a standard deviation of 7,196.

[17] This was stratified by weekday and hour, therefore, if a value was outside of 5-95% of values for that location, that day of the week and that hour of the day, it was removed.

In Section 4.2.3, the correlation between micro-locational characteristics and footfall is measured over time. To capture this, the footfall data was instead aggregated, per sensor, into hourly groups as follows:

00:00 – 03:00	Night
04:00 – 07:00	Early Morning
08:00 – 11:00	Morning
12:00 – 15:00	Afternoon
16:00 – 19:00	Early Evening
20:00 – 23:00	Evening

This process was completed separately for weekdays, Saturdays and Sundays to investigate how the relationship could also vary by day of the week. This aggregation minimises the impact of erroneous results, and it also streamlines the analysis so that it is more efficient. If the 13 micro-locational variables were compared to every hour of every weekday, 13 x 168 correlations would be calculated instead of the 13 x 18 aggregating in such a way that retains the key daily and hourly trends.

Figure 4-3 shows how the mean and standard deviation of footfall vary by the day of the week and the hourly groups. Each day of the week shows a similar pattern, with the afternoon being the busiest time and the evening and night being the quietest. However, there are subtle differences between the days. Weekdays and Sundays generally have a smaller footfall magnitude than Saturdays, yet the Early Evening time during Weekdays is comparable in magnitude to Saturday Early Evenings. Weekday nights are significantly quieter than the Friday into Saturday and Saturday into Sunday nights. Sundays also have a lower standard deviation and vary more in terms of footfall when compared to Weekdays and Saturdays.

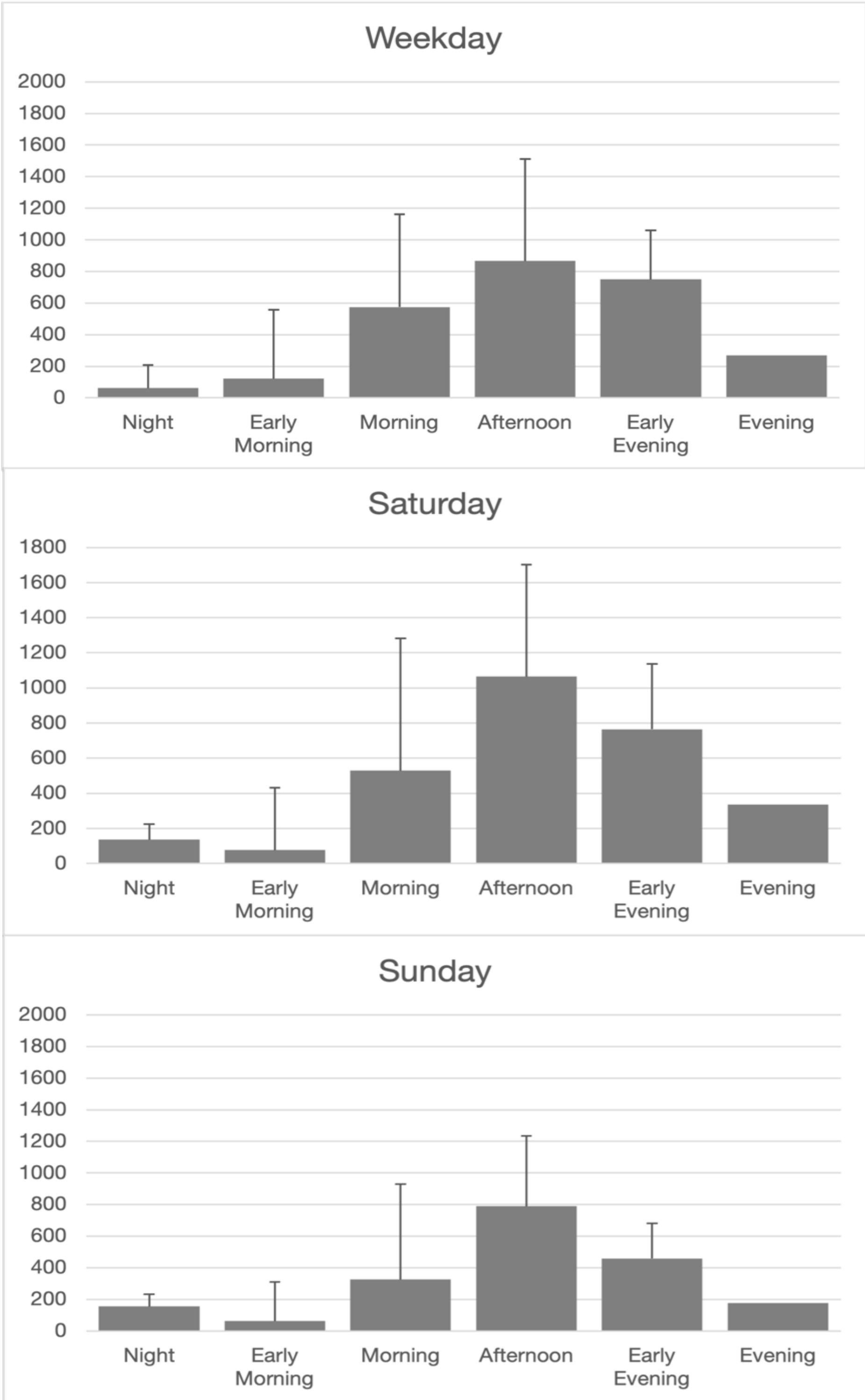


Figure 4-3 Mean and standard deviation for footfall by days of the week and hours of the day

4.2.2 Correlation between micro-locational characteristics and footfall

The correlations between micro-locational characteristics and mean daily footfall are relatively weak, with a mean absolute correlation of 0.16. However, this varied depending on the micro-locational characteristic. Most were significant (confidence interval of 95%), except for 'Distance to Entertainment Venue', 'Distance to Car Park' and 'Density of Units'. There is some uncertainty as to why this is. It could be that those characteristics do not significantly impact footfall, or that these features are not accurately represented in the data. Alternatively, they could only be significant at certain times of day or days of the week, but not for footfall overall. The correlations and statistical significance for each characteristic are given in Table 4-2 below.

Table 4-2 Correlation between mean daily footfall and micro-locational characteristics

	Correlation	P-value	
Workplace Population	0.35	0.000	***
Proportion of Night-time economy	0.18	0.000	***
Ratio of service to retail	-0.12	0.008	**
Proportion of Vacant stores	-0.12	0.010	**
Distance to Anchor Store	-0.13	0.003	**
Distance to Premium Store	-0.15	0.001	**
Proportion of Value stores	-0.19	0.000	***
Centrality of Street	-0.22	0.000	***
Proportion of Independent stores	-0.23	0.000	***
Distance to Transport Hub	-0.26	0.000	***
Distance to Car Park	0.08	0.075	
Distance to Entertainment Venue	-0.07	0.154	
Density of Units	0.03	0.485	

The highest correlation was 0.35 between 'Workplace Population' and footfall. 'Distance to Transport Hub' and 'Proportion of Independent stores' also produced significant negative correlations of -0.26 and -0.23, respectively. Therefore, according to the data, a location with a high workplace population, with more chain stores and close to a transport hub would experience a higher average daily footfall.

It should be noted that even the largest correlations calculated between footfall and these micro-locational factors are still weak. As a measure, footfall is very complex and can represent a large aggregation of causative factors, some which can be captured and others which are very transient. Therefore, can be challenging to draw any strong correlations between footfall and these factors using a general, top-down approach.

Interestingly, the correlation between vacancy rate and footfall is relatively weak. Vacancy rate and footfall are both key performance indicators for high streets; therefore, a high correlation between both would be expected. However, using this dataset, value stores have a stronger negative correlation with footfall than vacant stores. A proportion of those vacant units could have only been empty for a short period of time, and as a result, they are not indicative of decline. However, this pattern persists even when churn is taken into account. It could be due to the sample containing a high proportion of retail centres where footfall is particularly resilient to changes in vacancy rate. It could also suggest that vacancy and footfall are not unanimous indicators of decline. A high proportion of vacant units might not be a deterrent if there is demand for the occupied units or the environment is attractive. Another explanation could be that footfall and structural vacancy have a non-linear relationship, and that when the number of vacant units on a street reaches a certain proportion, it has a much stronger impact on footfall. It would make intuitive sense that there are not many locations like this in the sample, as the sensors tend to be placed in occupied units.

Conversely, a high footfall does not always indicate a solid retail location. Little is known about the relationship between footfall, spend and conversion rate. For example, the main route into a city centre with lots of vacant units could still have high footfall as many people pass through. The vacant units imply a decline, as the location has the footfall but cannot convert that into a sustainable income for businesses. The weak correlation between footfall and vacancy rate could be positive. It shows that they are strong performance indicators that capture different aspects of high street vitality and viability.

4.2.3 Correlation over time

When time is not considered, the correlation between micro-locational characteristics and mean daily footfall appears generally weak. However, this is less of the case when this relationship is explored over time.

The thirteen micro-locational characteristics were compared to Night, Early Morning, Morning, Afternoon, Early Evening and Evening times during weekdays, Saturdays and Sundays. There was significant variation found when exploring this relationship temporally. The strongest absolute correlations for each timeframe can be found in Table 4-3. 'Proportion of Night-time economy' had the strongest correlation with footfall during the Evening and Night. Whereas 'Distance to Transport Hub', 'Workplace Population' and 'Proportion of Independent Stores' were more strongly correlated during the Early Morning, Morning and Afternoon. The complete correlation tables for every timeframe and variable can be found in Appendix 4.5.

Table 4-3 Strongest correlations between micro-locational characteristics and footfall over time

	Night	**Early Morning**	**Morning**	**Afternoon**	**Early Evening**	**Evening**
Weekday	Proportion of Night-time economy (0.37 ***)	Distance to Transport Hub (-0.29 ***)	Workplace Population (0.31 ***)	Proportion of Independent stores (-0.39 ***)	Workplace Population (0.42 ***)	Proportion of Night-time economy (0.43 ***)
Saturday	Proportion of Night-time economy (0.40 ***)	Distance to Transport Hub (-0.27 ***)	Proportion of Independent stores (-0.38 ***)	Proportion of Independent stores (-0.44 ***)	Workplace Population (0.34 ***)	Proportion of Night-time economy (0.44 ***)
Sunday	Proportion of Night-time economy (0.40 ***)	Proportion of Night-time economy (0.32 ***)	Proportion of Independent stores (-0.28 ***)	Proportion of Independent stores (-0.40 ***)	Workplace Population (0.34 ***)	Proportion of Night-time economy (0.41 ***)

The correlation between micro-locational characteristics and footfall is more time-dependent for some variables than others. 'Proportion of Independent Stores' and 'Proportion of Night-time economy' vary considerably over time. During Evening, Night and Early Morning, the number of independent stores does not significantly correlate with footfall. Still, it is one of the strongest correlated factors during the daytime. The inverse is true with 'Proportion of Night-time economy'. It is one of the strongest predictors during the Evening and Night, yet the correlation becomes weak or insignificant during the daytime.

'Distance to Transport Hub' and 'Workplace Population' are the most consistently strongly significant variables. Both are always significant to the 90% confidence level. 'Distance to Transport Hub' ranges from -0.11 to -0.29 and 'Workplace Population' from 0.10 to 0.42. For all other variables, the strength, direction and significance of their relationship with footfall are hugely dependent on time.

'Distance to Entertainment Venue', 'Distance to Car Park' and 'Density of Units' were all not significant when only taking the average daily footfall. However, they are significant when looking at certain times and days. 'Distance to Entertainment Venue' has a weak negative correlation with Saturday footfall at Morning, Afternoon and Early Evening times, as well as Sunday afternoons (the closer to an entertainment venue, the higher the footfall). 'Distance to Car Park' has a weak positive correlation at Early Mornings on all days of the week. This trend is interesting, as the further the sensors were from a car park, the higher the footfall, which contradicts what would be expected. This could perhaps be due to large car parks being located outside city centres, or a higher proportion of early morning visitors travelling by public transport or foot as opposed to driving.

'Density of Units' has an unusual pattern as it is significantly correlated with footfall in Early Mornings and Afternoons but changes direction. In the Early Morning, the relationship is negative, and in the Afternoon it is positive. Therefore, in Early Mornings, there are more people in less dense retail environments, but by Afternoon this has switched. This indicates how the function or purpose of a retail centre might shift over time.

This section investigated how different micro-locational characteristics correlate with footfall magnitude and signature. The relationship between micro-locational characteristics and footfall is generally weak when time is not considered. However, when the time of the day and day of the week is taken into account, all micro-locational variables have a statistically significant correlation with footfall. The variables with the strongest and most consistent correlation with footfall are 'Workplace Population' and 'Distance to Transport Hub'. In contrast, variables such as 'Proportion of Independent stores' and 'Proportion of Night-time economy' have a strong relationship with footfall at certain times of the day. This section has demonstrated the importance of temporal context when evaluating the impact of the surrounding environment on footfall.

4.3 Classes of footfall micro-locational context – methodology

The micro-locational characteristics derived in Section 4.1 relate to footfall magnitude and signature in complex and multi-dimensional ways. The last section proved that the impact of a variable is highly dependent on context. For each of the 640 sensors, there are 13 functional and morphological micro-locational characteristics that could impact their footfall magnitude and signature to different extents and different ways. This density of data lends itself well to a case study investigation, exploring the different patterns and nuances within a particular micro-location (as is done in Chapter 5). However, to understand the general patterns that exist for Great Britain, data for all 640 sensor locations would have to be explored and analysed which would be a time-intensive task.

Instead, a clustering algorithm is applied. Cluster analysis is a technique that divides unlabelled data into a number of groups such that the points in the same group are similar to each other, and those in different groups differ. In this case, it would group each of the 640 locations into classes with similar functional and morphological features. Then, as opposed to exploring and analysing the individual patterns of 640 sensors, the analysis and exploration will only have to be completed once for each class.

Cluster analysis allows a classification of retail micro-locations to be generated, which can also be a valuable tool in aiding the understanding of retail centres. As discussed in Section 2.1.2, an appropriate classification system that can be easily conceptualised and understood is an efficient way of transferring knowledge about retail environments to retailers and decision-makers. It can be utilised in discussions into town centre function and consumer behaviour and allow comparisons between different locations. Therefore, the classifications established here to draw connections between micro-locational characteristics and footfall have more comprehensive applications.

In Section 4.3.1, different clustering algorithms are presented and discussed, and *K*-means clustering is chosen as the most appropriate method to cluster the micro-locational characteristics. Section 4.3.2 will detail the analytical approach and how the algorithm was implemented to produce the three classes of footfall context. Pen portraits to introduce and describe these classes will be in Section 4.3.3. Finally, Section 4.3.4 will discuss the limitations of these classes and the contextualises their application.

4.3.1 Cluster analysis methods

There are many algorithms designed for cluster analysis, including centroid-based clustering, hierarchical clustering, and density-based clustering. The choice of clustering algorithm depends on which would be most appropriate for the characteristics of the dataset.

The *K*-means algorithm was chosen for this analysis. *K*-means is a centroid-based clustering method first introduced in Lloyd (1982). It groups unlabelled data by minimising the distance between randomly generated cluster centres and nearby data points. With each iteration, the cluster centres adjust to minimise this distance further. When they cannot be minimised further, the cluster centres are stationary, and the algorithm has converged on a solution.

The advantages of *K*-means are that it is simple to conceptualise and implement, can be applied to datasets of different sizes, and is a commonly used and understood methodology in many fields, including geodemographic analysis (Burns et al., 2018; Spielman and Singleton, 2015). However, the *K*-means algorithm does have some limitations. Firstly, as it maximises the sum of squared distances, it is susceptible to outlying values; therefore, these should be removed before analysis. In addition, the result is also heavily dependent on the generation of the initial cluster centres, and vastly different results can be computed depending on this. It also requires *k* or the number of clusters to be known beforehand or chosen by the researcher.

An alternate method would be to use a hierarchical clustering or a density-based clustering algorithm.

Hierarchical clustering arranges the points in a hierarchy based on a measure of the distance in between them. There are two approaches: the top-down Divisive approach and the bottom-up Agglomerative approach. The Divisive approach assumes the data points all belong to one large cluster and begins to separate them based on the largest distance between the points. This occurs until a termination point is reached, either a pre-determined number of clusters or a measure of the minimum sum of squared errors. The Agglomerative approach assumes that every data point has an individual cluster and combines them iteratively with the data points that are closest to each other. Hierarchical clustering algorithms have many of the same disadvantages as *k*-means clustering. They are sensitive to outliers, highly dependent on the initial seed or order of the data, and the number of clusters has to be chosen by the researcher. In addition, it is more computationally expensive than *k*-means clustering.

However, a density-based clustering algorithm, such as DBSCAN (Moreira, Santos and Carneiro, 2005), avoids these disadvantages. It does not require the specification of cluster number and can filter out noise from outlying values. It takes a starting point and checks how many other points are within a user-specified distance. If it is greater than the minimum number of points for a cluster (defined by the user), then that point is marked as part of a cluster; else, it is marked as noise. The clustering procedure continues until each point is marked as within a cluster or as outliers.

DBSCAN was implemented on the dataset; however, the resultant clusters were not viable after parameter adjustment. Either they were too small and not representative of any larger patterns, or they weren't interpretable within the context of the data. Therefore, *K*-means clustering was chosen for this study.

4.3.2 Analytical Approach

In order to run the *K*-means algorithm, the features were standardised according to their mean and standard deviation. As *K*-means optimises the sum of squared distance, values with different units of measure, such as metres for distance and a 0 – 1 scale for proportion, could cause skew towards certain variables. In addition, outliers can have a significant impact on the results. Some locations were classed as outliers because they had unusually large or small values for some variables. For example, three sensors in Lymington were removed as they were over 18km from the nearest entertainment activity. Five additional sensors were removed iteratively throughout the clustering process, as they were the furthest point from any cluster centre. The resulting clusters were as compact and well-separated as possible without removing more outliers than necessary. The features were checked against each other to ensure there are no high correlations to avoid multicollinearity (see Appendix 4.3).

The clustering algorithm was run using $k = 3$. There was no prior indication from the data to suggest a value of k; therefore, a comparison of the average silhouette score was used. A silhouette score measures how well a certain point fits within the cluster it has been assigned. It ranges from +1, representing a point that fits perfectly in the generated cluster, to -1, which represents a point that poorly fits into the current cluster and would fit better in another. The average silhouette score is the mean silhouette score for every point in the clustering. The average silhouette score for different values of k, as shown in Figure 4-4 was used to determine that *k*= 3 provides the best separation and cluster results.

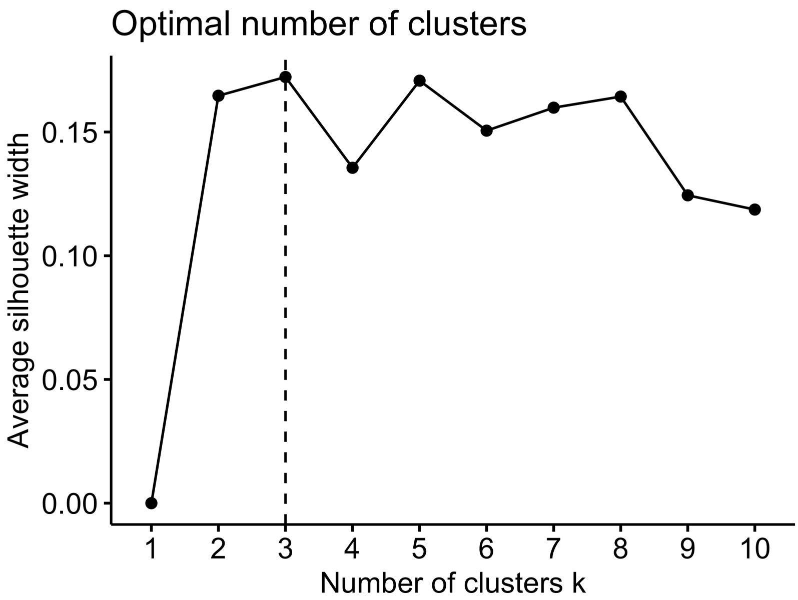


Figure 4-4 The change in average silhouette score for different values of *k*

One of the pitfalls of using the *K*-means algorithm is that it is a stochastic process (Lloyd, 1982). Therefore, if certain cluster centres were generated in an unfavourable position, it could lead to a poor result. To avoid this issue, the clustering was optimised. 10,000 iterations were run, each using unique and randomly generated starting centres from a static seed, and from the 10,000, the best clustering outcome was chosen.

The average silhouette score for the final clustering was 0.17. For a silhouette score, this is relatively low; however, this could be expected due to the ambiguous nature of boundaries between retail areas. It is rare to find a street or micro-location which only serves one purpose, and there is often qualities or retailers in a location that cater to a different function than others. In addition, even if there are streets that serve similar purposes, it is unlikely that they will also have the same structural qualities. Therefore, it is understandable that the clusters have a degree of overlap between them. Some methods tailor to this quality in datasets, such as fuzzy c-means clustering; however, they do not produce the clear-cut labels needed to define a classification system and compare each class to its footfall signature and magnitude.

4.3.3 Pen portraits

Cluster profiles, often referred to as 'Pen Portraits', were constructed so that the separate classes could be understood and conceptualised efficiently. These were obtained using the values of the cluster centres and through exploratory research into the characteristics of individual locations. This process is detailed in more depth in Appendix 4.6.

The three classes derived from applying the *k*-means clustering algorithm were given the following titles.

- ◊ Chain and Comparison Retail micro-locations [CCR]
- ◊ Business and Independent micro-locations [BI]
- ◊ Value-Orientated Convenience Retail micro-locations [VOCR]

Chain and Comparison Retail micro-locations was the largest cluster, containing 343 (54%) of the sensor micro-locations. Business and Independent micro-locations included 254 sensors (40%), and Value-Orientated Convenience Retail micro-locations was the smallest with 43 (7%) of the 640 sensors. Each class is described in the pen portraits across the following three pages.

Chain and Comparison Retail micro-locations [CCR]

Number of sensors: 343 (54%)

The CCR cluster was the most common of the three clusters, and almost every city or town in the sample had a sensor in this cluster. They are named after their predominantly comparison retail function and their dominance towards chain retailers. From the clustering features, these micro-locations had a low proportion of independent retailers, were close to anchor stores and premium retailers and had a bias towards retail outlets over services. As such, destination shopping locations fit well into this cluster, for example, **Oxford Street in London, Liverpool ONE in Liverpool** and **Queen Street in Cardiff**. These locations are designed for comparison goods shopping, with a range of chain stores catering to create a sizeable retail offer. These are sought after locations for retailers, often in the retail core of major cities.

Business and Independent micro-locations [BI]

Number of sensors: 254 (40%)

The BI cluster encompasses places with a tendency towards independent retail, often in financial and office-dominated districts. 212 (83%) of the sensors in this cluster are sensors in London, representing 70% of the total sensors in London. This cluster captures the employment areas and the destination for many commuters. These areas are common in larger cities, where people do not tend to live near where they work, explaining why this cluster is predominant in London. BI micro-locations have a high working population, are close to transport hubs, and have a high proportion of independent retailers in terms of the clustering features. Some examples of these places are **Holborn and the City of London, in London and NOMA** and **Spinningfields in Manchester**. This cluster also includes areas with a high proportion of night-time economy outlets, such as **Park Street in Bristol, Soho in London** and **Bold Street in Liverpool.**

A significant distinction of locations in this cluster is that they have 9% more restaurants than the average British high street, subsequently reflected in a near 1:1 ratio between service and retail outlets. This shows that this cluster has a more experience-based function than a comparison retail-based one. This is supported by their large distance from anchor stores and their small proportion of value retailers.

Value-Orientated Convenience Retail micro-locations [VOCR]

Number of sensors: 43 (7%)

The VOCR micro-locations cluster describes smaller, secondary centres of a larger urban area. These are residential areas with a high prevalence of budget convenience retailers, betting and charity shops. A higher proportion of value outlets, a considerable distance from premium stores and entertainment venues and a low workplace population also indicates VOCR micro-locations. These areas are the opposite of destination shopping areas; people visit these areas out of convenience. They exist due to their accessible location near residential areas so that consumers can gather their essentials without making an extended trip. VOCR micro-locations have few entertainment venues and night-time economy outlets, as these are things which people are willing to travel for. Some examples of locations that fit into this cluster are **Penge, Wood Green and Kilburn in London, Orpington, Shirley in Southampton,** and **Blatchington Road in Brighton.** VOCR micro-locations also have the most vacant units, suggesting they struggle to find retailers to fill stores. Another feature of this cluster is a distinctly higher proportion of charity shops. 5.9% of the nearest 25 stores to each sensor in this cluster were charity shops, compared to 1.8% in the CCR cluster and 0.6% in the BI cluster. This is 4.3% greater than the average for England and Wales of 1.6%.

These three classes, referenced by the acronyms CCR, BI and VOCR, are distributed in retail centres across Great Britain. As they are classified by individual micro-locations, one retail centre can be composed of different classes, depending on location. For example, thirteen of Liverpool's sensors are in CCR micro-locations, yet three are in BI micro-locations. Manchester is similar, with fourteen in CCR micro-locations and four in BI micro-locations. On the other hand, Plymouth has six sensors in CCR micro-locations and two in VOCR micro-locations. London, which has significantly more sensors than other cities in the dataset, has 212 sensors which are in BI micro-locations, 56 that are in CCR micro-locations and 23 that are in VOCR micro-locations. A town or city could be comprised of multiple different micro-locations, each with distinct micro-locational characteristics.

4.3.4 Limitations

There are some limitations that must be considered when examining and applying the results of this cluster analysis. Firstly, the sample size of 640 micro-locations for Great Britain is relatively small, biased towards London and the south of England. 52% of sensors are in the Greater London region, which has been shown to exhibit unique footfall patterns when compared to the nation as a whole (Mumford et al., 2021). Further, there are disproportionally fewer sensors in mid-sized centres and smaller centres, particularly in the north of England and Wales. Mid-sized retail centres and northern retail centres have been identified as the worst affected by unfavourable changes in the retail sector (Millington et al., 2015; Wrigley and Dolega, 2011). In addition, the sensors are predominantly located in city centre environments, as opposed to suburban high streets or district centres, which face unique challenges to their future retail vitality and viability (Griffiths et al., 2008). There are also few locations in the sample that experience atypical seasonal variations in footfall, such as through tourism. As such, the classification is skewed towards micro-locations in larger urban areas that tend to be more successful and sustainable retail destinations, potentially with lower vacancy rates and steady footfall and with a steady flow of visitors throughout the year.

Secondly, although this study has grouped each of the micro-locations into three clusters, they may not be as clearly delineated in reality. Cluster analysis is a well-established and widely used, however, its outputs are a representation determined by decisions made by the researcher. Another analysis which different parameters defined could produce alternate and equally valid results (Vickers and Rees, 2007). This quality, which is inherent to cluster analysis, means that these micro-locations are more complex than the cluster descriptions in reality. Therefore, it is reasonable to assume that any footfall signatures derived from them may be the same. Some of the sensors may have somewhat different footfall magnitudes and signatures than the average in their cluster, despite the overall similarity of a particular cluster's functional and morphological characteristics.

Finally, due to the aforementioned bias in the availability of footfall data, there are likely other identifiable micro-locations clusters in the wider country, which this study has not represented. For instance, in Mumford *et al.* (2021), four types of town were identified based on their monthly footfall patterns: comparison, holiday, speciality and multi-functional. It is apparent that our sample is biased towards Mumford et al.'s comparison centres overlooking the different micro-locational patterns that could exist in the remaining clusters, such as seasonal popularity, tourism and non-retail anchors (Mumford et al., 2021; Newing et al., 2018).

Although this classification has some restrictions as a result of data limitations, it is important to note its value. Sensor data, such as the footfall data used here, can provide quantitative insights into the relationship between footfall and the surrounding environment that are novel, in addition to reinforcing the qualitative and observational research that already exists. Even if this is currently only possible for a limited amount or type of retail centre, valuable insights can still be provided.

4.4 Classes of footfall micro-locational context – results

The last section introduced three classes of footfall micro-locational context: Chain and Comparison Retail micro-locations [CCR], Business and Independent micro-locations [BI] and Value-Orientated Convenience Retail micro-locations [VOCR]. This section will investigate the differences in footfall magnitude and signature between these three classes and relate these patterns to a micro-locational context, fulfilling this chapter's final objective.

Footfall measurements are often used as a proxy for retail centre vitality (Coca-Stefaniak, 2013; Millington et al., 2015); however, there is limited research quantifying how functional and morphological factors impact footfall magnitude and signature. By investigating the footfall patterns exhibited by these clusters built on functional and morphological characteristics, a greater understanding of variations in footfall magnitude and signature can be achieved.

Section 4.4.1 describe how the representative footfall signatures and magnitudes were derived for each class. Then, Section 4.4.2 identifies the key differences between the footfall measurements for each class, and Section 4.4.3 explores possible explanations for these differences and how they link to individual micro-locational characteristics.

4.4.1 Deriving representative footfall signatures and magnitudes

Footfall measurements from January 2017 until August 2018 were averaged across the locations in each cluster to investigate whether the different functions and characteristics of the micro-location impact footfall. Only the sensors with footfall data for 75% of an entire year were used to remove any bias from new or temporary sensors, which only have footfall data for potentially busier or quieter times of the year. This removed 12 sensors from the sample. Descriptive statistics were calculated for the average week (by the hour) and average weekday (by 5 minutes) for each cluster as shown in Table 4-4, Figure 4-5 and Figure 4-6 in the following section.

Table 4-4 Summary statistics for footfall (people per 5 minutes) across the three classes

Statistic		CCR micro-locations	BI micro-locations	VOCR micro-locations
Maximum:	Mon	94 @ 12:05	106 @ 17:10	55 @ 16:15
	Tues	95 @ 12:05	117 @ 17:10	61 @ 16:20
	Wed	96 @ 12:05	**121 @ 17:10**	61 @ 17:10
	Thurs	95 @ 12:05	119 @ 17:10	**62 @ 16:20**
	Fri	98 @ 12:05	113 @ 17:10	57 @ 16:20
	Sat	**116 @ 13:05**	92 @ 13:05	60 @ 13:25
	Sun	86 @ 13:05	71 @ 14:05	47 @12:05
Weekly Mean		37	49	27
Standard Deviation		32	31	19

4.4.2 Identifying trends in footfall signature and magnitude

Figure 4-5 shows that the BI micro-locations have higher footfall early in the morning on weekdays than CCR micro-locations. Although by 10:00, the CCR micro-locations are just as busy, and both rise in footfall until 12:05. This maximum weekday peak is consistent at 94-101 people per 5 minutes for CCR and BI micro-locations. Footfall in CCR micro-locations then decreases into the afternoon and evening, whereas footfall in BI micro-locations experiences a 14:00 lull before peaking again into the early evening. This is reflected through the consistent 17:10 maximum footfall values for BI micro-locations of 106-121 people per 5 minutes, shown in Table 2. This cluster is the busiest during the evening, keeping over 25 people per 5 minutes until past 22:00 and never dropping below five people per 5 minutes. BI micro-locations have a distinctive weekday footfall pattern consisting of three peaks at 8:00, 12:00 and 17:00.

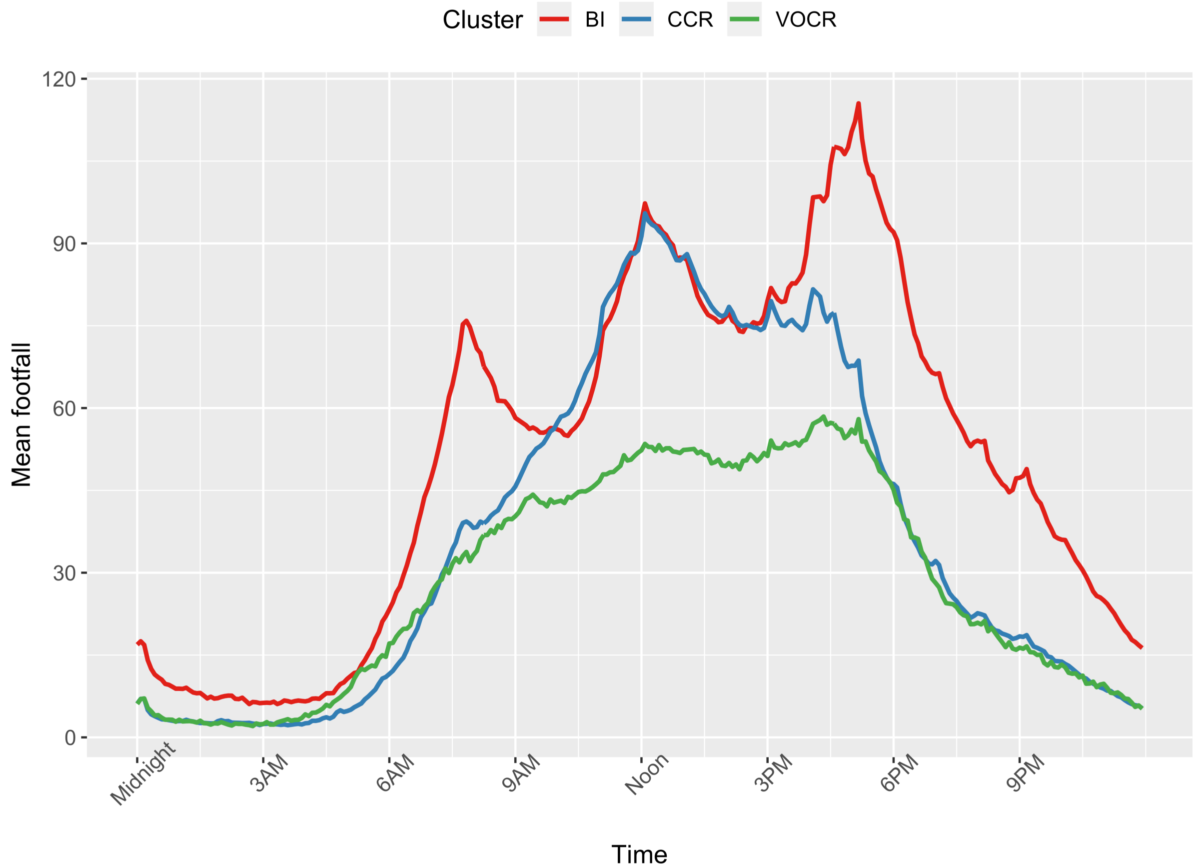


Figure 4-5 Average footfall distribution for each cluster for a weekday (Mon–Fri) to five-minute accuracy

The VOCR micro-locations have the lowest average footfall of all the clusters, and they are never the busiest. Their maximum value is 62 people per 5 minutes, which is just over half the size of the maximum values for the other clusters. The footfall signature of VOCR micro-locations is hump-shaped, slowly increasing from 5:00 to 16:15 – 17:10, where it peaks on weekdays. After then, footfall decreases exponentially to under ten people per 5 minutes by 22:30.

As visible in Figure 4-6, CCR micro-locations are significantly busier on Saturdays than the weekdays, with their maximum footfall of 116 people per 5 minutes at 13:05 that day. Although CCR micro-locations have the highest peak, BI micro-locations have the highest consistency, with a mean footfall of 49 people per 5 minutes, compared to 37 people per 5 minutes. However, VOCR micro-locations have the lowest standard deviation, showing that, although their average footfall is low, it is the most consistent throughout the day and the week.

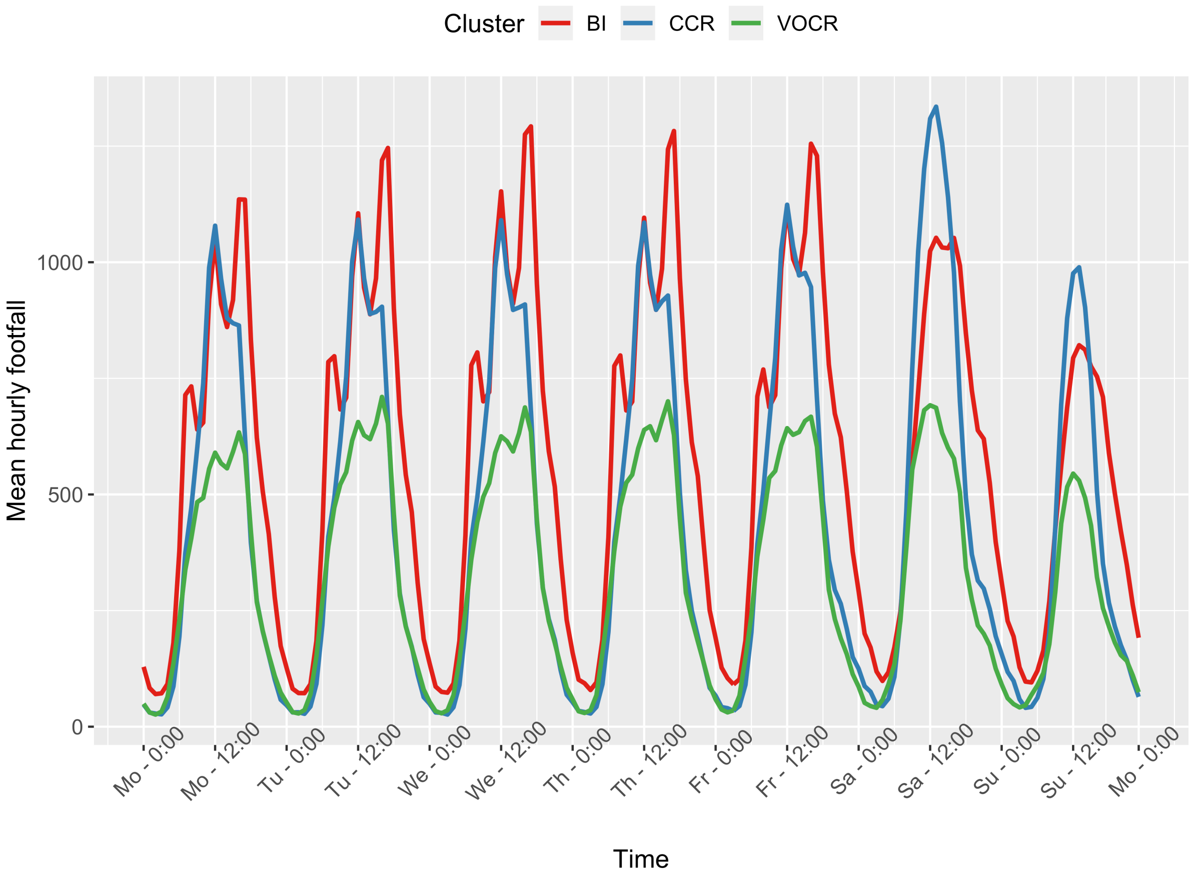


Figure 4-6 Average footfall across a week for each cluster to hourly accuracy

VOCR micro-locations have very similar footfall signatures during the weekend as the weekday. In contrast, BI micro-locations have very different footfall signatures. They have lower footfall at weekends, peaking at 92 people per 5 minutes at 13:05 and do not exhibit the three-peak structure previously observed. Instead, they show a peak in the early afternoon with a slow drop into the evening when they are the only cluster to retain significant footfall into the night. Friday and Saturday nights appear to be the busiest nights, staying at above 25 people per 5 minutes until after 00:00. In contrast, the other clusters have dropped below this threshold by 21:00. Sunday is the quietest day for every cluster, even the most consistent VOCR micro-locations exhibit a smaller peak on this day.

4.4.3 Discussion of findings

This study has produced three distinct clusters of retail micro-locations that vary in terms of their function and morphology: chain and comparison retail micro-locations [CCR], business and independent micro-locations [BI], and value-orientated convenience retail micro-locations [VOCR]. When these clusters' average weekly and daily footfall patterns were investigated, distinct patterns in signature and magnitude were evident. These differences in footfall signature and magnitude can be partially explained by various characteristics of the retail micro-location, essentially their form and function.

Firstly, the CCR micro-locations exhibited a footfall pattern with the busiest times on Saturdays and during daytime hours from late morning to early afternoon. This reflects this cluster's prominent comparison retail function indicated by its low service to retail ratio and the clustering's low proportion of independent stores. For the majority of people, Saturday is a day of leisure when they have ample free time. Comparison retail tends to be recreational and time consumptive (Guy, 1998), therefore supporting the link between this function and significant Saturday and daytime footfall. In addition, this cluster has the highest average density of retail units showing that the retail offer is more condensed in these micro-locations, therefore, increasing the overall footfall magnitude. Besides, a condensed retail offer can encourage linked trips, where consumers visit different locations on the same trip (Wrigley et al., 2009).

In comparison, the BI micro-locations have dominant weekday footfall with three peaks at 8:00, 12:00 and 17:00. This footfall pattern reflects commuting into and out of work, with an additional increase in footfall during a lunchtime break, similar to that observed in other studies (Berry et al., 2016; Lugomer and Longley, 2018). This is further supported by the large workplace population of the cluster and close proximity to transport hubs with many of the sensors located in central London - a destination for many public transport commuters (Lyons and Chatterjee, 2008). The absence of this pattern during the weekend confirms this interpretation and shows the extent to which the working population determines footfall in these locations. Furthermore, BI micro-locations retain footfall later into the evening than the other clusters. With a higher than average number of restaurants and bars, these micro-locations could also be viewed as attractive leisure and night-time economy destinations (Ravenscroft et al., 2000). However, the amount of footfall in the late evening is significantly less than during the day, demonstrating that, on average, this night-time economy function is supplementary to the workplace function.

The VOCR micro-locations are the quietest and steadiest in terms of footfall. Their convenience-based function could explain this constant and consistent flow of people as convenience retail is characterised by short and frequent trips (Guy, 1998). The VOCR micro-locations tend to be in residential areas that serve a local demand with a smaller catchment size, generating less footfall. The smaller magnitude of the footfall of these micro-locations could also be associated with larger distances to many footfall attractors such as anchor stores, transport hubs and entertainment activities.

However, not all of these footfall patterns can be explained by micro-location features. For example, in every cluster, Sundays saw 26-32% less footfall compared to the other days of the week, which can be explained by the reduced to 6 hours opening hours on this day for stores larger than 280 square metres, imposed by the 1994 Sunday Trading Act (GOV.UK, 2021). Research shows that these large stores can be key footfall attractors, and having these stores reduce their opening hours may deter people from visiting their high street on Sundays (Williamson et al., 2006).

These results help build a clear understanding of how and why footfall fluctuates throughout the day and week and better understand its relationship with micro-location characteristics. In general, these results show that footfall and, as an extension of that, retail vitality varies temporally and spatially on a micro-locational scale as a result of multiple external and internal influences. More specifically, this study shows some key drivers of footfall at a micro-location level: anchor stores, workplace population, the density of retail units and distance to transport hubs. However, it would be incorrect to assume that all retailers within a particular retail centre benefit equally from the increased footfall in terms of spend, as that depends on many other factors on a micro-location level (Millington et al., 2015). This supports strategies to increase high street vitality, which are holistic and consider this complexity of micro-locational factors within the wider retail centre. Footfall is often used as an indicator of high street vitality. Therefore, a better understanding of it, underpinned by reliable data and robust empirical analysis, is vital for business, academia and policymakers.

4.4.4 Implications of findings

The results of this chapter pertaining to variation in footfall magnitude, signature, and the function and form of a particular retail micro-location have many implications for various stakeholders. Firstly, it supports revitalisation and town centre strategies that consider the complexity of micro-locational influences within a retail centre. This chapter has shown the importance of these factors in determining footfall and retail centre vitality. This finding is particularly relevant as footfall is widely used to measure retail centre performance. Therefore, a clearer understanding of how and why it fluctuates would be beneficial. Understanding these factors can be valuable for retailers and planners in managing pedestrian flows, setting effective opening hours and investing in ideas that would be attractive to their target consumer. For example, BI micro-locations have a more significant daytime footfall than evening footfall, despite its night-time economy. This knowledge could be used to develop schemes to increase the dwell time of the daytime population and encourage them to support the night-time economy establishments, increasing the retail resilience of the area.

Secondly, these results have demonstrated the potential of using morphological and functional characteristics to predict footfall in areas with no sensors. Although these clusters are generalisations of micro-locations, they draw out patterns between specific characteristics and spatial and temporal footfall variations. With technological advancements increasing the wealth of data on urban characteristics and mobilities and the development of algorithms capable of processing this data, there is potential for these patterns to predict footfall for all retail areas. Through understanding the patterns and factors that influence footfall and quantifying their impact, it is possible to use that to predict footfall counts. This would be a valuable tool for benchmarking and location planning, managing pedestrian flows and business logistics such as opening hours and staffing and is the focus of the analysis in Chapter 6.

Thirdly, this work has contributed to a more comprehensive understanding of retail mobilities. Although many footfall determinants have been identified in the literature, how they impact footfall temporally is not always investigated or quantitatively shown. This chapter has demonstrated how different micro-locational characteristics impact footfall to 5-minute intervals throughout an average week, which provides new insight into footfall determinants and urban mobility as a whole.

4.5 Chapter Summary

Three objectives were established at the start of this chapter. These are summarised below:

The world around us	Investigate how different characteristics and contexts of the immediate environment impact footfall magnitude and signature
	Using characteristics of retail and footfall context, develop a classification that captures these main differences
	Identify how trends in footfall magnitude and signature differ between these different retail contexts

Section 4.1 and Section 4.2 investigated how a range of functional and morphological characteristics such as retail mix, vacancy and connectivity correlate

with footfall. They showed that patterns in the magnitude and signature of footfall data, and by extension retail vitality, can be explained by functional and morphological characteristics of the micro-location. In particular they have shown that crucial footfall attractors such as anchor stores and transport hubs can significantly drive footfall at certain times throughout the day and week.

K-means cluster analysis was used in Section 4.3 to define a classification on retail contexts. The results displayed three clear narratives of micro-location morphology, and function: Chain and Comparison Retail micro-locations, Business and Independent micro-locations and Value-Orientated Convenience Retail micro-locations.

Finally, Section 4.4 applied these classifications to gain a greater understanding of the interrelationship and patterns between retail context and footfall magnitude and signature. Three distinctive footfall signatures were defined for each of the three classes, demonstrating how the type of retail offer (comparison, convenience or recreational) and contextual factors can impact the magnitude and signature of footfall within the micro-location. In addition, it has shown how the relationship between micro-locational characteristics and footfall can be dependent on time and place.

This chapter presents the potential for functional and morphological characteristics of micro-locations as a predictor for footfall in locations where footfall is not measured. Future research will benefit from employing more footfall data to investigate monthly, annual and longer-term trends in footfall and how those could relate to functional and morphological characteristics. Modelling footfall for an entire retail centre could be invaluable for decision-making, urban planning and retail location planning.

5 What happens there

Exploring event-related temporary fluctuations in footfall magnitude and their relation to micro- and meso-scale characteristics

What happens there	
	Identify events which significantly impact footfall.
	Investigate how factors of both the immediate environment and in the wider context could influence this impact
	Explore the trends and similarites between footfall of different events in different locations and what they could imply about retail footfall

What happens there

Festivals and events can be vital for local footfall and the sustainability, viability and vitality of a retail centre itself. Event-based footfall surges, particularly festive footfall, are colloquially considered vital for economic and retail sustainability. In particular, the Christmas season and Black Friday are considered vital for the profit margins of retailers (British Retail Consortium, 2021c). In addition, the success of local festivals can be important for place-identity and resilience (Parker and Welsh Government, 2015).

In addition, the COVID-19 pandemic has starkly demonstrated how much retail footfall can depend on the wider context surrounding it. Consumers were forced towards online retail more than ever before, with many non-essential physical retail stores temporarily or permanently closing. The UK economy has shrunk, and consumers and businesses may need new incentives to return to their local high street.

Analysing retail footfall and events in the years preceding the pandemic could give key insights into the effectiveness of events for attracting footfall and assist in revitalising UK retail centres when they can safely reopen again. For businesses to grow their customer base, they rely on the potential of increased footfall. Therefore, it is valuable to have an insight into how the external influence of events may increase their chances. For business improvement districts or local planners, the local events that have proved successful in driving footfall in other locations in the past may help decide what nature of events to invest in in the future. Moreover, footfall is used as a key indicator of high street performance, and high street recovery, by multiple stakeholders. Understanding the extent to which national, regional and local events can influence it helps to form a fuller picture of what comprises footfall.

Events impact footfall – this is a well-documented occurrence. Local events and festivals are shown to drive footfall, and pedestrian counts or a relative increase in footfall are often used to gauge the impact of an event. For example, Naylor *et al.* (2016) estimated the Edinburgh festivals to attract more than 4.5 million people annually. Events do not have to be local to drive local footfall. There is an expectation for national events such as Black Friday and the festive season to draw people in to their local retail centre (BBC News, 2018b; Collinson, 2018; Sky News, 2018). Conversely, weather events such as Storm Deirdre in December 2018 are shown to have a negative impact are used to explain and account for drops in footfall and sales (BBC News, 2018a; Sky News, 2018). There are many temporary micro, meso and macro-scale influences that can be intrinsically linked to fluctuations in retail footfall magnitude from shopping events to sporting events, weather events, and festivals.

Events are known drivers of retail footfall and they're known to be important to the vitality and viability of a retail centre, however, there is a lack of research that explores this to any depth. Some Business Improvement Districts publish periodic footfall reports allowing stakeholders and the public insights into local and event-based trends. However, these are context-dependent, typically focusing on one location, such as Manchester or Cambridge BIDs (Cambridge BID, 2021; CityCo, 2021), with measures calculated by comparing the increase or decrease in footfall dependent on past values from the same location. These patterns aren't compared to other high streets in their region or nationally. Not much is known or researched about its characteristics of event-based footfall, for example the size of its impact, or how this can fluctuate from place to place and why.

This research gap could be justified by a lack of consistent or comparable data in the past. Manual footfall counts could be conducted during an event; however, a control count would be needed to ascertain its impact. Researching unpredictable events such as extreme weather events would prove difficult. However, advances in sensor technology have made it possible to gather frequent and consistent footfall measurements through permanently installed sensors. Using this data, we can investigate how national shopping and weather events impact different cities across the UK and examine the local events that may cause footfall to deviate from the norm.

Previous research utilising sensor collected footfall data removes temporary fluctuations through aggregation. This technique was used in the last chapter to investigate how micro-locational characteristics impact footfall magnitude and signature. It is also applied by Mumford *et al.* (2021) and Lugomer and Longley (2018) to create their respective footfall classifications. For the intended purpose to develop representative and applicable classifications, the removal of events and fluctuations in footfall is justified in these cases. However, these events are not insignificant. Which cities benefit from events, and which do not, and do certain micro-locations within a city experience temporary fluctuations in footfall differently from others and why? Do different events have a different impact? These are questions that will be investigated in this chapter.

This chapter aims to investigate the characteristics of event-related temporary fluctuations, including extreme weather events, shopping events such as Black Friday, local festivals and national holidays. It will explore their significance over two consecutive years through four case study micro-locations which were chosen to represent different meso- and micro-scale contexts.

Three objectives were established for this chapter. The first objective was to identify events which have a significant impact on footfall. This will be done through case study analysis, comparing the data from four different micro-locations to understand which events in 2017 and 2018 had the largest impact on footfall. Section 5.1 will introduce and describe these case study locations, justifying their selection and representability. Section 5.2 will document the method used to minimise error within the data and ensure the locations and times are comparable. A annual daily footfall ranking is derived for each case study location and this is used to identify the types of events that can drive temporary fluctuations in footfall.

The second objective was to investigate these fluctuations in the context of the case-study environment. In Section 5.3, the case study locations will be compared against each other and the micro- ,meso- and macro-scale factors that could compound with events, or mediate their effect, are identified.

The final objective was to explore these trends and similarities that occur in relation to events and discern their implication on retail footfall and its behaviour. Section 5.4 will discuss the observations made from the results, and the limitations of the study. The final section, Section 5.5, will summarise the chapter.

5.1 Deriving the case study micro-locations

In Chapter 2, a range of macro-, meso- and micro-scale characteristics were identified from the literature as impacting footfall. A summary of these can be found in Figure 4-1 in the previous chapter. The previous chapter focused primarily on micro-locational characteristics and exploring how the immediate environment surrounding a sensor could impact footfall. This chapter will examine factors at all scales, focusing on their impact on footfall during temporary events. The hypothesis is that every retail centre will be affected differently by National events, such as Black Friday or Christmas, Regional events, such as storms or hot weather, and Local events, such as festivals or sports games. Also, this difference could be explained by one of these factors.

Whereas the previous chapter used a top-down method, generalising retail centres into several classes based on their micro-locational characteristics, this chapter applies a bottom-up technique, focusing primarily on several key case study areas. This approach was chosen because some events are inherently contextual, such as local festivals or celebrations. It would be challenging to determine when these events occur for many locations and to explore their impact on footfall. The case studies chosen are representative of different micro-, meso- and macro-scale contexts. In this section, the case study locations will be derived and introduced. These are,

- ◊ Liverpool ONE in Liverpool
- ◊ Market Street in Manchester
- ◊ Old Town in Edinburgh, and
- ◊ New Town in Edinburgh.

The footfall data for those micro-locations must meet specific standards of quality. Section 5.1.1 will define how these data quality issues limited the selection of case study locations. Then, Section 5.1.2 will focus on Liverpool ONE in Liverpool and Market Street in Manchester. It briefly introduces the background and context of these cities before exploring their Intra-Regional and Micro-locational factors. These factors include population, distance and connectivity to other retail centres and the function and morphology of the micro-location. Section 5.1.3 repeats this process for Edinburgh New Town and Old Town, giving some background to the city and exploring contextual factors that could drive footfall. Finally, Section 5.1.4 evaluates these case study locations in a national context, assessing how representative they are of the UK.

5.1.1 Data quality considerations

As of 2020, there have been 1,212 sensors across 115 cities and towns in the UK. However, their distribution is not spatially or temporally even. Most are concentrated in London (403), and locations such as Edinburgh, Manchester and Leeds are also well-represented. The sensors are installed ad hoc, depending on the client. It is not uncommon for footfall sensors to be placed close to each other or for micro-locations within an otherwise well-represented city not to be captured.

In addition, although the SmartStreetSensor project has been active and collecting data since 2015, most sensors in their network were activated in late 2016 and early 2017. There is good coverage of footfall data during 2017 and 2018; however, the sensors began to be deactivated in 2019.

The distribution of the sensor network is explored in more depth in Chapter 3, and it was shown to be non-uniform. Some locations and times are well-represented, and others are not. This variation is a huge consideration when picking case study locations. To accurately and reliably understand the variations in retail footfall for one particular case study area, the data for that location needs to be:

- Complete: includes all 365 days of the year.
- Representative: has the same micro-locational context, such as the composition of surrounding units and the connectivity and morphology of the street.

Hypothetically, this could be achieved by a singular footfall sensor with consistent and complete data for any year. However, due to the data relying on appropriate installation, constant running and external influences, many sensors have noisy or missing data. Even for sensors that have complete data and appear accurate, there is no means to verify this. Therefore, the data within a case study area also needs to be:

- Verifiable: multiple sensors within the same micro-location allow the data to be cross-referenced.

For each case study location, there needs to be good spatial coverage (three or more sensors close to each other) and good temporal coverage (those sensors cover at least one year, with always two or more sensors running). Therefore, the first step in deriving case study locations is to narrow down the possible places that fulfil these criteria.

Temporal coverage

Figure 5-1 shows the temporal coverage for each sensor in the four case study micro-locations. They all have sufficient coverage for two whole years (2017 and 2018), ensuring that at least two sensors are collecting data for each month. Edinburgh New Town has the best coverage, with seven sensors within 300m of each other and running fairly consistently for the 24 months. Manchester Market Street and Edinburgh Old Town have the patchiest coverage, with only four sensors. Manchester Market Street has five months where only two sensors are running. Edinburgh Old Town has four. Nonetheless, these are micro-locations with some of the best coverage in the dataset. There is enough data to make solid inferences about the behaviour of footfall in these locations for both 2017 and 2018.

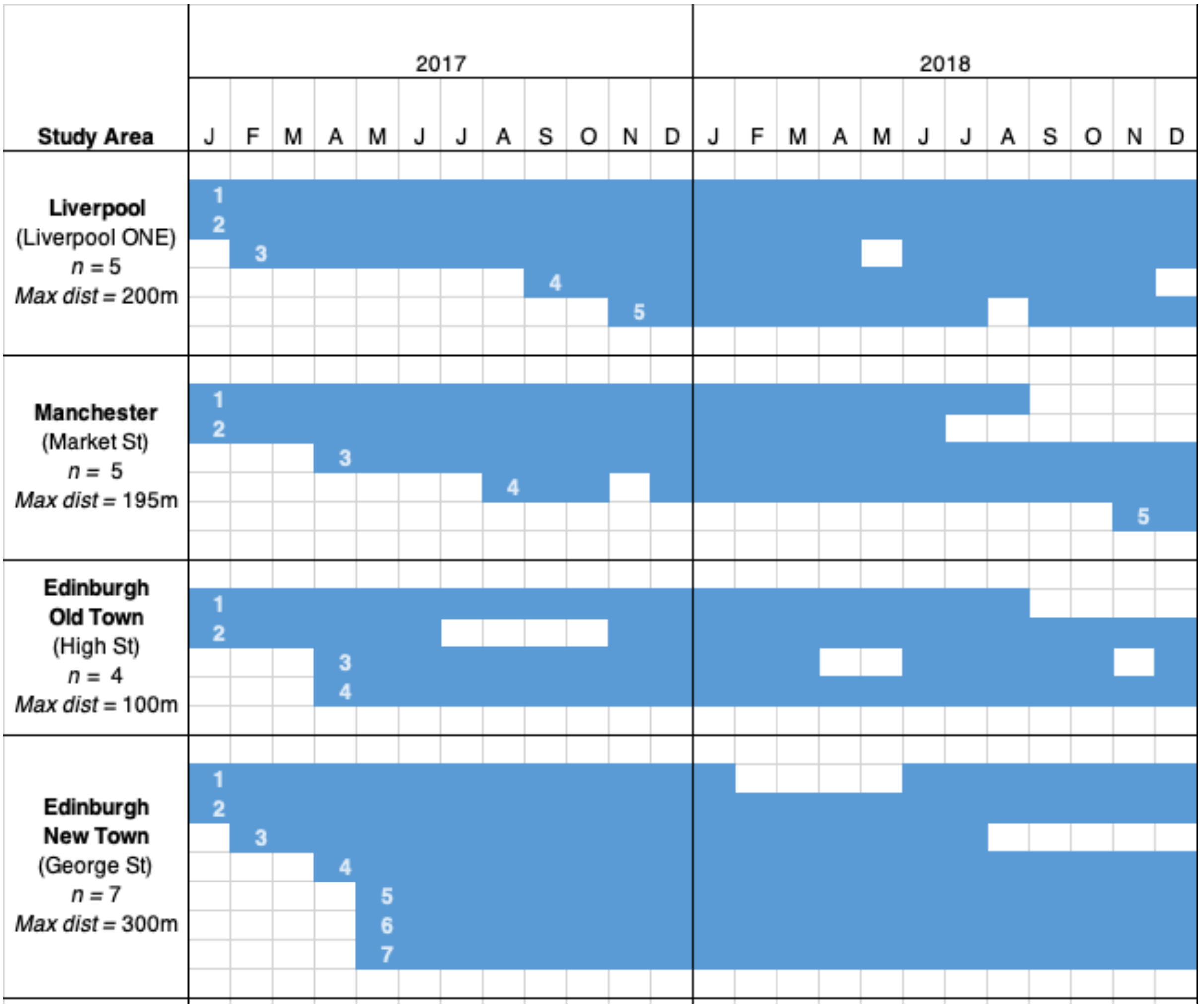


Figure 5-1 Temporal coverage of the sensors for the case study micro-locations

Spatial coverage

The selection of each representative sensor was based on the assumption that their proximity to other sensors and placement along the same throughway would result in a similar ranking of daily footfall magnitude, relative to each other. Every sensor in that micro-location would report the same busy days, regardless of the absolute count, which could be variable.

To minimise the error introduced by this assumption, the sensors selected were along the same street or close to the intersection of that primary street. Between 4 and 7 sensors represent each micro-location and these were all close in proximity – between 100 and 350 metres of each other. The distribution of sensors in the four micro-locations are given in Figure 5-2, Figure 5-3, Figure 5-4 and Figure 5-5.

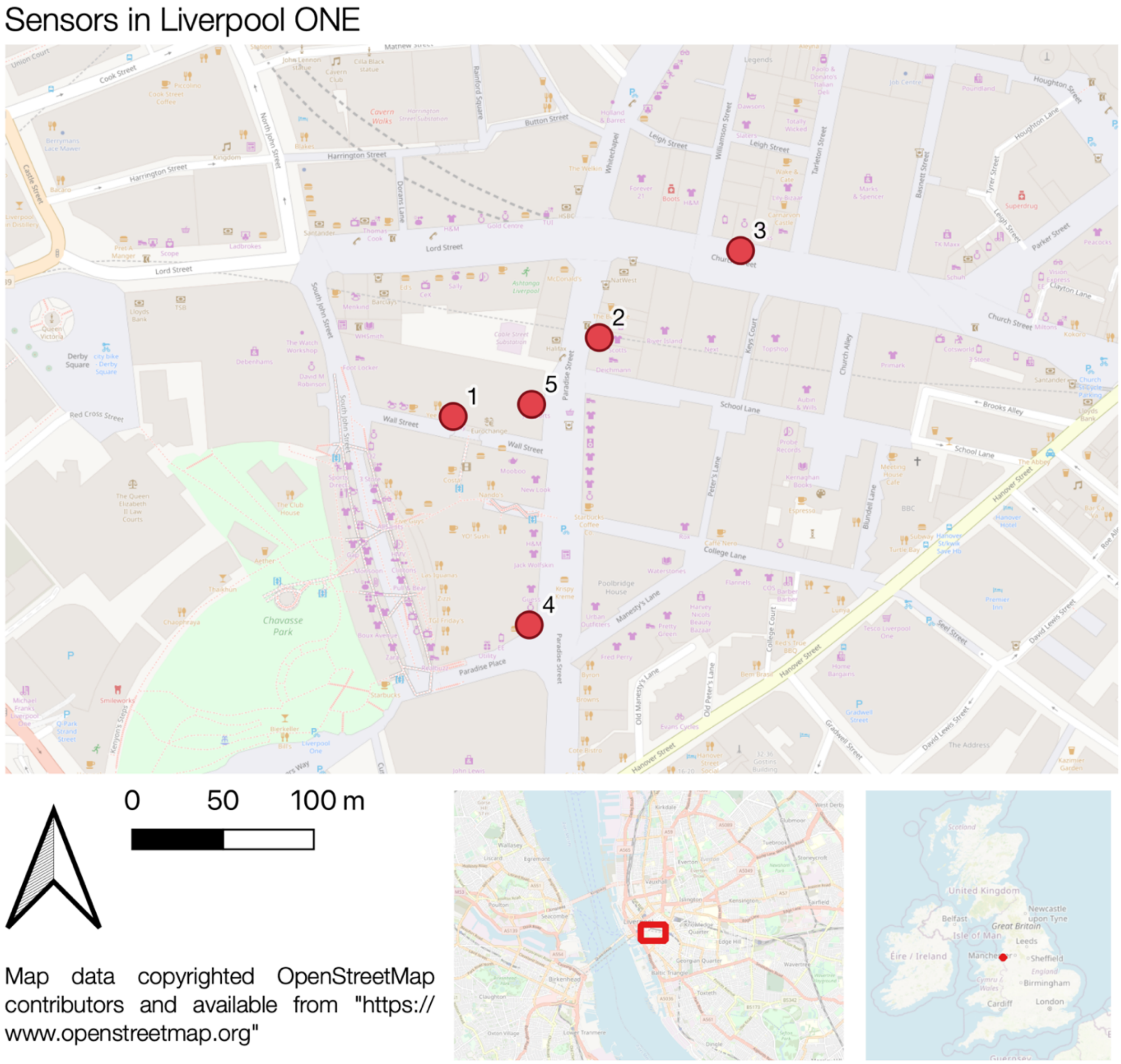


Figure 5-2 Sensors used in Liverpool ONE micro-location

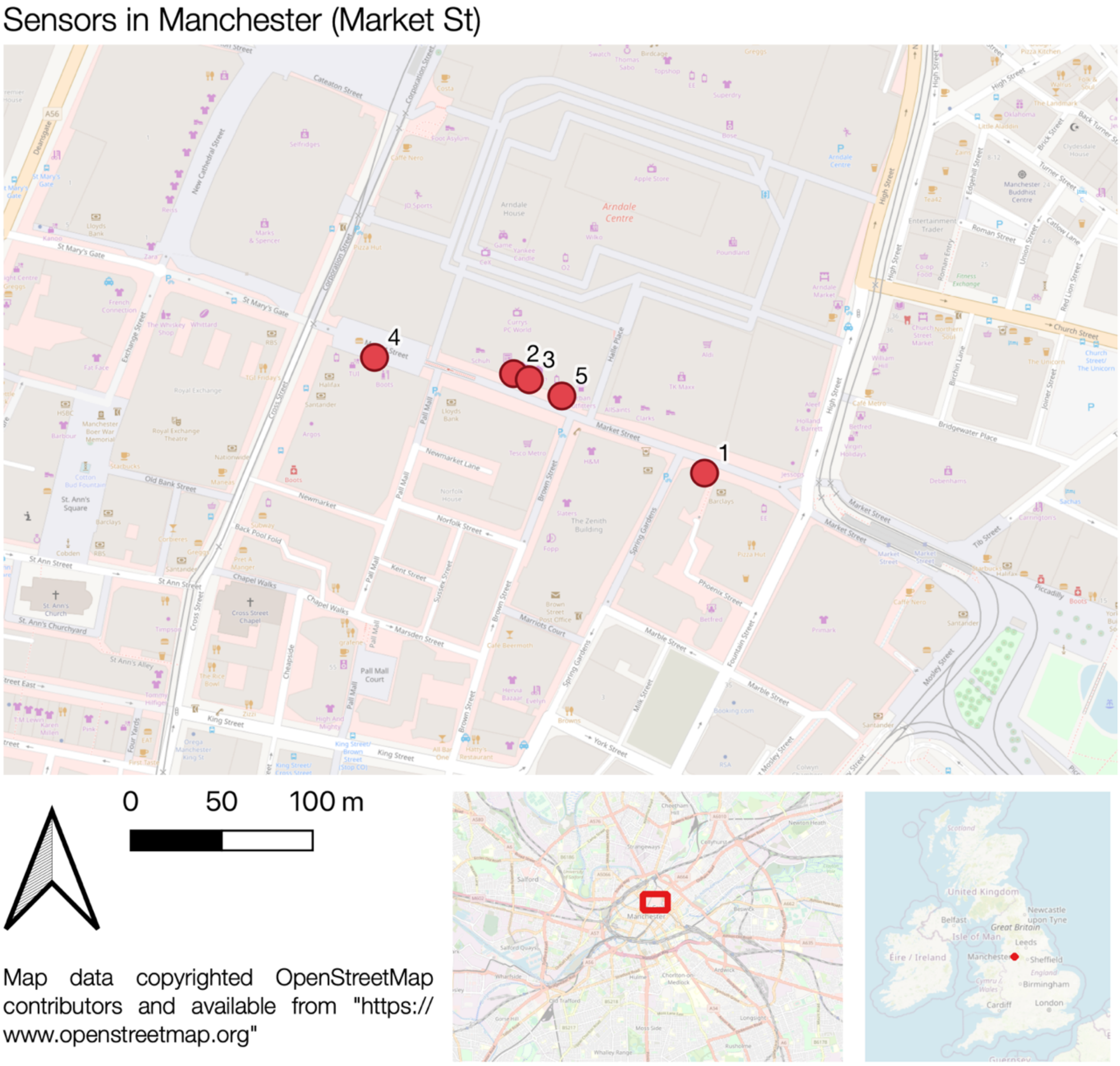


Figure 5-3 Sensors used in Manchester Market Street micro-location

Figure 5-4 Sensors used in Edinburgh Old Town micro-location

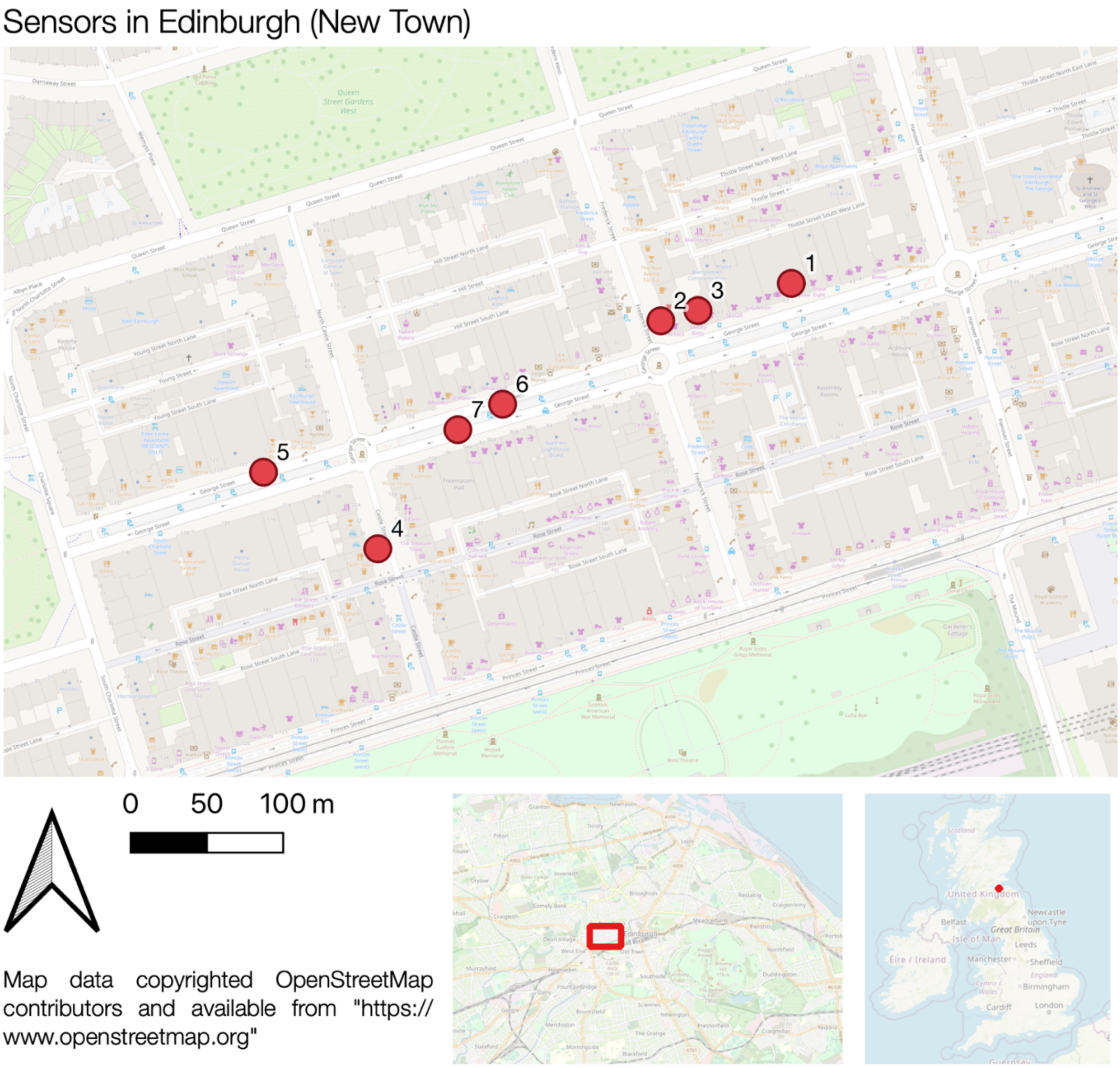


Figure 5-5 Sensors used in Edinburgh New Town micro-location

However, even though these sensors are close together, they do not have identical micro-locational contexts. For example, Edinburgh New Town and Liverpool ONE sensors are on separate roads. Even though these roads are connected, one road could be a throughway and the other a less popular side street. In addition, there are intersections between some of the sensors, which could interrupt a consistent pedestrian flow. Nonetheless, the sensors generally represent the same retail offer and function of a location, and they all contain valuable footfall data that is necessary for temporal completeness. Therefore, their inclusion is vital.

At the beginning of this sub-section, three priorities were identified for the sensors chosen to represent a micro-location. They had to be

- ◊ Complete (include all 365 days of the year)
- ◊ Representative (have the same micro-locational context) and
- ◊ Verifiable (include multiple sensors within the same micro-location allowing the data to be cross-referenced)

The temporal completeness, representativeness and verifiability were all validated for the four identified case studies. Liverpool ONE, Manchester Market St and Edinburgh Old Town and New Town all fulfilled the necessary level of data quality for this analysis. These case studies were handpicked based on their display of a range of intra- and inter-urban characteristics. The remainder of this section will present each of the locations, exploring the similarities and differences in their background, micro-locational and intra-regional characteristics and footfall.

5.1.2 Background – Liverpool ONE and Manchester Market Street

Liverpool city region is in North West England and has a population of 1.5 million people (ONS, 2018). In 2017, its retail spend potential was estimated to be £3.1bn (Harper Dennis Hobbs, 2017) – 6th in Britain – and it attracts both national and international tourism. In comparison, Greater Manchester, 30 miles east, has a population of 2.7 million people (ONS, 2019) and a retail spend potential of £3.5bn (Harper Dennis Hobbs, 2017), 4th in Britain. It also attracts national and international tourism, with both cities being well-connected by rail and home to an international airport.

Manchester and Liverpool have a history of decline in the 1980s, regeneration driven by popular culture and arts in the 1990s, and rapid redevelopment driven by investment in the 2000s (Haslam, 2004). Liverpool ONE, a £950m redevelopment of Liverpool's retail core, opened in 2008, while the extended and rebuilt Manchester Arndale reopened in 2006. The speed of redevelopment has left cities divided, with concentrated areas of wealth and investment next to pockets of deprivation (Daramola-Martin, 2009; Dolega and Lord, 2020) and with a varied yet similar retail offer.

Meso-scale characteristics

Figure 5-6 shows the retail centres which surround Liverpool and Manchester.

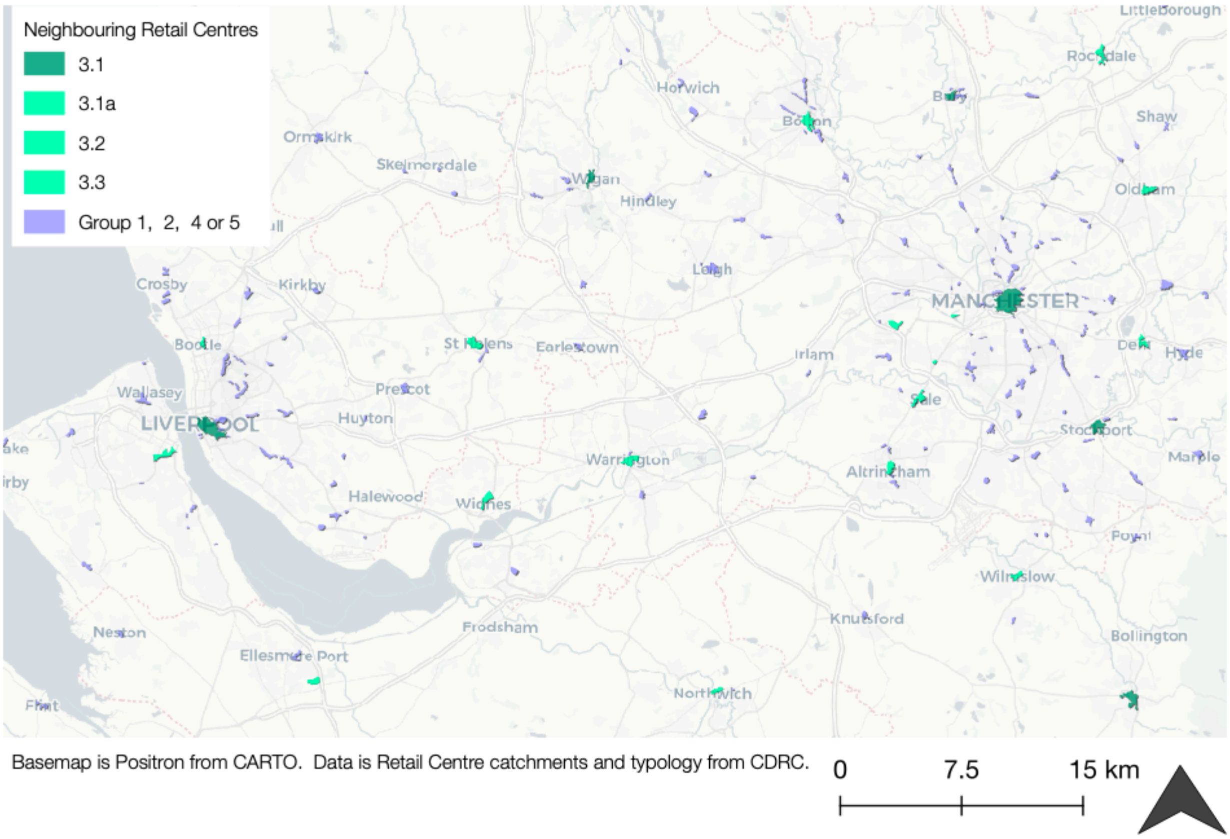


Figure 5-6 Map showing the surrounding retail centres of Liverpool and Manchester

The areas of Liverpool and Manchester that include the micro-locational case studies are both Subgroup 3.1 Premium shopping & leisure destinations of (semi) regional importance, as defined in the Retail Centre Typology introduced in Chapter 3. These are top regional, sub-regional destinations and affluent market towns that have a diverse and comprehensive offer of retail, services and leisure with many top national chains. Nearby, Wigan, Bury and Stockport offer similar Subgroup 3.1 locations; however, these are much smaller than those in Liverpool and Manchester. There are more retail centres that are also in Group 3 Leading Comparison & Leisure Destinations, including St Helens, Warrington, Bolton and Altrincham. In addition, Manchester has the Trafford Centre – a retail park outside the city which is in direct competition to the city centre.

Figure 5-7 shows the population distribution across Liverpool and Manchester. Most people are concentrated around these two cities. The geography of Liverpool limits population growth as it is bound by the coast, whereas areas of high population can surround Manchester. This also limits the impact of the competing retail centres. Although Liverpool and particularly Manchester have retail centres nearby that have a similar offer, the population density is such that it is well-served by its catchment.

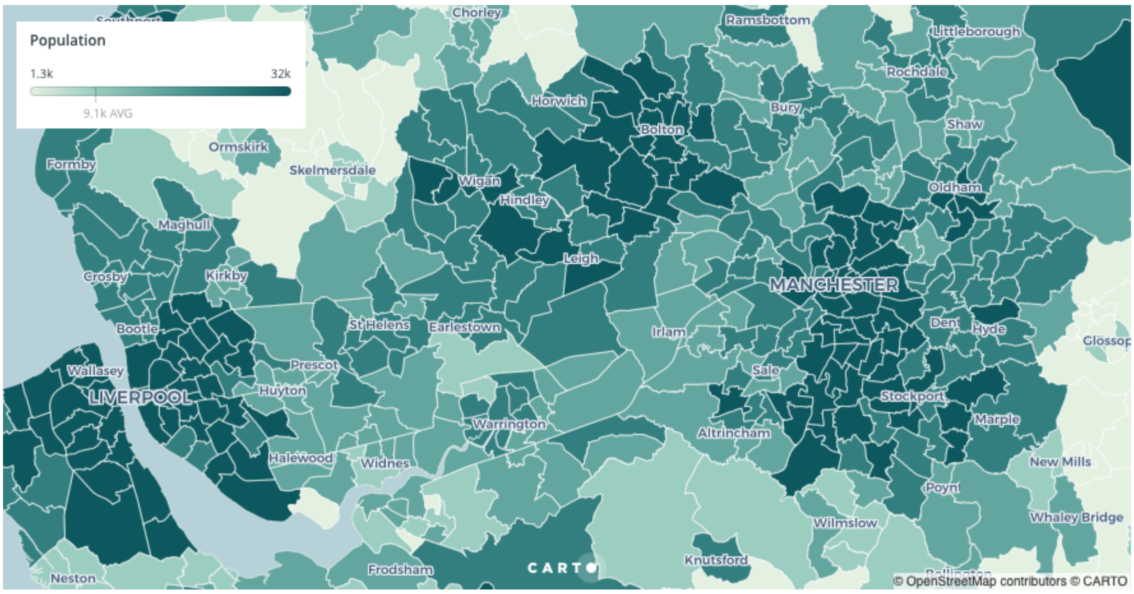


Figure 5-7 Map showing the surrounding population of Liverpool and Manchester

Data from the National Infrastructure Commission (2019) shows that Liverpool is more connected to other cities by public transport. Yet, Manchester is better connected by road when population and congestion are considered. Due to Manchester's location, roads pass through or by the city, whereas Liverpool is a terminus and more often, people will drive there with Liverpool as the destination in mind. Both Manchester and Liverpool have airports that connect them internationally. Both cities have multiple attractions, including shopping and retail, universities, home grounds of major football teams and places of cultural and historical significance. Nonetheless, Manchester is better connected to the wider country that Liverpool, with the exception of north Wales.

Liverpool is a smaller city in terms of population, and its coastal location limits population growth and connectivity to other cities, particularly by road. In contrast, Manchester is a larger city with more retail centres nearby that serve the periphery of its population-dense areas.

Micro-locational characteristics

Figure 5-8 shows the features that surround the Liverpool ONE micro-location. The sensors surround the shops, cinema and entertainment complex of Liverpool ONE, although none of them captures the footfall inside the covered shopping area. Stores on Paradise Street, where most of the sensors are located, include JD Sports, NatWest and New Look. Many chain retailers and popular services have large stores in Liverpool, capable of handling many customers. Paradise Street and Church Street are wide streets that are pedestrianised, and the principal train station is roughly 0.5 miles to the north.

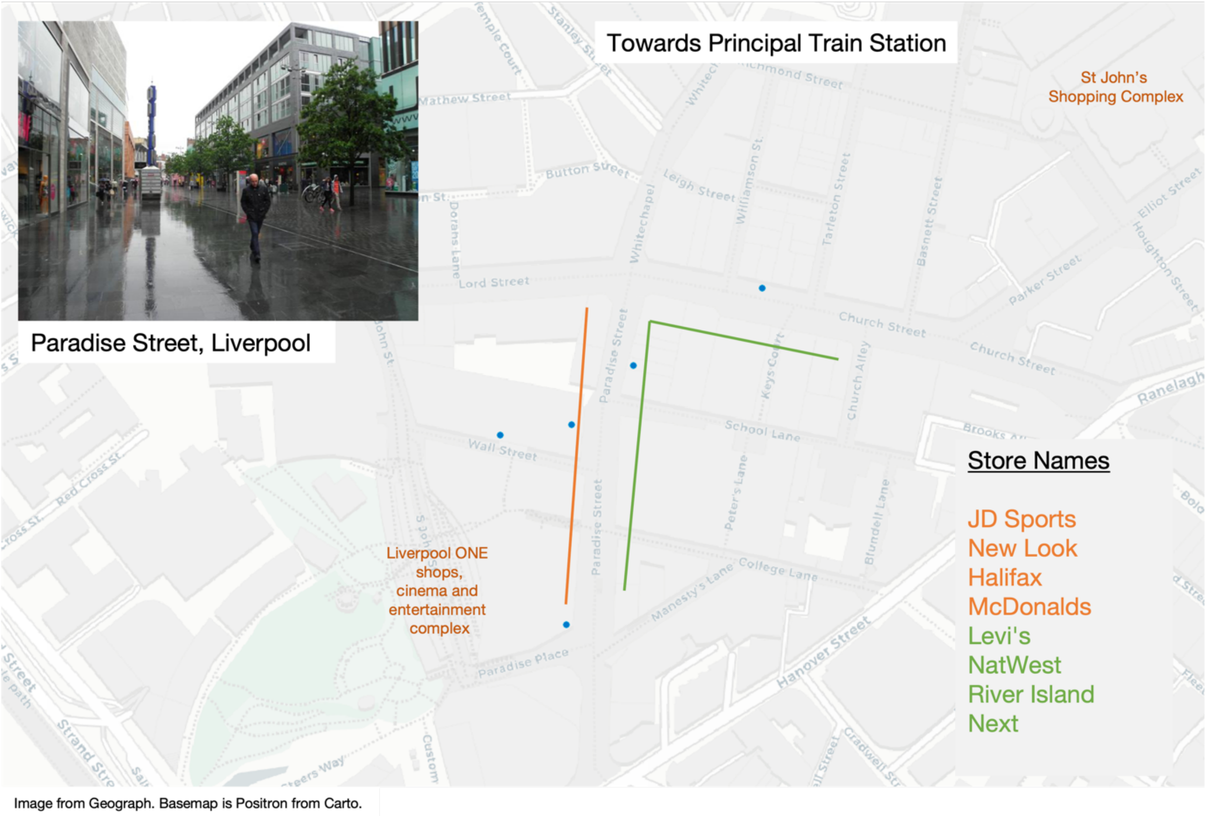


Figure 5-8 Map showing Liverpool ONE and the surrounding area

The features surrounding the Manchester Market Street micro-location are relatively similar to those in Liverpool. Figure 5-9 shows that the sensors along Market Street are close in proximity to the entrance to the Manchester Arndale, a shopping and entertainment complex that is similar to Liverpool ONE. Similar chain retailers and popular services line Market Street including TK Maxx, Sports Direct and Nationwide. Manchester Piccadilly train station is also an equal distance as Liverpool Lime Street is to that micro-location. Both are roughly 0.5 miles. However, Manchester has a tram line that can take passengers from the station right to the corner of Market Street and High Street, which can be beneficial for customers who are reluctant or unable to walk that distance. The area of Market Street where the sensors are located is also pedestrianised, similar to the Liverpool ONE micro-location.

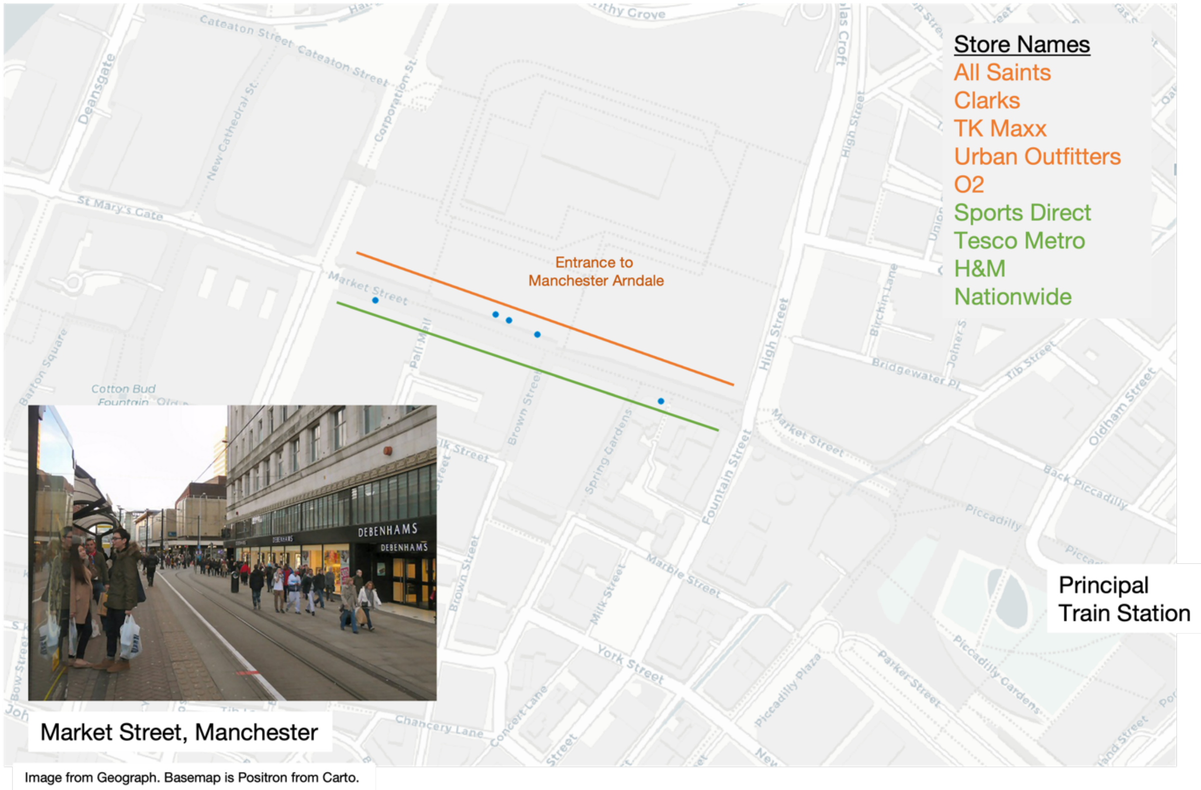


Figure 5-9 Map showing Manchester Market Street and the surrounding area

Table 5-1 further highlights the micro-locational similarities between Manchester Market Street and Liverpool ONE. It shows the retail composition of both case study micro-locations and was derived using the Retail Unit Address data for July 2018 (more information in Chapter 3). Both locations have a strong chain comparison retail presence. 86-87% of occupied units in both micro-locations are chain stores, and 69 retailers are present in both Manchester Market St and Liverpool ONE. Therefore, roughly a third of stores are the same.

Table 5-1 Retail composition of Liverpool ONE and Manchester Market St micro-locations. Derivation and expansion of "Other" Store Type in Appendix 5.1.

	Liverpool		Manchester	
Store Type	**Number of Stores**	**Perc**	**Number of Stores**	**Perc**
Fashion & General Clothing	49	23%	31	16%
Cafes & Fast Food	22	10%	20	11%
Vacant Property	11	5%	24	13%
Sports, Toys, Cycle Shops & Hobbies	20	9%	12	6%
Electrical Goods & Home Entertainment	14	7%	15	8%
Restaurants	17	8%	6	3%
Chemists, Toiletries & Health	14	7%	8	4%
Footwear	11	5%	10	5%
Jewellers, Clocks & Watches	8	4%	10	5%
Other (20 categories)	48	22%	53	28%
Total	**214**		**189**	

Cafes and restaurants complement the comparison retail presence. However, there are slightly more in Liverpool ONE, and both locations have very few discount and budget stores, accommodation and services. The vacancy rate is a standard indicator of the health of a retail area. In Table 5-1, Manchester Market St displays a much higher vacancy rate than Liverpool ONE of 13% compared to 5%, based off the data from 2018. However, as discussed in Wrigley and Lambiri (2014), not all vacancy is equal, and long-term structural vacancy is more detrimental than short-term churn, which is an essential element of retail centre adjustment. Out of the units vacant in the snapshot in Table 5-1, eight of the eleven were vacant six months prior. Six months later for Liverpool ONE and seven of the twenty-four for Manchester Market Street. This resulted in a structural vacancy rate of 4% for both locations, lower than the 6% average for the entire dataset.

When applying the classification derived in Chapter 4, both are Chain and Comparison Retail micro-locations. These micro-locations have a low proportion of independent retailers (11—14%), are close to anchor stores (61m for Liverpool ONE and 21m for Manchester Market St) and are bias to retail over services (9.3 in Liverpool ONE and 6.3 for Manchester Market St).

A strong chain and comparison retail offer drives footfall in both Liverpool ONE and Manchester Market Street. Both locations are close to public transport hubs and nearby to shopping and entertainment complexes that are major footfall attractions. They are the principal retail streets at the heart of two major cities.

Footfall

Both Liverpool ONE and Manchester Market St experience significant footfall. According to the average sum of the footfall data for each sensor in the micro-location, Liverpool ONE had 4.3 million visitors in 2017 and 5.3 million in 2018. In comparison, Manchester Market St had 7 million and 9.9 million visitors in 2017 and 2018. Daily, Liverpool ONE experiences a median footfall of 11,700 people, whereas Manchester Market St has almost double that at 22,600. This difference is roughly proportional to the different populations of the cities.

Despite its proximity to more comparable retail centres than Liverpool, Manchester Market Street appears to significantly attract consumers in that catchment. Many people are willing to travel to Manchester Market Street, whether for retail and recreation purposes or employment.

Figure 5-10 shows how the temporal distribution of footfall varies between Manchester Market St and Liverpool ONE. Generally, both micro-locations have similar patterns, with the busiest times being noon and the day being Saturdays. However, footfall for Liverpool ONE tends to be more variable, with steeper curves. While Liverpool ONE might experience peak footfall for an hour, Manchester Market St is more likely to peak for several hours during the day and maintain a more consistent level of footfall throughout the week. December is clearly the busiest month for Liverpool ONE, and although footfall increases through the year for Manchester Market St, instead November is the busiest month.

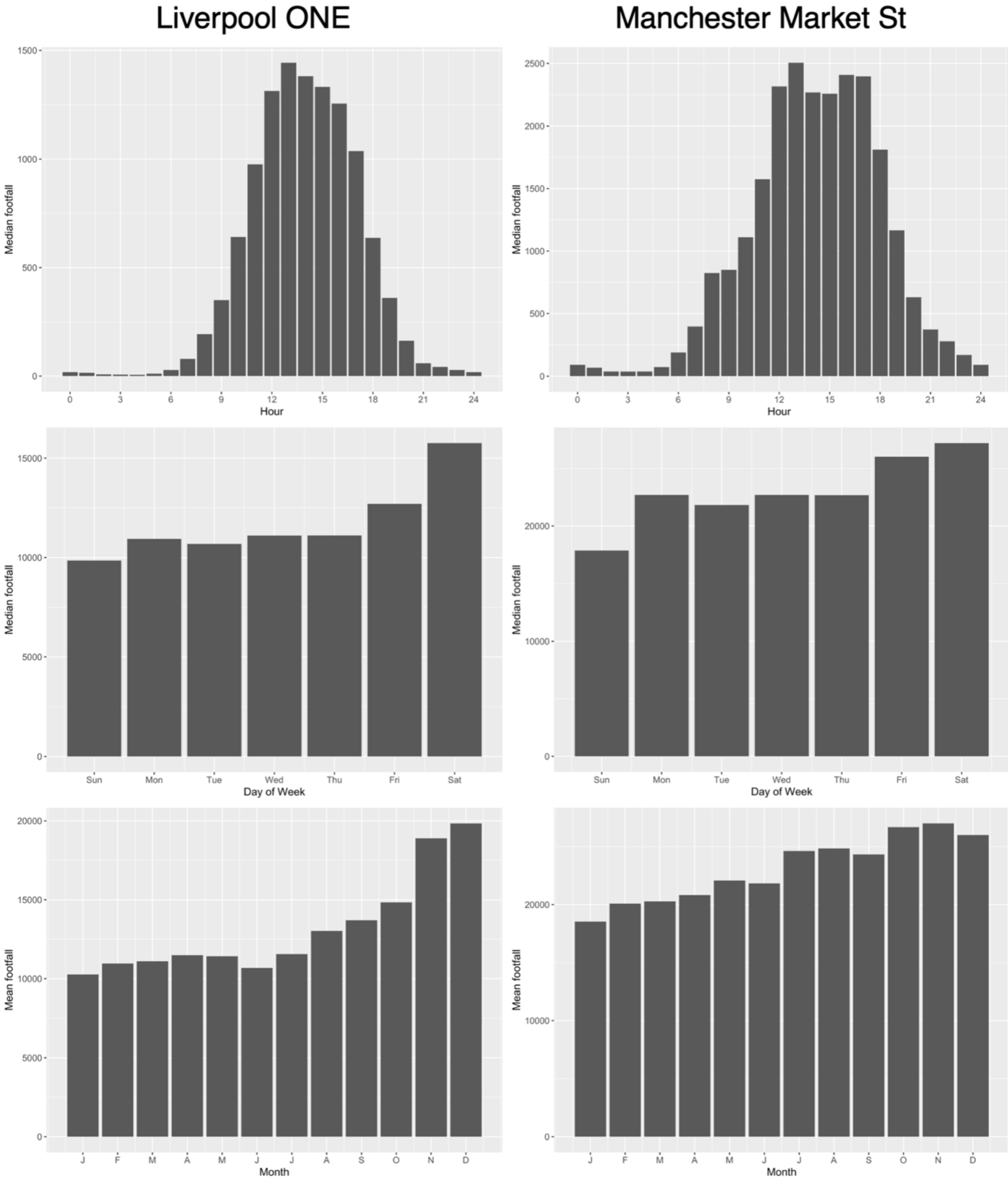


Figure 5-10 Temporal footfall distributions for Liverpool ONE and Manchester Market St micro-locations for the hour of the day, day of week and month of the year

In summary, Liverpool ONE and Manchester Market Street are two locations that have very similar micro-locational characteristics yet have different meso-scale qualities. Liverpool has a smaller population and catchment and is less well-connected to the rest of the country. In comparison, Manchester Market Street has a much larger catchment. Both are principal retail streets with a strong chain and comparison offer.

5.1.3 Background – Edinburgh Old Town and New Town

Edinburgh is a city with a population of 529,000 (National Records of Scotland, 2019), smaller than both Liverpool and Manchester. However, its cultural significance as a capital city draws in visitors, both nationally and internationally. Many of Edinburgh's historical attractions can be found in the south of the city, such as Edinburgh Castle and the Royal Mile. With its medieval architecture and independent shops, the Old Town area is particularly popular with tourists. Roughly half a mile to the north is Edinburgh New Town. The area is named as it is relatively newer than the old town, constructed in the 18th and 19th centuries. Although New Town has some tourist appeal, it principally comprises Edinburgh's principal shopping streets, including many national chain stores and leisure appealing to regular consumers as well as tourists.

Intra-regional characteristics

Figure 5-11 shows the retail centres that are close to Edinburgh. The map is to the same scale as Figure 5-6 which showed the same data for Liverpool and Manchester and comparing both, Edinburgh has significantly less competition from nearby retail centres. Retail centre distribution appears sparser in Scotland compared to North West England. Similar to Liverpool and Manchester, Edinburgh New Town is Subgroup 3.1 (premium shopping and leisure destinations) in the retail centre typology, whereas Edinburgh Old Town is in Subgroup 4.1 (Vibrant secondary urban destinations). There are no other Subgroup 3.1 retail centres within the map extent as direct competition, with the nearest being Glasgow, 42 miles from Edinburgh. However, there are other Subgroup 4.1 nearby Edinburgh Old Town. However, the attraction of Edinburgh Old Town is unique beyond its retail typology; therefore, these locations might not make a large impact.

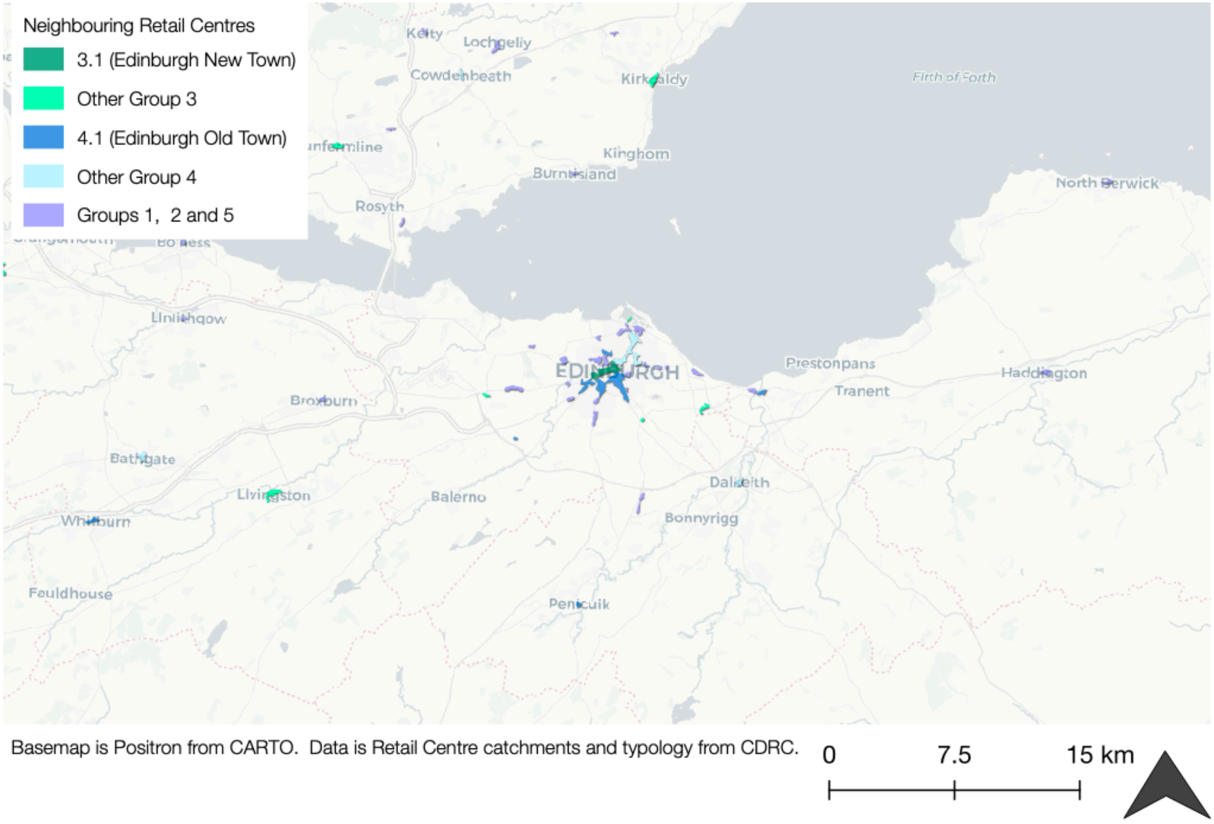


Figure 5-11 Neighbouring retail centres to Edinburgh

Figure 5-12 shows the population distribution around Edinburgh. Note that the areas are bigger than those used for Liverpool and Manchester due to different dataset availability for England and Scotland. Therefore, it is difficult to directly compare the two.

Edinburgh's immediate catchment is more densely populated than the surrounding areas, particularly to the east and south. Although there are few comparable retail centres in these areas, they are also sparsely populated. Therefore, there is little demand for retail centres in these locations. Edinburgh does suffer in terms of connectivity. There is a significant distance between Edinburgh and other major conurbations to the south. Much of the Scottish-English border is rural, with small towns and villages, and the closest area with a comparable population would be Newcastle, over 100 miles away.

Similarly, north of Edinburgh, there are a few smaller cities such as Dundee and Aberdeen, but the area is sparsely populated. Alike to Liverpool, Edinburgh is geographically constrained by a coastline, making connectivity to the city more challenging. Its biggest competition would be Glasgow, a city of a similar population to the west.

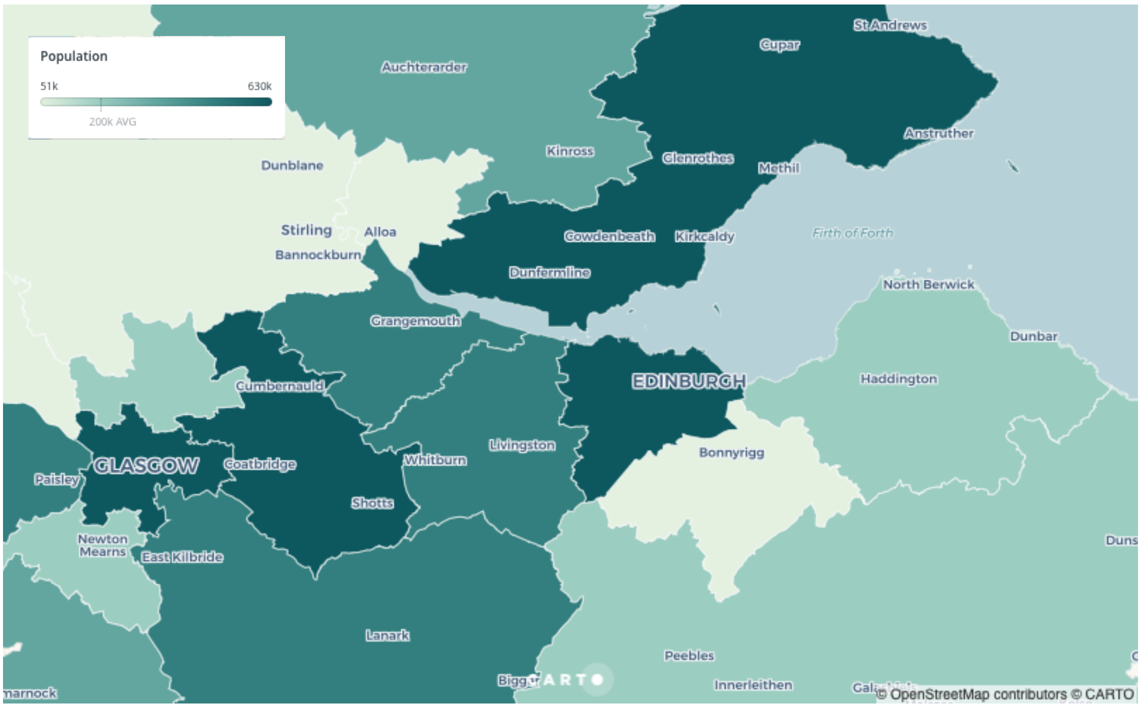


Figure 5-12 Map showing the population distribution around Edinburgh

While its distance from major conurbations can be a deterrent, Edinburgh is well-connected by train, road and air. It has a sparser population than Liverpool and Manchester, but also fewer competing retail centres.

Micro-locational characteristics

Figure 5-13 shows the Edinburgh Old Town micro-location and the features of interest in the surrounding environment. Edinburgh Old Town is the historic tourist area of the city, close to Edinburgh Castle and St Giles Cathedral. Retail units nearby appear to appeal to tourists, with specialist shops such as Balmoral Cashmere & Tweed and House of Edinburgh The Royal Mile and eateries such as The Albanach Bar and The Royal McGregor Bar. This offer is supplemented by some recognisable chains for example Starbucks and Bella Italia. The area is pedestrianised allowing people to move freely through the area and utilising for higher footfall. The principal train station, Edinburgh Waverly, is only 0.3 miles away.

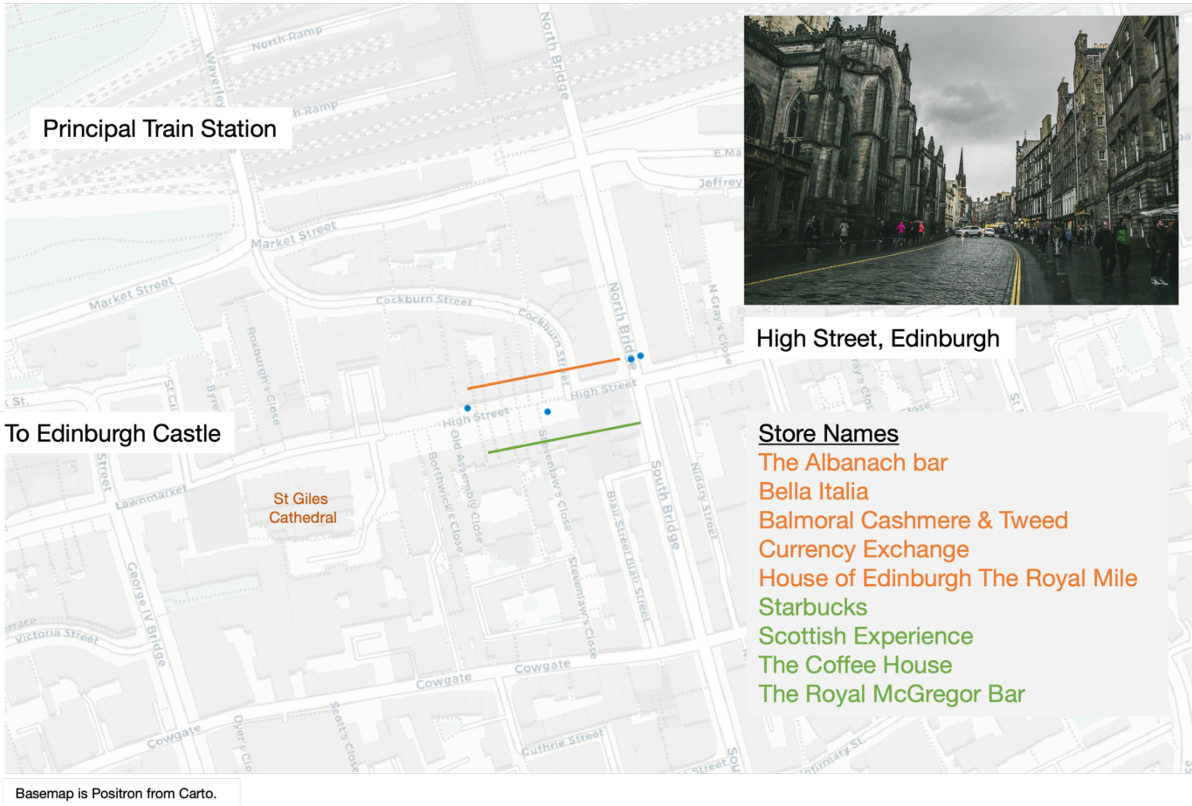


Figure 5-13 Annotated map showing the surrounding area of Edinburgh Old Town

Conversely, Edinburgh New Town presents a different offer, with premium clothing shops such as Joules and White Stuff and some chain coffee shops and restaurants. The offer may entice comparison shoppers who favour premium brands. George Street runs through the centre of the New Town area and is the street with the most footfall sensor coverage. It is not currently pedestrianised, but there are plans to open the street to pedestrians and cyclists in 2021. The principal train station is 0.5 miles away from George Street.

Figure 5-14 Annotated map showing the surrounding area of Edinburgh New Town

Table 5-2 shows the retail composition of Edinburgh's two case study micro-locations. Both areas have a strong leisure presence, with 36% of units in the Old Town and 28% in the New Town being restaurants, cafes, fast food outlets or bars, pubs and clubs. However, the nature and target consumer of these leisure units will be different in the New Town compared to the Old Town.

While the retail and leisure offer in the Old Town is very local and tourist-focused, in the New Town, it is more general, appealing to regular catchment shoppers. Both micro-locations have a low vacancy and structural vacancy rate. Edinburgh Old Town has a vacancy rate of just 5%, which decreases to 4% when considering properties that have been vacant longer than a year, and Edinburgh New Town is similar, with 7% dropping to 4% also.

Table 5-2 Retail compositions of Edinburgh Old Town and New Town micro-locations. Derivation and expansion of "Other" Store Type in Appendix 5.1.

	Edinburgh Old Town		Edinburgh New Town	
Store Type	**Number of Stores**	**Percentage**	**Number of Stores**	**Percentage**
Fashion & General Clothing	18	13%	54	21%
Restaurants	22	15%	31	12%
Gifts, China & Leather Goods	17	12%	4	2%
Cafes & Fast Food	15	11%	21	8%
Bars, Pubs & Clubs	14	10%	21	8%
Accommodation	10	7%	8	3%
Travel Agents & Tour Operators	7	5%	2	1%
Vacant Property	7	5%	18	7%
Other (24 categories)	32	23%	103	39%
Total	**142**		**262**	

Another key difference between the retail composition of the two Edinburgh micro-locations is the prevalence of accommodation, such as hotels or hostels, travel agents and gift shops in the Old Town compared to the New Town. The retail offer of the Old Town is targeted more towards visitors who are looking for a place to stay, things to do and souvenirs to buy, with 78% of the non-vacant units being independent retailers. In contrast, the New Town provides a more "typical" retail offer, similar to Manchester or Liverpool, with 52% chain store retailers.

Footfall

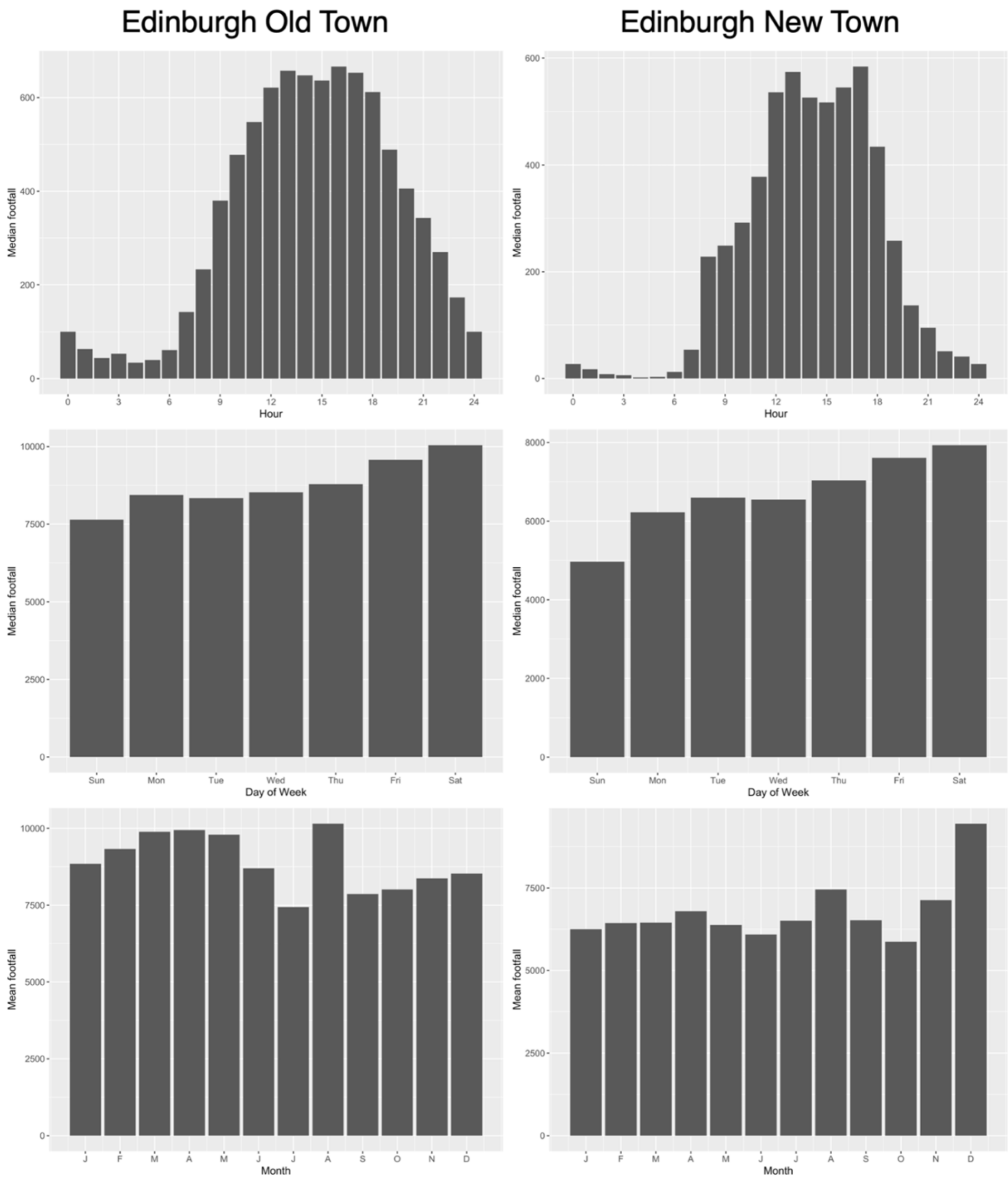


Figure 5-15 Temporal footfall distributions for both Edinburgh micro-locations for the hour of the day, day of week and month of the year

In 2017 and 2018, Edinburgh experienced an annual footfall of 3.0 – 3.5 million in the Old Town and 2.5 million for both years in the New Town micro-location. This number is less than the annual footfall of Liverpool ONE and Manchester Market St, but this is to be expected as Edinburgh has a smaller population. The daily median footfall for Edinburgh Old Town is 8,700 people, over 2,000 more than the daily median for New Town of 6,600.

The temporal distribution of footfall in Edinburgh Old Town and Edinburgh New Town is shown in Figure 5-15. During the day, the Old Town has a concave distribution of footfall, whereas the New Town displays a convex pattern, similar to that of Liverpool and Manchester city centres. The Old Town retains more footfall during the late evening and early morning, which is consistent with the higher proportion of night-time economy as shown in Table 5-2.

The Old Town also displays a unique annual footfall pattern, where March, April and August are the busiest months. November and December are the busiest months in the New Town, similar to Liverpool and Manchester.

Although they have identical intra-regional characteristics, Edinburgh Old Town and Edinburgh New Town vary in their micro-locational features. Edinburgh Old Town appeals to tourists with specialist leisure and independent retail geared towards visitors. In contrast, Edinburgh New Town is similar to Manchester Market Street and Liverpool ONE. It contains the chain and comparison retail locations that regular visitors to the city would expect to find.

5.1.4 Wider applications

The previous sections have introduced and described the four case studies used in this chapter, going into detail regarding the meso- and micro-locational context. The benefit of using case studies is the ability to go into contextual detail. However, one of the limitations is that any conclusions or observations may not be applicable on a wider scale. The following paragraphs will evaluate how representative these four-case study micro-locations are on a national context, and therefore how applicable conclusions might be on a larger scale.

Firstly, they are all in cities. Although Edinburgh, Liverpool and Manchester have somewhat different population sizes, they are all regional centres that draw consumers from a considerable catchment. Table 5-3 gives the retail centre typology group and supergroup for each micro-location. The case study micro-locations only represent two of the fifteen total groups found in Britain. Liverpool ONE, Manchester Market Street and Edinburgh New Town are all group 3.1, and Edinburgh Old Town is 4.1.

Table 5-3 Retail Centre Typology for case study micro-locations

Case Study Micro-location	Retail Centre Name	Supergroup	Group	Justification
Liverpool (Liverpool ONE)	Liverpool Central	3 – Leading Comparison & Leisure Destination	3.1 – Premium shopping & leisure destinations of (semi) regional importance	*North West England retail centres of a comparable size and offer, but individual local identities.*
Manchester (Market St)	Manchester Central			
Edinburgh New Town (George St)	Edinburgh Central			*Same city and local identity as Edinburgh Old Town but different retail offer*
Edinburgh Old Town (High St)	High Street, Old Town, Edinburgh	4 – Primary food & secondary Comparison	4.1 – Smaller vibrant urban destinations	*Same city and local identity as Edinburgh New Town but different retail offer*

Although the patterns drawn might not be applicable on a wide scale to all retail centres, any similarities between Liverpool ONE, Manchester Market Street and Edinburgh New Town might also apply to other centres in group 3.1. Examples of these include Leeds Central, Glasgow Central and Bristol.

As Edinburgh Old Town is primarily a tourist destination, its specific context is even more important. When and why people visit is individual to that location; therefore, the patterns observed not be directly comparable to other areas. However, the conclusions could match other locations more generally. For example, the relative impact of local, national and weather events could apply to other tourist locations in cities. Also, the comparison between Edinburgh Old Town as a tourist location and

Edinburgh New Town as a principal retail street could reveal patterns about other cities with these two areas.

The case studies used in this analysis have little generalisability to a wider application and there should be caution when using any results or observations to infer about other towns or cities. However, it is hoped that the patterns as they are related to factors of the meso- and micro- scale contexts might provide some avenues for further research.

5.2 Identifying temporary fluctuations in footfall

In the last section, four case study micro-locations were defined. These case study micro-locations were chosen to represent the variation of micro-locational and intra-regional factors, comparing and contrasting their retail offer and footfall contexts. First however, they were evaluated in terms of data quality, ensuring that the sensors within each micro-location could be combined to have complete 365 days of data, were representative of the same micro-location, and were verifiable. Each micro-location had several sensors near each other, where at least two of the sensors had data for any given month in 2017 and 2018. These data quality considerations were a vital foundation for this section.

In this section, an annual ranking of footfall is defined to identify unusually busy and quiet days throughout the year. Footfall magnitude is highly variable, even between sensors that are close together. A ranking is chosen because it removes the need to express footfall in terms of magnitude. In addition, it allows the sensors within a micro-location to be compared. Section 5.2.1 defines the methodology for how this annual ranking was determined for each micro-location.

Then, the temporary fluctuations of footfall have to be understood. Dates for potential national holidays and extreme weather events are defined, and the busiest days are labelled with events that occurred on the same day. These events could be national holidays and celebrations, weather events or local events and festivals. The definition of these and how they were identified and investigated is discussed in Section 5.2.2.

This section will utilise the four case study micro-locations and build an annual footfall ranking for the years 2017 and 2018 for each. This aids in identifying the temporary fluctuations in footfall that will be discussed in the remainder of the chapter.

5.2.1 Deriving an annual ranking

Although the representative sensors of a micro-location are close in proximity to each other, they can vary significantly in terms of footfall magnitude. For example, the average daily footfall for one of the sensors in Manchester Market Street is 27,000. However, for a sensor less than 150m away on the same street, footfall is measured at less than half at 11,000. It is impossible to tell whether these figures are valid measures of footfall variation or whether their variation is from sources of error in the data collection process[18]. Therefore, several analytical steps were taken to remove the variable magnitude of footfall and focus on ranking which days were busiest. These will be explained in the following paragraphs, and they are summarised in Figure 5-16.

[18] Upon writing this, I was not aware of the manual counts dataset that could have been used to validate or adjust these footfall sensors.

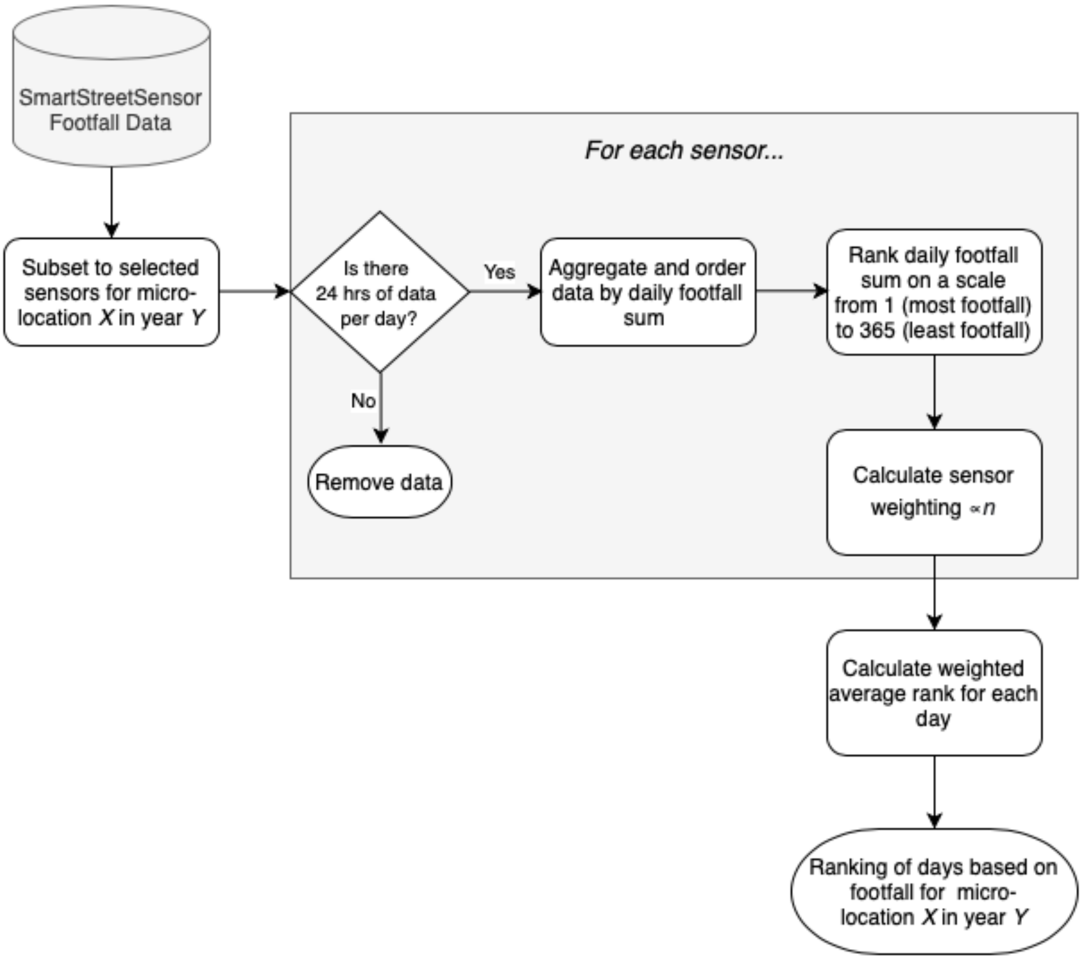


Figure 5-16 Flowchart of the method used to calculate the footfall ranking

First, the data is subset to a year (Y) and a micro-location of interest (X). For example, in deriving the ranking for Liverpool ONE in 2017, Y = 2017 and X = Liverpool ONE. Then, each sensor is validated to ensure that there are 24 hours of data before a daily sum is calculated. If there is not, that day is dropped.

Common methods of dealing with missing data, such as imputation, are not applicable in this analysis for two reasons. Firstly, there is not enough available data to make reliable inferences. As this dataset is a time series with an annual seasonal cycle, it is difficult to infer the measurement for a particular month without any past data. Secondly, finding the outlying or extreme values is fundamental to identifying which events impact a retail centre. Filling in the missing data with generated values risks minimising or removing these values of interest.

However, it still needs to be ensured that the outlying values are not false readings due to errors in the data collection process. The next step of this methodology aims to overcome this by supplementing the high temporal coverage data with sensors identified nearby (see Section 5.1.1) that have less coverage but can validate outlying values.

This supplementation was achieved using a ranking and weighting system. First, the daily footfall for the sensor from micro-location *X* and year *Y* is ranked on a scale of 1 (the day with the most footfall) to 365 (the day with the least footfall)[19]. This 1—365 scale is used even for the sensors with less than 365 days of data. For example, sensor 5 for the Liverpool ONE micro-location has 61 days of coverage for 2017 (*X* = Liverpool ONE, *Y* = 2017, *n* = 61). The days are first ranked from 1 to 61, then that ranking is scaled instead to 1 to 365 using the following equation.

$$\left(\left[\frac{rank\ -1}{n-1}\right] \times 364\right) + 1$$

This equation standardises the rankings, making the value of a rank between sensors comparable to one another. For each sensor, 1 denotes the busiest day, 365 the quietest, with the middle being roughly 182. However, this can have the impact of weighting sensors with a smaller *n* higher than a sensor with more data. The sensor where *n* = 61, the quietest day was ranked 61st, and now it is 365th, and because we only have 61 days of data, there is no way to tell whether it is genuinely the 61st busiest day, the 161st or the 365th.

To remove this impact, each sensor is given a weight proportional to *n*. As the sensors with full coverage (*n* = 365) have a complete picture of annual footfall, they will be weighted higher than those with less coverage. This weighting is calculated for each day. So, on a day where there is data from two sensors (*n* = 365 and *n* = 61), the weighting per sensor will be $n\ /\ (365 + 61)$. Therefore, the weighting is proportional to *n*. Suppose the next day had data from a third sensor where *n* = 90, the weighting per sensor would change to $n\ /\ (365 + 61 + 90)$. The scaled sensor

---

[19] Both 2017 and 2018 have 365 days. Neither are leap years.

ranks are then multiplied by the weights before being added together to calculate the final rank. This calculation ensures that inaccurate measurements will not skew the ranking because if a specific day ranks highly for one sensor but not for the others nearby, it will pull down its overall ranking.

To give an example of how this methodology works, mock data of rankings for two sensors for over two weeks were generated. The first sensor, Sensor 1, collected data for the entire two weeks, therefore, has a complete 1-14 ranking. Sensor 2, however, was only collecting data for the last week. Those days are ranked 1-7, and then that ranking is scaled to 1-14, and these are displayed as the dotted lines in Figure 5-17. As Figure 5-17 shows, the rankings calculated for the second week are between both sensors, taking both rankings into account. However, because there is more data for Sensor 1, the ranking point favours that observation. The resulting ranking is primarily informed by the sensor with the most data but can be influenced if another sensor has a significantly different ranking for that day.

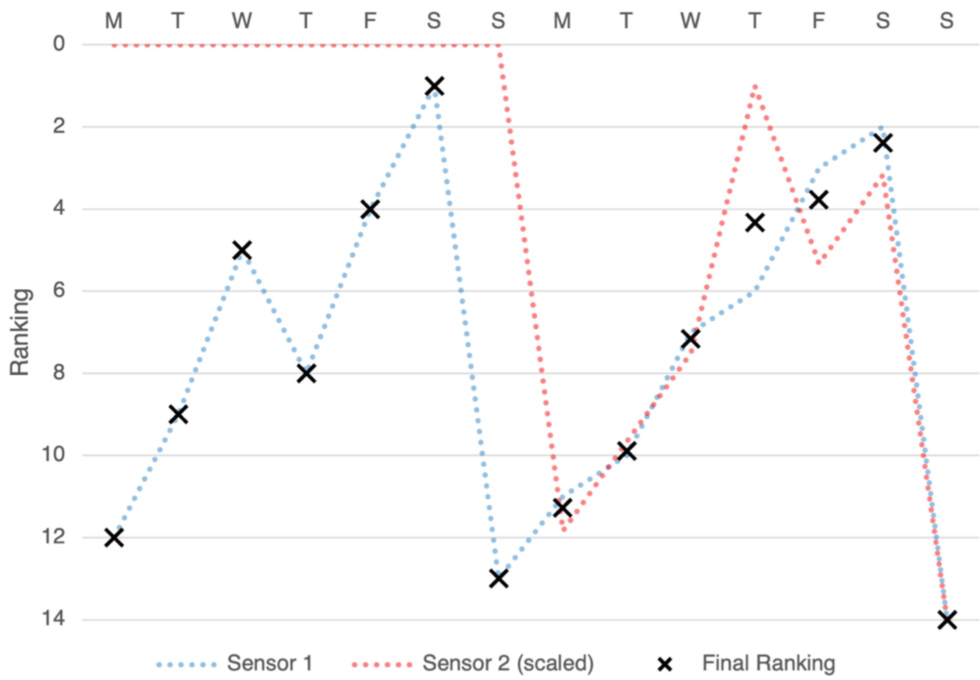


Figure 5-17 Example of the daily ranking methodology on mock data

A limitation of this method is that the final calculated ranks are decimals and not integers. As is visible in Figure 5-17, the rank for the second Thursday is 4.3. This decimal number poses an issue for interpretability, as a decimal can not be a ranking. Therefore, the ranking referred to in this chapter ranks those decimal numbers. For example, the calculated ranking of the second Thursday is 4.3, but that is 5th out of all the calculated rankings. Consequently, it will be referred to as the 5th busiest day.

5.2.2 Understanding fluctuations within the annual ranking

This method was applied for Liverpool ONE, Manchester Market Street, Edinburgh New Town and Edinburgh Old Town micro-locations for 2017 and 2018. Eight annual rankings were generated, which consisted of date and the rank from 1—365 from busiest to quietest day. The next step to understanding the footfall fluctuations within the case study micro-locations was to identify the potential causes for unexpectedly busy days and how these could relate to specific events.

National holidays and celebrations

The first hypothesised cause of temporary fluctuations in footfall is national holidays or shopping events. These include the festive season and run-up to Christmas, Black Friday, Boxing Day sales and January sales. These events are often expected to increase footfall as retailers lower their prices for a limited time, and many consumers would want to benefit from these offers. During the festive season and run-up to Christmas, many people purchase gifts for their families and friends or enjoy the experience and seasonal leisure-based activities. For example, Christmas markets or a seasonal light switch on.

National holidays such as Bank Holidays could also impact footfall. Consumers may want to enjoy a day of work and travel to an area with a recreational retail or leisure function, increasing footfall in these areas. However, it might have the opposite effect in the regions that heavily rely on working commuters to increase footfall.

National holidays and celebrations are simple to identify in the dataset as they are set dates and times which are recorded. The bank holidays and National holidays for

England and Scotland were easy to look up and identify. These can be found in Appendix 5.2.

Weather events

Another potential cause of temporary fluctuations in footfall is weather events. Storms and snow can negatively impact footfall as consumers may not want to travel to a high street location in poor weather. This impact of weather on footfall is supported in research by Parsons (2001), who found that temperature and rainfall have a negative effect on the number of visitors to a shopping location.

Unlike national events and holidays, weather events do not occur on pre-determined dates. Therefore, some secondary research was required to identify the significant weather fluctuations during 2017 and 2018 that could have impacted the case study locations. The Met Office keeps an online database the includes case studies of past severe weather events (Met Office, 2021). This database was used to identify six weather events during 2017 and 2018 that impacted the case study micro-location areas and could have had a significant impact on footfall. These are shown in Table 5-4.

Table 5-4 Weather events during 2017 and 2018 that impacted North West England and Scotland

Date	Event	Impacts related to footfall
16/10/2017	Ex-Hurricane Ophelia (Met Office, 2018a)	Heavy rain and winds cause unfavourable conditions keeping many indoors. Flights cancelled between Manchester and Edinburgh to the Republic of Ireland and Northern Ireland.
28/02/2018—01/03/2018	Snow and low temperatures, dubbed 'Beast from the East' (Met Office, 2018b)	Snow, low temperatures and strong winds cause unfavourable conditions keeping many indoors. Severe travel disruption with many roads, rail and airport services cancelled.
14/06/2018	Storm Hector (Met Office, 2018c)	Strong winds cause unfavourable conditions keeping many indoors, mainly in Edinburgh. Disruption to road and rail services.
06/2018—07/2018	Summer heatwave (Met Office, 2018d)	The summer was record-breaking for its warm, dry and sunny weather. Temperatures were often above 30°C. This could increase footfall as people want to enjoy the weather, but the warm weather can also deter some. This coincided with the Football World Cup, and many people attended pubs and bars to watch.
19/09/2018—21/09/2018	Storms Ali & Bronagh (Met Office, 2018e)	Travel disruption in Scotland. Businesses experienced power cuts and had to close. Strong winds cause unfavourable conditions, keeping many people inside.
15/12/2018	Storm Deirdre (Met Office, 2018f)	Strong winds and heavy rain keep people indoors. Limited travel disruption by air and road. This occurred on a Saturday during the run-up to Christmas.

There were disproportionately more extreme weather events in 2018 than in 2017. There was only one storm in 2017 on the Met Office records that impacted the case study locations and therefore could've impacted footfall. In 2018, there were four. During 2018, there was also an extreme spell of cold weather, nicknamed 'Beast from the East' in February and March, and a heatwave during June and July. This warm, dry weather spell coincided with the 2018 Football World Cup, where the England team made it to the semi-finals. Therefore, this could be a potential driver of footfall, especially in areas with pubs and bars in England, as people would want to watch, or areas in close proximity to supermarkets, which reported an increase of sales during this time (Smithers, 2018).

Local events or festivals

Another temporary factor theorised to impact footfall were local events. These are festivals, markets or events that are run and take place at a local level. An example could be a local food festival or celebratory parade. These can be more challenging to find records of than national or weather events. As this research takes place after the event happened, many of the exact details have been removed or replaced on resources, and a more thorough search can be necessary.

First, some events were already known to occur, such as the Edinburgh festivals. Known as a festival city, Edinburgh is home to 12 festivals annually. The largest of which is the Festival Fringe, which brings over 4.5 million visitors to the city (Naylor et al., 2016). The festivals have a significant economic and social impact. The 2015 Edinburgh Festival Impact Study (Naylor et al., 2016) found the festivals have an economic impact of £280 million for Edinburgh, with visitors spending £93.5 million on entertainment, shopping, food and drink and accommodation during their visit. The festivals generally have good reception from local residents also, with 78% of people believing that the festivals made Edinburgh a better place to live (Naylor et al., 2016).

After known events had been accounted for, the top 25 days for footfall for each micro-location were taken for 2017 and 2018. For each of the dates, a search of local news websites was undertaken. Local news websites were used as they often cover events for promotion and records. However, the coverage can be patchy, as not all events would be reported. For example, if a retailer had a promotional event or release, it may increase micro-locational footfall as many people crowd or line up outside the store. This promotional event would significantly increase footfall yet might not be reported on and might be hard to find information on after the occurrence.

Nonetheless, by using this method and researching the top 25 days for each year and micro-location, realistic explanations were found for all but ten days. Therefore, 90% of the highest footfall days could be explained by events. A full breakdown of these can be found in Appendix 5.3.

This method also picked out events and factors that had an unforeseen impact. For example, Halloween and local sporting events boosted footfall for some micro-locations, a factor that was not identified prior to analysis.

This section has detailed the methodology used to combine and validate the footfall data for the multiple sensors within a case study micro-location. Annual rankings were derived for 2017 and 2018. National holidays and festivals and weather events were identified, which have been hypothesised to impact footfall. In addition, secondary research was undertaken to explain the top 25 busiest days for each micro-location and year. 90% of these were event-related.

5.3 Results

As justified in the chapter introduction, it is clear that events such as national holidays, local festivals and extreme weather events can drive or impact footfall (BBC News, 2018a; Collinson, 2018; Naylor et al., 2016; Sky News, 2018). However, the extent to which that varies over different micro-locations and regional contexts has not been explored in depth. This section aims to fulfil the second research aim for this chapter by investigating how micro-locational and intra-regional differences can impact event-related footfall through case study examples.

First, Section 5.3.1 will explore Liverpool ONE and Manchester Market Street. As discussed in Section 5.1.2, Liverpool ONE and Manchester Market Street have similar micro-locational characteristics. Both are city centre principal retail streets close to a major entertainment and shopping complex that rely on chain and comparison retail. However, they differ in their intra-regional context. Manchester has a larger population and is better connected to the majority of the country, whereas Liverpool is smaller and comparably less well-connected. The busiest and quietest days for Liverpool ONE and Manchester Market Street will be explored using the rankings defined in the last section. In addition, the impact of national holidays and weather events will be compared. Drawing on the micro-locational and intra-regional features explored, this section will investigate possible explanations for the similarities and differences between the two micro-locations.

Then, Section 5.3.2 will inspect how event-related footfall varied during 2017 and 2018 for Edinburgh Old Town and Edinburgh New Town micro-locations. Unlike Liverpool and Manchester, Edinburgh Old Town and Edinburgh New Town have similar intra-regional contexts because they are within the same city. However, they differ in their micro-locational context, as explored in Section 5.1.3. Edinburgh Old Town caters to visitors and tourists to the city. Its proximity to tourist attractions such as Edinburgh Castle and independent stores and leisure outlets are curated to serve a tourist population. In contrast, Edinburgh New Town has a similar retail offer as Manchester Market Street and Liverpool ONE, with many chain and comparison retail outlets. This section will explore the footfall rankings and event-related footfall for Edinburgh New Town and Old Town, using the micro-locational variations introduced in Section 5.1.3 to examine possible explanations for these trends.

5.3.1 Liverpool ONE and Manchester Market Street

The annual ranking is visualised for Liverpool ONE in Figure 5-18 and Manchester Market Street in Figure 5-19. Red indicates the days with the highest footfall and blues the days with the lowest footfall.

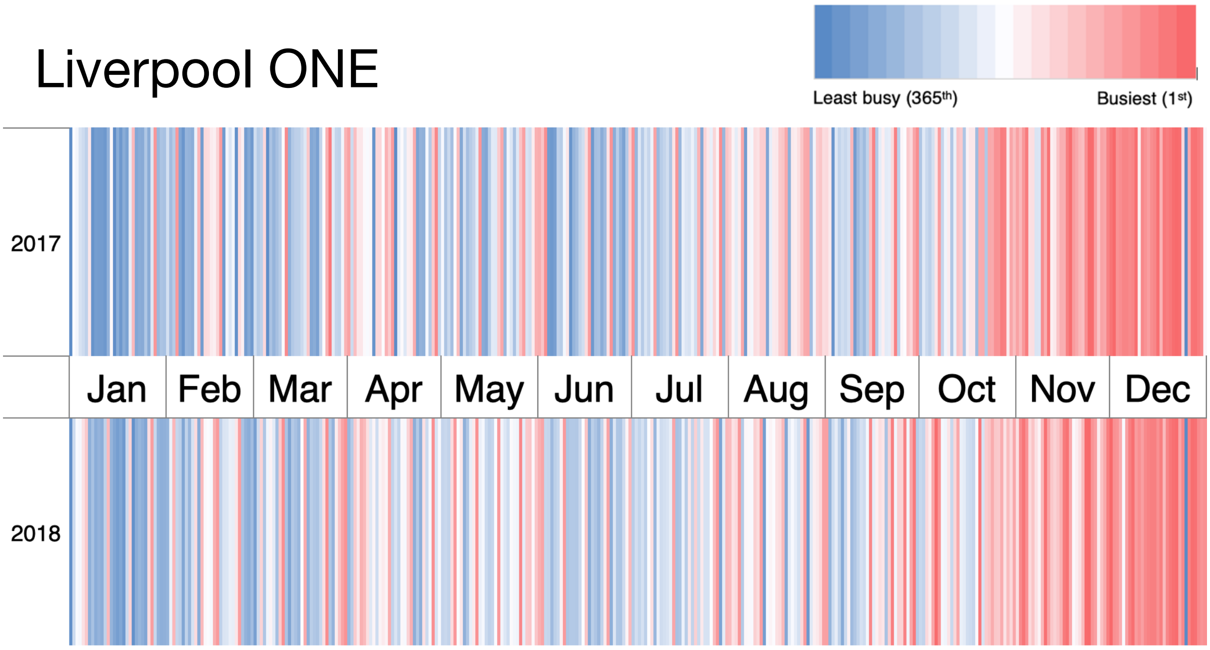


Figure 5-18 Liverpool ONE daily footfall ranking for 2017 and 2018

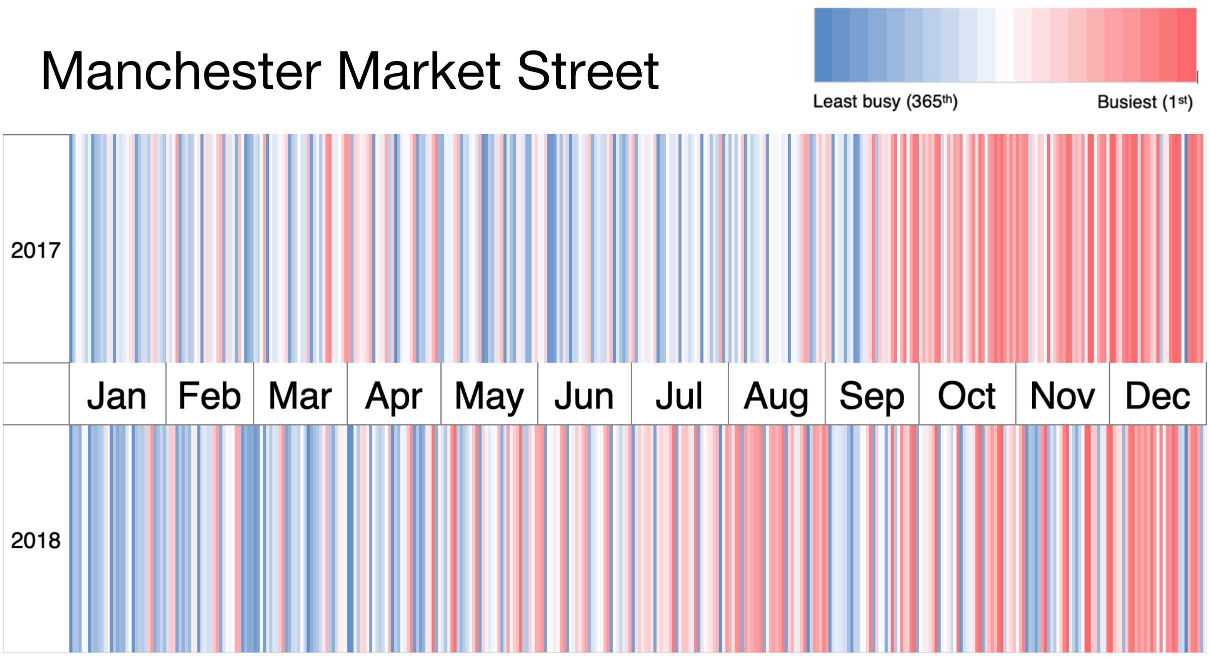


Figure 5-19 Manchester Market Street daily footfall ranking for 2017 and 2018

Both micro-locations are busiest in November and December; however, Liverpool is more consistently busy during that time. Liverpool experiences a median daily footfall of 25—29,000 people, double the median for the entire year. In comparison, Manchester's busier days were more evenly distributed throughout the year. This was particularly the case for 2018, when July and August are noticeably busy. Manchester Market St experienced a median daily footfall of 31,000 people in July 2018, compared to 22,000 the previous year.

National holidays

Table 5-5 contains the ranking for national holidays and bank holidays for Liverpool ONE and Manchester Market Street for 2017 and 2018. Bank holidays tended to rank higher for Liverpool ONE than Manchester Market Street. For example, the Saturday in Easter weekend ranked 34th and 47th for Liverpool ONE in 2017 and 2018, but only ranked 98th and 167th for Manchester Market Street. There were similar discrepancies for the Early May and Spring Bank Holidays. However, in August, the Summer Bank Holiday was an exception, where Manchester Market Street ranked higher than Liverpool ONE in 2018. In Figure 5-18, bands of higher footfall in red throughout the year also correspond with the school half-terms in February, May and October, and the Spring and Summer school holidays.

Table 5-5 Ranking for National Holidays for Liverpool ONE and Manchester Market Street

	Liverpool ONE		**Manchester Market Street**	
Name	**2017**	**2018**	**2017**	**2018**
New Year's Day	360th	360th	360th	359th
Good Friday	95th	52nd	286th	163rd
Saturday of Easter weekend	34th	47th	98th	167th
Easter Sunday	361st	323rd	361st	360th
Easter Monday	200th	307th	311th	358th
Early May Bank Holiday	171st	131st	318th	300th
Spring Bank Holiday	224th	114th	319th	282nd
Summer Bank Holiday	252nd	149th	214th	61st
Halloween (Friday)	23rd	82nd	12th	15th
Halloween (Saturday)	22nd	91st	44th	6th

Bank holidays drive footfall because people can have the day off work and decide to spend their recreation time in a retail location. Liverpool ONE and Manchester Market Street have a similar retail offer, yet perhaps consumers could perceive Liverpool ONE as a more attractive location to spend time.

On the other hand, as these rankings are relative to the location, this does not mean that one centre was busier than another during those times. It does mean that in 2017 and 2018, Bank Holidays were relatively more critical days for footfall for Liverpool ONE than they were for Manchester Market Street. Therefore, it could also mean that consumers visit Manchester Market Street consistently. A peak for Bank Holidays is relatively not as significant.

Halloween also drove footfall in both cities, though to a greater extent in Manchester Market Street. Some of the highest-ranking days of 2017 and 2018 were the weekend before Halloween. This difference could result from a more organised and extensive celebration effort by Manchester BID – Halloween in the City – which marks the season with art exhibits and parades.

Black Friday and festive footfall

The run-up to Christmas is the biggest driver of event-related footfall for Manchester Market Street and Liverpool ONE. Figure 5-20 shows how footfall ranking varied for both micro-locations starting from Black Friday and ending New Year's Eve.

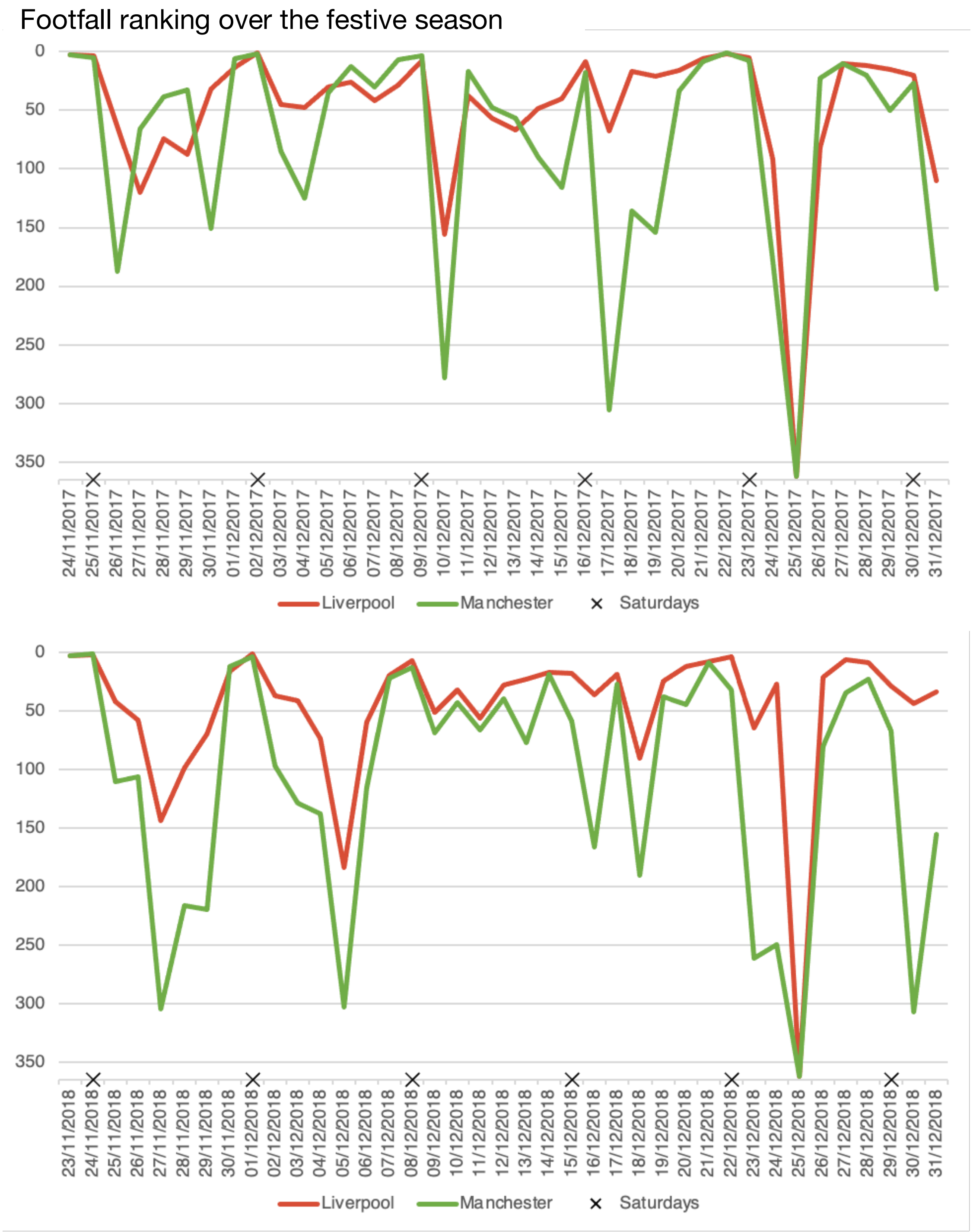


Figure 5-20 Footfall over the festive season for Liverpool ONE and Manchester Market Street

For Liverpool ONE, the busiest day of the year for 2017 and 2018 was the first Saturday of December (02/12/17 & 01/12/18). These days were also busy for Manchester Market Street, but not to the same extent. They ranked 2nd and 4th for

2017 and 2018. The busiest day for Manchester Market Street was 22/12/17 (the Friday before Christmas) and 24/11/2018 (the Saturday after Black Friday).

Generally, the Saturdays from Black Friday to Christmas rank highly for both micro-locations, although this is particularly the case for Liverpool ONE. Table 5-6 shows this ranking for Liverpool ONE and Manchester Market Street. The last Saturday before Christmas is often considered the busiest day of the festive run-up; however, in the case of these micro-locations, the last Saturday of November or the first Saturday of December might be more apt. This could be due to the commencement of Christmas markets or Christmas light switch on ceremonies during this time, driving footfall with a more experience-based purpose. In contrast, the last Saturday before December might be more significant in terms of sales.

Table 5-6 Ranking of Saturdays before Christmas for Liverpool ONE and Manchester Market Street

Saturdays before Christmas	**Liverpool ONE**		**Manchester Market Street**	
5 (25/11/2017 & 24/11/2018)	4th	2nd	5th	1st
4 (02/12/2017 & 01/12/2018)	1st	1st	2nd	4th
3 (09/12/2017 & 08/12/2018)	8th	7th	4th	13th
2 (16/12/2017 & 15/12/2018)	9th	18th	18th	59th
1 (23/12/2017 & 22/12/2018)	5th	4th	8th	32nd

An exception to Saturday ranking highly is the 15/12/2018. This day dropped ranks significantly from 2017 to 2018, particularly for Manchester Market Street. The second Saturday before Christmas went from 18th busiest in 2017 to 59th in 2018. This drop is likely the impact of Storm Deirdre, which brought strong winds and heavy rain. This extreme weather can impact footfall as many shoppers are deterred by the conditions and will postpone their trip or utilise alternatives such as online shopping (Badorf and Hoberg, 2020; Parsons, 2001). This impact will be discussed further in the later weather events section.

Interestingly, Manchester Market Street ranks Fridays in the run-up to Christmas highly, often within the top ten. They are not relatively as high for Liverpool ONE, yet the Sunday dips are less pronounced. The Boxing Day Sales on dates such as 27th and 28th of December also rank higher for Liverpool ONE than for Manchester Market Street. Liverpool's smaller catchment could explain this. When consumers are not travelling as far, they might be more willing to travel, particularly on days where there are limited opening hours, such as Sundays. A linked explanation could be that more people utilise public transport to visit Manchester Market Street. Often, services can be limited on Sundays or the period after Christmas, acting as a deterrent.

Weather events

Table 5-7 summarises the impact of major weather events identified in Section 5.2.2 on Liverpool ONE and Manchester Market Street during 2017 and 2018. The table uses dates to compare from the opposite year, accounting for the weekday.

Table 5-7 Impact of weather events on footfall ranking for Liverpool ONE and Manchester Market Street

Date	Event	Liverpool ONE	Manchester Market Street
16/10/2017	Ex-Hurricane Ophelia (Met Office, 2018a)	190th (2018: 190th)	123rd (2018: 244th)
28/02/2018—01/03/2018	Snow and low temperatures, dubbed 'Beast from the East' (Met Office, 2018b)	315th, 356th (2017: 238th, 286th)	319th, 346th (2017: 256th, 277th)
06/2018—07/2018	Summer heatwave (Met Office, 2018d)	Average Rank June 2018: 234 July 2018: 222 (June 2017: 235) (July 2017: 199)	Average Rank June 18: 166 July 2018: 159 (June 2017: 241) (July 2017: 223)
15/12/2018	Storm Deirdre (Met Office, 2018f)	18th (2017: 9th)	59th (2017: 18th)

Weather events significantly impact footfall ranking for both Liverpool ONE and Manchester Market Street. The 'Beast from the East' storm in February and March 2018 caused footfall to drop significantly in both micro-locations, resulting in one of the quietest days of the year.

The summer of 2018 was the warmest in the UK since 2006 and the driest since 2003 (Met Office, 2018d). Footfall in Manchester Market Street ranked, on average, significantly higher during this time than it had during 2017, with rank increasing by 64—75 places. The same impact was not seen in Liverpool ONE, where footfall ranked consistently for summer 2017 and 2018. The median footfall for July 2018 was 9,000, down from 10,000 in 2017.

Storm Deirdre also had a more significant impact on Manchester Market Street. The storm caused the second Saturday before Christmas to drop 41 places from its position in 2017. In contrast, it only dropped nine spots for Liverpool ONE.

In general, Manchester Market Street tends to be impacted by weather events more significantly than Liverpool ONE, in both good and bad weather. When considering intra-regional factors, Manchester is a larger conurbation than Liverpool, both in terms of population and size. On days with particularly bad weather, consumers would be more likely to opt for a smaller, local retail centre or use alternatives such as online shopping. In addition, consumers who would usually travel to Manchester city centre have the Trafford centre as an alternative, which is covered, indoor with a similar retail offer to the city centre.

Local events or festivals

Local events were also significant drivers of footfall in both Manchester and Liverpool. The finale of the Giants spectacular event in Liverpool in 2018 was the 5th busiest day of that year, and Manchester's Food and Drink Festival in late September was on the 19th most active day in 2017 and the 7th in 2018. As discussed earlier, the event Manchester BID hosted for Halloween also helped drive footfall.

For Liverpool, sporting events were also a driver of footfall. Several busiest days outside the Christmas period coincided with Liverpool FC home games. For example, the Liverpool Legends v Real Madrid Legends game on 25/03/2017 was ranked 18th for that year. None of the 20 busiest days for Manchester in 2017 and 2018 were on match days for Manchester United or Manchester City teams.

Public transport routes account for some of this difference. Manchester has direct access by train to stadiums for both teams. However, in Liverpool, visitors have to take a train to the city centre before taking a bus or taxi to the stadiums. This could encourage visitors to spend more time in the city centre, therefore increasing footfall.

Local events do appear to have a significant impact on footfall for both Liverpool and Manchester, although it is unknown what impact that would have had on sales and economic growth. Food festivals, such as Manchester's food festival, can have a negative impact on local food outlets who lose business for that day, and it is variable and difficult to predict how much people visiting Liverpool for events such as the

Giants' Spectacular or for a footfall match would spend in the local area. However, there is a non-economic benefit to these successful local event in that they encourage a local-identity and can be used to bring a community together, which in turn strengthens high street vitality and viability.

5.3.2 Edinburgh Old Town and Edinburgh New Town

The daily footfall ranking for Edinburgh Old Town and New Town are visualised in Figure 5-21 and Figure 5-22 respectively. Red represents the days with the highest ranked footfall and blue, the days with the lowest ranked footfall.

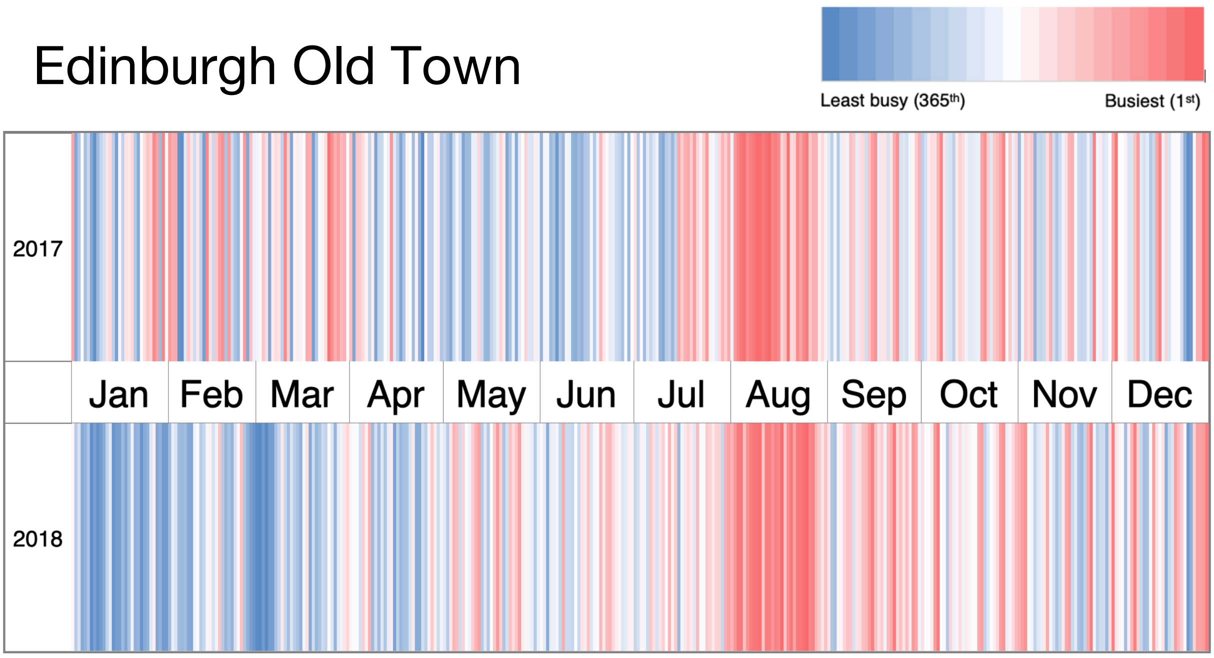


Figure 5-21 Edinburgh Old Town annual footfall rankings

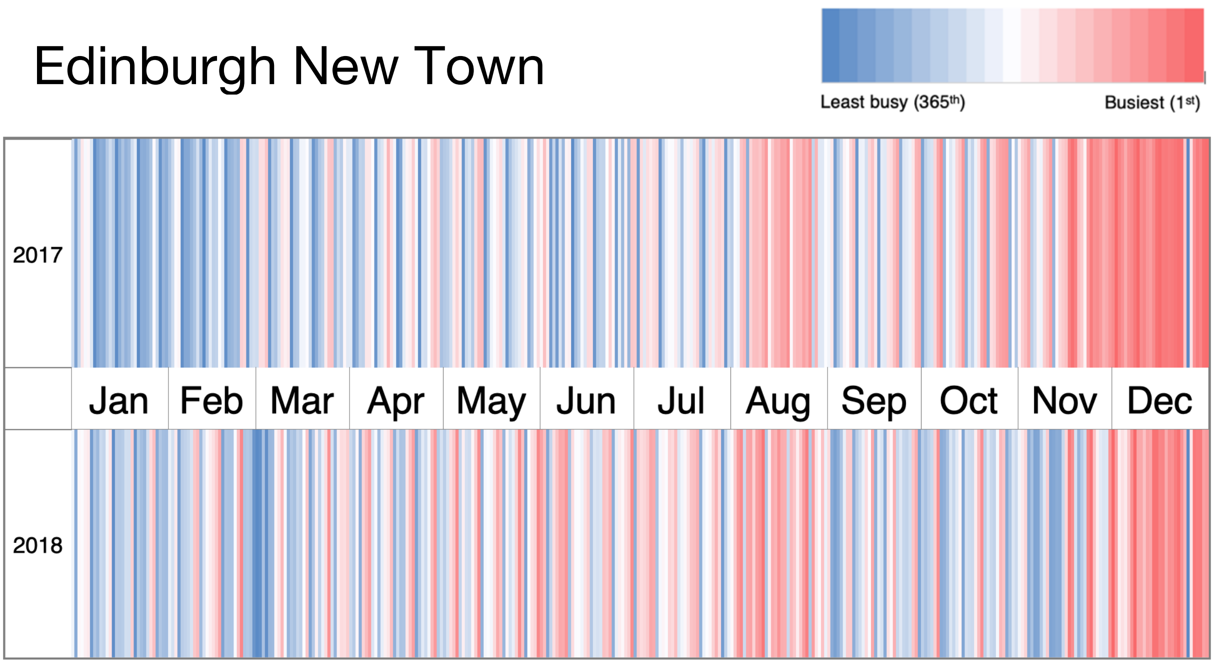


Figure 5-22 Edinburgh New Town annual footfall rankings

Edinburgh Old Town has a unique annual ranking, with 60-80% of its busiest days taking place in August. This coincides with the Edinburgh Fringe Festival. However, Edinburgh New Town shows a similar distribution to Manchester Market Street and Liverpool ONE. There are a high percentage of the busier days approaching the end of the year and Christmas. It was discussed that Edinburgh New Town has a very similar retail offer to the two other micro-locations. This indicates that retail offer could be more important than intra-regional factors in considering when a location is busiest. A micro-location such as Edinburgh Old Town that primarily appeals to tourists might well be less reliant on footfall over the Christmas period and be busier at other times of the year.

National holidays and festivals

Table 5-8 shows the footfall ranking for national and bank holidays for Edinburgh Old Town and Edinburgh New Town.

Table 5-8 National holidays and their footfall ranking for Edinburgh Old Town and New Town for 2017 and 2018

	Edinburgh Old Town		**Edinburgh New Town**	
Name	**2017**	**2018**	**2017**	**2018**
New Year's Day	54th	179th	245th	188th
2nd January	319th	275th	251st	330th
Good Friday	33rd	159th	134th	130th
Saturday of Easter	214th	133rd	200th	102nd
Easter Sunday	279th	177th	355th	245th
Easter Monday	344th	185th	328th	307th
Early May Bank Holiday	299th	117th	285th	172nd
Spring Bank Holiday	222nd	204th	293rd	191st
Summer Bank Holiday	14th	25th	112nd	52nd
St Andrew's Day	157th	115th	34th	32nd
Halloween (Friday)	15th	34th	43rd	76th
Halloween (Saturday)	216th	29th	40th	94th
New Year's Eve	6th	16th	4th	44th

Most national holidays appear to have a uniform impact on throughout Edinburgh. The Spring and Early May Bank Holidays rank similarly for the New Town and the Old Town, although the rankings are slightly higher in 2018 than 2017. An interesting exception is St Andrew's Day. St Andrew's Day on the 30/11 ranks highly for Edinburgh New Town, at 34th and 32nd. In contrast, it is unremarkable for Edinburgh Old Town. St Andrew's Day is a Bank Holiday that is specific to Scotland, and this difference could reflect the consumers each micro-location attracts. New Town targets a more local, Scottish catchment, while Old Town caters to international consumers.

Hogmanay or New Year's Eve celebrations in Edinburgh also attract visitors to the city, and celebrations span both micro-locations – the Concert in the Gardens and Street Party in the New Town and the Torchlight Procession and fireworks at Edinburgh Castle in the Old Town. In 2017, Hogmanay celebrations accounted for the 6th busiest day for the Old Town and 4th for the New Town making it a significant national and local event for both micro-locations.

Black Friday and festive footfall

Similar to Liverpool ONE and Manchester Market Street, Edinburgh New Town experiences its busiest days during the runup to Christmas. However, this is not the case for Edinburgh Old Town. The festival season that takes place in August accounts for the busiest days, dwarfing the Christmas footfall. Figure 5-23 is a graph that compares the two Edinburgh micro-locations.

Footfall ranking over the festive season

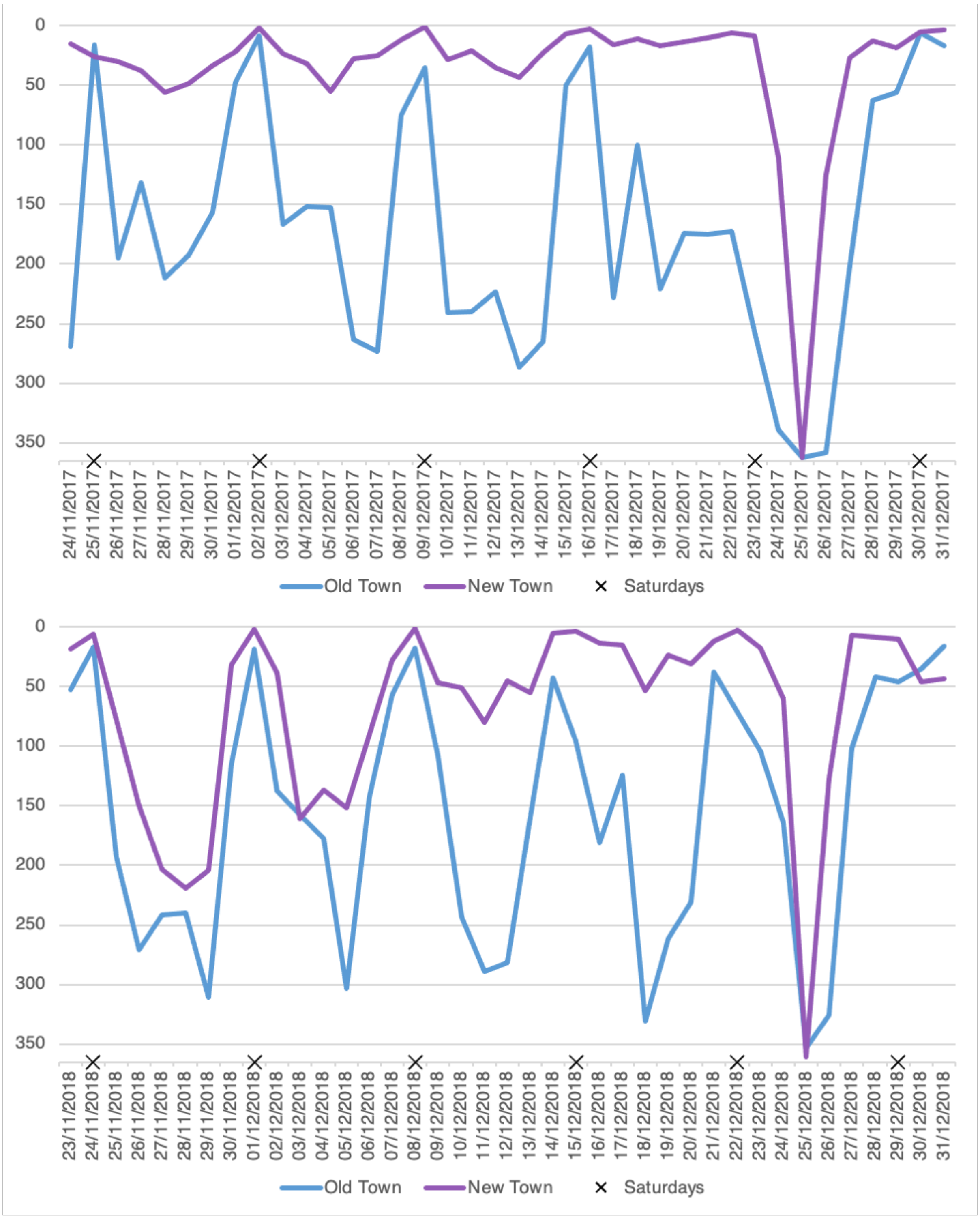


Figure 5-23 Footfall over the festive season for Edinburgh Old Town and Edinburgh New Town

Edinburgh New Town, which has a similar retail composition and offer as Liverpool and Manchester, is also dominated by footfall during the Christmas period. Again, Saturdays during December were the busiest days. However, Black Friday was not as ranked as highly. It was only the 15th busiest day in 2017 and the 17th in 2018. In comparison, it was 3rd busiest for Liverpool and Manchester.

Although Edinburgh New Town is busy on Black Friday, it is not a shopping event to the same magnitude as it is for the other cities. This could be a reflection of the willingness to participate in shopping events between the two regions, the presence of certain offers that might drive more footfall, or the cooler temperatures, with Black Friday 2017 having a high of 3°C, compared to 8°C and 7°C in Liverpool and Manchester respectively.

While Manchester and Liverpool don't attract significant footfall on Sundays, even during the Christmas period, some of Edinburgh New Town's busiest days were Sundays. This could be an impact of the Sunday Trading Act 1994 which restricts Sunday retail opening times in England and Wales, but not Scotland. The longer duration of time for consumers to be able to shop both keeps visitors in the retail centre for longer and is an attractor itself.

December is very consistently high ranking for Edinburgh New Town, particularly in 2017. Although Saturdays rank highest, the dip during the week is not pronounced and the ranking consistently stays above 75. In comparison, these days are ranked 250th and below for Edinburgh Old Town. As Edinburgh Old Town primarily caters to tourists, this could indicate an off-season, when there are fewer international visitors. Many Christmas events take place in Edinburgh during this time, largely in the Princes Street Gardens area which is closer to the New Town. Therefore, visitors to Edinburgh might prioritise the seasonal entertainment and retail shopping over the Old Town attractions.

Weather events

Table 5-9 shows the impact of different extreme weather events on footfall ranking for Edinburgh Old Town and New Town.

Table 5-9 Impact of weather events on footfall ranking for Edinburgh Old Town and New Town

Date	Event	Edinburgh Old Town	Edinburgh New Town
16/10/2017	Ex-Hurricane Ophelia (Met Office, 2018a)	200th (2018: 173rd)	209th (2018: 217th)
28/02/2018—01/03/2018	Snow and low temperatures, dubbed 'Beast from the East' (Met Office, 2018b)	356th, 361st (2017: 188th, 208th)	356th, 362nd (2017: 244th, 130th)
14/06/2018	Storm Hector (Met Office, 2018c)	283rd (2017: 312th)	290th (2017: 218th)
06/2018—07/2018	Summer heatwave (Met Office, 2018d)	Average Rank June 2018: 198 July 2018: 143 (June 2017: 265) (July 2017: 177)	Average Rank June 2018: 165 July 2018: 163 (June 2017: 225) (July 2017: 194)
19/09/2018—21/09/2018	Storms Ali & Bronagh (Met Office, 2018e)	191st, 63rd, 44th (2017: 219th, 111th, 49th)	337th, 139th, 147th (2017: 118th, 115th, 71st)
15/12/2018	Storm Deirdre (Met Office, 2018f)	96th (2017: 18th)	4th (2017: 3rd)

Some extreme weather events did not have much impact on footfall rankings for Edinburgh Old Town and New Town, such as Ex-Hurricane Ophelia or Storm Hector. Storms Ali & Bronagh negatively impacted Edinburgh New Town but not Edinburgh Old Town. In contrast, Storm Deirdre impacted Edinburgh Old Town but not Edinburgh New Town.

However, similar to Liverpool ONE and Manchester Market Street, the 'Beast from the East' decreased footfall significantly for both micro-locations. This impact lasted longer and was even more pronounced for Edinburgh, as can be seen through the corresponding bands of blue in Figure 5-21 and Figure 5-22. The period of bad weather reduced footfall in Liverpool and Manchester to 10-20% below the median, however, this was closer to 50% in Edinburgh, with both micro-locations experienced half the annual median for footfall. This could be due to Edinburgh's more northerly location, resulting in even cooler temperatures and snow, or the result of public transport cancellations deterring tourists and locals from visiting the city.

Local events or festivals

Figure 5-24 shows how footfall ranking varies during the month of August for Edinburgh New Town and Old Town. This coincidences with the main festival season for the city. Almost all the busiest days for Edinburgh Old Town were during this time, compared to 0-10% for Edinburgh New Town.

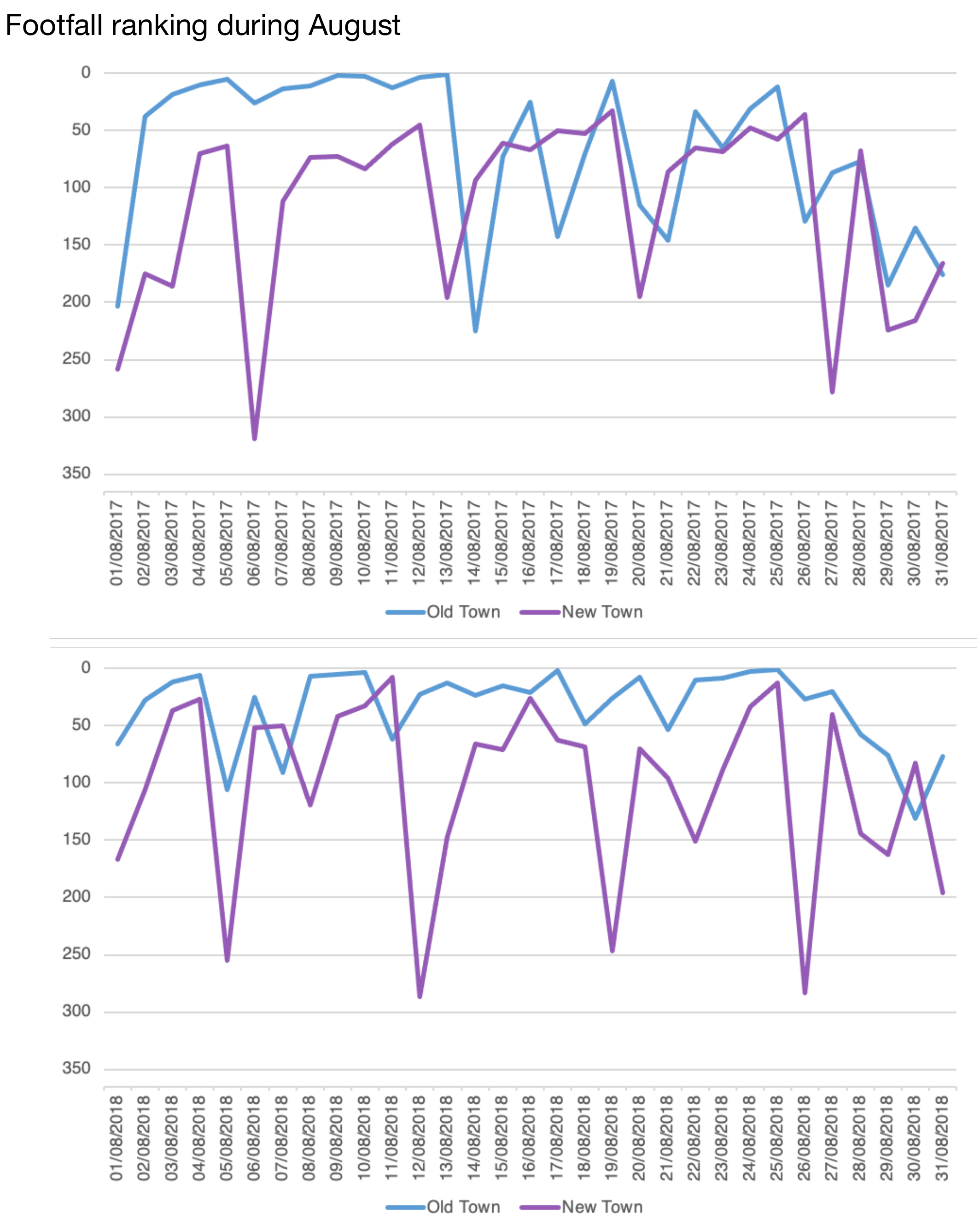


Figure 5-24 Footfall over the festival month of August for Edinburgh Old Town and Edinburgh New Town

The festival is located in areas around the city, but Outdoor stages for the Fringe Festival are located in the Old Town and in close proximity to the sensors. These would increase footfall and dwell time for the area significantly as visitors would stop by to enjoy the entertainment. The tourist appeal of the retail and leisure outlets in this area could keep visitors there. The festival also increases footfall for Edinburgh New Town, but this is not to the same extent.

5.4 Discussion

The last section described the results of the annual rankings for Liverpool ONE, Manchester Market Street, Edinburgh New Town and Edinburgh Old Town. The four case study micro-locations were uniquely impacted by national holidays, local festivals and extreme weather events. In some cases, this could be linked back to micro-locational and intra-regional characteristics, providing a possible explanation for similarities and differences between locations.

This section aims to fulfil the last research aim of this chapter by exploring trends and similarities between different events and what they imply about retail footfall.

Table 5-10 summarises the distribution of the top 20, or top 5%, of daily footfall for 2017 and 2018 for each case study micro-location in terms of the event or lack of event proposed as a significant driver of footfall on that day.

Table 5-10 Summary table of top 20 (top 5%) of the annual footfall rankings for each case study micro-location

Events	Liverpool ONE		Manchester Market St		Edinburgh Old Town		Edinburgh New Town	
	2017	2018	2017	2018	2017	2018	2017	2018
Christmas	17	16	17	9	3	1	18	15
Festivals & Local Events	0	1	3	4	14	17	2	2
School & Bank Holidays	1	2	0	4	1	0	0	0
Sporting Events	1	0	0	1	0	0	0	1
No Event	1	1	0	2	2	2	0	2

The Christmas period including Black Friday and Boxing Day Sales appeared in the top rankings for every case study micro-location. The dominance of Christmas as a footfall event is expected and highlighted further by the extended duration of the period, spanning a month from the end of November to the end of December. However, the nature of Christmas period footfall does vary between the case study micro-locations.

Edinburgh New Town and Liverpool ONE can be identified as the most reliant on Christmas period footfall. The chain and comparison retail offer attracts shoppers who are looking for a range of non-essential goods and gifts, and the presence of Christmas markets increases the potential for leisure or experience-based footfall. This is also the case for Manchester Market Street in 2017. However, in 2018, this impact was weaker likely as a result of poor weather and consumers choosing alternative options. Edinburgh Old Town tended towards Christmas period footfall the least, with the festival season in August pushing the traditionally busy Christmas days down in the ranking.

The prevalence of Black Friday footfall also appears to be locally unique, with Liverpool ONE and Manchester Market St experiencing very high footfall on these days, while it features lower down the ranking for Edinburgh New Town. Black Friday is still a relatively new concept in the UK, introduced in 2014, and there has been some criticism of the validity and value of the deals marketed towards consumers. It is yet to cement itself nationally as a day of high footfall across the country, but based on the data for 2017 and 2018, it has managed to create an impression on the chain and comparison retail centres of the North West.

Weather events are often used as an explanation for low footfall and sales by retailers, and the evidence found in this investigation provides some support of this. The unusually warm weather in the summer of 2018 did correlate with increased footfall in Manchester Market Street. Similarly, the cold spell in February/March 2018 decreased footfall significantly for these micro-locations. However, the impact is not always uniform. Manchester Market Street disproportionately suffered the effects of Storm Deirdre during December 2018, whereas Liverpool ONE and Edinburgh were less impacted.

Potentially a positive for organisers, town planners and business improvement districts is that local festivals and events can generate the same levels of footfall that Christmas does, throughout the year. The clearest example of this is Edinburgh Old Town, with the August festivals overtaking Christmas in terms of footfall and retaining high footfall for a prolonged amount of time. However, Manchester Market Street also manages to achieve this with its Food Festival and Halloween in the City events, and Liverpool ONE with its Giants Spectacular. Sporting events also correlated with high footfall days, particularly for Liverpool ONE. It is noteworthy that other micro-locations managed to have an event that correlated with many of their busiest days, demonstrating how important events can be in understanding the trends and fluctuations in footfall. Many of these popular events all take place in the latter half of the year, from July to October, taking advantage of school holidays and the chance of warmer weather.

Whereas this might not directly correlate to spend to the same extent as Christmas footfall, there are social and cultural benefits to higher footfall. Local events and festivals can help to market a retail centre and could help to create a sustainable and attractive image. There is a wide array of research and studies into place identity and place making that solidify a case for this (Derrett, 2003; Hartley, 2012; O'Sullivan and Jackson, 2002). A solid image and identity could help build a strong community, which is key to the vitality and viability of a retail centre (Portas, 2011).

5.4.1 Limitations

The analysis in this chapter has several limitations that must be considered. Firstly, it has a narrow spatial and temporal scope, which restricts the applicability of the results. Four uniquely contrasting case study micro-locations were included in the sample over two years – 2017 and 2018. However, the locations only represented two of the fifteen retail groups identified in the CDRC Retail Centre Typology. The window of two years limited the events and factors that could be investigated in this study. Unfortunately, these constraints were unavoidable due to data availability. Nonetheless, it leaves future research opportunities to explore other time periods and the impact of local, regional, and national events on locations with a different retail offer, such as market towns, tourist locations, or failing town centres.

Secondly, as discussed in the methodology section, to generate a consistent and reliable two-yearly footfall dataset, the readings of several sensors needed to be aggregated together. This aggregation minimised the impact of localised disturbances, such as crowds outside a bus stop or a particular retail outlet and fluctuations due to measurement error in one sensor. However, this analysis still assumes that they are generally capturing the same footfall trends, and a busy day for one sensor is also a busy day for another sensor nearby. As Figure 5-2 to Figure 5-5 show, the sensors chosen were close in proximity and network distance. Still, intersections and connecting streets could influence each sensor differently, and this complexity is not ingrained into the analysis.

Thirdly, to analyse the trends in the data, an understanding of influential events in 2017 and 2018 had to be achieved. This proved difficult in some instances. For example, some websites or news outlets may have taken down information regarding past local events or festivals, including their exact date. In addition, some events which may have a significant impact on footfall might not be documented in any traceable way, for example, individual retailer sales or political demonstrations. No one explanation could be found for a sizeable minority of top-ranking footfall days. A potential reason for this could be that footfall relies on human decision-making and that there is always an element of unpredictability.

Finally, although this study investigated the relationship between different events and footfall, this is not equivalent to spending. Whereas Black Friday and Christmas may generate a healthy correlation between footfall and spending, some events may be more economically beneficial to nearby retailers and the local economy than others. For example, a food festival may result in local restaurants or fast-food outlets losing income that day, as consumers spend their money at the event instead. While local events and festivals might increase social and cultural sustainability, drawing consumers in and providing an enjoyable experience, the immediate economic impacts could be limited.

5.5 Chapter Summary

Three objectives were established at the start of this chapter. These are summarised below:

What happens there	Identify events which significantly impact footfall.
	Investigate how factors of both the immediate environment and in the wider context could influence this impact
	Explore the trends and similarites between footfall of different events in different locations and what they could imply about retail footfall

The first objective was to identify events which have a significant impact on footfall. Using four case-study micro-locations, daily footfall was calculated and ranked annually for 2017 and 2018. The unusually busy and unusually quiet dates were investigated to determine if they were influenced by temporary events. National holidays, shopping events, weather events, sports events and local festivals were all found to drive footfall significantly across the case study locations.

Section 5.3 investigated how the context each micro-location was in may have influenced the relationship between events and footfall. Some trends were identified: for example, Manchester Market Street appeared less resilient to weather events which could be potentially related to its function as a retail destination, and locations with a chain and comparison retail mix experienced a higher peak of Christmas footfall. However, there generally seemed to be little correlation between the characteristics identified and the fluctuations observed in the limited sample.

It was a challenge to achieve the final objective and draw any meaningful conclusions from the data available. Some trends were identified; however, there are many limitations which challenge the generalisability and applicability of any conclusions. The data sample was limited in accuracy and scope, and it was a significant undertaking to find out which events did or did not take place on certain days in the past, and there was little to no information on that sentiment or reception of these events. These could potentially be significant factors in how events translate into footfall.

In conclusion, the impact of different local, regional and national events on retail footfall varies can relate to the character and size of the retail area. This study has produced four illustrative examples of the strength of this variation, how it can be linked to retail offer and high street identity and how retail centres may experience national events such as Black Friday and Christmas differently.

For many years, the temporal limitation on collecting footfall data has constrained understanding of footfall as a measure of high street sustainability and successfulness. However, the ability to collect continuous footfall data allows new insights into how prevalent and important events are in attracting people to retail areas. This research has proven the ability of Christmas to drive footfall and investigated how its influence differs from place to place. It has also shown how locally run events and festivals can produce similar levels of footfall throughout the year. It has also highlighted the prominence of Halloween and sporting events, drivers of footfall that may have been overlooked in the past. This opens avenues of further research into how event driven footfall impacts other retail centres across the UK and demonstrates some of the benefits and capabilities of continuous footfall data to strengthen and build sustainable high streets.

6 What remains unknown

Investigating the potential for a spatio-temporal prediction model for footfall data

What remains unknown	
	Define the criteria and use case for a footfall prediction model and identify appropriate methodologies to achieve it
	Create a preliminary model that predicts footfall that is location and time dependent
	Critique the performance of this model, identifying opportunities for improvement

What remains unknown

A measure of footfall for a retail centre or high street can be a powerful tool with many applications. As discussed in Chapter 2, footfall has many applications as an indicator of retail vitality, viability and performance (Wrigley and Lambiri, 2014b). It can be integrated into effective operational and marketing strategies for local businesses (e.g. Kirkup and Rafiq, 1999; Denison, 2005) and used to inform location planning of stores and to understand how pedestrians flow around a city. In addition, footfall can be applied to help determine charges such as business rates and rents to ensure that they are fair and comparable and to make informed decisions in retail policy. Footfall data has potential to provide useful insights with wide applications for a variety of stakeholders.

However, obtaining reliable and large-scale footfall data can be a challenge. Technological advances have made pedestrian counts more measurable and monitorable than ever before, but these methods can require a significant investment of time, money and skills before the data is applicable. Manual counts are a more accessible alternative, but they can only provide snapshots of footfall that may not be representative of other times or of other streets.

This chapter aims to explore one potential solution for this: applying machine learning methods to a database of previously collected footfall data from across the country to predict footfall for similar retail locations where it is not monitored. As has been discussed previously in this work, footfall is inherently complex and relies on human decision making, which is challenging to model and predict. However, results have also shown that elements of temporal and locational context can correlate with footfall, demonstrating that they may have some predictive potential which could be harnessed to provide a footfall benchmark for any retail centre in the UK which considers both locational and temporal context. This could be invaluable to retail centres that cannot regularly measure their own footfall. This chapter will investigate various options of how this could be achieved, providing preliminary insights into this novel and complex challenge.

It is challenging to review, compare or categorise the wealth of research on footfall and pedestrian prediction due to its heterogeneousness. Current research into footfall prediction and pedestrian modelling is wide and disparate, spanning across disciplines and exploring a variety of research goals. For example, some works focus on increasing pedestrian safety (e.g. Schneider *et al.*, 2012; Munira and Sener, 2017; Liu *et al.*, 2021). Others investigate the impact of different factors on footfall, such as weather or structural changes (e.g. Makkar, 2020; Cooper *et al.*, 2021; Martínez-de-Albéniz and Belkaid, 2021). Some research has a methodological focus, exploring the predictive capacity and accuracy of a certain technique for footfall prediction (e.g. Trasberg, Cheshire and Longley, 2018; Stavroulaki *et al.*, 2019; Chen and Zhou, 2020).

Research can also vary in terms of spatial scales, with some focusing on very small scales, while the majority focus on one city, either modelling locations within that city (e.g. Klapka *et al.*, 2013; Wang *et al.*, 2017; Chen and Zhou, 2020) or modelling footfall for an entire street network (e.g. Omer and Kaplan, 2017; Cooper *et al.*, 2021; Liu *et al.*, 2021). There are limited studies that instead focus on modelling footfall across several cities or countries.

Research into footfall analysis and modelling can also have different temporal focuses. Some works focus on making a singular prediction of footfall that is time independent, such as Klapka *et al.* (2013), whereas others apply time series analyses to make time dependent predictions for a location(s). Examples include Wang *et al.* (2017), Chen and Zhou (2020), Joshi *et al.*, (2021) and Liu *et al.* (2021).

The research goal, discipline, scale and availability and resolution of data can call for different approaches, methods and considerations for footfall modelling. A popular approach is to create a model based on pedestrian demand and features of the street or road network. In some transport planning works, these are called direct demand models (Cooper et al., 2021; Munira and Sener, 2017), however this naming is not consistent. Works by Stavroulaki *et al.* (2019), Bolin *et al.* (2021) and Sevtsuk (2021) are some examples that emulate this approach, but do not use this term.

These models aim to predict footfall using features of the network and popular origins or destinations, or measurable variables that might impact pedestrian demand over space. These are often applied within a certain context, successfully predicting footfall for one city or area of a city and using local knowledge to inform the choice of variables. However, that does limit the generalisability of these approaches to environments outside those investigated. For example, a footfall prediction model created for London may not be applicable to predict footfall for another city. Stavroulaki *et al.* (2019) is one of the few studies that tackles this issue, applying and comparing footfall prediction models for neighbourhoods in London, Stockholm and Amsterdam. They combined Space Syntax network analysis with elements of the environment, such as proximity to local schools, transport hubs and markets, and found that their chosen variables were significant in predicting footfall in all cities, showing that there could be potential for a generalisable footfall prediction framework[20].

Another methodology with a slightly different focus is time series modelling. Whereas direct demand and network analysis approaches put emphasis on predicting footfall over space, time series models aim to predict footfall accurately over time. In their simplest form, these models use past data collected for a location and use it to forecast what footfall might be in the future, however some can account for exogenous factors. For example, Chen and Zhou (2020) combined footfall time series data with weather variables to forecast footfall for nine Springboard sensor locations in Bath. Similar to direct demand and network analysis modelling, time series prediction models are not made to be generalisable. They aim to optimise the predictive capabilities for one single location from a wealth of data over time and directly applying the same model to another location may produce incorrect results.

In summary, footfall prediction research can generally be separated into two categories: those that predict across space and those that predict across time. Those

[20] It should be noted that (Stavroulaki et al., 2019) was not exclusively retail footfall focused.

that predict across space tend to focus on one city or area of a city during a set period. Those that predict across time tend to focus on one point location with a footfall sensor and historic data. As of writing, there has been few attempts to combine these predict footfall for both space and time in a real-world context. Recent works such as Lugomer and Longley (2018) and Mumford *et al.* (2021) have approached the challenge of spatial-temporal representation of footfall by developing novel and advantageous classifications of places based on time series data. While these present valuable insights into the temporal patterns of footfall across space, they have not yet been applied to a prediction task. Whipp *et al.* (2021) also acknowledges this research gap, evaluating conventional and novel data sources and their potential applications in estimating fluctuations in ambient populations in future research. Bolin *et al.* (2021) modelled footfall for London, Stockholm and Amsterdam and incorporated a time element within the model; however, this was based on limited data from three weeks during one month of a year.

This analysis in this chapter contributes to the literature and the relevant debates by developing a model that predicts footfall to address level over space and time and is generalisable across multiple retail centres in the UK. To our knowledge, this is the first attempt to design a model with this ambition on this spatial and temporal scale.

The original aspiration for this research was to design a model that could utilise the temporal resolution of the historic time series data to predict future footfall, adjusting the predictions as more data was made available. Unfortunately, this has not been possible to achieve in this case due to data limitations and the exceptional impact of the COVID-19 pandemic. The final model intends to utilise the existing footfall data, harnessing the temporal resolution and unparalleled spatial coverage to give insight into the distribution of footfall over time. Although an exact date or year cannot be entered into the model and accounted for in the prediction, the model does include a temporal dependent element while making footfall predictions which is novel when also predicting footfall over space.

This analysis also presents new insights into the generalisability of footfall prediction across towns and cities. The same technique is applied to predict footfall for retail centres across the UK, and how it performs in different cities and towns could give valuable insight into how generalisable these methods are. Additionally, three more models are built, each predicting footfall for one of the three micro-locational classes defined in Chapter 4: Chain and Comparison Retail micro-locations, Business and Independent micro-locations and Value-orientated Convenience Retail micro-locations. Chapter 4 proved that footfall magnitude and signature patterns varied between these classes. Therefore, the hypothesis is that models that are built to fit each class would be more accurate in predicting footfall than one generalised model.

Footfall can be challenging to predict as functional and morphological characteristics have different impacts depending on the spatial and temporal context. In addition, it is somewhat reliant on personal preference or choice. Attempting to model human behaviour will inevitably produce some error as people do not always behave in logical and predictable ways. It is estimated that only 60-70% of footfall is predictable, and this is based on studies that often focus on a snapshot of one city or neighbourhood (Jiang, 2009). A generalisable spatio-temporal model across multiple towns and cities, such as the one developed in this chapter, might be expected to have a higher error. This analysis will also provide valuable insights into the extent to which this is the case.

As discussed previously, measures of retail footfall have a vast number of applications for businesses, local planners and multiple stakeholders, however automated footfall data collection requires significant investments. A model such as the one investigated and introduced in this chapter allows retail locations that may not have the means to this investment some access to the benefits footfall data can provide. Although this figure might have limited accuracy, it provides a benchmark for footfall that is valuable in its coverage, novelty and accessibility.

Three objectives were established for this chapter. The first objective is to define the criteria and use case for a footfall prediction model and identify appropriate methodologies to achieve this. Section 6.1 will present the purpose and proposed operation of a footfall prediction tool, exploring the different methodological approaches which have been applied.

The second objective is to build a preliminary model. This model will be novel in its inclusion of both spatial and temporal features and predictions. It will also be designed to predict footfall for any retail address in the UK, which is novel in terms of generalisability. Section 6.2 will derive the data used for the model, and Section 6.3 will detail how the chosen method (random forest regression) will be applied to the data.

The final objective is to critique the performance of this model, identifying opportunities for improvement. This will be explored in Section 6.4, which will explore how well the model has performed on the data, the strengths and weaknesses and where the opportunities for future developments are in spatio-temporal footfall prediction.

6.1 Model design

Utilising the unprecedented coverage and amount of footfall data made available through the SmartStreetSensor project, this chapter will introduce different modelling approaches that can be applied to produce a time-dependent benchmark for footfall for retail addresses across the UK. The ideal outcome would be a model that is able to produce a reasonably accurate prediction of footfall, and which could provide insights that could be utilised in further research to build a complete tool or program. This objective has not been approached on this scale and for this purpose by any known literature, and the novelty presents a challenge when deriving a methodology but also an opportunity to establish a new framework.

This section will give an overview of the modelling and testing process. There are three sections, each determining the 'What?', 'Why?' and 'How?' of the model proposed in this chapter.

First, Section 6.1.1 will introduce the use case, guiding through how a user would interact with the final tool that this model aspires to. Then Section 6.1.2 will explore in greater depth the rationale behind this tool and applications it may have for potential users. Section 6.1.3 will then investigate the different methodological approaches that have been used for modelling footfall in a variety of studies and identifying which would be best for this purpose.

6.1.1 Use Case

Although this research is preliminary and focuses on investigating and presenting different methods that could be applied to construct a final tool, it is still important in development to consider how an end user might want to apply the result. In software development, this can be called a 'use case'. The use case details the 'what' of a potential model or program - what are the inputs, what are the processes and what are the outputs? Who are the users and what purposes may they have for applying this tool? These are all important considerations in the preliminary stages of development.

For this model, potential users could be business owners, decision makers or high street stakeholders who want to find out more about the footfall for a specific retail address. They would input this location and the model will give them an average hourly footfall count for that address. In addition, they might want to know this for a specific time of the day, week or year. When specifying this, the model will give them a prediction of footfall for their given retail address for that temporal context. These footfall benchmarks could help inform policy decisions, location planning, marketing approaches and many more applications.

To summarise, the model has four potential inputs,

- ◊ A retail address
- ◊ An hour of day
- ◊ A day of week
- ◊ A month of year.

It processes these inputs to give a resulting output: a footfall prediction that takes into account the spatial and temporal context. This will be fed back to the user. Further useful outputs may be graphical representations of the patterns of footfall they might expect to see in this location for that week or for months of the year, or heat maps that show how footfall varies within the nearby micro-locational context. This data will help the end user achieve and contextualise the footfall their location of interest is experiencing, compared to other locations and how this varies with time.

6.1.2 Rationale

A model that can give a time and location dependent benchmark for footfall has many practical applications for high street stakeholders. For example, it could be used as a key performance indicator to monitor the impact of any revitalisation strategy and the insights could applied to develop retail policies (Coca-Stefaniak, 2013; Graham, 2016; Hogg et al., 2004; Ministry of Housing, Communities & Local Government, 2014). As a measure of potential for a retail street, that could be used to inform business rates or rents and it could help business owners set opening times, staffing hours and plan marketing campaigns (Denison, 2005; Underhill, 2009; Yiu and Ng, 2010). It could also inform location planning for the location and opening of new stores. For local decision makers, it could provide an efficient and quick overview of their retail centre and how it compares to other retail centres nearby and consumers may be interested to know when certain areas might be busier in order to plan a shopping trip.

In addition to the potential practical applications of a model like this, it also contributes to the research and literature into footfall and pedestrian flow prediction. The analysis in this chapter is novel in two ways.

Firstly, it tests the generalisability of footfall modelling. The majority of analyses in current literature focus primarily on one location or retail environment and design a methodology which incorporates the factors that are deemed necessary for that specific analysis. Very few then investigate if their model would be accurate when applied to other locations and contexts. By applying the same methodology and model across multiple retail environments across the country, this chapter will test whether footfall prediction models could be generalisable.

Second, the model incorporates temporal prediction, adjusting the footfall value dependant on factors of the time of the day, the day of the week and the month of the year. This was incredibly difficult to include in studies prior to the technological advances in automated footfall collection, which have facilitated the collection of high-resolution data.

Many footfall prediction studies since have utilised sensor-collected footfall data, however these often focus either on predicting footfall across space or predicting footfall across time, instead of both. For example, Stavroulaki *et al.* (2019) created a model that could predict footfall in different spatial and neighbourhood contexts, but not over time, using just a snapshot of three weeks of high resolution footfall data. On the other hand, works such as Chen and Zhou (2020) apply time series analysis to predict footfall over time for nine different locations in Bath, utilising the wealth of historic data available to learn trends and patterns. However, their models are individual to the locations the time series models were made for and one model couldn't necessarily be applied to another location. The research presented in this section is novel as it combines elements of time series analysis with spatial prediction to make footfall predictions that are both temporally and spatially dependent.

6.1.3 Methodological approach

As discussed in the introduction to this chapter, there is no agreed consensus on an approach to take when modelling footfall. For many studies, the approach is tailored to the specific research question, dataset or context being investigated. In addition, very few studies incorporate both spatial and temporal elements into their analyses, or apply the analysis to multiple cities or locations as this chapter aims to do. This is often due to data restrictions as the pedestrian counts are completed manually or over a short window of time, but as the continuous automated collection of footfall data becomes more popular, the demand for spatio-temporal analysis and predictive frameworks might grow in future.

This section explores some potential options for methodological approaches, critiques their application for this use case and identifies some challenges and pitfalls that are presented when these are applied to the SmartStreetSensor dataset. It is unknown whether these insights would also apply to similar footfall databases, such as Springboard UK, but this a potential route for future investigation.

Agent-based modelling

Agent-based modelling is a model that simulates the decisions, actions and interactions of autonomous agents. Applied for the purpose of footfall prediction, it would model each individual person and the predicts decisions they might make as they travel around a simulation of a retail centre. The number of times the agents pass a certain location would be counted and that would result in the footfall prediction. Often, an agent-based modelling approach is applied on a very small scale, such as an intersection or a shopping area (Hoogendoorn, 2004; Kitazawa and Batty, 2004). This approach can work well in this context because it can be account for the complexities, nuances and details that are important to consider when focusing in on a small scale. However, it may prove exceptionally challenging to construct an agent-based model that covers multiple different retail centres across the country and accounts for the time. Therefore, it will not be applied in this instance.

Time series analysis

With technological advances making continuous footfall data collection more accessible, the impact of time on pedestrian counts can be studied and understood in greater detail. Several studies have utilised time series modelling to understand the footfall patterns for a certain location and used that to forecast future footfall for those locations. In their simplest form, these models use past data collected for a location and use it to forecast what footfall might be in the future, however some can account for exogenous factors.

Time series analysis and modelling has many powerful applications. As many footfall measuring devices, such as those installed by Local Data Company, are used by the client who owns the shop, being able to understand footfall in-depth for their location is a major advantage for investing in a sensor. Time series forecasting holds many benefits for this client, as it learns the past footfall data of their store and uses it to give a well-informed prediction of future footfall. However, time series models are not designed to be generalisable. They aim to optimise the predictive capabilities for one single location from a wealth of data over time and directly applying the same model to another location may produce incorrect results.

There is a lot of potential in time series analysis for applications in footfall prediction. However, for a use case that requires generalisability and a model that can be applied across many locational contexts, time series analysis is not easily suited. In investigating the different methodologies that could be applied to construct this model, time series was thoroughly explored. From constructing separate ARIMA and SARIMA models for each location, to using the principles of time series decomposition[21] to try and detect patterns between different sensors, time series analysis was applied exhaustively to this dataset. However, it was clear that it, in this case, it would not produce generalisable and applicable results. This is thought to be a result of two major factors: the quality and resolution of the data source, and the variability of footfall itself.

Firstly, the limited tenure of a large proportion of the sensors presented many limitations for time series analysis. Depending on the method applied, between two and six complete cycles are needed, at minimum to accurately capture seasonal trends (Hanke and Wichern, 2005). In this case of annual cycles, this would require at least two years of complete data, which many of the sensors did not have.

Secondly, the changes to MAC address randomisation and security discussed in Chapter 3 were detrimental to the comparability of the data over time, and the

---

[21] Time series decomposition breaks down a time series into several components that are representative of underlying patterns. These are commonly the trend, seasonal and remainder components, however holiday effects can be included also. The trend component represents the secular variation of the time series that shows the long-term change of the average value. The seasonal component represents the effect of periodic fluctuations, for example, time of the day or month of the year. The remainder component is the noise that is left over once the trend and seasonal components are removed.

sustainability of the project. The removal and decommissioning of the sensors throughout 2019 and 2020 not only impacted the temporal range of data available, but also presented a question for the long-term viability of the project using Wi-Fi based methods. There was limited benefit in creating a footfall model that could forecast future footfall based on recent data if there would be less recent data.

Thirdly, the COVID-19 pandemic has had an exceptional impact on footfall. From March 2020 onwards, the way UK consumers shopped fundamentally changed to accommodate the national lockdown. Events were cancelled, travel was restricted, and non-essential retailers were forced to close their stores. Across 2020 and 2021, retail footfall was highly controlled by government policy and social attitudes, both negatively and positively. For example, the Eat Out to Help Out policy during August 2020 was put in place to encourage people back to the high street to help boost the economy and retail footfall (Kollewe, 2020). At time of writing, it still remains to be seen how the COVID-19 pandemic might impact footfall in the long-term, but in the short-term, it makes the task of forecasting future footfall significantly more challenging. This is just one example of the huge variability of footfall which limits the application of time series analysis methods on a generalisable scale. Due to this, and other limitations mentioned, time series analysis was not applied in this chapter. However, there are avenues for future research in using this method, combined with spatial modelling methods.

Space Syntax

Space Syntax is a network analysis technique which is developed specifically for the purpose of modelling cities and urban environments for a range of planning and architecture applications. It was developed by Hillier and Hanson, (1984) and has been applied to predict footfall in a range of cities from the Netherlands to South Korea (Lee et al., 2020; Read, 1999). The method works by representing the street network through axial maps, which plot the visual sightlines of a pedestrians across open space. Network analysis metrics of these maps have been shown to be good predictors of footfall. In a literature review that collated Space Syntax investigations of individual cities, Jiang (2009) showed that 60-70% of pedestrian flows could be predictable using this approach.

When modelling footfall, space syntax is typically applied to a much larger area than agent-based modelling, to the neighbourhood or city level, which makes it more fitting for the use case defined in this chapter. However, the process of creating these axial maps is time-consuming and often needs reliable data and good insight into the on-the-ground environment. This doesn't present an issue when looking at one location, or three cities as in Stavroulaki *et al.*, (2019) but can be an obstacle for analysis that includes significantly more than three cities, as this analysis aims to. It would be a time-consuming and computationally expensive task. In addition, for the long-term aim of this model as a tool to predict footfall for any retail address, the process has to create an axial map for a user-entered location reasonably quickly. Innovative new research is underway to explore ways to make the process of generating axial maps easier and more automated, (e.g. Liu and Jiang, 2012), so, although a space syntax approach is challenging to apply currently, it could be more achievable in the future.

Another consideration with a space syntax approach is that it there is no intuitive way to incorporate time. Space syntax would have to be used alongside a regression model, such as a direct demand model, to achieve this.

Direct demand modelling

As discussed in the introduction, direct demand models involve combining network analysis methods with factors of demand to predict transport flows such as footfall. This is typically done through a regression model where the outcome is the flow (e.g. footfall) and the input is a range of demand factors and network measures that have proven to be good predictors. These network measures could be calculated through space syntax or through other forms of analysis, and the demand factors could be informed by the context or more general, for example population or proximity to a store.

This approach was used in Stavroulaki *et al.* (2019) and in subsequent work by Bolin *et al.* (2021). Despite the wealth of literature on prediction of pedestrian counts, these primarily focus on one location and there are very few studies that attempt to achieve prediction across different cities. One of those is Stavroulaki *et al.* (2019). Aiming to test whether footfall prediction models were generalisable across cities, they modelled footfall in London, Stockholm and Amsterdam applying a direct demand approach to all three. Their model used measures from space syntax with three demand factors (distance to markets, distance to schools and distance to transport hubs) and incorporated time variable controls into their model. Their results showed that this approach was able to model footfall accurately across these cities, proving that footfall prediction could be generalisable across three different cities.

As this research also aims to predict footfall using the same model across different cities, the methodology successfully applied in Stavroulaki *et al.* (2019) would make an appropriate starting point. However, the analysis by Stavroulaki *et al.* (2019) and Bolin *et al.* (2021) does differ from this analysis in several ways.

Firstly, it considers neighbourhoods within three cities that are relatively similar in terms of hierarchy. London, Stockholm and Amsterdam are all well-connected capital cities of international importance. Conversely, the sample of towns and cities in the dataset used in this analysis ranges from major cities such as London and Manchester, regional centres such as Wakefield and Gloucester, to smaller towns. It is unknown whether the insights found in Stavroulaki *et al.* (2019) will also ring true when attempting to model cities of different sizes and importance, but the analysis here will test this. Population is included in the model to attempt to control for the different sizes in retail centres.

Secondly, the neighbourhoods surveyed by Stavroulaki *et al.* (2019) contained both retail and non-retail environments. Their chosen demand factors reflect this, for example, selecting proximity to schools as a major demand. This analysis solely focuses on retail areas, and the demand factors are adjusted accordingly

Stavroulaki *et al.*, (2019) also use network analysis methods from space syntax axial maps. As discussed previously, space syntax becomes less viable as an approach with more locations as more axial maps need be generated and this is a time consuming, computationally expensive and subjective task. The steps involved would be similar to,

1. Acquire data that includes the shape of the road network and buildings around it
2. Generate an axial map of this area by finding the sightlines
3. Apply network analysis to these sightlines to calculate the centrality (the number of links a node has) and betweenness (the number of times a node is along the shortest path between two other nodes)

While this is achievable when looking at a few areas, it gets increasingly harder the more locations are analysed, and for the final use case, which could involve calculating these measures on the fly, it is not the most fitting approach.

An alternative would be to use betweenness or centrality of the street network. This differs from angular betweenness and angular centrality as it does not consider the angles of the streets and does not need the generation of an axial map. Instead, it simply relies on a generation of the street network and a calculation of how connected that street is to others around it. The limitation of this is that it removes one of the key properties which make angular measures effective: lines of visibility. The theory that underpins this is that pedestrians prefer to minimise angles in their route planning and will likely choose straight routes above winding ones where they cannot see their destination. However, by including the betweenness centrality of a street will provide some insight into how connected that street is to others around it and capture some elements of the network that could be used to understand footfall.

This chapter will use a direct demand approach adapted from Stavroulaki *et al.*, (2019) to model footfall across the UK.

6.2 Deriving the data

The first step in constructing a model is to derive the data that will be used in training. In this case, it would include the independent variables (population, network measures, demand factors and temporal factors) and the outcome to be predicted (footfall). As the final use case for this model is to be applied as a tool that can predict footfall for any location, it is important that the independent variables are from datasets with good spatial coverage and that they are efficient and quick to calculate. The datasets used to derive these variables are all introduced and discussed in Chapter 3 and this section will detail their derivation as follows.

Section 6.2.1 will detail the derivation of the independent variables including population, network measures, the factors of demand and temporal factors. Then, Section 6.2.2 will describe the cleaning and adjustment process applied on the footfall data as the outcome or dependent variable.

6.2.1 Independent variables

Sixteen independent variables were selected for the direct demand model, under four categories: population, network measures, demand factors and temporal factors. A description for each is given in Table 6-1.

Table 6-1 Variables used, description and justification

Variable	Category	Description	Rationale
Population	Population	Population of the retail centre	To account for differences in retail centre size (e.g. Huff, 1963)
Betweenness/Centrality	Network measure	Measure that determines the importance of that street compared to others in the network	Shown to correlate well with footfall in other approaches (e.g. Stavroulaki *et al.*, 2019; Bolin *et al.*, 2021)
Workplace population	Demand factors	Captures number of daytime commuters	Locations with a high concentration of employers are shown to drive footfall (e.g. Swinney and Sivaev, 2013; Berry *et al.*, 2016)
Anchor Stores		Captures proximity to anchor stores	Anchor stores have been shown to attract and drive footfall for themselves and nearby stores (e.g. Pullens, 2018)
Premium Stores		Captures a certain retail function	Factor for determining retail centre function (e.g. Guy, 1998; Millington *et al.*, 2015)
Entertainment		Captures the proximity to entertainment and leisure venues	Entertainment venues have been linked with resilient retail centres (e.g. Hart *et al.*, 2014)
Vacant Stores		Captures the number of unoccupied stores	Symptom of a lack of demand (Swinney and Sivaev, 2013)
Independent Stores		Captures the ratio of chain stores to independent stores	Factor for determining retail centre function

			(e.g. Guy, 1998; Millington *et al.*, 2015)
Value Stores		Captures a certain retail function	Factor for determining retail centre function (e.g. Guy, 1998; Millington *et al.*, 2015)
Night-time Economy locations		Captures elements of the night-time economy	Strong night time economies have been linked to resilient retail centres (e.g. Hart *et al.*, 2014)
Density of stores		Captures the availability of stores within an area	Size and number of stores and retail outlets can be linked to attractiveness, which can drive footfall (e.g. Huff, 1963)
Transport Hubs		Captures the connectivity via public transport	Locations close to transport hubs have been shown to drive footfall (e.g. Berry *et al.*, 2016)
Car Parks		Captures the connectivity via car	Having good car park access has shown to be a demand of consumers (Coca-Stefaniak, 2013; Tyler et al., 2012)
Hour of day, day of week, month of year	Temporal factors	Adds a temporal aspect which can be user-defined	Previous analyses within this thesis (Chapter 4) have shown time is crucial in how different characterisitcs interact with footfall
Holiday		Considers whether or not the day is a holiday or not (omitted due to lack of data)	Previous analyses within this thesis (Chapter 5) have shown public holidays can have an

			extreme impact on footfall

These were all derived from datasets introduced in Chapter 3 and the methodology for this is detailed in the following sub-sections.

Population

Population was included in the model as a measure of potential visitors to a retail centre, in addition to controlling for the different sizes of retail centre within the sample. The measures of population were derived from ONS mid-2018 populations estimates for Output Areas in England and Wales and the mid-2018 National Records of Scotland population estimates for local authorities in Scotland. These sources have good data coverage and are kept up to date.

However, there a challenge in this as, though both the ONS and National Records of Scotland data cover the same time period, the area size for the population estimates are very different. The output areas as used for England and Wales are significantly smaller than the local authorities available for Scotland. There are over 180,000 output areas in England and Wales and just 32 Scottish local authorities.

To overcome this, retail catchments were drawn for each retail centre, and the percentage of output area or local authority that overlaps this area is used to calculate the population. This does rely on the assumption that the population is spread evenly among an area, which remains a significant assumption to make; however, this allows the population estimates from the ONS for England and Wales and the ones from the National Records of Scotland to be comparable.

There are several different methods for determining what the size of the retail catchment should be. Applying Retail Gravitation Theory, a catchment should encapsulate the population which that retail centre is the nearest, with the proximity of surrounding retail centres determining the spread of the catchment. Research that has built on that theory states that attraction of a retail centre is a key element to be considered and that the specific offer and size of a retail centre may impact the catchment size. However, these theories do not consider the today's omnichannel retail context; a consumer may not travel a significant distance to buy a product if they could purchase it online from their own home.

To determine the size of a retail catchment, and therefore the population served by that retail centre, six different measures of distance were tested. The best measure was determined to be the one that correlated best with footfall. The retail centre boundaries and the Group and Subgroup typology were from Dolega *et al.* (2021). The measures that were tested were:

- ◊ Zero distance — The population within the boundary of the retail centre.
- ◊ Walk distance — The population within 'walkability' distance of retail centre boundary (400m).
- ◊ Drive distance — Population within 20 minutes' drive of the retail centre (assuming 30mph) (16km).
- ◊ Competition distance — Population within distance to the nearest retail centre/2.
- ◊ Group competition distance — Population within distance to the nearest centre of the same Group/2.
- ◊ Subgroup competition distance — Population within distance to the nearest centre of the same Subgroup/2.

Using correlations can be a crude method of feature selection, as it can mask elements of a variable that might only show as useful when included within the model itself. A more comprehensive method would be to construct multiple models, each with a new combination of features and then test which was the most accurate and reliable. For ease of analysis, and as this is more a preliminary and investigative task, correlations were deemed satisfactory for the scope of this paper. However, when generating a final instance of a model such as this, testing each iteration within a model would be an appropriate method to use.

The correlations between each measure and footfall are given in Table 6-2.

Table 6-2 Correlations between different population catchment measures and mean footfall

Catchment Measure	Correlation with footfall	P-value
Zero distance	15.3%	<0.01
Walk distance	14.5%	<0.01
Drive distance	5.2%	0.17
Competition distance	12.0%	<0.01
Group distance	12.6%	<0.01
Subgroup distance	31.7%	<0.01

The population catchment size that correlated most with footfall was one which considered the distance to the nearest retail centre within the same subgroup of the typology outlined in Dolega *et al.* (2021). Although still relatively weak, this correlation was significantly stronger than the other catchments, therefore subgroup distance was chosen to calculate the population catchments used in the model.

Network measures

The network measures that were considered for this model were betweenness and degree centrality. As explained previously, degree centrality is the simplest measure of centrality, and captures how many nodes are connected to one node. Betweenness centrality instead captures how often a node is used in the shortest path between other nodes. These were the same measures as were used successfully in Stavroulaki *et al.* (2019) and Bolin *et al.* (2021). The key difference is that, instead of calculating these on an axial map generated of the area of interest, these are created from the road and street centrelines.

The road and street network used was from OpenStreetMap and was brought into python using a library called OSMNX (Boeing, 2017). The network analysis was completed using NetworkX and the entire python code was integrated into the R code used for the rest of the model using 'reticulate'. This means that, although different coding languages and programs were used to generate the model, they can both be integrated into each other. To make this process as streamlined as possible, these two measures were combined in Python before being transferred back to model in R. Therefore, it the model results they are combined together under the title 'centralities'.

Demand factors

Estimating and understanding the demand for a retail centre can be complicated and nuanced. The purpose and intent of an individual trip is difficult to predict, and individual retail centres or micro-locations may have their own unique factors of demand. Factors of demand can also be dependent on a mix of qualities within a retail centre, for example, if a consumer wanted to go to a cinema and have dinner afterwards, they are looking for a location to visit that has both entertainment and restaurants and not one or the other (Wrigley et al., 2009).

Hart *et al.* (2014) conducted a qualitative investigation into how consumers interact with retail environments which gives insight into the complexity of these decisions. They state that around 60% of journeys through a retail centre are habitual, with many people following a similar route, but the reasoning behind this varies from person-to-person, giving "parking, purchase convenience, familiarity, safety concerns and … the need to shelter from bad weather" as some examples. As it is complex and context-dependent, retail centre demand can be difficult to capture in data and to generalise across different towns and cities.

Many factors of retail demand have already been explored and defined in previous chapters. In Chapter 2, a range of characteristics on different scales were identified from the literature as impacting footfall. Chapter 4 defined different micro-locational influences on footfall such as transport hubs, anchor stores and working population. In addition, Chapter 5 explored population and connectivity as mediating factors on the impact of events and festivals. This builds off the principles of works such as Reilly's Retail Gravitation Theory (1931) and Huff's Model of Trade Area attraction (1963), who defined retail centre attractiveness, population and distance to competing retail centres (which are already included in the model) as key determinants of the number of visitors to a retail centre. A list of demand variables included within this model and their justification can be found in Table 6-1, and many of these demand variables, such as vacancy rate and density of stores can also be a proxy for attractiveness.

Many of the demand variables were derived from the same dataset – Local Data Company's Retail Unit Address data – the exception being workplace population, which was derived using the Workplace Zones and Daytime population dataset and the method to derive it was the same as was used in Chapter 4.

As this is the case, this limits the applicability of the model to the retail units included within Local Data Company's Retail Unit Address data. Many of the variables that will be needed to determine footfall are not able to be calculated for retail addresses outside of this dataset. As discussed in Chapter 3, the Retail Unit Address data has good coverage, however it is not exhaustive and locations that are not in this dataset will not be able to be inputted into this model. This could be improved in any future iterations by improving coverage of this dataset.

This limitation does also have a benefit in terms of processing time. The final model will have to quickly derive these features for a given location, and if that location is always within the Retail Unit Address dataset, this will speed up analysis and remove room for error as the location does not need to be geocoded prior to processing.

The variables derived from the Retail Unit Address data also bear much similarity to those derived in Chapter 4. There has been some adaption to the categories defined to take into account name changes and store closures. These can be found in Appendix 6.1.

To ensure that these variables were derived in the best way to predict footfall, different representations were tested for each variable and the one chosen to include in the model was the one which correlated the best with footfall. Each variable was tested as a 'distance to' variable, which took the distance between the location of interest (either a sensor location in the training dataset or a user inputted location in the use case) and the nearest instance of the variable. They were also tested as proportions, using a straight line buffer from the location of interest and calculating how many of the units inside that buffer were units that reflected the variable of interest. The different buffer sizes tested were 100m, 150m, 200m, 250m, 300m and 350m. 50m was also tested, but for many of the locations, there were not enough units within 50m to make any inferences.

Pearson correlation was used to test which representation produced the variables with the strongest correlation to footfall. Table 6-3 shows the results.

Table 6-3 Correlations between footfall and demand variables using different representations

Variable	Correlation with footfall	Representation
Vacancy	-12%	100m
Value	-5%	150m
Independent	-29%	100m
Night-time economy	12%	350m
Density of Units	34%	350m
Car Parks	5%	Distance
Transport Hubs	-4%	Distance
Anchor Stores	-26%	Distance
Entertainment Venues	-12%	Distance
Premium	-19%	Distance

As Table 6-3 shows, many of the variables showed the highest correlation when the Euclidean distance was applied to measure the shortest distance between the sensor and the unit measured (e.g. anchor store, entertainment venue, transport hub). Many of these are negative correlations showing that the closer the unit is, the higher the footfall can be. Five variables are derived using buffers. For some variables, smaller buffers of 100m produced the highest correlation with footfall, while density of unit and measuring the proportion of night-time economy locations, a wider buffer of 350m was more beneficial. It is possible that the most appropriate method to represent the different demand variables might also vary from retail centre to retail centre, and with time.

There is scope to extend and improve the data derivation process by determining which representation could be best for each locational and temporal context by calculating the correlations of different representations of a variable with footfall at a certain time of day, or hierarchy of retail centre, making the selection of buffer size and distance measure a lot more variable and fluid. Although this could improve the predictions of the overall model, it will likely not be to an amount which is proportional to the complexity, and it could increase the risk of overfitting. Therefore, this will not be explored further in this analysis.

Temporal factors

The temporal factors included within the model are hour of day, day of week and month of year. A previous iteration of this analysis included holidays as a factor, but found there was not enough data to make reliable predictions. From Chapter 5, holidays have been shown to have a huge impact on footfall. So as not to skew or disrupt any model predictions of non-holiday days, holidays were removed. The full list of these days is given in Appendix 6.2. With continuous and automated collection of footfall data becoming more and more accessible, it is hoped that holiday effects can also be modelled in the future.

6.2.2 Footfall data

Before any analysis can take place using the footfall data, it must first be cleaned, pre-processed and adjusted to account for the sources of error. These sources of error are discussed in more detail in Chapter 3, however they can be summarised under two main headings: sources of overcounting and sources of undercounting. As shown in Figure 3-10, sources of overcounting include smart devices that are counted that don't indicate a pedestrian, such as someone carrying multiple Wi-Fi capable devices, printers and computer in shops or staff and resident's devices, and issues from Mac Address Randomisation that make it difficult to account for these. Pedestrians who pass the sensors and are not registered present a source of undercounting. This could be due to a lack of device ownership, or if they did not have a device on their person, or infrequent probe request so that their device isn't measured by the sensor. Research by Freudiger (2015) showed that the average device sends out 55 probe requests an hour, just less than once a minute. This presents a source of undercounting as the pedestrian's device might not emit a Wi-Fi probe request for the sensor to pick up as they pass by. In order to mitigate the impact of these sources of error, the footfall data were adjusted to reflect manual counts.

Manual counts

In Chapter 3, the manual counts dataset was also introduced and compared to the footfall counts data to show that the footfall sensors tended towards undercounting. 72% of sensors significantly undercounted footfall (>10%), with most being less than half of the recorded manual count. However, this was not a uniform pattern. 20% of the sensors overcounted footfall compared to the manual counts, therefore only 8% of sensors measured footfall that was within 10% of the manual footfall counts. The average absolute error of the sensor counts against the manual counts was found to be 66.7%.

Therefore, to improve the validity of the footfall magnitude data, it will be adjusted according to the manual counts. This was done by calculating an adjustment factor using the following equations.

$$\text{median}\left(\frac{\text{Sensor Count}_{tdl}}{\text{Manual Count}_{tdl}}\right) = \text{Adjustment factor}_{dl}$$

$$\text{Footfall}_{dl} \times \frac{1}{\text{Adjustment factor}_{dl}} = \text{Adjusted footfall}_{dl}$$

Where:

$t =$ Time

$d =$ Device

$l =$ Location

For each device-location pairing, a median of the ratio between the sensor count and the manual count was calculated as the adjustment factor. This is then applied to all sensor collected footfall data for that device-location pairing. Out of 1,378 device-location pairings in the full footfall data, 876 had manual counts measurements and adjustment factors, therefore were kept.

Some potential causes for these errors were investigated, including proximity[22] to computer, electrical and phone shops, proximity to leisure outlets, proximity to transport stops and the time of day, day of week and yearly quarter. If these were significantly related to the error between manual and sensor counts in a location, it could have allowed for a more appropriate or accurate adjustment factor. However, none of these were found to have a significant relationship ($p > 0.1$).

This adjustment factor has several assumptions and limitations that should be taken into account. Firstly, it assumes that the manual counts are the true values of footfall, and any inconsistent counting or human error is negatable. It also assumes that the counts taken by the sensor that day are typical of that sensor in that location. A quality check was completed to ensure that the sensor counts were not <5% or >95% of counts for that location at that weekday and hour, therefore filtering out any extreme values. However, it is impossible to tell with no doubt if the sensor was behaving typically or atypically that day.

A second major assumption is that the adjustment factor and the error of the sensor remains constant over time. Therefore, the adjustment factor that is calculated as an average of the observations of manual counts for a certain device in a certain location is an accurate representation of the adjustment factor for that device in that location at any time during its tenure.

This is a significant assumption based off a minimum of three observations. The median was chosen to aggregate the adjustment factor for multiple observations to reduce the impact of any outlying values. However, in some cases, the calculated adjustment factor was quite significant. For the sensors that undercounted most severely, the footfall was increased by a factor of 21, and for the sensors which overcounted footfall was reduced by up to ten times.

---

[22] Close proximity was defined as 25m. As the range of probe requests is difficult to quantify due to factors such as obstacles, materials of the environment and the make and model of device, 25m was selected as a generous distance to encompass any influential factors.

Figure 6-1 gives an example of how the adjustment factor changed the footfall count for one sensor, with the black lines showing the original footfall count, and the red showing the adjusted counts. Due to the manual counts showing the sensor significantly undercounting in 2017 and in one observation in 2016, the adjustment factor increased the adjusted footfall count relative to the sensor footfall count. Although the adjustment makes footfall closer to the manual counts in the sample of data from 15th May 2017, it has the opposite impact on two out of the three counts completed on 12th July 2016. It is unknown which of the two manual counts is correct. This shows that, although the adjustment factor is largely beneficial, it does make some unavoidable, yet significant assumptions about the data.

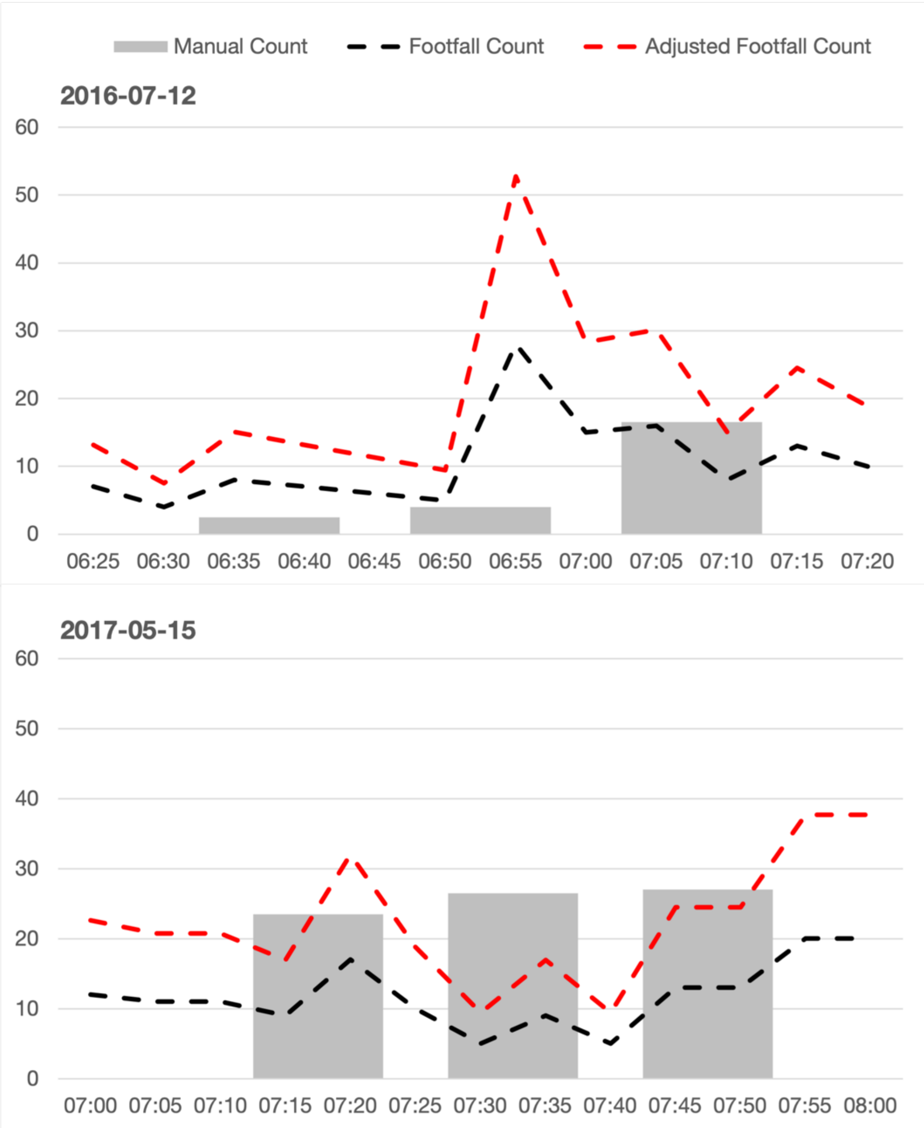


Figure 6-1 Differences between manual count, footfall count and adjusted footfall count

Cleaning

In total, the footfall data spanned 53 months, from July 2015 until November 2019. Although there was data until May 2020, these were removed due to the impact of the COVID-19 pandemic and the deactivation of sensors. The data for December 2019 was also removed due to significant inconsistencies and missing data.

The data was then cleaned and adjusted as follows:

- Duplicate days were removed when the device in a location changed
- The data was transformed to accurately reflect the time zone in Britain (UTC to GMT/BST)
- Outliers were removed using `tsoutliers()` function in R (Chen and Liu, 1993; López-de-Lacalle, 2019). For ease and due to lack of data, holiday days were removed for this analysis. The complexity of determining how events such as Christmas, Bank Holidays and Black Friday impact footfall is difficult to predict and largely unknown, as shown in the analysis in Chapter 5. An interesting avenue for further research would be to attempt to model and predict footfall for these key days. However, currently there is not enough data to be able to reliably generalise their effects. A full list of holiday days can be found in Appendix 6.2.

Firstly, the data was aggregated to hourly sums. This was done to avoid the noise that can occur when using the five-minute counts, both due to the cleaning procedure LDC uses to reduce overcounting and contextual fluctuations, for example, bus times for a sensor near to a bus station. Additionally, it reduces the size of the dataset by roughly eleven times, from 176 to 16 million rows, which is more efficient computationally. The data was cleaned at hourly level.

However, for the purpose of training and testing different methodologies, analysing 16 million rows remains a computationally expensive task (including all the features, combined the dataset could be up to of 16 million rows x 19 columns). The analysis was attempted on the full dataset and found to be incredibly time consumptive to construct just one model, rendering optimisation or comparison of different approaches difficult. In addition, at an hourly level the dataset continued to contain a significant amount of noise that made modelling challenging.

Solutions

Several solutions will be proposed to mitigate this issue and to make modelling this footfall data feasible.

The first is daily aggregation. By aggregating the footfall data to daily counts as opposed to hourly counts, it reduces the dataset size to roughly 3% of its former size, to just 475,000 rows. This would be a lot more realistic in computational terms, and would also remove much of the variation introduced by hourly counts that can prove challenging to account for within a model.

The disadvantage of this solution is that it loses the insight and value of high temporal resolution data. This can result in a model that does not meet the use case defined in the previous section and weakens the number of potential applications. For this reason, this solution was not chosen.

A second solution is to create a two-step modelling process. Much of the complication and size of the dataset is added with the temporal variables, with the other independent variables only varying with location. Therefore, when footfall and the temporal variables are removed many of the rows in the resulting dataset would be repeats. This leaves a dataset of the 668 unique locations, which is significantly more manageable.

The two-step process would be structured as followed. First, there is a location model that considers the independent variables and predicts an aggregate of footfall which is not time dependent. Then, there is a temporal model which models the temporal variation. This model will be a lot simpler than the location model, in order to make the process manageable. The location-based prediction of footfall is combined with the temporal model to create a time and location dependent benchmark for footfall.

Time series decomposition was explored while attempting to construct this model, however this methodology was not chosen for two reasons. Firstly, as previously mentioned, the footfall time series were not complete or reliable enough to detect meaningful patterns in seasons, particularly annual seasons. Secondly, it does not consider how the demand factors interact with time. In Chapter 4, it was shown through correlation that some factors, such as night-time economy and transport hubs, varied greatly in their significance depending on the time of the day or day of the week. Therefore, a proportion of the explanatory power of these variables could be lost by selecting this method.

The final method, and the one chosen for this analysis, was to group the footfall data, but not to the extent that the hourly accuracy is lost and produce individual models for each time group. Therefore, the data is not aggregated to these time series groups, and the predictions made are still hourly, but the process is constructed of 32 smaller models, all considering a subset of the data which is roughly 500,000 rows. The time groups were summarised as in Table 6-4. This solution was chosen as it minimises the size of the dataset being modelled, while continuing to preserve the hourly accuracy. In addition, it considers how the different independent variables, in particular the demand variables, might impact footfall differently depending on time.

Table 6-4 Groupings of temporal factors

Month	Quarter	Weekday	Weekend	Hour	Hour group
January		Monday		0	
February	1	Tuesday		1	
March		Wednesday	F	2	Night
April		Thursday		3	
May	2	Friday		4	
June		Saturday	T	5	
July		Sunday		6	
August	3			7	
September				8	Morning
October				9	
November	4			10	
December				11	
				12	
				13	
				14	Afternoon
				15	
				16	
				17	
				18	
				19	
				20	Evening
				21	
				22	
				23	

6.2.3 Summary

The dataset used in this analysis consists of 668 unique locations, described by 16 independent variables which consider retail centre demand, accessibility and elements of its spatial representation, in addition to temporal factors. These will be used to model the dependent variable – footfall – through a machine learning regression model.

6.3 Model application

The aim of this analysis, as defined in the use case in Section 6.1.1, is to create a model that can predict footfall while considering the impact of spatial and temporal factors. A direct-demand approach was chosen, which applies network analysis and features of demand in a regression model where the dependent variable is footfall. Temporal characteristics, such as season of the year, day of the week and time of the day will also integrated into the regression model.

This following section will focus on the application of the model to the data derived in the previous section, Section 6.2. The first section, Section 6.3.1, will introduce the different regression algorithms that were considered for this analysis, evaluating their use for this purpose and justifying the selection of a random forest algorithm.

When building a predictive model, it is essential to know how accurate the output is. This can be achieved by comparing the footfall the model predicts for a location with a sensor to the actual value measured however this does not indicate how well the model predicts data that it has not seen before. As briefly discussed in the use case, it is vital that this model predicts unseen data, such as a user entered location and time, with accuracy. Therefore, the model must be tested on data that has not been used in the training dataset. This will be achieved through a train-test split, which will be discussed, in addition to other methods of model evaluation, in Section 6.3.2.

6.3.1 Regression algorithms

There is a plethora of different regression algorithms that could be used to fulfil this task. The choice of regression model can often depend on qualities and quantities of the data and the strengths and weaknesses of each approach. Several regression approaches were considered for this task, and these were,

- Linear regression
- Decision Trees
- Random Forest

- XGBoost, and
- Neural Networks.

Linear regression

Linear regression is the most commonly used, widely understood and simplest regression algorithm. It uses a straight line to model the relationship between an independent variable(s) and the dependent variable. It is quick to calculate, which is a huge benefit, however, it is limited to only modelling linear relationships, or relationships that can be easily transformed into linear relationships. This presents a challenge when analysing this dataset, as we are looking at the relationship between 16 different variables and footfall, in a combination of 32 different temporal contexts. It would be a huge assumption that these are all linear and a significant task to transform each individual feature to make this so. For these reasons, linear regression was not used in this analysis.

Decision Trees

A Decision Tree algorithm predicts the value on unknown data based on learned rules and decisions about the known data. Figure 6-2 gives an example of how a simple decision tree built on this data could look. Each node represents a rule based on a feature of the data, and each branch splits the data based on that rule. The final outcome, or leaf, is the estimated value for footfall.

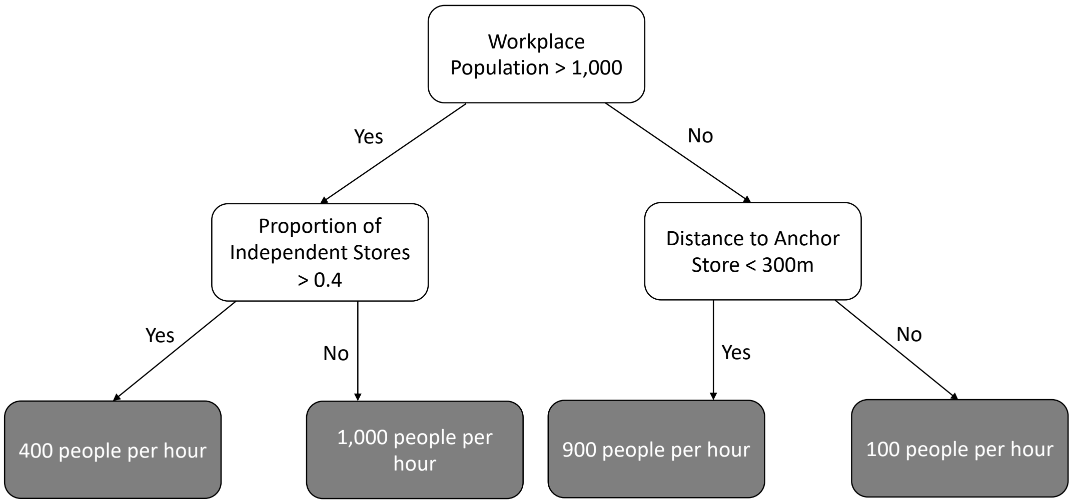


Figure 6-2 Example of a simple decision tree model

The example given in Figure 6-2 is a very simple tree, with the data only being split twice. This would be very quick to run but would likely over-generalise the data, missing the more complex patterns and relationships. On the other hand, if the tree has too many nodes and branches it might predict the seen data better, but it could end up being overfit and less accurate on unseen data. In addition, it is more complex to implement and run. The optimal tree size defines a tree that is as simple as possible, while still retaining a good accuracy score.

However, a decision tree algorithm will always be limited by the fact that it has just one tree. As an algorithm, it has low bias as the predictions are based on real data, but high variance and can be overfit to patterns in the training data. For this reason, a singular decision tree was not used in this analysis.

Random Forest & XGBoost

Ensemble methods can be used to mitigate the impact of the high variance that decision trees are prone to. Instead of building a singular decision tree which is used to classify the data, these methods build multiple different trees that classify an unlabelled point many times, each using a different set of rules and decisions. The ensemble method then aggregates the results to produce one final result, or class.

There are two main types of decision tree ensemble algorithms: bagging algorithms, such as Random Forest, and boosting algorithms, such as XGBoost.

A Random Forest algorithm works by bagging the data into smaller, random samples and building a decision tree on subsets of the data. In addition, only a subset of the features is considered at each split. For example, there are 13 features that will be used in this classification, but a random selection of only four of these might be considered at each split. This number is defined in the algorithm as mtry. The random forest algorithm generates a large number of trees in parallel to each other, relying on the bagging process to create a range of different combinations of rules. This mitigates the risk of overfitting to the training data, while helping to preserve the variance.

A boosting algorithm creates trees iteratively, with each new tree attempting to correct the errors made by the last tree. Individually, these trees might be weak models, but when combined they can create strong model. XGBoost is currently a popular and widely used algorithm that applies a machine learning technique called gradient boosting to solve regression problems. It starts with a default value and uses the decision tree to predicting the error of the starting value.

Both Random Forest and XGBoost were applied to this dataset and Random Forest was found to result in the best performance and was therefore chosen for the modelling in this thesis.

Neural Networks

This section on neural networks will be brief as this model was not tested on the data, but the potential it could have for data sources like this justifies being mentioned. Neural networks are powerful algorithms that can be used to learn and model data through the derivation and processing of big data and a large number of features. For example, a Long Short Term Memory model, or LSTM (a model that uses neural network architecture) can analyse hundreds and thousands of time series, analysing each point and the difference between each point in order to predict future observations. The limitation of neural networks and the reason why they were not used in this case, is that they require a large amount of reliable data, and can be very computationally expensive. However, as data sources such as LDC's SmartStreetSensor database grow and become more robust and reliable, neural network models, such as LSTM may be avenues to consider for future research.

6.3.2 Model evaluation

When building a predictive model, it is essential to know how accurate the output is. This can be achieved by comparing the footfall the model predicts for a location with a sensor to the actual value measured through a process such as cross-validation, however this does not indicate how well the model predicts data that it has not seen before. As briefly discussed in the use case, it is vital that this model predicts unseen data, such as a user entered location and time, with accuracy. Therefore, the model must be tested on data that has not been used in the training dataset. This can be achieved through a train-test split.

A train-test split is a technique that involves dividing the dataset used to construct the model into two subsets: the training set and the testing set. The training set is used to fit the model while the testing set is kept aside during construction. Then, when the model is fit to the training data, it is evaluated using the testing data, and as the model does not see the testing set during its construction, it can be handled the same way a user inputted data might be. The predicted values can be compared to the actual values and an accuracy given.

Throughout this chapter, models are evaluated using the root mean squared error, or RMSE, which measures how close the predicted values are to the actual values through the equation below.

$$\text{RMSE} = \sqrt{\overline{(p - o)^2}}$$

Where

p = Predicted values

o = Observed values

The RMSE is equivalent to the standard deviation of the model residuals and is a commonly used measure of performance for predictive modelling. It is in the same units as the dependent variable, which in this case is hourly footfall. Therefore, the RMSE is mean absolute difference of the number of people per hour the model predicts are the number of people per hour measured in the dataset.

The R^2 figure is also quoted in the results for completeness. The R^2 of coefficient of determination denotes how much of the variation within the data can be explained by the model, as a percentage. It is calculated by taking the proportion of sum of squares from the regression over the total sum of squares.

The train-test split is configured by determining the percentage of the original data that is marked for training and the percentage marked for testing. This differs depending on the problem and the available data. Using a train-test split is an appropriate method of model evaluation when the dataset is of sufficient size that both the training and testing datasets give a suitable representation of the original dataset. That includes enough observations to cover all common combinations of input variables. The training subset of the dataset will also have to be large enough for any specific methods and algorithms used.

After pre-processing and cleaning of the footfall data, 668 unique locations remained across Great Britain. Employing the classification derived in Chapter 4, 430 of these were Chain and Comparison Retail micro-locations, 198 were Business and Independent micro-locations and 40 were Value-Orientated Convenience micro-locations. A 80:20 training testing split was used, which was stratified on location and cluster, so 80% of locations were used to train the model and 20% were kept aside to evaluate performance on unseen data. These were tested to validate that they were representative of the sample, and cross-validation was also employed in the modelling process to further minimise any overfitting.

6.4 Results & Discussion

As discussed in the previous section, the regression model was fitted 32 times to reflect the different seasons of the year, days of the week and times of the day, the details of which can be found in Table 6-4. The R package 'caret' (Kuhn, n.d.) was used to implement the random forest models and 3-fold cross validation was used to select the best parameters for each by minimising the variance. This section will explore the results of the model, and discuss its performance and future developments.

Section 6.4.1 will give the results for each of the models, including the out-of-bag RMSE, the RMSE on the test data and the R^2 values. Section 6.4.2 will explore the differences between the models in relation to their temporal context and Section 6.4.3 will investigate how the model performs on unseen data. An example of the model application is presented in Section 6.4.4. Then, the limitations and future developments will be discussed in Section 6.4.5 and Section 6.4.6.

6.4.1 Model results

Table 6-5 gives the results of each model. The out-of-bag or OOB RMSE gives the root mean squared error of the model on the training data, showing how well it fits the data which it has seen. The RMSE shows how well the model predicts the testing data which it has not seen. In general, the model performs reasonably well on the training data and the model captures 81% of the variation within the data, on average. This ranges from an R^2 of 67% for weekday nights in quarter 4 to 89% for weekday mornings in quarter 2. On average, the model predicts within 383 people per hour for seen data, and 851 people per hour on unseen data.

Table 6-5 Evaluation metrics of model

Quarter	Weekday/Weekend	Hour	OOB RMSE	RMSE	R Squared
1	Weekday	Night	95.89	180.72	77%
1	Weekend	Night	125.40	223.73	75%
1	Weekday	Morning	331.72	780.25	88%
1	Weekend	Morning	342.22	793.42	88%
1	Weekday	Afternoon	545.95	1365.50	86%
1	Weekend	Afternoon	631.29	1635.27	86%
1	Weekday	Evening	296.22	687.53	88%
1	Weekend	Evening	372.07	834.33	84%
2	Weekday	Night	98.63	186.61	77%
2	Weekend	Night	121.13	233.63	76%
2	Weekday	Morning	340.10	842.01	89%
2	Weekend	Morning	365.02	841.86	88%
2	Weekday	Afternoon	604.60	1419.37	85%
2	Weekend	Afternoon	675.62	1649.88	85%
2	Weekday	Evening	338.59	775.24	86%
2	Weekend	Evening	402.13	877.89	82%
3	Weekday	Night	105.36	173.57	72%
3	Weekend	Night	122.77	200.78	71%
3	Weekday	Morning	353.41	774.70	86%
3	Weekend	Morning	381.50	797.03	85%
3	Weekday	Afternoon	611.32	1450.64	83%
3	Weekend	Afternoon	681.69	1578.24	83%
3	Weekday	Evening	348.48	731.93	82%
3	Weekend	Evening	409.08	835.68	77%
4	Weekday	Night	136.20	200.34	67%
4	Weekend	Night	159.21	254.04	69%
4	Weekday	Morning	404.85	837.47	85%
4	Weekend	Morning	446.88	883.01	84%
4	Weekday	Afternoon	681.69	1571.85	83%
4	Weekend	Afternoon	821.16	1854.35	82%
4	Weekday	Evening	418.69	798.01	80%
4	Weekend	Evening	500.98	961.65	76%

6.4.2 Differences between temporal factors

Table 6-6 show the model metrics averaged by quarter. The model appears to perform best on quarter 1 data (January, February and March) and worst on quarter 4 data (October, November and December). It is to be expected that quarter 4 might prove most challenging to predict, as predictions are based on generalisations and this period of the year tends to have more holidays and anomalous events such as Christmas, Black Friday and Halloween. These have all been shown to have a significant impact on footfall in Chapter 5.

Table 6-6 Model metrics by quarter

Quarter	Average of OOB RMSE	Average of RMSE	Average of R Squared
1	342.60	812.59	84%
2	368.23	853.31	83%
3	376.70	817.82	80%
4	446.21	920.09	78%

In addition, the RMSE and OOB RMSE values will be proportional to the level of footfall at that time. Quarter 4 tends to be the busiest time of the year for footfall and Quarter 1 tends to be the quietest. The same margin of error for both would look more significant for Quarter 4 due to the larger magnitudes of footfall.

Variable importance was also calculated for each of the models. This was done using feature permutation. This measures how much the prediction error increases if the variable is excluded from the model, therefore the higher the permutation, the more important the variable is assumed to be.

There was no significant difference between the importance of variables between quarters, with the most important features being night-time economy locations (largely driven by the impact of the evening and night models), independent stores, the hour of day, the distance to anchor stores and the workplace population. The variables which reduced the model errors the least were the month of the year, the proportion of value stores and the proportion of vacant stores. Table 6-7 gives the average feature permutation across all models.

Table 6-7 Mean feature permutation across models

	Feature Permutation
Night-Time Economy locations	75.7
Independent Stores	62.9
Hour	54.8
Anchor Stores	44.7
Workplace Population	44.2
Car Parks	35.5
Density of Stores	33.6
Population	28.4
Entertainment	25.0
Betweenness/Centrality	22.8
Weekday	21.2
Premium Stores	21.0
Transport Hubs	19.7
Vacant Stores	16.8
Value Stores	4.7
Month	0.0

Although the variable importance does not differ significantly between quarters, it does differ between different times of day and days of the week. Table 6-8 shows the average feature permutation for Nights (12am-6am), Mornings (6am-12pm), Afternoons (12pm-6pm) and Evenings (6pm-12am) as well as for weekdays and weekends. Many variables vary in importance from model to model. For example, the presence of night-time economy locations are better predictors for the evening and night models than they are for the morning and afternoon models. This makes sense as the restaurants, bars, pubs and clubs that may be in these locations are likely to have more of an influence on footfall past 6pm. Similarly, workplace population has more of an impact on the weekday model in comparison to the weekend model, as the weekday model would include the working week. This demonstrates that there is value added by training separate models based on time.

Table 6-8 Mean feature permutation by time of the day and day of week

	Morning	Afternoon	Evening	Night	Weekday	Weekend
Hour	98.8	5.5	55.5	59.3	59.0	50.5
Month	0.0	0.0	0.0	0.0	0.0	0.0
Transport Hubs	19.3	14.7	18.0	26.9	21.5	17.9
Premium Stores	24.2	24.2	13.3	22.2	22.1	19.8
Entertainment	21.5	19.2	19.8	39.6	22.9	27.1
Anchor Stores	65.1	46.7	38.6	28.6	49.1	40.4
Workplace Population	45.0	46.6	50.2	34.9	49.9	38.5
Betweenness/Centrality	13.0	19.3	15.0	43.9	24.0	21.6
Population	19.0	21.9	18.0	54.9	29.5	27.3
Vacant Stores	11.7	12.7	16.6	26.4	17.9	15.8
Independent Stores	76.1	100.0	47.2	28.4	64.1	61.8
Value Stores	4.9	5.8	5.8	2.2	4.8	4.5
Density of Stores	27.7	51.9	30.0	24.6	32.1	35.0
Night-Time Economy locations	59.2	43.5	100.0	100.0	81.6	69.7
Car Parks	38.6	41.6	26.8	34.8	36.4	34.5
Weekday	18.4	4.6	6.6	55.2	30.5	11.8

However, these time periods may be limited in their granularity. The influence of different demand variables could change significantly within these time periods – from Saturday to Sunday, or from 2pm to 5pm. Transport hubs are not a particularly strong predictor in any of the models, despite previous analyses showing it was a strong predictor of footfall at 9am and 5pm. The use of hourly groupings could mask this impact.

Table 6-9 shows the model metrics when averaged across the days of the week and the hours of the day. The weekday model generally performs better on seen and unseen data in comparison to the weekend model, which is logical as it has more data to learn patterns from (five days of the week compared to two). In addition, due to the impact of the Sunday Trading Day Act, there may be more difference between Saturday and Sunday compared to the difference between the days from Monday to Friday.

Table 6-9 Model metrics by day of the week and hour of the day

Row Labels	Average of OOB RMSE	Average of RMSE	Average of R Squared
Weekday	357.0	798.5	82%
Weekend	409.9	903.4	81%
Morning	370.7	818.7	87%
Afternoon	656.7	1565.6	84%
Evening	385.8	812.8	82%
Night	120.6	206.7	73%

When comparing the R^2 values, the morning model captures the most variation in the data, at 87% and the night model captures the least. This could be explained through night-time footfall (12am-6am) being sparse and not necessarily as connected to factors of the environment as footfall during other times of the day. Interestingly, although the evening model as a lower R^2 value than the morning model, it performs better on unseen data, as shown through the lower average RMSE. This could partially be explained through a generally smaller magnitude of footfall, however, the OOB RMSE for the evening models are, on average, bigger than the morning models. It could be that evening footfall is slightly more generalisable than morning footfall, therefore the model can produce more accurate predictions. This is a possible explanation as there could be many different demand factors at play to drive footfall during the morning – commuting, retail, leisure – and comparably fewer between 6pm and midnight – only select units are open, people are generally at home more, the work day is over.

This difference in RMSE between training and validation data is very common and expected, as models perform better on data, they have learned on rather than data that they haven't seen. However, this difference in RMSE is significant and it is clear that the model is overfitting to the training data. There are several solutions to prevent overfitting and produce generalisable and accurate models, and these will be discussed in Section 6.4.5.

6.4.3 Performance on unseen data

When the models are used to predict values on the test, or unseen, data, this can provide insight into which locations and situations the model might predict successfully and where it might over- or under-estimate.

Figure 6-3 shows the difference between the actual footfall data and the model predictions for the unseen data. The model shows high variance and bias in predictions. The high variance is shown by the dispersion of the points. For example, for the actual observations that were close to 2,500 people per hour, the model predicted these anywhere from a few hundred to 5,000.

The bias is shown through the gradient of the line. A model with no bias would have a line with a gradient of 1 on the graph in Figure 6-3, showing that for places where actual footfall is, for example, 1,500 people per hour, the prediction is also, on average, 1,500 people per hour. The model predictions appear relatively unbiased for observations where footfall is measured at below 2,000 people per hour, or less – the trend line passes through the same variables for predicted and actual. However, for busier observations, the model does seem to bias towards underpredicting footfall. For example, observations that were measured at 5,000 people per hour were predicted, on average, at just over 2,500 people per hour.There also appears to be an upper limit on model predictions of around 7,500 people per hour, whereas real data can exceed this number. This is likely the result of a lack of data and measurements collected for these busier locations and times. They are unusual and rare. Only 23 locations in the sample of 668 averaged higher than 2,500 people per hour, which is just over 3%. These are areas such as Tottenham Court Road, Manchester Arndale and Liverpool ONE – major retail cores of large cities which may not be generalisable to other locations in the sample.

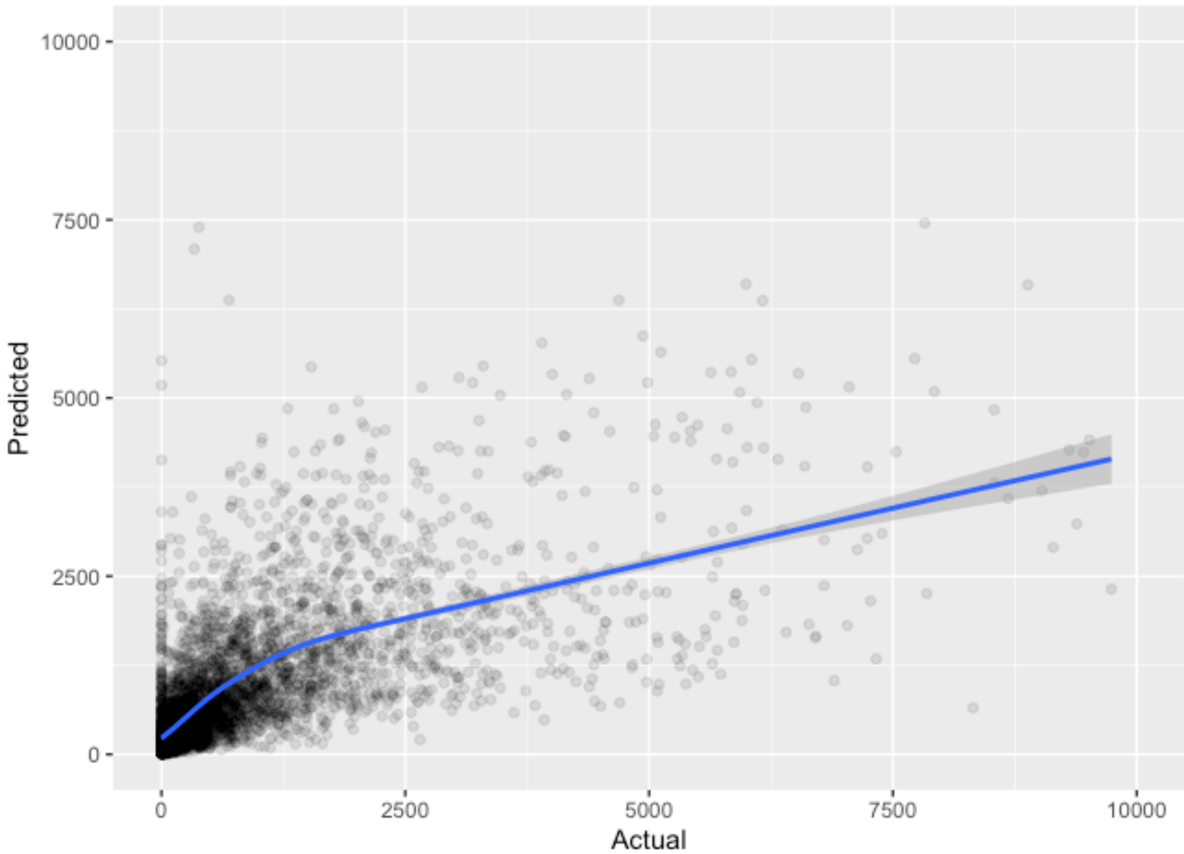


Figure 6-3 Graph showing the prediction to actual values in the unseen data

To further evaluate the model for any bias, the performance on the individual locations was also compared. This was done to see if there were any features in common for locations that the model consistently under- or over-predicted. Correlation tests were undertaken between the percentage error for a location and the thirteen spatial variables used in the model. The results are given in Appendix 6.3 and show that none of these correlations were significant.

This shows that, although the model has some bias, this is not detectable in the features that were derived. This could be because the independent variables that were calculated for the major retail cores of large cities were similar to those of many other locations with lower footfall, such as regional centres. With no features to differentiate them, the model could end up with a bias towards underprediction.

Unfortunately, it is impossible to tell how much this bias is exacerbated by the sources of error within the data itself. For example, another location close to Manchester Arndale has an average hourly footfall of 700 people per hour, and sensors along Oxford Street measured an average hourly footfall of 1,700 people per hour – significantly lower than the measurements for nearby locations. If a feature was added to label major retail cores, it might not solve this bias as the variation in measurements due to the sources of error explored in Chapter 3 are significant in themselves that any prediction based on this data could be inaccurate[23].

6.4.4 Example – Fouberts Place, London

An example from the test set was taken at random to demonstrate what a model prediction would look like. Fouberts Place is in London, near Regent Street, as shown in Figure 6-4. The average hourly footfall is 943 people per hour and, in the classification derived in Chapter 4, it was classified as a Chain and Comparison Retail micro-location. The location is 118m from an anchor store, 131m from a transport hub and 145m from an entertainment venue. There are no vacant units or value units nearby, and the proportion of independent stores is 40% and night-time economy locations is 20%.

Figure 6-4 Map showing example location- Fouberts Place

[23] This is discussed further in Section 6.4.4

Figure 6-5, Figure 6-7, Figure 6-7 and Figure 6-8 on the following pages show the actual and predicted footfall for Fouberts Place, for March (Q1), June (Q2), September (Q3) and November[24] (Q4) respectively. Although, the RMSE of this location was, on average 507 people per hour (which represents a 55% error), the graphs show that accuracy can vary significantly depending on the time the prediction is trying to capture. Nights and Mornings are consistently more accurate than the predictions for Afternoons and Evenings, where the model appears to underpredict and the 5-6pm peak is not as pronounced in the predictions as it is in the actual data.

The model appears to make the most accurate predictions in Q3. This is reflected through the lower RMSE of 352 people per hour and echoes the relatively low RMSE for the Q3 model identified in Table 6-6. Upon data exploration, this is not a result of imbalance in the data, as the four quarters have a similar number of observations. A potential explanation is that footfall is more closely linked with the features identified during July, August and September than it is during other months of the year. A future avenue of research could be to look if this is related to other factors, for example the more pleasant weather during this time of year, or the coincidence with school holidays.

Even though the magnitude of footfall was generally low, the model did pick up the footfall signature reasonably well, replicating the three peaks daily pattern on weekdays and the smoother curve on weekends. This variation is likely captured by the inclusion of hour as a feature, which was also shown to be one of the most important variables.

[24] The other months were chosen as the last months in the quarter. However, November was chosen instead of December to remove the influence of holiday days.

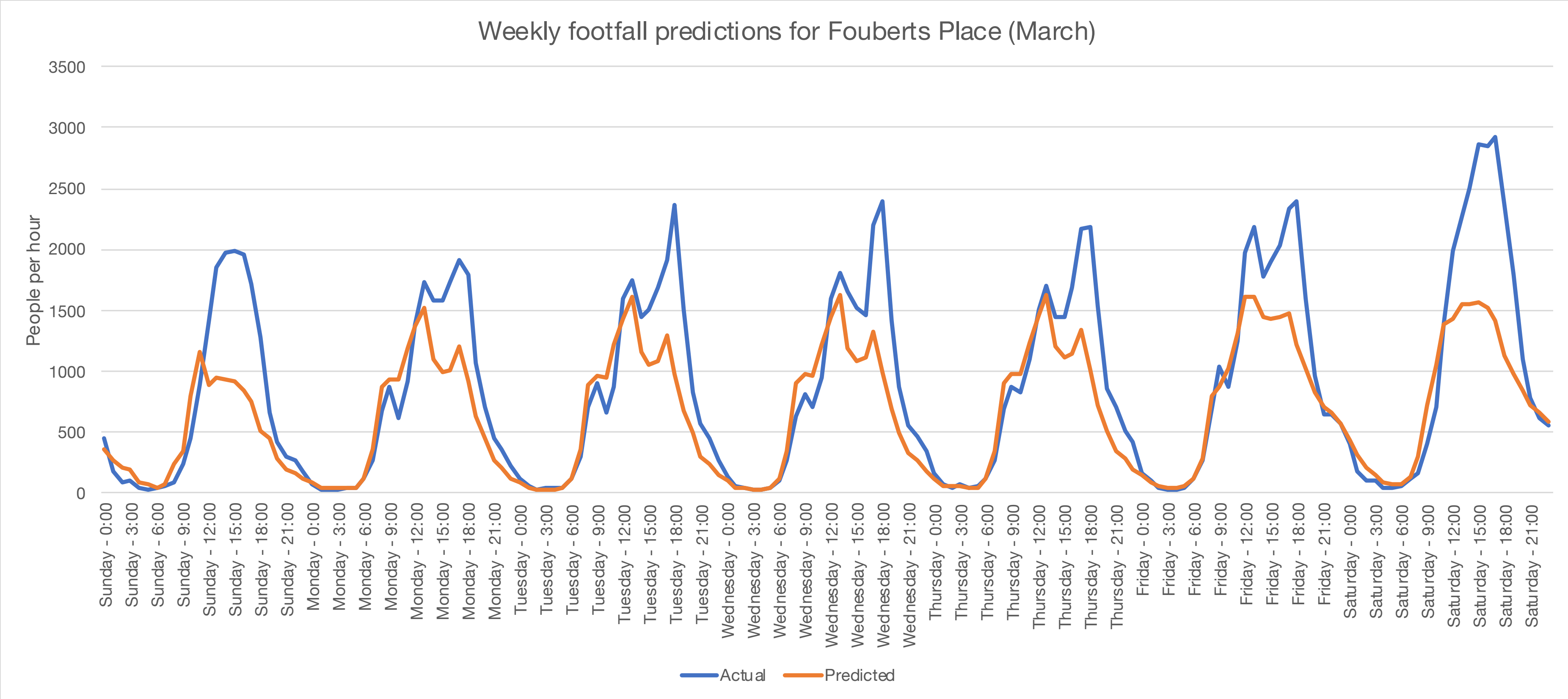


Figure 6-5 Graph showing predictions for Fouberts Place hourly footfall for March

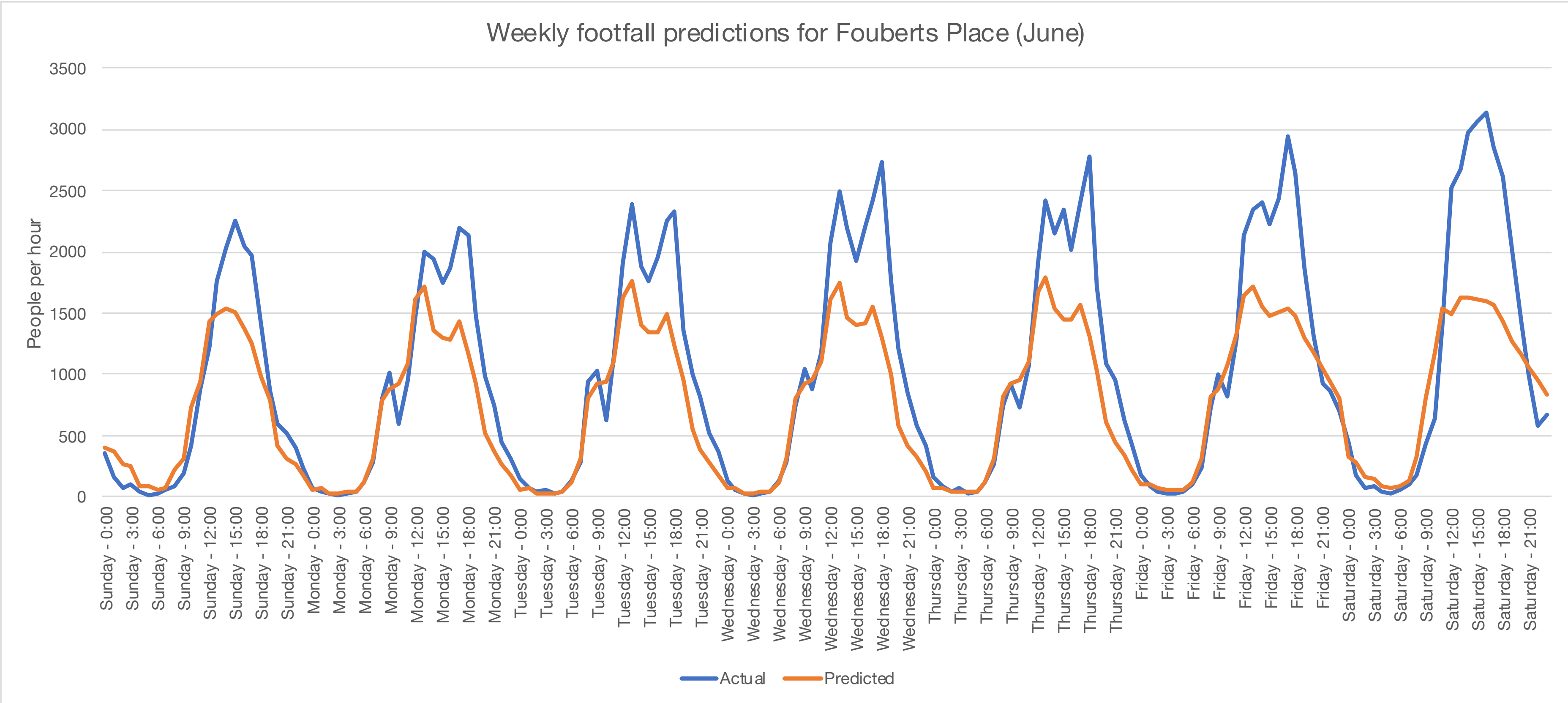


Figure 6-6 Graph showing predictions for Fouberts Place hourly footfall for June

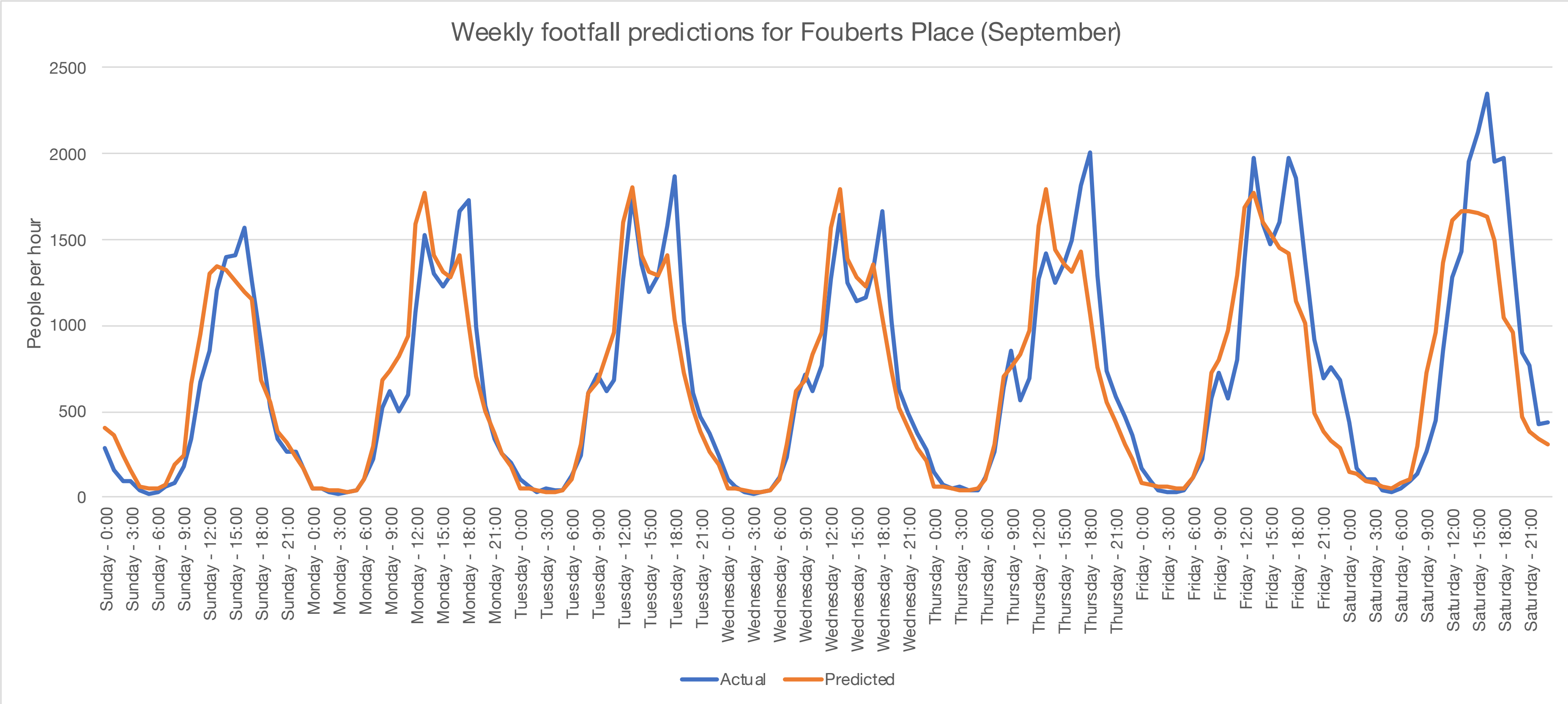


Figure 6-7 Graph showing predictions for Fouberts Place hourly footfall in September

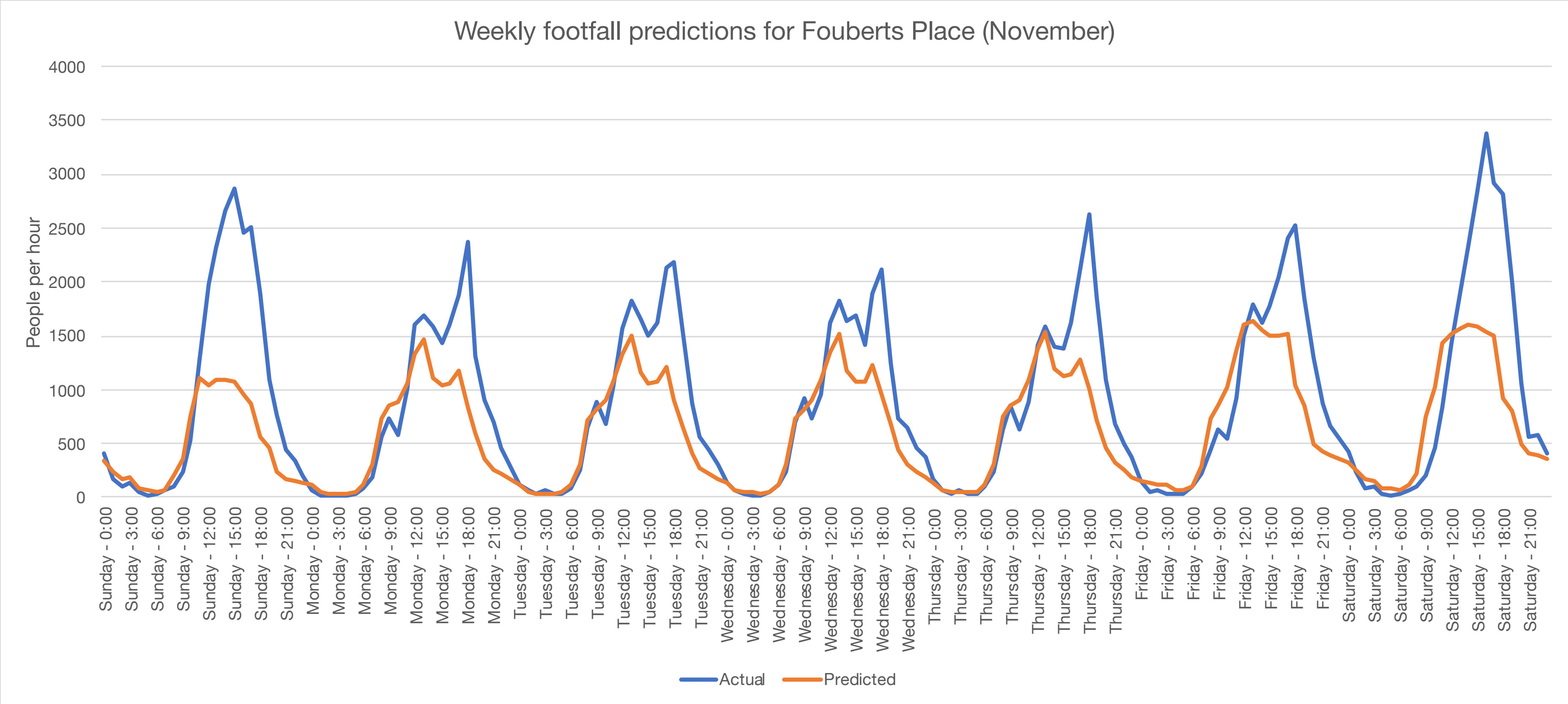


Figure 6-8 Graph showing predictions for Fouberts Place hourly footfall for November

This section has applied the random forest regression methodology to the SmartStreetSensor footfall data to make spatio-temporal predictions of hourly footfall. Although work still need to be undertaken before the results themselves are applicable, the conclusions to this preliminary analysis are positive. The methodology appears to capture some of the patterns within the data, however, struggles to make accurate predictions for busier locations and during busier times, generally underpredicting the people per hour significantly. The models also seem to overfit the data, with the RMSE for the test dataset being significantly higher than the out-of-bag RMSE. The limitations and future developments which could resolve these issues will be discussed next.

6.4.5 Limitations

Unfortunately, this analysis has had many limitations, some of which have already been discussed in this chapter. For example, there is significant error within the data that heavily limits its transferability. Although these footfall counts have been filtered and adjusted in line with manual counts, there still exists inconsistencies that cannot be accounted for. For example, the increasing challenge which MAC address randomisation has posed has made the collection and treatment of the data significantly different from year to year.

The COVID-19 pandemic also hindered the scope of this project, as it was an unprecedented event that impacted high street footfall in such a significant way that a model based on past data would find it incredibly difficult to understand. The analysis was adjusted to be independent of year to account for this; however, that meant modelling data from 2015-2019 with the assumption that year did not have an impact, when it likely has done.

In addition, there are many retail settings that are not represented in the training data in a significant way. As discussed in Chapter 3, sensor locations were largely decided by clients of LDC, therefore there can be large concentrations of sensors in particular environments, whereas other locations are significantly underrepresented.

For example, there are proportionally more sensors in Central London when compared to other major cities such as Birmingham. In addition, the data the sensors measure are not directly comparable to each other. The different context of the sensor, the environment and the devices it is measuring could mean that two sensors on the same street record significantly different counts.

There are also limitations in the availability of data to capture features. There are many factors that can influence retail centre demand, and although this analysis has captured a significant representation of these, there are many more that data is not available for. For example, attractiveness of units, seasonal population fluctuations (whether tourists or students), weather conditions, perception of the retail space, purpose of visit – all of these could impact footfall, however, were not taken into account in the model.

A final limitation is with the random forest algorithm used for this regression. One of the qualities of this model is that it cannot predict values that it hasn't seen before. Therefore, if the busiest were not in the training set, the model could not possibly accurately predict values that high. This could potentially explain why the model had a bias and would not predict above 7,500 people per hour, when the test set had observations of 10,000 people per hour.

6.4.6 Future developments

To achieve the final use case outlined in Section 6.1.1 and to produce a model than can make reasonably reliable spatio-temporal predictions for footfall, more data is needed. The addition of more training data can be beneficial in many machine learning applications, as it exposes the model to more real-life circumstances which helps to improve predictions. The data needs to cover a large and representative distribution of the retail environments that would want to be predicted, ensuring that the dataset does not have any bias towards certain cities or streets. To further improve the quality and accuracy of predictions, the data needs to be collected consistently in the same location for a longer amount of time. Although the SmartStreetSensor data has unprecedented coverage and resolution, due to the increase of MAC address randomisation and other limitations in data collection constancy, it does not enough reliable data to create a dependable model. In addition, the current dataset does not contain enough observations for the machine learning to separate the noise from the underlying trends and patterns. In order to make this model applicable, collecting more footfall data would be the top priority.

Another future development would be to source and derive more data that could be used as features in the model. Space syntax methods have been very successful when applied in other studies. Future research could develop a database of axial maps for multiple retail centres which would facilitate the use of space syntax features within the model. As mentioned previously, there are also many features of demand that could be included within the model, but there is little to no current data regarding these. This model could improve by including more of these features. In addition, model performance could also improve by dropping features that are low variable importance, for example, month of the year or proportion of value stores. These can cause the model to overfit to noise within the training data and harm performance.

The method presented in this chapter could also be improved upon. Different regression algorithms could be tested to see if they yield better results, such as long short-term memory (LSTM) models. The method used to determine time splits could also be developed as the six-hour hourly groups appeared too coarse, demonstrated by the most important variable for all the models being the hour of day. Therefore, it would be interesting for future research to explore whether training a different model on each hour would further improve the model. However this would likely result in more overfitting if applied to the current dataset and would require additional data.

Finally, future research could explore developing different models for different retail contexts. This is a challenge to achieve with the SmartStreetSensor dataset as many of the sensors are in similar retail environments. With intentional collection of future data, it may be possible to have different models for the busiest city centre retail cores, then for regional centres, for market towns and for other unique retail contexts.

6.5 Chapter Summary

At the start of this chapter, three objectives were established. These were,

What remains unknown	Define the criteria and use case for a footfall prediction model and identify appropriate methodologies to achieve it
	Create a preliminary model that predicts footfall that is location and time dependent
	Critique the performance of this model, identifying opportunities for improvement

Section 6.1 defined the resulting model to be a tool usable by high street stakeholders and decision makers to determine a benchmark for footfall for any retail address in the UK. This benchmark would also be time dependent, taking into account the hour, day of the week and month of the year before making a prediction. Research into footfall modelling and prediction is heterogenous and there are a range of different approaches that have been applied across different research areas for different purposes. Direct demand models were identified as the most appropriate for this use case.

Section 6.2 and Section 6.3 detailed the creation of the model, from the derivation of the different features (population, network measures, demand factors and temporal factors) to the different algorithms that could be applied within a direct demand framework, to how the resulting model would be evaluated. A random forest algorithm was chosen, and this was applied 32 times to different temporal subsets of the data. 20% of the locations with data were left out of the modelling process in order to test the performance on unseen data.

Section 6.4 evaluated the performance of the model, identifying the limitations and assumptions that it relies on and suggesting areas for future improvements. The model performed reasonably well – however, the metrics showed that it was overfit to the training data. This is usually due to the training data not being fully representative and can also be indicative of a noisy, unpredictable dependent variable. Both are likely factors in the performance of the model. The fundamental development that needs to take place to create the most accurate model would be to collect more footfall data. The dataset would need to be representative of different retail centres, continuously collected over at least two years and the collection methodology would have to be consistent and sustainable. With technology becoming increasingly accessible, the hope is that more footfall data will be collected also.

In addition to the three objectives, this chapter also posed two novel propositions: to create a footfall model that takes into account both spatial and temporal factors, and to create a footfall model that is generalisable across different retail centres. Although there is room for improvement for the latter, as the model tended to underpredict footfall for unseen locations, the resulting model is evidence that both of propositions are plausible and achievable, to reasonable accuracy.

In conclusion, this chapter has presented a novel tool for footfall prediction that could have many valuable applications to high street stakeholders. Although further development and more data is needed before it can be applicable, this analysis presents a framework that could be followed by when attempt future footfall prediction task, highlighting considerations that should be made about the data and its context.

7 Discussion and conclusion

Footfall serves a key role in high street vitality and viability. It is a unique key performance indicator which is reflective of both the economic resilience and community and social strength of a retail centre. It is used in town management schemes (C. Parker et al., 2016), in event management (Naylor et al., 2016), in business and marketing (Denison, 2005) and in crisis recovery (Harding and Powell, 2011).

Despite its wide usage and potential, there are relatively few studies and literature which explore footfall, how it varies from place to place and over time and what factors determine it. Technological advances in automated data collection from sensors have made footfall more measurable and it is now possible to collect footfall data continuously, from locations all over the country. Due to this, the field of footfall analytics and urban analysis is growing quickly and making the most of the new insights this technology can provide.

The importance of understanding footfall is more pertinent the ever as the UK makes tentative steps towards recovery from the coronavirus pandemic. When government mandated lockdowns forced non-essential physical retailers to close and strongly encouraged people to stay at home, it forced footfall down to nothing. Many people began working from home or using online retailers who may have never considered using them before. The retail environment and the priorities of the consumer have changed, and it is up to the high street to adapt. Although the last few decades have not been smooth for UK high streets, with online shopping, out-of-town shopping centres and convenience culture reducing demand (Wrigley et al., 2015), it is hoped that understanding trends in footfall from the past might help establish a way forward to recovery for the future.

Chapter 2, the literature review explored the evidence base on high street vitality and viability, footfall and how these two topics intersect. There has been a myriad of attempts to revitalise high streets over the previous decades, and particular emphasis was given to the task in the wake of the 2008/9 financial crisis (Grimsey et al., 2013; Portas, 2011; Swinney and Sivaev, 2013). Despite this, many schemes struggled to make a long-term impact, and there was an identified need for more efficient methods of transferring knowledge from academic to high street stakeholders and decision makers, which could be through classifications, data or tools – such as key performance indicators.

As previously mentioned, footfall is a widely used key performance indicator for high street vitality and viability (Hogg et al., 2004), however the knowledge base for footfall and how it behaves and varies is small, segregated and often based on empirical qualitative evidence. Footfall is not only a key performance indicator, but has been coined as the 'lifeblood of the high street' (Birkin et al., 2017). If footfall is strong, then the high street itself is strong. Therefore, gaining a greater understanding of footfall and how it relates to the world around it is paramount.

The intent of this thesis is to contribute to this field by using footfall data to explore patterns and relationships between pedestrian behaviour and the form and function of retail centres. The LDC SmartStreetSensor dataset was the core dataset of this analysis, and was combined with retail, demographic and network data to learn more about footfall and how it varies over space and time. This was done through three novel analytical chapters. The next section, Section 7.1 will review each analysis, summarising the key results, implications and contributions. Then, Section 7.2 will discuss the broad conclusions from these analyses and the recommendations that could be taken forward for future analyses. Section 7.3 will give some final concluding remarks.

7.1 Summary of contributions

There were three analytical chapters in this thesis, with each exploring a different research question related to footfall and the retail environment. These were entitled 'The world around us', 'What happens there' and 'What remains unknown'. The first analytical chapter focused on the relationships between footfall and its immediate environment and context. The second chapter, 'What happens there' investigated temporal fluctuations in footfall and how they could be related to micro- and meso-scale factors. The final analysis presented a novel footfall prediction model, which attempted to utilise this data to create a tool which could estimate footfall for any location, even where it is not known. Each chapter was rooted in an identified literature gap and provided novel, valuable insights for future research.

7.1.1 The world around us – summary

In Chapter 4, 'The world around us - Quantifying temporal variations in footfall in relation to micro-locational characteristics' the relationship between footfall and the immediate context was explored. Many micro-locational characteristics such as vacancy rate, retail mix and proximity to transport hubs have been connected to footfall. Although some quantitative analyses exist that do explore the specific relationships, these are largely limited in scope, focusing on a certain city or context. This chapter identified a need for more data-driven analysis which quantifies the interrelationships between footfall and micro-locational characteristics and how these can vary over time and space. It aimed to create an evidence base which would be applicable to different retail environments and for it to be accessible through deriving a classification.

Objectives and results

In order to achieve this, three objectives were established. These were:

The world around us	Investigate how different characteristics and contexts of the immediate environment impact footfall magnitude and signature
	Using characteristics of retail and footfall context, develop a classification that captures these main differences
	Identify how trends in footfall magnitude and signature differ between these different retail contexts

Firstly the relationship between footfall and functional and morphological characteristics was explored through correlation tests. Using a range of quantitative factors, such as distance to anchor stores, proportion of night-time economy locations and proximity to transport hubs, this analysis investigated how they relate to footfall, and how this can change based on temporal factors. It demonstrated how factors such as workplace population, proportion of independent/chain stores and distance to transport hub can correlate with footfall and showed that patterns in the magnitude and signature of footfall data, and by extension retail vitality, can be explained by functional and morphological characteristics of the micro-location.

It also demonstrated how the relationship between footfall and these characteristics can also be dependent on time. For example, proportion of independent stores was not significantly correlated with footfall during the evening, night or early morning, but was one of the strongest significant correlations during the day.

To investigate how this can vary between retail centres, and to make this information more transferable and applicable, a classification was created which grouped retail locations by these micro-locational characteristics. *K*-means clustering was applied to distinguish three clusters of footfall micro-locational context. These were Chain and Comparison Retail micro-locations, Business and Independent micro-locations and Value-Orientated Convenience Retail micro-locations.

Chain and Comparison Retail micro-locations was the biggest class, encompassing 54% of the locations. They were named after their primarily comparison retail functionality and dominance of chain retailers. Chain and Comparison Retail micro-locations included Oxford Street in London, Queen Street in Cardiff and Liverpool ONE in Liverpool.

Business and Independent micro-locations formed the second largest class, with 40% of the locations. These are locations with more of a tendency towards independent retail and high working populations. These are areas such as Holborn and the City of London. There appeared to be a micro-class within this class which also had a high proportion of night-time economy locations and restaurants, encompassing areas such as Soho in London and Bold Street in Liverpool.

The final micro-location class was Value-Orientated Convenience micro-locations, named after their higher proportion of budget convenience retailers. These were often in locations more accessible to residential areas with a convenience retail function, such as Wood Green and Kilburn in London. This was the smallest class, with only 7% of the data.

The last objective was to explore how footfall magnitude and signature varies between these three classes. The Chain and Comparison Retail micro-locations exhibited a footfall pattern with the busiest times on Saturdays and during daytime hours from late morning to early afternoon, reflective of the cluster's prominent comparison retail function. The Business and Independent micro-locations have dominant weekday footfall with three peaks at 8:00, 12:00 and 17:00. This footfall pattern reflects commuting into and out of work, with an additional increase in footfall during a lunchtime break. The Value-Orientated Convenience micro-locations were the quietest and steadiest in terms of footfall, with the convenience function explaining a constant and consistent flow of short, frequent trips.

Limitations

This analysis faced several limitations that must be considered. Firstly, the sample size is small and may not accurately represent the breath of retail environments in the UK. There are disproportionately fewer sensors in mid-sized and smaller areas, and retail locations which are outside of city centres. The dataset is biased towards London, which is known to exhibit different footfall patterns that the rest of the country. As a result of this bias within the sample, the classification is skewed towards micro-locations in cities.

Secondly, due to this data bias, there may be many other identifiable micro-locational classes in the country which this analysis has not represented. For example, there was no 'holiday' or 'tourist' micro-locations, as were present in the classification by Mumford *et al.* (2021). It is apparent that this sample is biased towards Mumford et al.'s comparison centres and overlooks the different micro-locational patterns that could exist in the remaining clusters.

In addition, although the study groups each of the micro-locations into three clusters, these cluster may not be easily delineated in reality. There was not strong definition between the classes, and on some variables they significantly overlap. This is common for cluster analyses that are based on real-world data, however it should be made clear that the classification is a generalisation and many micro-locations will be more complex in reality. Therefore, it is reasonable to assume that any footfall signatures derived from them may be the same, and some of the sensors may have somewhat different footfall magnitudes and signatures than the average in their cluster, despite the overall similarity of a particular cluster's functional and morphological characteristics.

Finally, the move towards work-from-home and hybrid working due to the COVID-19 pandemic could significantly alter the patterns observed in each of these micro-locations for the future. For example, with fewer people going into offices, the peaks of Business and Independent locations could be less pronounced, and Value-orientated Convenience micro-locations which are close to residential areas might experience a boost in footfall due to people working from home.

Contributions

The results of this analysis provided novel, quantitative information about how footfall relates to the surrounding environment. A classification was also created to ensure these results are accessible and applicable to high street stakeholders and decision makers. It has demonstrated how important it can be to consider a wealth of different factors in determining footfall and in understanding high street vitality.

These results can also have applications for retailers and planners. This knowledge could be used to better understand pedestrian flows and consumer behaviour within a micro-location. For example, Business and Independent micro-locations have a more significant daytime footfall than evening footfall. This knowledge could be used to develop schemes to increase the dwell time of the daytime population and encourage them to support the night-time economy establishments, increasing the retail resilience of the area.

This analysis also established a foundation for the footfall prediction analysis undertaken in Chapter 6 by drawing patterns between different micro-locational factors and footfall. These insights could also be applied in future footfall analysis and prediction.

These results have also contributed through a more comprehensive understanding of retail mobilities. The analysis presented here, and in the paper published in Applied Spatial Analysis and Policy, provide new insights into footfall determinants and the relationship between them and urban mobility.

Future research could benefit from employing more footfall data to investigate monthly, annual and longer-term trends in footfall and how those could relate to functional and morphological characteristics. Modelling footfall for an entire retail centre could be invaluable for decision-making, urban planning and retail location planning. It would also be valuable to revisit this classification in a post-COVID retail context.

7.1.2 What happens there – summary

'What happens there – Exploring event-related temporary fluctuations in footfall magnitude and their relations to micro- and meso-scale characteristics' was the title for the second analytical chapter. It explored which temporary events impact footfall, their effects and how these are mediated by micro- and meso-scale factors. Events impact footfall – this is something that is inherently known in discussions regarding high streets. Footfall is used to measure the success of local events, such as the Edinburgh Fringe Festival and events such as Christmas and Black Friday are considered vital times economically for retailers.

However, there was little to no literature which identified which events impacted footfall and compared their impact across different retail contexts. The relative impact of events is difficult to capture without consistent key performance monitoring and this would have to exist across several locations to allow these to be compared. The SmartStreetSensor data allowed for this novel analyses to take place, using quantitative analysis to identify which events have significant impacts on footfall, the size of their impact and how this relates to the micro- and meso-scale context.

Objectives and results

Three objectives were established in order to complete this aim. These were,

What happens there	Identify events which significantly impact footfall.
	Investigate how factors of both the immediate environment and in the wider context could influence this impact
	Explore the trends and similarites between footfall of different events in different locations and what they could imply about retail footfall

Firstly, the different events which significantly impact footfall were identified. Due to limitations with the quality, coverage and quantity of data, a case study approach was deemed the most appropriate way to collect this information. Four case study micro-locations were identified based on their availability of complete, representative and verifiable data, and their representation of different micro- and meso-scale factors. Liverpool ONE and Manchester Market Street were chosen to demonstrate similar micro-locational contexts – both retail cores of major UK cities – yet different meso-scale contexts, as Manchester is a larger and more connected conurbation. Two micro-locations in Edinburgh was also selected with the same meso-scale contexts, yet different micro-locational contexts, with New Town representing the comparison retail core and Old Town representing a tourism function.

An annual daily footfall ranking was established to identify temporary fluctuations in footfall and which events caused this. Weather events, such as Beast from the East and the summer 2018 heatwave, national events, such as Black Friday, Bank Holidays and Christmas and local events, such as festivals and sports matches were all demonstrated to boost footfall in these locations.

The impact of these events were then compared to the micro- and meso-scale characteristics of the case study locations. Manchester Market Street appeared to be more sensitive to weather events than Liverpool ONE, which could potentially be related to its larger catchment and more consumers having to travel further to reach the retail destination, whereas Liverpool ONE has a smaller catchment that was less impacted by weather events. In addition, the locations with a high comparison retail and high proportion of chain retailers appeared more reliant on Christmas footfall, whereas this was not the case for Edinburgh Old Town, where footfall was highest during festival season. However, Manchester Market St and Liverpool ONE both had local events which resulted in some of the busiest days of the year, such as the Halloween events in Manchester and the Giants Spectacular in Liverpool. The impact of Black Friday was also found to vary locally, with Manchester and Liverpool consumers taking part in the shopping event and Edinburgh consumers not. This did not appear to relate to any factors where there is data for, and is potentially a result of a difference in consumer attitudes, although no research can be found detailing this.

Limitations

The limitations to this analysis may have had a significant impact on these results. Firstly, the data was limited in terms of quality. Due to error and fluctuations in accuracy, a ranking had to be used, as opposed to real numbers, as the real numbers were not consistent. This tremendously limited the insights that could be made from this analysis, as footfall could only be compared relatively and not absolutely.

Secondly, it was a significant undertaking to find specific details regarding events that took place in the past. A case study analysis was appropriate for this investigation, as it was recognised that contextual and local factors could have a large impact; however, it was very difficult to find any contextual and local information through secondary sources. Factors such as sentiment or reception of events by local consumers, as well as rudimental facts such as the date an event took place on was difficult to find. In addition, some events which may have a significant impact on footfall might not be documented in any traceable way, for example, individual retailer sales or political demonstrations. No one explanation could be found for a sizeable minority of top-ranking footfall days.

Finally, the use of a case study analysis on limited data means that the generalisability and applicability of any conclusions are limited. The locations used as case studies only represented two of the fifteen retail groups identified in the CDRC Retail Centre Typology, and the window of two years (2017 & 2018) limited the events and factors that could be investigated in this study. Unfortunately, these constraints were unavoidable due to data availability.

Contributions

The ability to collect continuous footfall data over multiple locations at the same time creates opportunities for novel analyses. This chapter aimed to use this data to build a consensus on what events have a temporary impact on footfall, the significance of this impact, and how it could relate to micro- and meso-scale factors. This research has established a wide range of different events that impact footfall in four case study micro-locations in 2017 and 2018 and explored how this impact can differ between micro-locations. In itself, this makes a significant contribution as there are no known works that compare footfall fluctuations over time and over space.

Due to the data limitation discussed, the conclusions from this analysis are limited in their real-world applicability. The largest contribution of these results and their conclusions would be to lead and direct future research to explore footfall fluctuations and their relationship to events in more depth. This analysis has shown how the largest fluctuations in footfall are often tied to events, therefore in order to understand more about footfall, it would be beneficial to understand in greater depth how it relates to events. It leaves future research opportunities to explore other time periods and the impact of local, regional, and national events on locations with a different retail offer, such as market towns, tourist locations, or failing town centres.

Future research could also benefit from a footfall database where data collection is more sustainable, and the methodology is comparable over space and time. In addition, benefits could be gained by collecting resources and feedback regarding events as they happen, as this analysis has found this information challenging to collect after the event.

Finally, future research could potentially incorporate detail on spending. It could be posited that, while Black Friday and Christmas may generate significant spending, an event such as a food festival may result in local restaurants or fast-food outlets losing income that day as consumers spend their money at the event instead. While local events and festivals might increase social and cultural sustainability, drawing consumers in and providing an enjoyable experience, the immediate economic impacts could be limited.

7.1.3 What remains unknown – summary

The third analytical chapter was entitled 'What remains unknown – Investigating the potential for a spatio-temporal prediction model for footfall data'. It explored how the relationships between footfall and the factors identified in previous chapters could be used to predict footfall. Obtaining reliable and large-scale footfall data can be expensive and challenging for many high street stakeholders. Therefore, by applying machine learning methods to a database of previously collected footfall data from across the country, a model could be created that could predict footfall for similar retail locations where it is not monitored.

There exists research into footfall modelling and prediction, however this is heterogeneous, and the approach and methodology chosen heavily depends on the available data, the scope and the research question. However, two key gaps were identified that this analysis aimed to fill: creating a footfall prediction model that takes into account space and time and testing whether a footfall model could be generalisable to the entire country. In both these ways, the resulting model is novel for footfall analysis and prediction.

Objectives and results

Three objectives were established in order to design and construct this footfall model. These were,

What remains unknown	Define the criteria and use case for a footfall prediction model and identify appropriate methdologies to achieve it
	Create a preliminary model that predicts footfall that is location and time dependent
	Critique the performance of this model, identifying opportunities for improvement

Firstly, the use case, rationale and approach for the model was decided. The resulting final product would be a model that could give a reliable benchmark for footfall for any retail address in the UK. This benchmark would also be time-dependent, taking into account the hour of the day, the day of the week and the month of the year. This model would have applications for the monitoring and designing of revitalisation strategy, for retailers in store location planning and it could also be used as a measure of potential for a street, to determine where footfall could be increased. Although data limitations do mean that this final model is not achievable with the SmartStreetSensor dataset, this analysis serves as a proof of concept that this model could be possible in the future.

A direct-demand approach was chosen for this use case, which involves creating a regression model where the dependent variable is footfall and the independent variables are factors of population, demand and network analysis. In order to consider temporal variation, temporal factors were also added as independent features in the model, and 32 different models were created to reflect how the relationship between demand variables and footfall can also be dependent on time.

The model used 16 independent variables and applied a random forest regression model to the data.

The resulting combined model had an out-of-bag RMSE of 383 people per hour, and a RMSE on unseen data of 851 people per hour. Given the inherent unpredictability of footfall and the data quality issues, these results are fairly positive. The difference between the two scores do indicate that the model is currently overfitting to the noise in the training data, and usually this is resolved by the inclusion of more data. The model also shows high variance and a significant bias towards underpredicting footfall, particularly with busier sensors.

Several opportunities for development were identified. Firstly, more data will be needed to achieve the use case outlined. The data needs to be collected using a sustainable methodology that can be compared over time and space. In addition, more historic data is needed to understand annual and seasonal fluctuations. Another development would be to include more features of footfall demand, such as area attractiveness or space syntax measures, which could better model the trends in the data. Finally, the application of different methods could be explored to see if they would yield better results, such as LSTM models.

Limitations

This analysis had many limitations, with the most detrimental being the significant error within the data. This heavily limits the transferability of these results as the patterns identified in the model could merely reflect the noise or inconsistencies within the dataset, as opposed to real life trends. Although these footfall counts have been adjusted in line with manual counts, this does not appear to be a viable solution to ensure these counts reflect reality as the error itself is not consistent or predictable. Therefore, the only conclusions outside this analysis that could be applied would be theoretical and methodological, as opposed to footfall values themselves.

Another limitation is the spatial distribution of the sample. Sensor locations were largely decided by clients of LDC, therefore there can be large concentrations of sensors in particular environments, whereas other locations are significantly underrepresented. Therefore, when building a model using this data, the model ends up bias towards locations it knows well, for example London, and consequently makes poor predictions for other locations. Some of the London sensors could have been removed to balance the dataset, however the dataset was limited that it was important to ensure as many sensors as possible could be utilised.

The model also does not take into account every temporal factor. Predictions are dependent on the hour of day, the day of the week and the month of the year, but the model does not allow the user to enter a date or year. The original aspiration for this analysis did desire this to be included, but the lack of available data covering two years or longer, in addition to the influence of the COVID-19 pandemic meant this was impossible.

Contributions

A model that can give a time and location dependent benchmark for footfall has many practical applications. It could be applied as a key performance indicator to monitor the impact of any revitalisation strategy and the insights could be used to develop retail policies (Coca-Stefaniak, 2013; Graham, 2016; Hogg et al., 2004; Ministry of Housing, Communities & Local Government, 2014). As a measure of potential for a retail street, itcould help business owners set opening times, staffing hours and plan marketing campaigns (Denison, 2005; Underhill, 2009; Yiu and Ng, 2010), and it could also inform location planning for the location and opening of new stores. It provides an overview of a location and how it compares to locations around it. Although the model in this chapter should not be applied until it can be developed using more data, it presents a framework that could be applied in the future, when more data is collected.

In addition to the potential practical applications of a model like this, it also contributes to the research and literature into footfall and pedestrian flow prediction through the novel creation of a spatio-temporal footfall model.

Firstly, it tests the generalisability of footfall modelling. The majority of analyses in current literature focus primarily on one location or retail environment and design a methodology which incorporates the factors that are deemed necessary for that specific analysis. Very few then investigate if their model would be accurate when applied to other locations and contexts. By applying the same methodology and model across multiple retail environments across the country, this chapter will test whether footfall prediction models could be generalisable.

The model also incorporates temporal prediction, adjusting the footfall value dependant on factors of the time of the day, the day of the week and the month of the year. This was incredibly difficult to include in studies prior to the technological advances in automated footfall collection, which have facilitated the collection of high-resolution data. The research presented in this section is novel as it combines elements of time series analysis with spatial prediction to make footfall predictions that are both temporally and spatially dependent.

Future research could build on this model, applying footfall data that is more comparable and more representative of different areas to reduce overfitting and improve predictions. This model could also be developed by the integration of new features of demand and updating the features of demand that exist within the current model. This analysis also introduced other algorithms and approaches that could be applied in footfall prediction. For example, with complete and accurate time series information LSTM models could be applied and this could yield better results.

7.2 Discussion

The analysis chapters presented a novel, data-driven exploration into footfall and how it can behave and relate to the world around it. Each chapter focused on a specific research question, applying quantitative data analysis methods to achieve the answer with the data available. However, these answers often overlapped with each other, and allowed for wider insights on footfall data, how it is collected and applied, and its role in retail geography.

This section will focus on those insights, detailing the key messages that can be taken from this thesis and how they could be applied. Section 7.2.1 will critique the quality of the data and the viability of the collection method. Section 7.2.2 will focus on footfall context and to what extent the environment can determine footfall. Section 7.2.3 discusses the future applicability of any conclusions in the light of the COVID-19 pandemic. The recommendations for future research will be summarised in Section 7.2.4.

7.2.1 Data quality and viability

The lack of data quality and consistency has presented a significant challenge for these analyses. When compared with manual counts, it was found that the average absolute error of the sensors was 66.7%, which is a significant error margin. Multiple different sources of error were presented in Chapter 3. Some of these were related to phone ownership, some to logistical issues and some to technological barriers; however, the amalgamation of these sources of error result in data that is unreliable and inconsistent.

An attempt was made in Chapter 6 to adjust the counts to be closer to the manual counts collected. However, the error did not seem to be consistent over time, and the same sensor count appear to overcount one day and undercount on another. When some these sources of error were quantitatively investigated (e.g. through correlation with proximity to electric/phone shops, demographic & population statistics etc), again, no consistent pattern was found. Within the dataset, there were sensors that were within 100m of each other, measuring significantly different footfall counts.

Some of the results and conclusions from the analysis on this data appear to be in agreement with previous theories and literature on this topic. However, some results have been less intuitive. It is impossible to say whether any of the practical conclusions or quantitative patterns and relationships observed are reflective of reality. There is so much that is unknown about the data and its true accuracy. Therefore, I cannot recommend that any of the conclusions from the analyses are applied in real life based on this research as sole evidence. These analyses can still have purpose, as they are novel and the first to try to quantify and capture a lot of these interrelations between footfall and specific factors of form and function. However, the results can only be as reliable as the data used.

Another quality of the data that has been repeatedly highlighted in these analyses is the temporal coverage. Although the analyses included data from 2015-2020, many sensors were not running continuously for that time, with very few collecting more than two full years of data. This is unavoidable for LDC due to their methodology and business model which relies on client participation; however, this made it challenging to complete any time series analysis or to analyse the long-term changes and shifts in footfall.

Beyond the SmartStreetSensor data, there were also limitations in terms of supplementary data sources. The data sources used in this analysis were a combination of the LDC retail unit address data, the 2017 retail boundaries, the 2011 workplace zones and boundaries, the 2014 NapTAN data and the 2015 car parks dataset. Although some of these are regularly updated, workplace zones, NapTAN and car parks data are not. This was acceptable for this analysis as the footfall data covered 2015-2019, however as time goes on these datasets will become more inaccurate.

In addition, there were also factors that would have been interesting to explore, but were excluded from analysis due to lack of data. For example, tourist numbers could be a key footfall driver that was difficult to quantify or capture from supplementary data. Local factors, such as new housing estates being built which could increasing the population, and a calendar of events would also be useful data to collect. It could also be beneficial for the application of footfall prediction models if axial maps could be generated for retail areas in the UK. The analyses were limited in terms of measures of centrality and network connectivity as the resource with the best openly available coverage was road centrelines. It did not include the information needed to generate sightlines or axial maps, and to obtain this information and generate axial maps for all locations in the sample would be a significant undertaking.

Recommendations

There are two clear recommendations that could be taken from the data quality issues faced in these analyses.

Firstly, that the SmartStreetSensor data is not applied to draw any real-life conclusions without consideration of the error and assumptions that are made. The sensors are not comparable over space, with sensors next to each other reporting different results, and with the introduction of MAC address randomisation, the counts are not comparable over time. Although this may not be a concern for LDC's clients who just look at the data for their sensor, when comparing a network of sensors, the fact that there is no set range or proximity in which a person has to be to be counted by the sensor leaves means that treating each sensor as equal in terms of functionality and what they measure might be a large assumption. The error appears unpredictable and inconsistent; therefore, it is impossible to methodically remove. The Wi-Fi based methodology employed by LDC is unsustainable due to these external influences.

The second recommendation is that future footfall data collection methodologies put a high priority on sustainability and regularity of methods, if they would intend to pursue analyses similar to those in this thesis. As footfall has an annual seasonal component, it is important that at least two or three years of data is collected before this can be realistically taken into account. In order to discern the true impact of a certain factor on footfall, it would also be useful to analyse how footfall changes when a certain factor changes in a location. For example, monitoring how footfall may rise with population increase, or decrease with a rise in structural vacancy. These insights are only possible when a historic database of footfall for one location is collected.

7.2.2 Footfall context

Many of the bodies of work the present recommendations to improve high street vitality and viability (e.g. Grimsey et al., 2013; Parker et al., 2014; Portas, 2011; Swinney and Sivaev, 2013; Wrigley et al., 2015) highlight how a 'one approach fits all' mentality is not appropriate for high street planning, and that understanding is needed of the processes in a certain location. This philosophy somewhat contends the analyses completed in this thesis which have attempted to generalise and draw similarities between different places as opposed to establishing knowledge on how they are unique. Although these analysis have highlighted patterns and drawn conclusions, none of them have been unanimous and there were always locations that were exceptions to the rule. For example, for some locations, workplace population is incredibly important in fuelling footfall. In others, it does not seem to draw any correlation. Some areas can be generalised, whereas others seem to have their own rules where contextual knowledge is needed to understand.

In the same manner as a 'one approach fits all' mentality should not be directly applied to high street revitalisation, the results and conclusions observed with footfall and its relationship to other characteristics does not apply to every location. The classification created in Chapter 4, the trends observed in Chapter 5 and the model designed in Chapter 6 should not be directly applied to a location because every location is different. Footfall is influences by a complex collection of factors, some that may be quantifiable and others which are not. The strength of the influence of each factor can also change depending on the time of day, the month of the year, the political context and many other factors. As a result of this complexity, correlations which would assumed to be strong prior to analysis quite weak in reality. An example of this would be the lack of strong correlation between footfall and vacancy. The results have highlighted how one-sided and limited an insight footfall can give.

Footfall remains to be one factor which determines high street vitality and viability. It is the most commonly used indicator, and it is often quoted in media and articles as a barometer for how the high street is faring, but there are many other quantitative and qualitative measures that can be employed to give a more holistic view and footfall is best applied in combination with them.

However, locations which may be most in need of data and KPIs might not be those who can access the equipment needed to collect this data or employ the analysts to interpret it. Many of the locations that have the ability to collect and analyse footfall data are large, reasonable wealthy cities with funding, for example Manchester, Cambridge and Edinburgh.

It was hoped that some clear conclusions drawn from data driven research might help pave a way forward, providing insights that might be applicable for different towns and cities across the country. Succinctness and an easy transfer of knowledge was identified as a real barrier to progress in Chapter 2 and the classification in Chapter 4 and the model in Chapter 6 were designed to fulfil this purpose. However, due to the data limitations, the applicability of these output to real-world locations is limited.

The point should be made that these models are only generalisations and should be used to support an approach tailored to the local context. A conclusion from these analyses that is clear is that footfall is complex and impacted by many different, contextual factors, as are any insights from it. There does not seem to be any shortcuts or easy answers when it comes to high street revitalisation and any observations and conclusions that have been drawn from this analysis are not strong or consistent enough to be generalisable or transferable to other locations.

Recommendations

There are two recommendations that can be applied to future analyses. Firstly, there is a need to collect and monitor footfall and another KPIs in areas that need revitalisation or are outside of city centres. Conclusions from analysis using the

SmartStreetSensor data would be challenging to apply to the locations most in need of revitalisation and insights as these locations are not represented within the data.

The second recommendation is to undertake further research and analysis into how footfall correlates with spend. This research has shown how unpredictable footfall can be, therefore more needs to be known about how specifically footfall and spend relate to each other. Due to lack of available data, this was the only research gap identified in the literature review that this thesis could not contribute to. This is particularly important if a strategy intends to increase footfall in the hopes that it will increase economic strength, or if footfall will be applied to determine business rates or rents. There is an underlying assumption that more footfall means higher spend but there is little publicly available analysis that validates this. It is a strong recommendation that retailers and businesses look to share their sales data with researchers who can undertake this research, particularly through secure data providers such as the Consumer Data Research Centre.

7.2.3 Future applicability and paradigm shifts

During the writing on this thesis, we've witnessed a monumental change in how retail centres are used and interacted with. The forced closure of non-essential retail due to the pandemic and how social distance policies and fears about the virus have had significant impacts on communities and how people interact with each other. Many people used online retail who might've never done so before. This was already a trend pre-COVID, but the pandemic has accelerated these changes. Physical retail and footfall are less and less synonymous with the retail industry, how much people are spending and the strength of the retail sector.

The observations, results and conclusions in this thesis were all based on pre-COVID data, and the changes in the last few years have highlighted how quickly the standard and assumptions can change. For example, one of the strongest correlations observed was between footfall and workplace population. As more and more workplaces adopt permanent hybrid working models where their employees work from home for a portion of time, it will be interesting to see if that correlation weakens. Areas which relied on commuters and workers to buy products might find themselves suffering. The increase of people working from home could also be beneficial for local, convenience high streets as they might be utilised more, but there could also be marked changes in rural areas, as people relocate from expensive cities. The use of physical retail centres has shifted dramatically since the start of this thesis research and analysis whether these conclusions will still be true or relevant in a post-COVID world is yet to be seen.

COVID-19 has presented physical retail with a monumental shift, comparable in scale of the forces of change such as the rise of out-of-town retail and the 2008/09 financial crisis. It has also catalysed the impact of online shopping. In previous decades, the preservation and adaption of the high street have become a political priority, both due to the structural economic and social importance of the high street and due to public interest in sustainable high streets. It is yet to be seen if these revitalisation attempts will receive a new wind of public funding, or whether the managed decline of some high streets is becoming the more preferable option. In the last decade, the growth of online shopping has been astronomical, and has impacted each retail sector differently. However, for high street comparison retailers such as department stores, clothing stores or electronics stores, it is rare to find any company who can maintain a purely high street presence with no online store[25]. Many retailers who struggled to adapt post-financial crisis would have similarly struggled through the pandemic. The high street does not have the economic share that it once did, which could have implications for how wealth is geographically distributed.

High streets are also community and social hubs, and one key argument to preserving these environments is for local identity. However, the internet had also had somewhat of an impact in this also, as local online groups and noticeboards can foster a similar sense of community around a place without the necessity for the physical place. In the wake of the pandemic, the popularity of online forms of communication for social and working purposes is growing and becoming more engrained in everyday life. With innovations such as the metaverse predicted for the

[25] A notable exception is Primark, who maintain a purely physical retail presence with an emphasis on value clothes and homewares.

future, there would also be growing interest into what extent online community could subsidise physical interactions and community tied to physical place. Would the high street still serve this integral purpose in twenty or thirty years if online communities continue to grow?

To return to the present day, it is clear that the next few years will be a vital time for the sustainability of many high street environments. The pandemic has catalysed many economic and social changes including the rise of online retail and the adaption of hybrid working. The funding and focus, or lack of, that high street environments receive over the next few years could be highly influential to their future existence. Physical retail environments will need to adapt. Fortunately, the knowledge base on how this can be achieved is significantly larger than it was ten or fifteen years ago and many of the tools are available but may require smart application to be successful.

Recommendations

In the wake of the COVID-19 pandemic and in the move towards recovery for many high streets, consistent monitoring of key performance indicators will be more vital than ever. Measurements should be consistent and regular, ensuring that there is a variety of hard and soft indicators, and quantitative and qualitative evidence collected. Collecting this data could help decision makers understand more about their current context and identify and monitor improvements and trends.

A second recommendation is for the continuation of research that explores the future role of the high street as a retail destination, an economic hub and as a social and community centre. These should include recognition that the purposes and functions of the high street today may be different to those of the past and may continue to shift in the future.

7.2.4 Summary of recommendations

This discussion touched on three thematic discussion points which arose from the analyses within this thesis. The first explored the data quality issues within the dataset, recommending that the SmartStreetSensor footfall data is not applied without considerations of the significant error within the data. This could impact work published prior to the discovery of these issues, for example the research of Lugomer, 2019 and Lugomer and Longley, 2018. The second explored the conflict between the top-down and generalisable approach of footfall analyses and the bottom-up, contextual approach recommended in research on retail revitalisation strategy, noting the merit for both but acknowledging the barrier which can exist when gaining access to the required data. The third discussion point explored the future applicability of footfall research in a rapidly changing context, highlighting that the need for understanding high street performance is greater now than ever before. To summarise, the recommendations made are as follows:

1. The SmartStreetSensor footfall data should not be directly applied without considerations for the sources of error. Comparisons between sensors and with one sensor looking historically can be flawed.
2. Future footfall data collection methods should prioritise sustainability and measurement periods of two years or more if further insights on the behaviour of footfall want to be gained.
3. Increased funding for failing high streets who could benefit from collection of footfall and key performance indicator data.
4. Openness from retailers to share spend data so that it could be combined with footfall data to better understanding how the two can relate.
5. The consistent monitoring and collection of a combination of different key performance indicators by town centre decision makers.
6. The continuation of research exploring the future role and purpose of the high street, especially in a post-COVID world.

These recommendations have impacts both for policy and decision-makers and for future research.

Recommendations to policy and decision-makers

The key recommendation for policy makers and decision makers is that the integration of consistent and accurate footfall data can be incredibly valuable to monitoring and understanding the economic and social health of a retail centre. The value of footfall data compounds over time, as the more historic data exists, the better trends can be understood. Therefore, sustainability should be a priority when implementing footfall data collection methods.

Footfall data can be used to evaluate the successfulness of policies or events that might take place in a retail centre. Alternatively, it can be used to monitor the impact of negative events, such as future lockdowns or weather events. Footfall can give a numeric quantity which is comparable to historic levels and can be used to quickly translate these impacts to other stakeholders.

In addition, footfall data can be applied to identify opportunities for growth and development. Areas with low footfall could benefit from an analysis of the issues which are impacting that area and decision makers can implemented targeted policies to combat them, for example investing in infrastructure, reducing crime or increasing positive promotion. There is also opportunity to identify areas which currently have high footfall but where retail presence is underutilised. These areas can have high potential for growth and could be marketed to potential retailers as proposed locations for business.

Footfall data can also provide insight into the habits and behaviours of consumers, which stores they are attracted to, and at what times of day and week. This can be combined with surveys and qualitative research to understand what kind of function the retail centre might be serving and how this might shift over time. This can be used to inform the future vision of an area, which types of businesses to encourage and the infrastructure and amenities to provide.

Although footfall can provide powerful new insights for policy makers, some consideration should also be given to the barriers to successful policy implementation which have impacting the resilience of many retail centres in the UK. Firstly, the amount to which the implementation of retail policy is a priority to decision makers. There is an argument that, as economic and social hubs, the sustainability and improvement of physical retail centres should always be a priority to local decision makers, however at times there can be more urgent issues. Despite this, limited funding and lack of prioritisation of the high street from politicians at local and national scales can inhibit progress.

Coordination and cooperation between stakeholders can also be a barrier to successful implementation of retail policy. It might be that the long-term gain of a sustainable high street might have some short-term costs to some groups that they are unwilling to pay. For example, policy makers could decide to request retailers pay a fee to install footfall sensors within their premises, which retailers might not want to pay in advance of the benefit of those insights. There could also be trepidation that any change will disrupt a status quo that is working well for them. The installation of footfall sensors could reveal that their store is more profitable than the property owner thought, and result in rent increases. Alternatively, if a business does not seem to have a place in the proposed future vision of a high street, they might not be motivated to collaborate. On the other hand, cooperation can be achieved between policy makers, business holders and property owners without the consideration of the current residents and consumers, driving the gentrification of an area. While this might lead to a successful high street in an economic sense, the high street as a source of identity or community hub could be dramatically changed or destroyed in some cases.

Footfall data can be a valuable tool for policy and decision-making but without the foundation of stakeholder collaboration, a strong and inclusive vision for the retail centre and adequate funding, the successful implementation of retail policy can be challenging.

Recommendations for future research

For researchers working with sensor-collected footfall or pedestrian count data, the verifiable accuracy of the data source is paramount. Small differences can make big impacts when comparing footfall numbers, and the validity of conclusion greatly depend on the reliability of the data source.

There are several methods of checking the accuracy of footfall data. These include validating the data through manual counts, comparing sensors to the counts and variation of nearby sensors, comparing the data collected to historic data for that sensor and comparing the collected footfall data to data from other sources. It would be best practise for future researchers to pursue as many of these as feasible and report the results in their research.

There are many opportunities for future research to harness the qualities of footfall data to learn more about how people interact with physical retail locations and the wider implications of this. The link between footfall and spend is widely unknown, and a key element when it comes to understanding the economic relevance of footfall. Additionally, there is opportunity for research to build off the foundation of data built by organisations such as LDC and to compare future data with historic to understand how changes such as technology advances, retail strategies and remote working impact retail footfall. Finally, there is still more that can be learned and understood about how footfall relates to other morphological and functional properties of the surrounding environment, and if these could be used to predict and model pedestrian flows.

7.3 Concluding remarks

When this research commenced, there was very little known about footfall, how it relates to other factors and the generalisability of these relationships. This thesis has combined a vast amount of data from multiple sources to gain a greater understanding of footfall and these results could be best summarised by stating that footfall is complex.

Footfall is connected to many different characteristics of the environment, some which can be quantitively represented and tested, and others which cannot. The nature of these relationships can also vary from place to place – on micro-, meso- and macro-scales – and over time – from hour to hour, to the day of the week, to the time of year. The SmartStreetSensor dataset from LDC is one of the largest available footfall resources in terms of spatial and temporal representation, yet it could not cover enough unique locations over enough time for trends and patterns to be observed through the noise. The key contribution this research makes is to identify that future research will require more data, it will require consistent data and it will require better representation of different types of retail environments.

To gather and analyse footfall data takes investment of both time and money. Footfall encompasses many different processes - both tangible and intangible - into a singular responsive measure. There's somewhat of an irony to the fact that the qualities of footfall that make it so challenging to analyse and explain are the same qualities that make investments into understanding it so valuable.

With all limitations considered, it can be said that there does appear to be correlations between footfall and quantifiable characteristics of demand. Further, there is potential to harness these relationships to make plausible predictions for locations where data is not collected. This in itself is a strong and novel contribution that presents exciting opportunities for development. As technological advancements look set to continue capturing retail processes in novel and interesting ways, future research has the capability to understand the high street and consumers in more depth than was ever possible before.

At time of writing, the high street is emerging from the lockdowns of the COVID-19 pandemic. The international and national economic outlook appears uncertain and so smart adaption to shifts in consumer demand will become a greater necessity for UK high streets. There is indeed competition from online retail, but the increase in hybrid and remote working has the potential to shift the paradigm on consumer demand, and, for the first time in decades, favour local high streets.

What is known is that the value of footfall data compounds over time, and the longer reliable measurements have been taken for, the more understanding can be gained. Therefore, retailers and local decision makers should consider integrating technology and data collection into their physical stores, sooner rather than later. After all, you cannot make more time.

8 References

Antipov, G., Baccouche, M., Berrani, S.-A., Dugelay, J.-L., 2017. Effective training of convolutional neural networks for face-based gender and age prediction. Pattern Recognit. 72, 15–26. https://doi.org/10.1016/j.patcog.2017.06.031

Badorf, F., Hoberg, K., 2020. The impact of daily weather on retail sales: An empirical study in brick-and-mortar stores. J. Retail. Consum. Serv. 52, 101921.

Barron, C., Neis, P., Zipf, A., 2014. A comprehensive framework for intrinsic OpenStreetMap quality analysis. Trans. GIS 18, 877–895.

Batty, M., Desyllas, J., Duxbury, E., 2003. Safety in numbers? Modelling crowds and designing control for the Notting Hill Carnival. Urban Stud. 40, 1573–1590.

BBC News, 2021. Shopper numbers still almost 30% below pre-pandemic levels. BBC News.

BBC News, 2018a. Bad weather hits Christmas shopping on the High Street - BBC News [WWW Document]. URL https://www.bbc.co.uk/news/business-46583424 (accessed 10.14.21).

BBC News, 2018b. Boxing Day sales: Footfall down for third year, analysts say. BBC News.

Berry, T., Newing, A., Davies, D., Branch, K., 2016. Using workplace population statistics to understand retail store performance. Int. Rev. Retail Distrib. Consum. Res. 26, 375–395.

Birkin, M., Clarke, G., Clarke, M., 2017. Retail location planning in an era of multi-channel growth. Routledge.

Bishop, G., Welch, G., 2001. An introduction to the kalman filter. Proc SIGGRAPH Course 8, 41.

Boeing, G., 2017. OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Comput. Environ. Urban Syst. 65, 126–139.

Bolin, D., Verendel, V., Berghauser Pont, M., Stavroulaki, I., Ivarsson, O., Håkansson, E., 2021. Functional ANOVA modelling of pedestrian counts on streets in three European cities. J. R. Stat. Soc. Ser. A Stat. Soc.

Bonné, B., Barzan, A., Quax, P., Lamotte, W., 2013. WiFiPi: Involuntary tracking of visitors at mass events, in: 2013 IEEE 14th International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM). IEEE, pp. 1–6.

Boyle, M., 2019. Mobile internet statistics - 2020 [WWW Document]. Finder UK. URL https://www.finder.com/uk/mobile-internet-statistics (accessed 9.14.21).

Bras, N., Koster, H., Tang, C.K., van Ommeren, J.N., 2021. Covid-19 and Shopping Streets.

Brewin, D.G., 2020. The impact of COVID-19 on the grains and oilseeds sector. Can. J. Agric. Econ. Can. Agroeconomie 68, 185–188.

British Retail Consortium, 2021a. 2020 sees footfall down over 40% [WWW Document]. Br. Retail Consort. URL https://brc.org.uk/news/corporate-affairs/2020-sees-footfall-down-over-40/ (accessed 7.23.22).

British Retail Consortium, 2021b. Retail, Rates and Recovery: How business rates reform can maximise retail's role in levelling up.

British Retail Consortium, 2021c. Retail at Christmas.

British Retail Consortium, 2017. VACANCY RATE REMAINS STABLE THOUGH FOOTFALL DECLINE DEEPENS [WWW Document]. Br. Retail Consort. URL https://brc.org.uk/news/2017/vacancy-rate-remains-stable-though-footfall-decline-deepens (accessed 10.4.21).

Brown, S., 1991. Retail location: the post hierarchical challenge. Int. Rev. Retail Distrib. Consum. Res. 1, 367–381.

Bucklin, L.P., 1971. Retail gravity models and consumer choice: a theoretical and empirical critique. Econ. Geogr. 47, 489–497.

Burns, L., See, L., Heppenstall, A., Birkin, M., 2018. Developing an individual-level geodemographic classification. Appl. Spat. Anal. Policy 11, 417–437.

Burt, S., 2010. Retailing in Europe: 20 years on. Int. Rev. Retail Distrib. Consum. Res. 20, 9–27.

Cambridge BID, 2021. Footfall [WWW Document]. Camb. BID. URL https://www.cambridgebid.co.uk/footfall (accessed 12.19.21).

Campos, C., Dent, A., Fry, R., Reid, A., 2011. Impact of the recession. Reg. Trends 43, 1–69.

CDRC, 2019. Data Profile: SmartStreetSensor Footfall Data. [WWW Document]. CDRC. URL https://data.cdrc.ac.uk/system/files/LDC6_Profile.pdf (accessed 9.14.21).

CDRC, n.d. Data Services | CDRC Data [WWW Document]. CDRC. URL https://data.cdrc.ac.uk/data-services (accessed 9.14.21).

CEMBa, n.d. What are the levels of evidence? – Center for Evidence Based Management. URL https://cebma.org/faq/what-are-the-levels-of-evidence/ (accessed 10.13.21).

Centre for Retail Research, 2020a. List of Major Retail Failures 2008 to end-2019.

Centre for Retail Research, 2020b. Who's Gone Bust in Retail? The Centre For Retail Research [WWW Document]. Cent. Retail Res. URL https://www.retailresearch.org/whos-gone-bust-retail.html#bycompany (accessed 10.14.21).

Chen, C., Liu, L.-M., 1993. Forecasting time series with outliers. J. Forecast. 12, 13–35.

Chen, H., Zhou, Y., 2020. Forecasting high-street footfall in real time. Proc. GISRUK 2020.

CityCo, 2021. Manchester Monthly Footfall Trends. CityCo. URL https://cityco.com/report/monthly-footfall/ (accessed 12.19.21).

Coca-Stefaniak, A., 2013. Successful town centres: Developing effective strategies.

Collinson, P., 2018. Black Friday will shift sales online and cut pre-Christmas footfall. The Guardian.

Cooper, C.H., Harvey, I., Orford, S., Chiaradia, A.J., 2021. Using multiple hybrid spatial design network analysis to predict longitudinal effect of a major city centre redevelopment on pedestrian flows. Transportation 48, 643–672.

Cunche, M., 2014. I know your MAC address: targeted tracking of individual using Wi-Fi. J. Comput. Virol. Hacking Tech. 10, 219–227.

Cunche, M., Matte, C., 2016. On Wi-Fi Tracking and the Pitfalls of MAC Address Randomization. Nouv. Défis L'Internet Obj. Interact. Homme-Mach. Facteurs Hum. 18–18.

Dalgeish, R., 2021. Retail sales, Great Britain - Office for National Statistics February 2021 [WWW Document]. URL https://www.ons.gov.uk/businessindustryandtrade/retailindustry/bulletins/retailsales/february2021 (accessed 7.23.22).

Daramola-Martin, A., 2009. Liverpool One and the transformation of a city: Place branding, marketing and the catalytic effects of regeneration and culture on repositioning Liverpool. Place Brand. Public Dipl. 5, 301–311.

Deloitte, 2022. Sustainability & Consumer Behaviour 2022 [WWW Document]. Deloitte U. K. URL https://www2.deloitte.com/uk/en/pages/consumer-business/articles/sustainable-consumer.html (accessed 2.25.23).

Demir, L., Cunche, M., Lauradoux, C., 2014. Analysing the privacy policies of Wi-Fi trackers, in: Proceedings of the 2014 Workshop on Physical Analytics. pp. 39–44.

Denison, T., 2005. Breaking from convention: Bringing rigour to retail adtracking. J. Target. Meas. Anal. Mark. 13, 234–238.

Department for Communities and Local Government, 2009. Extent of retail development taking place in England, 2007.

Derrett, R., 2003. Making sense of how festivals demonstrate a community's sense of place. Event Manag. 8, 49–58.

Desyllas, J., Duxbury, E., Ward, J., Smith, A., 2003. Pedestrian demand modelling of large cities: an applied example from London.

Dolega, L., Lord, A., 2020. Exploring the geography of retail success and decline: A case study of the Liverpool City Region. Cities 96, 102456.

Dolega, L., Pavlis, M., Singleton, A., 2016. Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. J. Retail. Consum. Serv. 28, 78–90.

Dolega, L., Reynolds, J., Singleton, A., Pavlis, M., 2021. Beyond retail: New ways of classifying UK shopping and consumption spaces. Environ. Plan. B Urban Anal. City Sci. 48, 132–150. https://doi.org/10.1177/2399808319840666

Ebeling, D., Luker, Z., Pacheco, S., Payne, A., Rae, N., 2018. Validating self-reported trends using WiFi tracking, in: International Conference on Human-Computer Interaction. Springer, pp. 233–237.

Erath, A., van Eggermond, M.A., Ordóñez, S.A., Axhausen, K.W., 2017. Introducing the pedestrian accessibility tool: walkability analysis for a geographic information system. Transp. Res. Rec. 2661, 51–61.

Fenwick, S., 2017. Shopping choices shrink in Portas Pilot towns. BBC News.

Findlay, A., Sparks, L., 2014. Secondary retail during economic crisis and austerity, in: Evolving High Streets: Resilience and Reinvention-Perspectives from Social Science. ESRC. University of Southampton.

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. Sociometry 35–41.

Freudiger, J., 2015. How talkative is your mobile device? An experimental study of Wi-Fi probe requests, in: Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks. pp. 1–6.

Genecon LLP, 2011. Understanding High Street performance 147.

Giantspectacular.com, 2019. Giant Spectacular Liverpool's Dream [WWW Document]. Liverp. Dream. URL https://www.giantspectacular.com/ (accessed 10.14.21).

Goddard, E., 2021. The impact of COVID-19 on food retail and food service in Canada: A second assessment. Can. J. Agric. Econ. Can. Agroeconomie 69, 167–175.

GOV.UK, 2021. Trading hours for retailers: the law [WWW Document]. GOV.UK. URL https://www.gov.uk/trading-hours-for-retailers-the-law (accessed 10.14.21).

GOV.UK, 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. [WWW Document]. URL https://www.legislation.gov.uk/eur/2016/679/contents (accessed 9.14.21).

Graham, C., 2016. The relationship between high street footfall, attraction and conversion, in: Business in a Dynamic World Conference.

Graham, C., Khan, K., Ilyas, M., 2019. Estimating the value of passing trade from pedestrian density. J. Retail. Consum. Serv. 46, 103–111.

Graham, C., Peleg, A., 2017. Will a smart city have a high street?, in: Sustainable Smart Cities. Springer, pp. 143–157.

Griffiths, S., Vaughan, L., Haklay, M., Emma Jones, C., 2008. The sustainable suburban high street: a review of themes and approaches. Geogr. Compass 2, 1155–1188.

Grimsey, B., Hopkinson, M., Hood, N., Pascoe, E., Shellard, C., Sadek, J., Cassidy, K., Dehullu, V., Baker, M., 2018. The Grimsey Review 2. London.

Grimsey, B., Turner-Mitchell, P., Shellard, C., Pascoe, E., Baker, M., Vasili, S., Sadek, J., 2013. The Grimsey review: An alternative future for the high street. London.

Guy, C., 2006. Retail productivity and land-use planning: negotiating 'joined-up'retail planning policy. Environ. Plan. C Gov. Policy 24, 755–770.

Guy, C.M., 1998. Classifications of retail stores and shopping centres: some methodological issues. GeoJournal 45, 255–264.

Haklay, M., 2010. How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. Environ. Plan. B Plan. Des. 37, 682–703. https://doi.org/10.1068/b35097

Hanke, J.E., Wichern, D.W., 2005. Business forecasting. Pearson Educación.

Harding, A.J.M., Powell, F., 2011. Variations in pedestrian traffic count in christchurch due to the september 2010 darfield (canterbury) earthquake, in: Proceedings of the Ninth Pacific Conference on Earthquake Engineering: Building an Earthquake-Resilient Society. pp. 14–16.

Harper Dennis Hobbs, 2021. Vitality Rankings 2021 - Beaconsfield named as Britain's leading retail destination [WWW Document]. Harper Dennis Hobbs. URL http://hdh.co.uk/retail/vitality-rankings-2021-beaconsfield-named-britains-leading-retail-destination/ (accessed 9.23.21).

Harper Dennis Hobbs, 2017. 2017 Vitality Rankings – Top British Centres.

Hart, C., Stachow, G.B., Rafiq, M., Laing, A., 2014. The customer experience of town centres.

Hartley, R., 2012. Managing High Street change through silliness. J. Urban Regen. Renew. 6, 164–171.

Haslam, D., 2004. Manchester England.

Hillier, B., Hanson, J., 1984. The Social Logic of Space, Cambridge Univ.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., Xu, J., 1993. Natural movement: or, configuration and attraction in urban pedestrian movement. Environ. Plan. B Plan. Des. 20, 29–66.

Hogg, S., Medway, D., Warnaby, G., 2004. Town centre management schemes in the UK: marketing and performance indicators. Int. J. Nonprofit Volunt. Sect. Mark. 9, 309–319.

Hood, N., Clarke, G., Clarke, M., 2016. Segmenting the growing UK convenience store market for retail location planning. Int. Rev. Retail Distrib. Consum. Res. 26, 113–136.

Hoogendoorn, S.P., 2004. Pedestrian flow modeling by adaptive control. Transp. Res. Rec. 1878, 95–103.

Hristov, L., Reynolds, J., 2015. Perceptions and practices of innovation in retailing: Challenges of definition and measurement. Int. J. Retail Distrib. Manag. 43, 126–147. https://doi.org/10.1108/IJRDM-09-2012-0079

Huff, D.L., 1963. A probabilistic analysis of shopping center trade areas. Land Econ. 39, 81–90.

Hwang, I., Jang, Y.J., 2017. Process mining to discover shoppers' pathways at a fashion retail store using a WiFi-base indoor positioning system. IEEE Trans. Autom. Sci. Eng. 14, 1786–1792.

Institute for Government, 2021. Timeline of UK coronavirus lockdowns, March 2020 to March 2021.

Islam, S., Jones, S.L., Dye, D., 2014. Comprehensive analysis of single-and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama. Accid. Anal. Prev. 67, 148–158.

Javare, P., Khetan, D., Kamerkar, C., Gupte, Y., Chachra, S., Joshi, U., 2020. Using Object Detection and Data Analysis for Developing Customer Insights in a Retail Setting, in: Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST).

Jiang, B., 2009. Ranking spaces for predicting human movement in an urban environment. Int. J. Geogr. Inf. Sci. 23, 823–837.

Jiang, P., Zhang, Y., Fu, W., Liu, H., Su, X., 2015. Indoor mobile localization based on Wi-Fi fingerprint's important access point. Int. J. Distrib. Sens. Netw. 11, 429104.

Joshi, V., Jha, K., Jain, M., Kulkarni, S., 2021. Tourism Footfall Forecasting and Recommendation System, in: 2021 International Conference on Communication Information and Computing Technology (ICCICT). IEEE, pp. 1–5.

Kalikova, J., Krcal, J., 2017. People counting by means of wi-fi, in: 2017 Smart City Symposium Prague (SCSP). IEEE, pp. 1–3.

Kerr, D., 2013. Russian police spy on people's mobile data to catch thieves - CNET [WWW Document]. CNet. URL https://www.cnet.com/tech/services-and-software/russian-police-spy-on-peoples-mobile-data-to-catch-thieves/ (accessed 9.14.21).

Kirkup, M.H., Rafiq, M., 1999. Marketing shopping centres: challenges in the UK context. J. Mark. Pract. Appl. Mark. Sci.

Kitazawa, K., Batty, M., 2004. Pedestrian behaviour modelling.

Klapka, P., Erlebach, M., Král, O., Lehnert, M., Mička, T., 2013. The footfall of shopping centres in Olomouc (Czech Republic): an application of the gravity model. Morav. Geogr. Rep. 21, 12–26.

Kleinman, Z., 2020. Covid contact-tracing app not sharing data with police. BBC News.

Kniffin, K.M., Narayanan, J., Anseel, F., Antonakis, J., Ashford, S.P., Bakker, A.B., Bamberger, P., Bapuji, H., Bhave, D.P., Choi, V.K., 2021. COVID-19 and the workplace: Implications, issues, and insights for future research and action. Am. Psychol. 76, 63.

Knight Frank, 2019. UK Retail Monitor Q1 2019 2.

Kobsa, A., 2014. User acceptance of footfall analytics with aggregated and anonymized mobile phone data, in: International Conference on Trust, Privacy and Security in Digital Business. Springer, pp. 168–179.

Kollewe, J., 2020. Eat out to help out scheme increases UK high street footfall | Eat out to help out | The Guardian [WWW Document]. The Guardian. URL https://www.theguardian.com/business/2020/aug/10/eat-out-to-help-out-scheme-increases-uk-high-street-footfall-covid-19 (accessed 10.14.21).

Komninos, A., Dunlop, M.D., Wilson, J.N., 2021. Combining Infrastructure Sensor and Tourism Market Data in a Smart City Project—Case Study 1, in: Intelligent Computing for Interactive System Design: Statistics, Digital Signal Processing, and Machine Learning in Practice. pp. 197–233.

Kuhn, M., n.d. Package "caret" - Classification and Regression training.

Lee, S., Yoo, C., Seo, K.W., 2020. Determinant Factors of Pedestrian Volume in Different Land-Use Zones: Combining Space Syntax Metrics with GIS-Based Built-Environment Measures. Sustainability 12, 8647.

Lerman, Y., Rofè, Y., Omer, I., 2014. Using space syntax to model pedestrian movement in urban transportation planning. Geogr. Anal. 46, 392–410.

Liu, M., Li, L., Li, Q., Bai, Y., Hu, C., 2021. Pedestrian flow prediction in open public places using graph convolutional network. ISPRS Int. J. Geo-Inf. 10, 455.

Liu, X., Jiang, B., 2012. Defining and generating axial lines from street center lines for better understanding of urban morphologies. Int. J. Geogr. Inf. Sci. 26, 1521–1532.

Lloyd, S., 1982. Least squares quantization in PCM. IEEE Trans. Inf. Theory 28, 129–137.

Lo, R.H., 2009. Walkability: what is it? J. Urban. 2, 145–166.

López-de-Lacalle, J., 2019. Package "tsoutliers" documentation.

Lugomer, K., 2019. The Composition and Functioning of Retail Areas (Doctoral). Dr. Thesis UCL Univ. Coll. Lond. UCL (University College London).

Lugomer, K., Longley, P., 2018. Towards a comprehensive temporal classification of footfall patterns in the cities of Great Britain, in: Leibniz International Proceedings in Informatics, LIPIcs. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, p. 43.

Lugomer, K., Soundararaj, B., Murcio, R., Cheshire, J., Longley, P., 2017. Understanding sources of measurement error in the Wi-Fi sensor data in the Smart City, in: Proceedings of GISRUK 2017. GIS Research UK (GISRUK).

Lyons, G., Chatterjee, K., 2008. A human perspective on the daily commute: costs, benefits and trade-offs. Transp. Rev. 28, 181–198.

MacDonald, J., Singleton, A., Dolega, L., Ballantyne, P., 2022. Data Profile: Retail Centre Boundaries.

Makkar, G., 2020. Real-Time Footfall Prediction Using Weather Data: A Case on Retail Analytics, in: Data Management, Analytics and Innovation. Springer, pp. 529–542.

Marr, B., 2021. 10 Best Examples Of Augmented And Virtual Reality In Retail [WWW Document]. Forbes. URL https://www.forbes.com/sites/bernardmarr/2021/09/13/10-best-examples-of-augmented-and-virtual-reality-in-retail/ (accessed 2.25.23).

Martin, J., Mayberry, T., Donahue, C., Foppe, L., Brown, L., Riggins, C., Rye, E.C., Brown, D., 2017. A study of MAC address randomization in mobile devices and when it fails. ArXiv Prepr. ArXiv170302874.

Martínez-de-Albéniz, V., Belkaid, A., 2021. Here comes the sun: Fashion goods retailing under weather fluctuations. Eur. J. Oper. Res. 294, 820–830.

Matte, C., Cunche, M., Rousseau, F., Vanhoef, M., 2016. Defeating MAC address randomization through timing attacks, in: Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks. pp. 15–20.

Mazumdar, S., Bagheri, N., Chong, S., Cochrane, T., Jalaludin, B., Davey, R., 2020. A Hotspot of Walking in and around the Central Business District: Leveraging Coarsely Geocoded Routinely Collected Data. Appl. Spat. Anal. Policy 13, 649–668.

Met Office, 2021. Past weather events [WWW Document]. Met Off. URL https://www.metoffice.gov.uk/weather/learn-about/past-uk-weather-events (accessed 12.17.21).

Met Office, 2018a. Ex-Hurricane Ophelia 16 October 2017.

Met Office, 2018b. Snow and low temperatures February to March 2018.

Met Office, 2018c. Strong winds and heavy rain from Storm Hector.

Met Office, 2018d. Summer 2018.

Met Office, 2018e. Strong winds and heavy rain from storms Ali and Bronagh.

Met Office, 2018f. Storm Deirdre.

Mikušová, M., Janečková, V., 2010. Developing and implementing successful key performance indicators. World Acad. Sci. Eng. Technol. 42, 969–981.

Millington, S., Ntounis, N., 2017. Repositioning the high street: evidence and reflection from the UK. J. Place Manag. Dev.

Millington, S., Ntounis, N., Parker, C., Quin, S., 2015. Multifunctional Centres: a sustainable role for town and city centres. Manch. Inst. Place Manag.

Millington, S., Ntounis, N., Parker, C., Quin, S., Roberts, G., Steadman, C., 2018. High street 2030: Achieving change. Manch. Metrop. Univ. Inst. Place Manag.

Ministry of Housing, Communities & Local Government, 2019. National non-domestic rates collected by councils in England: 2017 to 2018 [WWW Document]. GOV.UK. URL https://www.gov.uk/government/statistics/national-non-domestic-rates-collected-by-councils-in-england-2017-to-2018 (accessed 2.25.23).

Ministry of Housing, Communities & Local Government, 2014. Town centres and retail [WWW Document]. GOV.UK. URL https://www.gov.uk/guidance/ensuring-the-vitality-of-town-centres (accessed 10.14.21).

Moreira, A., Santos, M.Y., Carneiro, S., 2005. Density-based clustering algorithms–DBSCAN and SNN. Univ. Minho-Port. 1–18.

Moritz, S., Bartz-Beielstein, T., 2017. imputeTS: time series missing value imputation in R. R J 9, 207.

Mumford, C., Parker, C., Ntounis, N., Dargan, E., 2021. Footfall signatures and volumes: Towards a classification of UK centres. Environ. Plan. B Urban Anal. City Sci. 48, 1495–1510.

Mumford, C.L., Parker, C.R., Ntounis, N., Dargan, E., 2020. Supplementary Material for Footfall Signatures and Volumes: Towards a Classification of UK Centres.

Munira, S., Sener, I.N., 2017. Use of Direct-Demand Modeling in Estimating Nonmotorized Activity: A Meta-analysis 49.

Nakanishi, M., Cooper, L.G., 1974. Parameter estimation for a multiplicative competitive interaction model—least squares approach. J. Mark. Res. 11, 303–311.

Naseri, R.N.N., 2021. Issues And Challenges Of Online Shoppingactivities On The Impact Of Corona Pandemic: A Study On Malaysia Retail Industry. Turk. J. Comput. Math. Educ. TURCOMAT 12, 7682–7686.

National Infrastructure Commission, 2019. Data measuring the connectivity within (intra-urban), and between (inter-urban), different places [WWW Document]. NIC. URL https://nic.org.uk/data/all-data/transport-connectivity/ (accessed 12.19.21).

National Records of Scotland, 2019. Population Estimates. [WWW Document]. URL http://citypopulation.de/en/uk/cities/ua/ (accessed 12.19.21).

Naylor, R., Branzanti, C., McLean, B., Lonie, D., 2016. Edinburgh Festivals Impact Study.

Newing, A., Clarke, G., Clarke, M., 2018. Applied spatial modelling for retail planning in tourist resorts. Int. J. Retail Distrib. Manag.

Ntounis, N., Parker, C., 2017. Engaged scholarship on the High Street: the case of HSUK2020. J. Place Manag. Dev.

O'Dea, S., 2021. UK: smartphone ownership by age 2021 | Statista [WWW Document]. Statista. URL https://www.statista.com/statistics/271851/smartphone-owners-in-the-united-kingdom-uk-by-age/ (accessed 9.14.21).

O'Dea, S., 2020. Children: tablet and smartphone ownership by age UK 2019 | Statista [WWW Document]. Statista. URL https://www.statista.com/statistics/805397/children-ownership-of-tablets-smartphones-by-age-uk/ (accessed 9.14.21).

Oliveira, L., Schneider, D., De Souza, J., Shen, W., 2019. Mobile device detection through WiFi probe request analysis. IEEE Access 7, 98579–98588.

Omer, I., Kaplan, N., 2017. Using space syntax and agent-based approaches for modeling pedestrian volume at the urban scale. Comput. Environ. Urban Syst. 64, 57–67.

ONS, 2021. Retail Sales Index time series - Office for National Statistics [WWW Document]. Off. Natl. Stat. URL https://www.ons.gov.uk/businessindustryandtrade/retailindustry/datasets/retailsales (accessed 10.14.21).

ONS, 2020a. High streets in Great Britain - Office for National Statistics March 2020 [WWW Document]. Off. Natl. Stat. URL https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/highstreetsingreatbritain/march2020 (accessed 10.14.21).

ONS, 2020b. Overview of the UK population - Office for National Statistics [WWW Document]. URL https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/2020 (accessed 2.25.23).

ONS, 2020c. Impact of the coronavirus (COVID-19) pandemic on retail sales in 2020 - Office for National Statistics [WWW Document]. Off. Natl. Stat. URL https://www.ons.gov.uk/economy/grossdomesticproductgdp/articles/impactofthecoronaviruscovid19pandemiconretailsalesin2020/2021-01-28 (accessed 10.14.21).

ONS, 2020d. Coronavirus and shielding of clinically extremely vulnerable people in England - Office for National Statistics [WWW Document]. Off. Natl. Stat. URL https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronavirusandshieldingofclinicallyextremelyvulnerablepeopleinengland/28mayto3june2020 (accessed 10.14.21).

ONS, 2020e. Population estimates for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics [WWW Document]. URL https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/latest (accessed 10.8.21).

ONS, 2019. United Kingdom: Countries and Major Urban Areas - Population Statistics, Maps, Charts, Weather and Web Information [WWW Document]. URL http://citypopulation.de/en/uk/cities/ua/ (accessed 12.19.21).

ONS, 2018. Data from 2017's mid-year population estimates for UK, England and Wales, Scotland and Northern Ireland.

Ordnance Survey, 2023. OS Open Roads | Data Products [WWW Document]. Ordnance Surv. URL https://www.ordnancesurvey.co.uk/business-government/products/open-map-roads (accessed 2.24.23).

O'Sullivan, D., Jackson, M.J., 2002. Festival Tourism: A Contributor to Sustainable Local Economic Development? J. Sustain. Tour. 10, 325–342. https://doi.org/10.1080/09669580208667171

Parker, C., Ntounis, N., Millington, S., Quin, S., Castillo-Villar, F.R., 2017. Improving the vitality and viability of the UK High Street by 2020: Identifying priorities and a framework for action. J. Place Manag. Dev.

Parker, C., Ntounis, N., Quin, S., Grime, I., 2014. High Street research agenda: identifying High Street research priorities. J. Place Manag. Dev.

Parker, C., Ntounis, N., Quin, S., Millington, S., 2016. Identifying factors that influence vitality and viability. Inst. Place Manag.

Parker, C.R., Ntounis, N., Quin, S., Millington, S., 2016. High Street UK 2020 Project Report.

Parker, D., Welsh Government, 2015. Evaluation of Food Festivals Supported by the Welsh Government for the Financial Period of April 2014–2015 127.

Parsons, A.G., 2001. The association between daily weather and daily shopping patterns. Australas. Mark. J. AMJ 9, 78–84.

Partington, R., 2020. Big fall in Black Friday UK retail sales despite online spending boom. The Guardian.

Pavlis, M., Dolega, L., Singleton, A., 2018. A modified DBSCAN clustering method to estimate retail center extent. Geogr. Anal. 50, 141–161.

Philp, S., Dolega, L., Singleton, A., Green, M., 2021. Archetypes of Footfall Context: Quantifying Temporal Variations in Retail Footfall in relation to Micro-Location Characteristics. Appl. Spat. Anal. Policy 1–27.

Porta, S., Strano, E., Iacoviello, V., Messora, R., Latora, V., Cardillo, A., Wang, F., Scellato, S., 2009. Street centrality and densities of retail and services in Bologna, Italy. Environ. Plan. B Plan. Des. 36, 450–465.

Portas, M., 2011. The portas review. Indep. Rev. Future Our High Str.

Pullens, B.B., 2018. Footfall as rent-determinant.

Raford, N., Ragland, D., 2006. Pedestrian volume modeling for traffic safety and exposure analysis: The case of Boston, Massachusetts.

Ravenscroft, N., Reeves, J., Rowley, M., 2000. Leisure, property, and the viability of town centres. Environ. Plan. A 32, 1359–1374.

Read, S., 1999. Space syntax and the Dutch city. Environ. Plan. B Plan. Des. 26, 251–264.

Redondi, A.E.C., Sanvito, D., Cesana, M., 2016. Passive classification of Wi-Fi enabled devices, in: Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. pp. 51–58.

Reilly, W.J., 1931. The law of retail gravitation. WJ Reilly.

Reynolds, J., 2000. eCommerce: a critical review. Int. J. Retail Distrib. Manag.

Sanderson, D.C., Edwards, V.M., 2014. What Tenants Want: UK occupiers' requirements when renting commercial property and strategic implications for landlords 27.

Scheurer, J., Porta, S., 2006. Centrality and connectivity in public transport networks and their significance for transport sustainability in cities, in: World Planning Schools Congress, Global Planning Association Education Network,.

Schimpl, M., Moore, C., Lederer, C., Neuhaus, A., Sambrook, J., Danesh, J., Ouwehand, W., Daumer, M., 2011. Association between walking speed and age in healthy, free-living individuals using mobile accelerometry--a cross-sectional study. PloS One 6, e23299. https://doi.org/10.1371/journal.pone.0023299

Schneider, R.J., Henry, T., Mitman, M.F., Stonehill, L., Koehler, J., 2012. Development and application of volume model for pedestrian intersections in San Francisco, California. Transp. Res. Rec. 2299, 65–78.

Sevtsuk, A., 2021. Estimating pedestrian flows on street networks: revisiting the betweenness index. J. Am. Plann. Assoc. 87, 512–526.

Sharma, K., 2011. Impact of recession on buying behavior of Indian consumers. Int. Rev. Bus. Res. Pap. 7, 381–392.

Shen, Y., Karimi, K., 2016. Urban function connectivity: Characterisation of functional urban streets with social media check-in data. Cities 55, 9–21.

Sidla, O., Lypetskyy, Y., Brandle, N., Seer, S., 2006. Pedestrian detection and tracking for counting applications in crowded situations, in: 2006 IEEE International Conference on Video and Signal Based Surveillance. IEEE, pp. 70–70.

Singleton, A.D., Dolega, L., Riddlesden, D., Longley, P.A., 2016. Measuring the spatial vulnerability of retail centres to online consumption through a framework of e-resilience. Geoforum 69, 5–18.

Sky News, 2018. Shoppers shun high street on peak Christmas weekend [WWW Document]. Sky News. URL https://news.sky.com/story/shoppers-shun-high-street-on-peak-christmas-weekend-11582520 (accessed 11.22.21).

Smithers, R., 2018. World Cup and hot weather bolster supermarkets sales. The Guardian.

Soundararaj, B., Cheshire, J., Longley, P., 2020. Estimating real-time high-street footfall from Wi-Fi probe requests. Int. J. Geogr. Inf. Sci. 34, 325–343.

Spielman, S.E., Singleton, A., 2015. Studying neighborhoods using uncertain data from the American community survey: a contextual approach. Ann. Assoc. Am. Geogr. 105, 1003–1025.

Springboard, n.d. High Streets | Retail Data Analytics | Store Performance. URL https://www.spring-board.info/high-streets/ (accessed 9.15.21).

Srinivasan, S.R., Sivakumar, S.N.V., 2011. Strategies for retailers during recession. J. Bus. Retail Manag. Res. 5, 94–104.

Sruthi, M.S., 2019. Iot based real time people counting system for smart buildings. Int. J. Emerg. Technol. Innov. Eng. 5.

Stavroulaki, G., Bolin, D., Pont, M.B., Marcus, L., Håkansson, E., 2019. Statistical modelling and analysis of big data on pedestrian movement, in: Proceedings Sof the 12th Space Syntax Symposium. pp. 1–24.

Stevens, B., 2018. Poundworld: What went wrong (Part I) - Retail Gazette [WWW Document]. URL https://www.retailgazette.co.uk/blog/2018/06/poundworld-went-wrong-part-i/ (accessed 7.23.22).

Su, J., Jin, Y.-X., 2007. In doors location technology research based on WLAN, in: Proc. 6th WSEAS Int. Conf. Instrum., Meas., Circuits Syst. pp. 109–114.

Sun, X., 2012. An urban morphological study on Swedish cities from a topological perspective.

Sun, Y., Zhang, M., Sun, Z., Tan, T., 2018. Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers. IEEE Trans. Pattern Anal. Mach. Intell. 40, 332–351. https://doi.org/10.1109/TPAMI.2017.2669035

Swinney, P., Sivaev, D., 2013. Beyond the High Street. Why our city centres really matter.

Teller, C., Alexander, A., 2014. Store managers–the seismographs in shopping centres. Eur. J. Mark.

Theodoridis, C., Ntounis, N., Pal, J., 2017. How to reinvent the High Street: evidence from the HS2020. J. Place Manag. Dev.

Thurstain-Goodwin, M., Unwin, D., 2000. Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations. Trans. GIS 4, 305–317.

Trasberg, T., Cheshire, J.A., Longley, P., 2018. Integrating New Measures of Retail Unit Attractiveness into Spatial Interaction Models, in: Proceedings of GISRUK 2018. GIS Research UK (GISRUK).

Tucker, I., 2019. Why you should worry if you have a Chinese smartphone. The Observer.

Turner, A., Penn, A., 2002. Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment. Environ. Plan. B Plan. Des. 29, 473–490.

Tyler, S., Semper, G., Guest, P., Fieldhouse, B., 2012. The relevance of parking in the success of urban centres, A review for London Councils.

Underhill, P., 2009. Why we buy: The science of shopping–updated and revised for the Internet, the global consumer, and beyond. Simon and Schuster.

Unwin, J., Symonds, P., Laike, T., 2017. Does lighting affect pedestrian flows? A pilot study in Lund, Market Harborough and Dublin, in: Proceedings of the Lux Europa 2017 European Lighting Conference-Lighting for Modern Society. Lighting engineering society of Slovenia, Ljubljana, Slovenia.

URBED, 1997. Town Centre Partnerships: A survey of good practice and report of an action research project.

Üsküplü, T., Terzi, F., Kartal, H., 2020. Discovering activity patterns in the city by social media network data: a case study of Istanbul. Appl. Spat. Anal. Policy 13, 945–958.

Vickers, D., Rees, P., 2007. Creating the UK National Statistics 2001 output area classification. J. R. Stat. Soc. Ser. A Stat. Soc. 170, 379–403.

Wang, X., Liono, J., McIntosh, W., Salim, F.D., 2017. Predicting the city foot traffic with pedestrian sensor data, in: Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, MobiQuitous 2017. Association for Computing Machinery, New York, NY, USA, pp. 1–10. https://doi.org/10.1145/3144457.3152355

Warnaby, G., Yip, K.M., 2005. Promotional planning in UK regional shopping centres: an exploratory study. Mark. Intell. Plan.

Weltevreden, J.W., 2007. Substitution or complementarity? How the Internet changes city centre shopping. J. Retail. Consum. Serv. 14, 192–207.

Weppner, J., Bischke, B., Lukowicz, P., 2016. Monitoring crowd condition in public spaces by tracking mobile consumer devices with wifi interface, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. pp. 1363–1371.

Whipp, A., Malleson, N., Ward, J., Heppenstall, A., 2021. Estimates of the Ambient Population: Assessing the Utility of Conventional and Novel Data Sources. ISPRS Int. J. Geo-Inf. 10, 131.

Williamson, B., Hargreaves, J., Bond, J., Lay, H., 2006. The economic costs and benefits of easing Sunday shopping restrictions on large stores in England and Wales. Rep. Dep. Trade Ind.

Wood, S., Reynolds, J., 2012. Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. Geoforum, Themed issue: Spatialities of Ageing 43, 1076–1087. https://doi.org/10.1016/j.geoforum.2012.06.014

Wood, Z., 2019. UK high streets "in downward spiral" with one in 10 shops empty. The Guardian.

Wright, G., 2021. Footfall down 4.2% last week despite central London seeing returns to office - Retail Gazette [WWW Document]. Retail Gaz. URL https://www.retailgazette.co.uk/blog/2021/09/footfall-down-4-2-last-week-despite-central-london-seeing-returns-to-office/ (accessed 9.15.21).

Wrigley, N., Dolega, L., 2011. Resilience, fragility, and adaptation: new evidence on the performance of UK high streets during global economic crisis and its policy implications. Environ. Plan. A 43, 2337–2363.

Wrigley, N., Lambiri, D., 2014a. Convenience culture and the evolving high street.

Wrigley, N., Lambiri, D., 2014b. High street performance and evolution: a brief guide to the evidence. University of Southampton.

Wrigley, N., Lambiri, D., Astbury, G., Dolega, L., Hart, C., Reeves, C., Thurstain-Goodwin, M., Wood, S., 2015. British high streets: From crisis to recovery? A comprehensive review of the evidence.

Wrigley, N., Lambiri, D., Cudworth, K., 2009. Linked trips and town centre viability. Town Ctry. Plan. 78, 433–438.

Yiu, C.Y., Ng, H.C., 2010. Buyers-to-shoppers ratio of shopping malls: A probit study in Hong Kong. J. Retail. Consum. Serv. 17, 349–354.

Yuill, C., 2009. The credit crunch and the high street:'coming like a ghost town.' Sociol. Res. Online 14, 97–102.

Yuo, T.-T., Crosby, N., Lizieri, C.M., McCann, P., 2003. The management of positive inter-store externalities in shopping centres: some empirical evidence.

Zhou, Q., Wang, S., Liu, Y., 2022. Exploring the accuracy and completeness patterns of global land-cover/land-use data in OpenStreetMap. Appl. Geogr. 145, 102742. https://doi.org/10.1016/j.apgeog.2022.102742

9 Abbreviations

BI — Business and Independent Micro-locations
CCR — Chain and Comparison Retail Micro-locations
CDRC — Consumer Data Research Centre
LDC — Local Data Company
RMSE — Root Mean Squared Error
VOCR — Value-Orientated Convenience Retail Micro-locations

Appendix

Appendix 3.1 – Datasets used

SmartStreetSensor footfall data (aggregated) (2015–2020)

Available from the CDRC at: https://data.cdrc.ac.uk/dataset/local-data-company-smartstreetsensor-footfall-data-–-research-aggregated-data

SmartStreetSensor footfall data (non-aggregated) (2015–2020)

Available from the CDRC at: https://data.cdrc.ac.uk/dataset/local-data-company-smartstreetsensor-footfall-data

Not used in this thesis

Retail Unit Address data (2015–2019)

Available from the CDRC at: https://data.cdrc.ac.uk/dataset/local-data-company-retail-type-vacancy-and-address-data

Retail Centre Boundaries and Typology (2017)

Openly available from the CDRC at: https://data.cdrc.ac.uk/dataset/historic-retail-centre-boundaries

Department for Communities and Local Government Retail Centres (2004)

Openly available from the CDRC at: https://data.cdrc.ac.uk/dataset/retail-centre-locations-dclg-version

Retail Centre Boundaries and Typology (Updated 2021)

Openly available from the CDRC at: https://data.cdrc.ac.uk/dataset/retail-centre-boundaries

Not used in this thesis

Workplace Zones and Daytime population (2011)

Openly available from UK Data Service at:

https://statistics.ukdataservice.ac.uk/dataset/economic-activity-daytimeworkday-population-england-northern-ireland-and-wales-2011

For England, Wales and Northern Ireland only

National Public Transport Access Nodes (NaPTAN) (2014)

Openly available from the Department for Transport at: https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan

UK Car Parks (2015)

Openly available from the Department for Transport at: https://data.gov.uk/dataset/7e7ef556-4173-4dcb-8eef-8ddde4e3824d/car-parks

OpenStreetMap (Accessed in 2018 for analysis in Chapter 4 & 2021 for analysis in Chapter 6)

Openly available to explore at: https://www.openstreetmap.org/

Data downloaded using OSMnx Python package. Information at: https://osmnx.readthedocs.io/en/stable/

Journal article: (Boeing, 2017)

Springboard Footfall Data

Available from Urban Big Data Centre at: https://www.ubdc.ac.uk/data-services/data-catalogue/commercial-and-retail-data/springboards-footfall-data/

Benchmarks available as subscription from: https://www.spring-board.info/benchmarking/

Appendix 3.2 – Metadata for LDC's Retail Unit Address Data

Variable (2017)	Variable (2018/2019)	Description
Unit	Unit	For address, e.g. Unit 1
Building	Building	For address, e.g. Manchester Arndale
StreetNo	StreetNo, StreetNo2, StreetLetter	For address. Expanded into different variables. The 2017 dataset had street numbers such as '1-3' changing to '1-Mar'. In 2018/9, StreetNo would be 1, and StreetNo2 would be 3.
Street	Street	For address (e.g. High Street, Market Street)
Town	Town	Town or city of unit
County	*NA*	County of the unit. Removed by 2018.
Postcode	Postcode	Postcode of the unit.
Region	*NA*	Region of the unit. Removed by 2018.
Latitude	Latitude	Latitude of the unit.
Longitude	Longitude	Longitude of the unit.
PremiseId	PremiseId	Unique identifier of building
OccupierId	OccupierId	Unique identifier for the occupier of the building
OccupierName	ShopName	Name of retailer or business, likely renamed as it is not linked to OccupierId.
MultipleID	*NA*	Unique identifier for chain retailers
MultipleName	*NA*	Name of chain retailer. Removed by 2018.
Classification	*NA*	Comparison, Convenience, Leisure, Service, Non-Retail or Misc. Removed by 2018.
NA	BusinessTypeId1	Unique identifier for a category.
Category	BusinessType1	Category of the store within the classification.
NA	BusinessTypeId2	Unique identifier for sub-category.
Subcategory	BusinessType2	Subcategory of a store within a category.
PremiseStatus	*NA*	'Live' or 'Vacant' for vacant properties. Information in 'BusinessType1' for 2018 onwards.
NA	Concession	If a retailer is a concession within another store '1' else '0'
NA	ShopWithinShop	If a shop is inside another shop '1' else '0'

NA	CareOfName	Name of the main retailer that oversees concession or the shop within shop.
NA	BusinessSicId	Unique identifier for store descriptions
NA	BusinessSicName	Description of goods sold

Appendix 3.3 – A non-exhaustive list of businesses that collect footfall data and the method they employ.

Company	**Method**
Local Data Company	Wi-Fi-based methods counting smartphones
Blix	
Proximity Futures	
Euclid Analytics	
Springboard	Camera (either AI or time-of-flight)
Brite yellow	
Ipsos Retail Performance	
Retail Sensing	
Footfall Cam	
Terabee	
RetailNext	
Prism Skylabs	
Axper	
BT	Mobile data
Hystreet.com	Laser scanner
Huq	GPS through software from mobile app partners
PFM Footfall Intelligence	Offer multiple sensing technologies
Sensormatic	Unclear on Website
Wireless Social	
Tamoco	
Parallax	

Appendix 4.1 – Archetypes of Footfall Context: Quantifying Temporal Variations in Retail Footfall in relation to Micro-Location Characterisitcs

Abstract

The UK retail sector is constantly changing and evolving. The increasing share of online sales and the development of out-of-town retail provision, in conjunction with the 2008-09 economic crisis, have disproportionately impacted high streets and physical retail negatively. Understanding and adapting to these changes is fundamental to the vitality, sustainability and prosperity of businesses, communities and the economy. However, there is a need for better information to support attempts to revitalise UK high streets and retail centres, and advances in sensor technology have made this possible. Footfall provides a commonly used heuristic of retail centre vitality and can be increasingly estimated in automated ways through sensing technology. However, footfall counts are influenced by a range of externalities such as aspects of retail centre function, morphology, connectivity and attractiveness. The key contribution of this paper is to demonstrate how footfall patterns are expressed within the varying context of different retail centre architypes providing both a useful tool for benchmarking and planning; but also making a theoretical contribution to the understanding of retail mobilities. This paper integrates a range of contextual data to develop a classification of footfall sensor locations; producing three representations of sensor micro-locations across Great Britain: *chain and comparison retail micro-locations, business and independent micro-locations* and *value-orientated convenience retail micro-locations*. These three groups display distinct daily and weekly footfall magnitudes and distributions, which are attributed to micro-locational differences in their morphology, connectivity and function.

Keywords

Retail, footfall, town centre micro-locations, cluster analysis

Introduction

The retail landscape in the UK is constantly evolving. In 2019, 19.2% of retail sales were made online; an increase of 13% over 10 years (ONS, 2020). The increasing share of online sales, in conjunction with the 2008-09 economic-crisis, the development of out-of-town shopping retail space provision and shifting consumer behaviour, are major drivers for retail industry change, and physical retail has suffered disproportionately as a result (Parker et al. 2016a; Portas 2011; Burt 2010; Wrigley et al. 2015). This period of retail upheaval has had significant consequences, especially for those businesses who have failed to adapt to changed consumer purchasing behaviour and online competition. Recent examples include Clintons and Forever 21 who, along with 41 other retail chains, went into administration in 2019 (Centre for Retail Research 2020). With what were once household names disappearing from the high street, concern has cultivated within media, public opinion and government on what this means for the future of the retail industry and the UK economy.

There is a consensus that data driven empirical evidence is needed to support high street performance and revitalisation strategies (Portas 2011; Wrigley & Dolega 2011). In particular, footfall, often cited as the '**lifeblood**' of a high street vitality and viability (Birkin et al., 2017), is a key measure for the successfulness of these strategies and a widely used proxy for their economic performance (Coca-Stefaniak 2013; Millington et al. 2018). Footfall can be defined as the count of people travelling through a shopping area at a given point in time (Lugomer et al. 2017). As a measure, footfall is responsive to both characteristics of the macro-scale environment, such as broader economic trends, catchment population or weather conditions (Dolega et al. 2016; Makkar 2020), and the micro-scale environment, referred to as the micro-location. Micro-location analysis recognises the influence of the immediate environment on footfall, for example, the mix of retailers along a street or walkability (Brown 1993) as such, larger retail centres can encompass multiple micro-locations. There is limited research detailing or quantifying the relationships between footfall and qualities of the micro-location, resulting in low understanding of the opportunities and pitfalls footfall data may present. This can have implications for decision makers, who may use footfall as a primary measure of high street vitality and viability, and for the understanding of retail mobilities as a whole. As such this paper uses quantitative data to investigate the relationship between patterns in footfall and the function, morphology and connectivity of retail micro-locations by fulfilling three key objectives: i) create a classification of the micro-locations based on the functional and morphological properties; ii) identify the key characteristics of these micro-location clusters and iii) examine how the temporal footfall patterns vary across different micro-location clusters.

This paper continues as follow. In Section 2, the importance of footfall as an indicator for retail centre vitality is discussed in addition to identifying retail centre qualities which determine footfall. Section 3 concerns the data collection, derivation and analytical approach used to cluster the 640 micro-locations across Great Britain into three representative clusters. These clusters are investigated in terms of their different attributes and their average footfall distributions in Section 4 and in Sections 5 and 6 the different processes behind these results and their implications are discussed.

Retail centre vitality and footfall

Retail centre vitality is a term used to reflect the liveliness of a retail centre and is measured by its busyness both across space and time (Parker et al. 2016b). There has been a wide range of normative studies into retail centre vitality, though, as a result of the negative impact of recent changes in the retail sector, there has been an emergence of more critical research (Parker et al. 2016a). Efforts by the government and private sector have aimed to understand the challenges which high streets are facing, and how they can adapt to succeed in the future (Portas 2011; Coca-Stefaniak 2013; Parker et al. 2016b; Grimsey 2018). There is a general consensus that sustainability and prosperity can be found through cooperation of stakeholders towards a clear and accountable vision (Portas 2011; Coca-Stefaniak 2013; Grimsey 2018). However, evidence suggests that there are limited examples of successful application of these practices (Parker et al. 2016a; Wrigley et al. 2015). To establish sustainable retail environments for the future, it is key to understand what impacts vitality (Coca-Stefaniak 2013; Parker et al. 2016a). Retail centres can be viewed as complex economic systems, and as such their vitality is driven by a number of internal and external factors, such as attractiveness, diversity and accessibility (Parker et al. 2016b).

There is also a plethora of research that investigates various measures of retail centre economic performance. A common measure is vacancy rate (Wrigley et al. 2015) and its derivatives such as vacancy rate change, structural vacancy and spatial clustering of vacant units. Retail offer and commercial rents are also commonly used for finer-scale performance insights (Wrigley et al. 2015). Another commonly used heuristic in academia, industry and in government for vitality and sustainability of a retail centre is footfall (Coca-Stefaniak 2013; Millington et al. 2018). Footfall was identified as the most influential factor for high street vitality and viability by Parker et al. (2016a) as a result of consulting 22 retail experts for their insights. Research suggests that this could be in part due to the positive correlation between footfall and potential spend (Graham 2017; Koster et al. 2019; Warnaby & Yip 2005), which in turn, can be linked to high return on investment for stakeholders, consequently attracting future investment and creating economically viable retail centres (Graham, Khan & Ilyas 2019).

Footfall is also a proxy for the vitality of a retail centre beyond consumer spend. It can be used to capture the attractiveness of a location as a community hub, workplace or other social and communal functions which a retail centre can provide to its consumers (Millington et al. 2015). A clear example of this is Edinburgh, a city ranked 3rd in the UK for footfall, however only 12th in terms of actual spend (Millington et al. 2015). This shows that there is a proportion of Edinburgh's footfall that does not translate into spend. The utility of footfall as a measure that encompasses many different influences and processes of the retail environment makes it a beneficial and useful indicator of retail vitality and viability.

Determinants of footfall

Footfall is determined by a multitude of factors on different spatial and temporal scales, visualised in Figure 1. Here, these determinants are summarised under three main headings: functional, morphological and other. The factors which influence footfall are interrelated, complex and can be difficult to quantify. This comprehensiveness can present a problem when trying to explain temporal and spatial variations in magnitude and signature. The magnitude of footfall can be defined as the amount of people measured in a certain set time period and the signature refers to the variation of footfall magnitude over time.

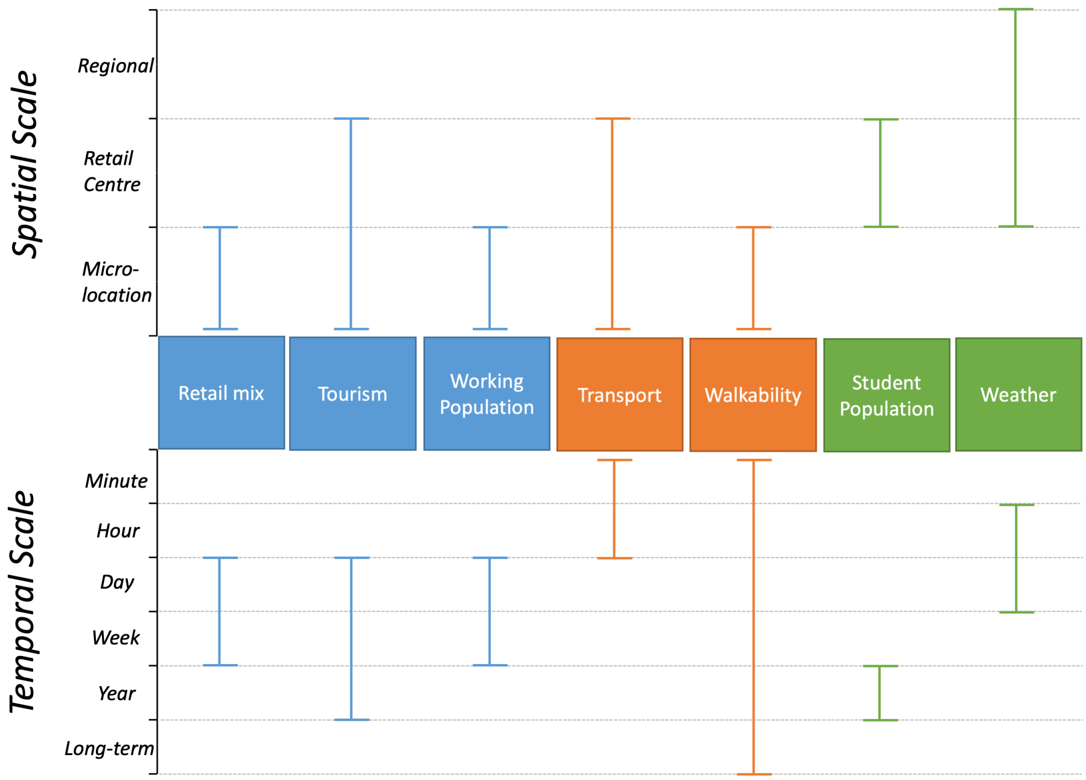


Fig. 1 Diagram summarising the spatial and temporal impacts of different footfall determinants as discussed in Section 2.1

Function

As shown in Figure 1, footfall is influenced by a multitude of factors on different temporal and spatial scales. Working population, retail mix and tourism all influence micro-location footfall to a daily, weekly or seasonal temporal scale and relate to the retail centre function. The function of a retail centre is the purpose which it serves to users and most retail centres are multi-functional, simultaneously performing several purposes (Millington et al. 2015). Characteristics such as the presence of anchor stores or the tendency towards premium or value goods can all indicate the retail centre identity, who it may appeal to, and consequently, when they may visit (Guy 1998).
The function of a retail centre or micro-location impacts footfall in several ways. Firstly, having a varied and cohesive retail mix has been shown to boost retail centre vitality and attractiveness (Millington et al. 2015; Tyler et al. 2012). The better the ability of the micro-location retail offer to match consumer demand of the consumers, the busier it can become, increasing the magnitude of footfall (Portas 2011; Parker et al. 2016a).
Secondly, research shows that the function of a retail centre is closely aligned to both diurnal and other periodic patterns of use. For example, retail centres in locations with a high concentration of employers and businesses typically have higher daytime footfall (Berry et al. 2016; Swinney & Sivaev 2013). Such relationships have been shown to drive footfall and sales during weekdays, especially in the early morning, at midday and in the early evening (Berry et al. 2016). On a seasonal scale, tourist destinations such as Cornwall can see grocery retail demand double during on-season (Newing et al. 2018) with tourists that are likely to spend more than local customers (Newing et al. 2014). In addition, event-based tourism can drive footfall on a more short-term basis. For example, the Giant Spectacular Liverpool's Dream event drew in 1.3 million people over 4 days in October 2018 (giantspectacular.com 2019).
Thirdly, studies which have investigated temporal change in footfall signature and magnitude have explained their results by primary retail centre function. Mumford et al. (2017) identified four distinctive annual footfall distributions for the UK, attributing their differences to four functions: comparison retail, holiday destinations, speciality retail and a multifunctional purpose. Similarly, in Lugomer & Longley (2018), footfall data was clustered based on the hour of the day, resulting in nine different patterns which were partly explained by different primary functions.

Walkability and morphology

Another factor which influences footfall is walkability, impacting micro-locational footfall over multiple temporal scales (see Figure 1). There are many contesting definitions of walkability however, in this case, walkability can be defined as the attractiveness of a street to a pedestrian. This can pertain to physical characteristics, security, network connectivity and transport connectivity (Lo 2009). Indeed, certain morphological properties of streets have been shown to increase their walkability, such as wide streets with gentle slopes that are well lit have been shown to be the most attractive (Erath et al. 2017; Unwin et al. 2017).

Additionally, how the street is situated within the wider network has proven to be a reliable indicator of pedestrian counts (Hillier et al. 1993; Raford & Ragland 2006). In particular, well-connected streets tend to have higher footfall as it is often the shortest route from their origin to their destination. This can be determined by various measures of centrality including closeness and betweenness, which respectfully capture the closeness of a node to other nodes and the prominence of a node as a bridge between other nodes (Freeman 1977; Porta et al. 2009).

As such, they can be used to predict busy intersections, or nodes. The added benefit of betweenness centrality, as opposed to closeness centrality, when investigating pedestrian flows is that it can be calculated for the edges, or streets, as well as the nodes.

Streets can also have high walkability if they are close to access points for other forms of transport, such as train stations, car parks or bus stops (Mazumdar 2019). As popular origins and destinations, these features can concentrate footfall to particular micro-locations (Scheurer & Porta 2006). Anchor stores, restaurants and entertainment venues have demonstrated footfall attraction in a similar fashion (Hart et al. 2014; Koster et al. 2019; Teller and Alexander 2014; Üsküplü 2020; Yuo et al. 2003). The proximity of stores to major transport hubs has been shown to increase their footfall and sales, particularly at commuting times (Berry et al. 2016). Having good access to car parking is a demand of retail areas and many consumers will avoid using public transport in favour of the convenience of their own vehicle. Therefore, the proximity of a retail area to a public car park can influence the quantity of visitors and impact footfall for the entire retail centre (ATCM 2014).

Additional factors

In addition to walkability and function, there are numerous other factors which have been proposed to influence the magnitude and distribution of retail centre footfall, for example weather, with rain and snow drastically reducing daily pedestrian counts (Makkar 2020).

Although extreme weather is typically a dynamic and short-term influence, it can have significant consequences, particularly if it coincides with planned periods of high expected retail, such as the Christmas season.

Academic literature points to many functional and morphological influences on footfall, however, to our knowledge, no literature exists which quantifies the impact of a combination of these influences. Therefore, a data driven exploration of footfall spatial and temporal patterns will add quantifiable evidence to the existing evidence base in this research area, in particular to observed relationships between footfall and the characteristics of the surrounding micro-location.

Methodology

Footfall data

Footfall data were provided by the ESRC Consumer Data Research Centre / Local Data Company [LDC] whose sensors use probes from Wi-Fi enabled devices to estimate the number of smart phone devices passing by as a proxy for footfall. The device sends an individual MAC address to the sensor, which is anonymized and used to determine which kind of device the signal came from. Devices which are not smart phones are filtered out, as are duplicate counts from residents or staff nearby by filtering out MAC addresses that appear in several chronological time periods. The counts are aggregated to 5-minute intervals.

The approach relies on some assumptions which may limit its accuracy (Lugomer et al. 2017; Soundararaj et al. 2020). Firstly, a count of smart phones is not a perfect count of people as not everyone owns one or has one with them as they travel around a retail centre. Secondly, as the battery of a phone gets lower, it does not send out the wi-fi probes as far or as often, making it less likely to be picked up by the sensor than if it was on full battery. Thirdly, if a pedestrian has their Wi-Fi switched off, depending on the model of the device, the sensor may not register them. Fourthly, due to increased phone security implemented within newer phone operating systems, MAC addresses are scrambled a lot more frequently, making it harder to filter out repeat counts. Furthermore, there can be practical issues which cause measurement inaccuracies such as power cuts, sensors being mistakenly switched off or differences in positioning and orientation of the sensor.

A number of measures have been taken to overcome these problems. Before any analysis was run on the footfall counts the measurements were compared to manual counts. These manual counts take place for every location at a range of times throughout the day, month and year to ensure that the footfall counts are adjusted as reliably as possible (Soundararaj et al. 2020).

As of August 2018, LDC had sensors in 840 locations in 88 towns and cities across the UK (LDC, 2018). Due to data availability restraints, the study used 640 sensors from 40 high street retail locations in Great Britain. The distribution of sensors is particularly biased towards London (n=291), with 45% of the sensors, as well as larger cities such as Manchester (n=18), Liverpool (n=16) and Cardiff (n=8). Excluding London, the number of sensors per location ranges from n=20 in Kingston-upon-Thames to n=1 in Gateshead and in Windsor. Although the majority of sensors in the sample are in larger cities, some regional centres and market towns are also represented, such as Taunton (n=6) and Market Harborough (n=13). The full geographical distribution of the sample can be found in Appendix A.

Derivation of footfall descriptors

Drawing on previous work identified from the literature review, we can draw a series of broad micro-locational influences on footfall that are related to: 'functionality' and 'morphology and connectivity'. Within each category, there are a range of potential variables that can be assembled to differentiate between the footfall sensor micro-locations. By understanding the differences in footfall descriptors between the footfall sensor locations, elements of their footfall magnitude and signature can be better inferred. The descriptors used are not an exhaustive list of footfall influencers, therefore this analysis relies on the assumption that the impact of other influencers is negatable.

A summary of the variables within their category and their specification are shown in Table 1. The Functionality category captures aspects of context that may attract people to a retail area. The purpose for patronage of a retail area is logically linked to a temporal factor, for example, food outlets will attract more people during mealtimes and an area rich with bars and restaurants, would attract people in the evenings aligned to opening hours.

The morphology and connectivity category encompasses features of walkability and attractiveness such as transport accessibility, density of units and the centrality of the street within the retail centre network.

For several descriptors, a 100m circular buffer[26] around the sensor was used to select the stores close enough to be considered within the immediate retail environment of the sensor. 100m was chosen as it encompasses a reasonable sample of stores to derive a full picture of the retail environment but is not so large as to remove the micro-locational variation of interest. This relies on the assumption that there is a dense concentration of retail units around the store the sensor is based in, and that the circular shape can appropriately capture this. Sensors with fewer than 5 units within the buffer area (total of 5 sensors) were removed from the sample as there are not enough stores to get a representative understanding of the proportions within the retail environment. The resulting number of stores in the buffer ranged from 7 to 189, which was used to define the density of stores variable. This was combined with the number of features such as independent and value stores to calculate proportions to represent these characteristics. Also, a proportion of vacant units was calculated within each buffer to obtain vacancy rate for each micro-location.

---

[26] From a methodological standpoint, a walking network distance would be more appropriate for this analysis than a Euclidean distance. However, due to the COVID-19 pandemic limited access to on-campus resources, the computing capabilities needed to use this measure were unavailable. Preliminary data exploration and the relatively short distances would indicate that using network distance would have negligible overall impact.

A Euclidean distance, as opposed to a proportion, was calculated for some features, such as anchor stores and premium stores, as they appear in most retail centres, though not in multitude. When a proportion was calculated for these features, they returned measures with more constrained variation. As such, distance was deemed to be a more appropriate measure. Table 1 below provides a summary of the variables, their specification. The correlation coefficients between these variables are shown in Appendix B.

Table 1: Key features of the functionality and morphology and connectivity variables used as micro-location footfall descriptors

Category	Variable	Specification
Functionality	Distance to the nearest anchor store	Euclidean distance (metres) to nearest anchor store, identified by their brand name (e.g. John Lewis, Primark, Debenhams, full list in Appendix C)
	Distance to the nearest premium store	Euclidean distance (metres) to the nearest premium store, identified by their brand names (e.g. The White Company, Burberry, full list in Appendix C)
	Distance to the nearest entertainment activity	Euclidean distance (metres) to the nearest venue which offers an entertainment activity (e.g. Cinemas, Arcades, Museums). These were identified using the LDC's (2017) survey sub-categorisation (full specification in Appendix C)
	Proportion of vacant stores (vacancy rate)	The proportion of vacant store identified using the LDC's (2017) survey within a 100m straight line buffer of the sensor
	Proportion of value stores	The proportion of stores identified as value stores by their brand name (e.g. Aldi, Home Bargains, full list in Appendix C) within a 100m straight line buffer of the sensor
	Proportion of independent stores	The proportion of stores identified as independent by the singular instance of their store name in the dataset within a 100m straight line buffer of the sensor

	Proportion of night-time economy locations	The proportion of locations within a 100m straight line buffer of the sensors which offer a typical evening appeal (e.g. bars, clubs, restaurants, fast food) identified using LDC's (2017) survey categorisation (full specification in Appendix C)
	Workplace population	The average of the daytime population densities of the workplace zone in which the sensor falls into, and those which border it (ONS, 2017).
	Ratio of service to retail	The ratio of the locations within a 100m straight line buffer of the sensor which are identified as service locations by LDC's (2017) survey classifications to those identified as comparison retail and food retail (e.g. grocery stores, butchers, confectioners, further specifics in Appendix A)
Morphology and Connectivity	Distance to the nearest transport hub	Euclidean distance (metres) to the nearest group of bus stops or train station as identified in the NaPTAN dataset (Department for Transport, 2014).
	Distance to the nearest car park	Euclidean distance (metres) to the nearest car park as identified by the Department for Transport (2015)
	Density of stores	The number of store units within a 100m straight line buffer of the sensor
	Centrality of the street	The street centrality measure was calculated from networks generated by the OSMnx python library. OSMnx uses data from Open Street Map to generate a network graph of a road structure within a boundary. The CDRC retail centre boundaries (Pavlis, Dolega & Singleton, 2017) were used to generate the pedestrian network around a sensor. The edge betweenness centrality of the street

		which the sensor was on was is then calculated to give the street centrality measure. Edge betweenness was chosen as the centrality measure because it can be applied to streets instead of intersections, where most of the footfall measurements are taken from. This captures the prominence of a street as a pass-through route.

Analytical Approach

Understanding how the footfall descriptors derived in Section 3.2 relate to the footfall magnitude and signature for their sensors is a complex and multi-dimensional task. For each of the 640 sensors, there are 13 functional and morphological descriptors which could impact their footfall magnitude and signature at different times of day and days of week. Although this density of data would be beneficial for a case study analysis, it is too noisy and condensed for this investigation. Therefore, a methodology was derived to reduce the dimensionality of the data so that it represented the key trends for the footfall descriptors.

K-means clustering is an unsupervised algorithm that groups unlabelled data into similar clusters based on their features. It was chosen for this study as it summarises the data so that the main variations in footfall descriptors are still maintained yet reduces the dimensionality so that it is more manageable for comparison with footfall data. Other potential methods, such as creating an aggregate measure, could result in the loss of information from the different footfall descriptors which could be key for explaining a footfall trend. In addition, *K*-means clustering is a commonly used and understood methodology in many fields including geodemographic analysis (Burns et al. 2018; Spielman & Singleton 2015).

The algorithm attempts to minimize the sum of squared Euclidean distance between randomly generated cluster centres and nearby data points (Lloyd 1982). When the sum of squared distance cannot be minimized and the cluster centres are stationary, the algorithm has converged on a solution. The best solution for a *k*-means clustering is one which generates well-separated and compact clusters which are interpretable within the context of the data.

In order to run the *k*-means algorithm, the features were standardised according to their mean and standard deviation. As *k*-means optimises the sum of squared distance, outliers can have a large impact on the results. Some locations were classed as outliers because they had unusually large or small values for some variables. For example, three sensors in Lymington were removed as they were over 18km from the nearest entertainment activity. A further five sensors were removed iteratively throughout the clustering process, as they were the furthest point from any cluster centre. The resulting clusters were as compact and well-separated as possible without removing more outliers than necessary.

The features were then checked against each other to ensure there are no high correlations to avoid multicollinearity (see Appendix B).

The clustering algorithm was run using $k = 3$. There was no prior indication from the data to suggest a value of k therefore a comparison of average silhouette score was used. A silhouette score is a measure of how well a certain point fits within the cluster it has been assigned. It ranges from +1 which represents a point which fits perfectly in the generated cluster, to -1 which represents a point which poorly fits into the current cluster and would fit better in another. The average silhouette score is defined as the mean silhouette score for every point in the clustering. The average silhouette score for different values of k, as shown in Figure 2, were used to determine that $k= 3$ provides the best separation and cluster results.

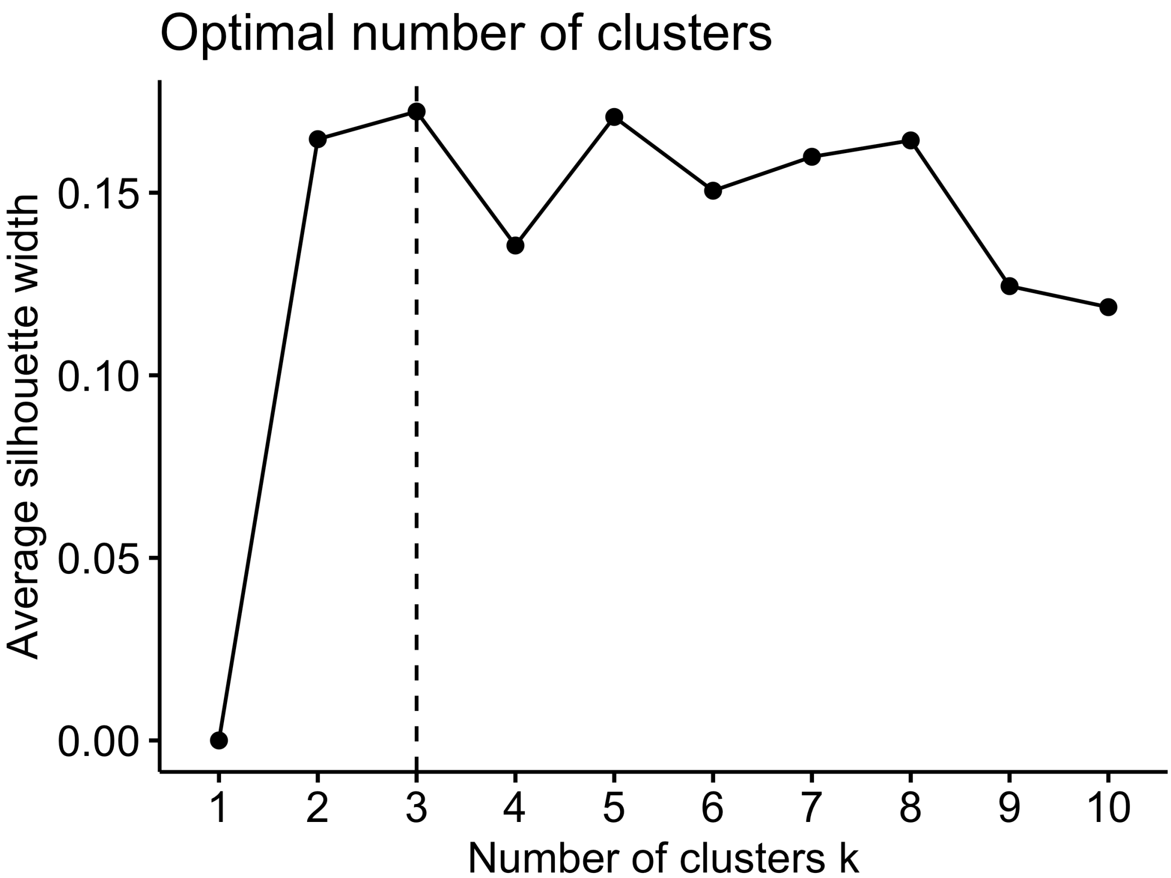


Fig. 2 The change in average silhouette score for different values of k in k-means clustering algorithm

One of the pitfalls of using this method is that it is a stochastic process. Therefore, if certain cluster centres were generated in an unfavourable position then it could lead to a poor result. To avoid this issue, the clustering was optimised using 10,000 runs with different randomly generated starting centres to find the best clustering outcome.
The average silhouette score for the final clustering was *0.17*. Although this is quite low, this is a result of the ambiguous nature of boundaries between retail areas. It is rare to find a street or micro-location which only serves one purpose and there is often qualities or retailers in a location which cater to a different function than others. In addition, even if there are streets which serve similar purposes, it is unlikely that they will also have the same structural qualities. Therefore, it is understandable that the clusters have a degree of overlap between them. There are methods which tailor to this quality in datasets, notably fuzzy c-means clustering, however they do not produce the clear-cut labels which will be useful when comparing the clusters to their average footfall signature.

Results

Cluster derivations

Cluster profiles often referred to as 'Pen Portraits' were then obtained based on values of the cluster centres and exploratory research into individual locations (see Appendix D). The values for the cluster centres and the within sum of squares can be found in Appendix D. The three clusters derived in our analysis were titled *chain and comparison retail micro-locations, business and independent micro-locations* and *value-orientated convenience retail micro-locations*.

Chain and comparison retail micro-locations [CCR]

Number of sensors: 343 (54%)
The CCR cluster was the most common of the three clusters and almost every city or town in the sample had a sensor in this cluster. They are named after their predominantly comparison retail function and their dominance towards chain retailers. From the clustering features, these micro-locations had a low proportion of independent retailers, were close to anchor stores and premium retailers and had a bias towards retail outlets over services. As such, destination shopping locations fit well into this cluster, for example, Oxford Street in London, Liverpool ONE in Liverpool and Queen Street in Cardiff. These locations are designed for comparison goods shopping, with a range of chain stores catering to create a large retail offer. These are sought after locations for retailers, often in the retail core of major cities.

Business and independent micro-locations [BI]

Number of sensors: 254 (40%)

The BI cluster encompasses places with a tendency towards independent retail, often in financial and office-dominated districts. 212 (83%) of the sensors in this cluster are sensors in London, representing 70% of the total sensors in London. This cluster captures the employment areas and the destination for many commuters. These areas are common in larger cities, where people do not tend to live near where they work, explaining why this cluster is predominant in London. In terms of the clustering features, BI micro-locations have a high working population, are close to transport hubs and have a high proportion of independent retailers. Some examples of these places are Holborn and the City of London, in London and NOMA and Spinningfields in Manchester. This cluster also includes places which also have a high proportion of night-time economy outlets such as Park Street in Bristol, Soho in London and Bold Street in Liverpool. A significant distinction of locations in this cluster is that they have 9% more restaurants than the average British high street, subsequently reflected in a near 1:1 ratio between service and retail outlets. This shows that this cluster has a more experience-based function than a comparison retail-based one. This is supported by their large distance from anchor stores, and their small proportion of value retailers.

Value-orientated convenience retail micro-locations [VOCR]

Number of sensors: 43 (7%)

The VOCR micro-locations cluster describes smaller, secondary centres of a larger urban area. These are more residential areas with a high prevalence of budget convenience retailers and betting and charity shops. They are defined by their higher proportion of value outlets, their larger distance from premium stores and entertainment venues and their low workplace population. These areas are the opposite of destination shopping areas; people visit these areas out of convenience. They exist due to their accessible location near to residential areas so that consumers can gather their essentials without making a longer trip. VOCR micro-locations have few entertainment venues and night-time economy outlets, as these are things which people are willing to travel for. Some examples of locations which fit into this cluster are Penge, Wood Green and Kilburn in London, Orpington, Shirley in Southampton, and Blatchington Road in Brighton. VOCR micro-locations also have the most vacant units, suggesting that they struggle to find retailers to fill stores. Another feature of this cluster is a distinctly higher proportion of charity shops. 5.9% of the nearest 25 stores to each sensor in this cluster were charity shops, compared to 1.8% in the CCR cluster and 0.6% in the BI cluster and 4.3% greater than the average for England and Wales of 1.6%.

Cluster footfall signature and magnitude

Footfall measurements are often used as a proxy for retail centre vitality (Coca-Stefaniak, 2013; Millington et al. 2018), however there is limited research quantifying how functional and morphological factors impact footfall magnitude and signature. By investigating the footfall patterns exhibited by these clusters built on functional and morphological characteristics, a greater understanding of variations in footfall magnitude and signature can be achieved. Footfall measurements from January 2017 until August 2018 were averaged across the locations in each cluster to investigate whether the different functions and characteristics of the micro-location impact footfall. Only the sensors with footfall data for 75% of a full year were used to remove any bias from new or temporary sensors which only have footfall data for potentially busier or quieter times of the year. This removed 12 sensors from the sample. Descriptive statistics were calculated for the average week (by hour), and average weekday (by 5 minutes) for each cluster as shown in Table 2 and Figures 3 and 4.

Table 2: Summary statistics for footfall (people per 5 minutes) across the clusters

Statistic		**CCR micro-locations**	**BI micro-locations**	**VOCR micro-locations**
Maximum:	Mon	94 @ 12:05	106 @ 17:10	55 @ 16:15
	Tues	95 @ 12:05	117 @ 17:10	61 @ 16:20
	Wed	96 @ 12:05	**121 @ 17:10**	61 @ 17:10
	Thurs	95 @ 12:05	119 @ 17:10	**62 @ 16:20**
	Fri	98 @ 12:05	113 @ 17:10	57 @ 16:20
	Sat	**116 @ 13:05**	92 @ 13:05	60 @ 13:25
	Sun	86 @ 13:05	71 @ 14:05	47 @12:05
Weekly Mean		37	49	27
Standard Deviation		32	31	19

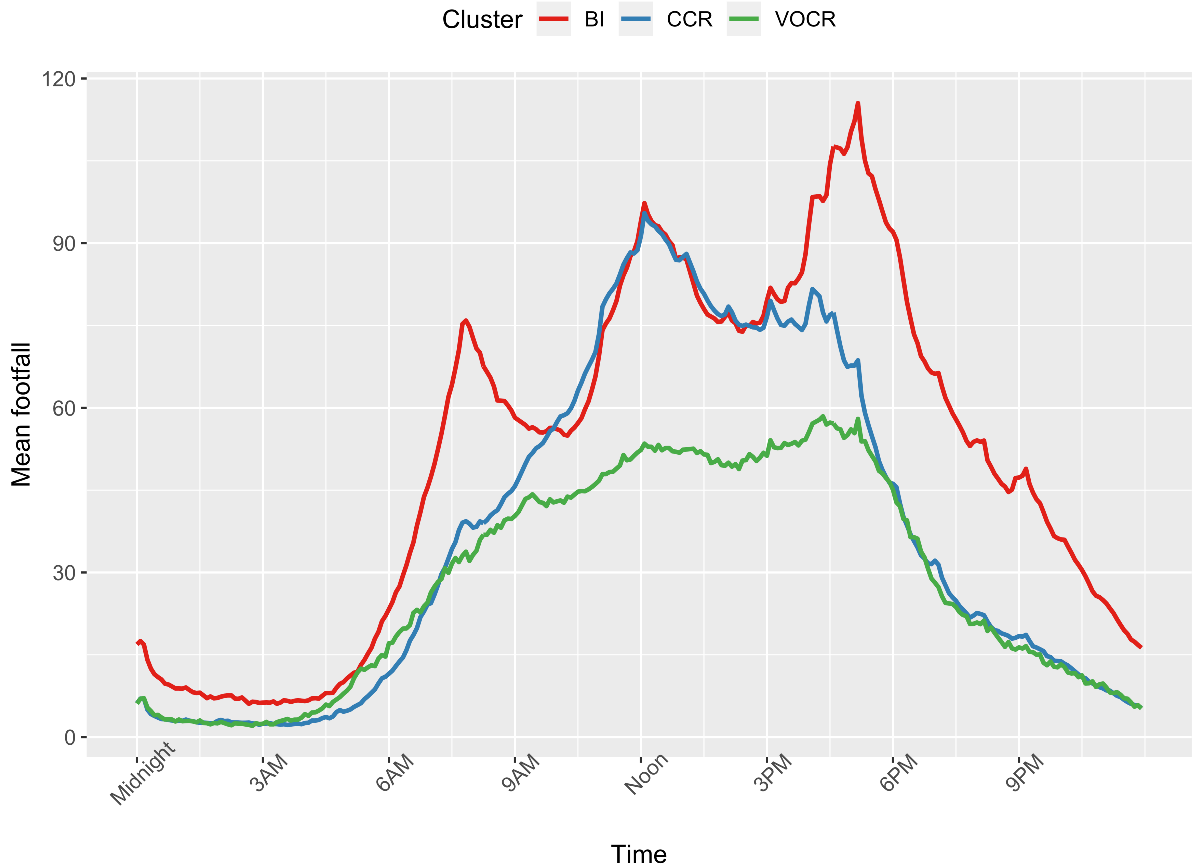


Fig. 3 Average footfall distribution for each cluster for a weekday (Monday to Friday) to 5-minute accuracy

Figure 3 shows that early in the morning on weekdays, the BI micro-locations have higher footfall than CCR micro-locations. Although by 10:00, the CCR micro-locations are just as busy, and both rise in footfall until 12:05. This maximum weekday peak is consistent at 94-101 people per 5 minutes for CCR and BI micro-locations. Footfall in CCR micro-locations then decreases into the afternoon and evening, whereas footfall in BI micro-locations experiences a 14:00 lull before peaking again into the early evening. This is reflected through the consistent 17:10 maximum footfall values for BI micro-locations of 106-121 people per 5 minutes, shown in Table 2. During the evening, this cluster is the busiest, keeping over 25 people per 5 minutes until past 22:00 and never dropping below 5 people per 5 minutes. BI micro-locations have a distinctive weekday footfall pattern consisting of three peaks at 8:00, 12:00 and 17:00.

The VOCR micro-locations have the lowest average footfall of all the clusters, and they are never the busiest. Their maximum value is 62 people per 5 minutes, which is just over half the size of the maximum values for the other clusters. The footfall signature of VOCR micro-locations is hump shaped, slowly increasing from 5:00 to 16:15 – 17:10, where it peaks on weekdays. After then, footfall decreases exponentially to under 10 people per 5 minutes by 22:30.

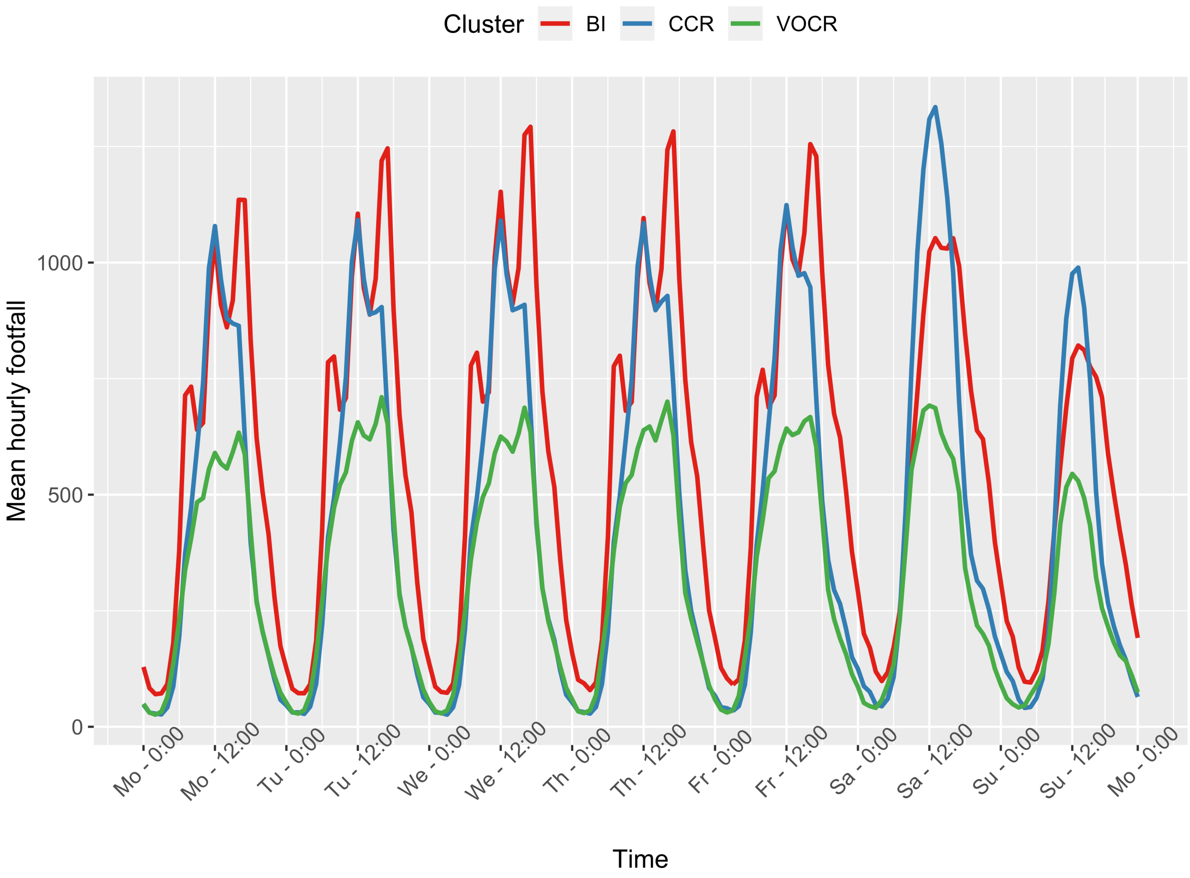


Fig. 4 Average footfall across a week for each cluster to hourly accuracy

As visible in Figure 4, CCR micro-locations are significantly busier on Saturdays compared to the weekdays, with their maximum footfall of 116 people per 5 minutes at 13:05 that day. Although CCR micro-locations have the highest peak, BI micro-locations have the highest consistency, with a mean footfall of 49 people per 5 minutes, compared to 37 people per 5 minutes. However, VOCR micro-locations have the lowest standard deviation, showing that, although their average footfall is low, it is the most consistent throughout the day and throughout the week.

VOCR micro-locations have very similar footfall signatures during the weekend as the weekday, in contrast BI micro-locations have very different footfall signatures. They have lower footfall at weekends, peaking at 92 people per 5 minutes at 13:05 and do not exhibit the three peak structure previously observed, instead showing a peak at early afternoon with a slow drop into evening when they the only cluster to retain significant footfall into the night. Friday and Saturday nights appear to be the busiest nights, staying at above 25 people per 5 minutes until after 00:00. In contrast, the other clusters have dropped below this threshold by 21:00. Sunday is the quietest day for every cluster even the most consistent VOCR micro-locations, exhibit a smaller peak on this day.

Discussion

This study has produced three distinct clusters of retail micro-locations which vary in terms of their function and morphology: chain and comparison retail micro-locations [CCR], business and independent micro-locations [BI], and value-orientated convenience retail micro-locations [VOCR]. When the average weekly and daily footfall patterns of these clusters were investigated, distinct patterns in signature and magnitude were evident. These differences in footfall signature and magnitude can be partially explained by various characteristics of the retail micro-location, essentially their form and function.

Firstly, the CCR micro-locations exhibited a footfall pattern with the busiest times on Saturdays, and during daytime hours from late morning to early afternoon. This reflects this cluster's prominent comparison retail function indicated by its low service to retail ratio and the low proportion of independent stores in the clustering. For the majority of people, Saturday is a day of leisure when they have ample free time. Comparison retail tends to be recreational and time consumptive (Guy 1998), therefore supporting the link between this function and significant Saturday and daytime footfall. In addition, this cluster has the highest average density of retail units showing that the retail offer is more condensed in these micro-locations, therefore, increasing the overall footfall magnitude. Besides, a condensed retail offer has the capacity to encourage linked trips, where consumers visit different locations in the same trip (Wrigley et al. 2009).

In comparison, the BI micro-locations have weekday dominant footfall with three peaks at 8:00, 12:00 and 17:00. This footfall pattern reflects commuting into and out of work, with an additional increase in footfall during a lunch time break, is similar to that observed in other studies (Berry et al. 2016; Lugomer and Longley 2018). This is further supported by the large workplace population of the cluster and close proximity to transport hubs with many of the sensors located in central London - a destination for many public transport commuters (Lyons & Chatterjee 2008). The absence of this pattern during the weekend confirms this interpretation and shows the extent to which working population determines footfall in these locations. Furthermore, BI micro-locations retain footfall later into the evening than the other clusters. With a higher than average number of restaurants and bars, these micro locations could be also viewed as attractive leisure and night-time economy destinations (Ravenscroft et al. 2000). However, the amount of footfall in the late evening is significantly less than during the day, demonstrating that, on average, this night-time economy function is supplementary to the workplace function.

The VOCR micro-locations are the quietest and steadiest in terms of footfall. This constant and consistent flow of people could be explained by their convenience-based function as convenience retail is characterised by short and frequent trips (Guy 1998). The VOCR micro-locations tend to be in residential areas which serve a local demand with a smaller catchment size, therefore generating less footfall. The smaller magnitude of footfall of these micro-locations could be also associated with larger distance to many footfall attractors such as anchor stores, transport hubs and entertainment activities.

However, not all of these footfall patterns can be explained by features of the micro-location. For example, in every cluster Sundays saw 26-32% less footfall compared to the other days of the week, which can be explained by the reduced to 6 hours opening hours on this day for stores larger than 280 square metres, imposed by the 1994 Sunday Trading Act (Gov.UK 2019). Research shows that these large stores can be key footfall attractors and having these stores reduce their opening hours may deter people from visiting their high street on Sundays (Williamson et al. 2006).

These results help to build a clear understanding of how and why footfall fluctuates throughout the day and week and better understand its relationship with micro-location characteristics. In general, these results show that footfall and, as an extension of that, retail vitality, vary temporally and spatially on a micro-locational scale as a result of multiple external and internal influences. More specifically, this study shows some key drivers of footfall at a micro-location level: anchor stores, workplace population, density of retail units and distance to transport hubs. However, it would be incorrect to assume that all retailers within a particular retail centre benefit equally from the increased footfall in terms of spend, as that depends on many other factors on a micro-location level (Millington et al. 2015). This supports strategies to increase high street vitality which are holistic and consider this complexity of micro-locational factors within the wider retail centre. Footfall is often used as an indicator of high street vitality therefore a better understanding of it, underpinned by reliable data and robust empirical analysis is vital for business, academia and policy makers.

Implications

The results of this study pertaining to variation in footfall magnitude, signature and the function and form of the particular retail micro-location have a number of implications for various stakeholders. Firstly, it supports revitalisation and town centre strategies which consider the complexity of micro-locational influences within a retail centre, as this study has shown the importance of these factors in determining footfall and retail centre vitality. This is particularly relevant as footfall is widely used as a measure for retail centre performance, therefore having a clearer understanding of how and why it fluctuates would be beneficial. Understanding these factors can be valuable for retailers and planners in managing pedestrian flows, setting effective opening hours and investing in ideas which would be attractive to their target consumer. For example, BI micro-locations have a significantly bigger daytime footfall than evening footfall, despite its night-time economy. This knowledge could be used to develop schemes to increase the dwell time of daytime population and encourage them to support the night-time economy establishments, increasing the retail resilience of the area.

Secondly, these results have demonstrated the potential of using morphological and functional characteristics to predict footfall for areas where there are not sensors. Although these clusters are generalisations of micro-locations, they begin to draw out patterns between certain characteristics and spatial and temporal footfall variations. With technological advancements increasing the wealth of data on urban characteristics and mobilities and the development of algorithms capable of processing this data, there is potential for these kinds of patterns to be used to predict footfall for all retail areas. This would be a useful tool for benchmarking and location planning, managing pedestrian flows and business logistics such as opening hours and staffing.

Thirdly, this study has contributed to a more comprehensive understanding of retail mobilities. Although many footfall determinants have been identified in literature, how they impact footfall temporally is not always investigated or quantitatively shown. This paper has demonstrated how different micro-locational characteristics impact footfall to 5-minute intervals throughout an average week which provides new insight into footfall determinants and urban mobility as a whole.

Limitations

There are some limitations which have to be considered when examining and applying the results of this study, in addition to the data limitations discussed previously. Firstly, the sample size of 640 micro-locations for Great Britain is relatively small, with a bias towards London and the south of England. 52% of sensors are in the Greater London region, which has been shown to exhibit unique footfall patterns when compared to the nation as a whole (Mumford et al. 2017). Further, there are disproportionally fewer sensors in mid-sized centres and smaller centres, particularly in the north of England and Wales. Mid-sized retail centres and northern retail centres have been identified as the worst affected by unfavourable changes in the retail sector (Millington et al. 2015; Wrigley & Dolega 2011). In addition, the sensors are predominantly located in city centre environments, as opposed to suburban high streets or district centres, which face their own unique challenges to their future retail vitality and viability (Griffiths et al. 2008). As such the data sample is skewed towards micro-locations in larger urban areas that tend to be more successful and sustainable retail destinations, potentially with lower vacancy rates and steady footfall.

Secondly, although this study has grouped each of the micro-locations into three clusters, they may not be as clearly delineated in reality. Cluster analysis is a well-established and widely used form of analysis, however its outputs are a representation determined by decisions made by the researcher, which, if made differently would produce alternate and yet still valid results (Vickers & Rees 2007). This inherent quality of clustering techniques means that these micro-locations are more complex than the cluster descriptions. This is evident through the variation of footfall signatures within each cluster. The distributions shown in Figures 3 and 4 are the averages for all the sensors within that cluster and they may not reflect all micro-locations in that cluster. Some of the sensors may have somewhat different footfall magnitudes and signatures compared to the average in their cluster, despite overall similarity of a particular cluster functional and morphological characteristics.

Finally, due to the aforementioned bias in the availability of footfall data, it is likely that there are other identifiable micro-locations clusters in the wider country which have not been represented by this study. For instance, in Mumford et al. (2020) four types of town were identified based on their monthly footfall patterns: comparison, holiday, speciality and multi-functional. It is apparent that our sample is biased towards Mumford et al.'s comparison centres overlooking the different micro-locational patterns that could exist in the remaining clusters, such as seasonal popularity, tourism and non-retail anchors (Mumford et al. 2020; Newing et al. 2018).

Conclusion

In conclusion, this study has provided a novel application of sensor data to better understand retail behaviours and footfall. It has shown that patterns in the magnitude and signature of footfall data, and by extension retail vitality, can be to an extent, explained by functional and morphological characteristics of the micro-location. In particular, the ability of key footfall attractors such as anchor stores and transport hubs to significantly drive footfall at certain times throughout the day and week. This paper has also demonstrated the importance of the type of retail offer, comparison, convenience or recreational, on the magnitude and signature of footfall within the micro-location. The results display three clear narratives of micro-location morphology, function and footfall distribution, which aid greater understanding of the interrelationship and patterns that exist between them. Although the value added by this study is clear, it needs to be highlighted that the identified clusters are merely a representation of the more complex real world and any application of these narratives to a unique micro-location should consider the different functions which that place represents (Millington et al. 2015). Finally, future research will benefit from employing more footfall data to facilitate investigation into monthly, annual and longer-term trends in footfall and how those could relate to functional and morphological characteristics. In this study we present the potential for functional and morphological characteristics of micro-locations as a predictor for footfall in locations where footfall is not measured. Being able to model footfall for an entire retail centre could be invaluable for decision-making, urban planning and for retail location planning.

References

Association of Town & City Management [ATCM], (2014). In-Town Parking: What works? Innovative practices in parking provision. Online Report. ATCM. https://thegreatbritishhighstreet.co.uk/pdf/GBHS-What-Works.pdf. Accessed November 2019.

Berry, T., Newing, A., Davies, D. and Branch, K., (2016). Using workplace population statistics to understand retail store performance. *The International Review of Retail, Distribution and Consumer Research,* 26(4), 375-395.

Birkin, M., Clarke, G. and Clarke, M., (2017). Retail Location Planning in an Era of Multi-Channel Growth. Routledge.

Brown, S., (1993). Micro-scale Retail Location: Cinderella or Ugly Sister? *International Journal of Retail & Distribution Management,* 21(7) DOI https://doi.org/10.1108/09590559310046013

Burns, L., See, L., Heppenstall, A. and Birkin, M., (2018). Developing an Individual-level Geodemographic Classification. *Applied Spatial Analysis and Policy*, 11(3), 417-437.

Burt, S. (2010). Retailing in Europe: 20 years on. *The International Review of Retail, Distribution and Consumer Research*, 20(1), 9-27.

Centre for Retail Research, (2020). Who's Gone Bust in UK Retailing 2019. Online Report. Centre for Retail Research. https://www.retailresearch.org/whos-gone-bust-retail.html#bycompany. Accessed March 2020.

Coca-Stefaniak, A., (2013). Successful Town Centres - Developing effective strategies. Online Report. ATCM. http://thegreatbritishhighstreet.co.uk/pdf/Successful-Town-Centres.pdf?2. Accessed November 2019.

Department for Transport, 2014. *NaPTAN: National Public Transport Access Nodes*, Electronic Dataset, Data.Gov.UK, Accessed: February 2019

Department for Transport, 2015. *Car Parks*, Electronic Dataset, Data.Gov.UK, Accessed: February 2019

Department of the Environment, (1996). *Planning Policy Guidance No .6: Town Centres and Retail Developments*, Revised PPG6. London. https://www.nrpf.org.uk/PDF/PPG%206%20-%20June%201996%20.pdf. Accessed January 2020.

Dolega, L., Pavlis, M. and Singleton, A., (2016). Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. *Journal of Retailing and Consumer Services,* 28, 78-90.

Erath, A., van Eggermond, M.A., Ordóñez, S.A. and Axhausen, K.W., (2017). Introducing the pedestrian accessibility tool: walkability analysis for a geographic information system. *Transportation Research Record: Journal of the Transportation Research Board,* 2661(1), 51-61.

Freeman, L.C., (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35-41.

Giantspectacular.com. 2019. 'Giant Spectacular Liverpool's Dream'. Online News. https://www.giantspectacular.com. Accessed December 2019

Gov.uk, 2019. 'Trading hours for retailers: the law'. Online Guidance. Gov.UK. https://www.gov.uk/trading-hours-for-retailers-the-law. Accessed December 2019.

Graham, C, 2017. 'Footfall, attraction and conversion; a retail empirical generalisation.' Conference Report. *Academy of Marketing. University of Hull 03 - 06 Jul 2017.* London South Bank University. https://openresearch.lsbu.ac.uk/item/86yw8. Accessed December 2019.

Graham, C., Khan, K. & Ilyas, M. (2019). Estimating the value of passing trade from pedestrian density, *Journal of Retailing and Consumer Services,* 46, 103-111.

Grimsey, B. (2018). The Grimsey Review 2. Online Report. The Grimsey Review. http://www.vanishinghighstreet.com/wp-content/uploads/2018/07/GrimseyReview2.pdf. Accessed November 2019.

Griffiths, S., Vaughan, L., Haklay, M. and Emma Jones, C., 2008. The sustainable suburban high street: a review of themes and approaches. *Geography Compass*, *2*(4), pp.1155-1188.

Guy, C.M., 1998. Classifications of retail stores and shopping centres: some methodological issues. *GeoJournal,* 45(4), 255-264.

Hart, C.A., Stachow, G.B., Rafiq, M. and Laing, A., (2014). The customer experience of town centres. Online Report. Loughborough University. https://repository.lboro.ac.uk/articles/The_customer_experience_of_town_centres/9494594. Accessed November 2019.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., Xu, J., (1993), Natural movement: or configuration and attraction in urban pedestrian movement, *Environment & Planning B: Planning & Design,* 19, 29-66

Koster, H.R., Pasidis, I. and van Ommeren, J., (2019). Shopping externalities and retail concentration: Evidence from Dutch shopping streets. *Journal of Urban Economics*, 114, 103-194.

Lloyd, S.P. (1982). Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), 129-137.

Lo, R.H., (2009). Walkability: what is it?. *Journal of Urbanism*, *2*(2), 145-166.

Local Data Company [LDC], 2017. *Retail Unit Address Data,* Electronic Dataset, Consumer Data Research Centre, Accessed: November 2018

Local Data Company [LDC], 2018, *SmartStreetSensor Footfall Data,* Electronic Dataset, Consumer Data Research Centre, Accessed: November 2018

Lugomer, K. and Longley, P., (2018). Towards a Comprehensive Temporal Classification of Footfall Patterns in the Cities of Great Britain. *10th International Conference on Geographic Information Science (GIScience 2018)*. DOI: 10.4230/LIPIcs.GISCIENCE.2018.43

Lugomer, K., Soundararaj, B., Murcio, R., Cheshire, J. and Longley, P., (2017). Understanding sources of measurement error in the Wi-Fi sensor data in the Smart City. Conference Paper. *In Proceedings of GISRUK 2017. GIS Research UK (GISRUK).* https://discovery.ucl.ac.uk/id/eprint/10062790/1/GISRUK_2017_paper_95.pdf. Accessed December 2019.

Lyons, G. and Chatterjee, K., 2008. A human perspective on the daily commute: costs, benefits and trade-offs. *Transport Reviews*, *28*(2), pp.181-198.

Makkar G. (2020). Real-Time Footfall Prediction Using Weather Data: A Case on Retail Analytics. Conference Paper. In: *Sharma N., Chakrabarti A., Balas V. (eds) Data Management, Analytics and Innovation*. Advances in Intelligent Systems and Computing, vol 1042. Springer, Singapore.

Mazumdar, S., Bagheri, N., Chong, S., Cochrane, T., Jalaludin, B. and Davey, R., (2019). A Hotspot of Walking in and around the Central Business District: Leveraging Coarsely Geocoded Routinely Collected Data. *Applied Spatial Analysis and Policy*. https://doi.org/10.1007/s12061-019-09321-7

Millington, S, Ntounis, N, Parker C and Quin, S. (2015). Multifunctional Centres: A Sustainable Role for Town and City Centres, *Institute of Place Management, Manchester Metropolitan University*.

Millington, S, Ntounis, N, Parker C, Quin, S, Roberts, G, Steadman, C. (2018). High Street 2030: Achieving Change, *Institute of Place Management, Manchester Metropolitan University*.

Mumford, C.L., Parker, C.R., Ntounis, N. and Dargan, E., 2017. 'A clustering study to verify four distinct monthly footfall signatures: a classification for UK retail centres.' Online Technical report. Place version 5. *Institute of Place Management, Manchester Metropolitan University.* Available at: https://placemanagement.org/media/99861/monthlysignatureclusteringv45withlondon.pdf. Accessed November 2019

Mumford, C., Parker, C., Ntounis, N. and Dargan, E., Footfall signatures and volumes: Towards a classification of UK centres. *Environment and Planning B: Urban Analytics and City Science.* Available at: doi:10.1177/2399808320911412. Accessed October 2020

Newing, A., Clarke, G. and Clarke, M., (2014). Exploring small area demand for grocery retailers in tourist areas. *Tourism Economics,* 20(2), 407-427.

Newing, A., Clarke, G. and Clarke, M., (2018). Applied spatial modelling for retail planning in tourist resorts. *International Journal of Retail & Distribution Management,* 11/12.

ONS, 2017. *Workplace Zones: A new geography for workplace statistics*, Electronic Dataset, Data.Gov.UK, Accessed: February 2019

ONS, 2020. *Retail Sales Index time series,* Electronic Dataset, https://www.ons.gov.uk/businessindustryandtrade/retailindustry/datasets/retailsales, Accessed: April 2020

Parker, C.R., Ntounis, N., Quin, S. and Millington, S., (2016a). High Street UK 2020 Project Report. *Institute of Place Management, Manchester Metropolitan University.* Available at: https://placemanagement.org/media/57742/HSUK2020-End-of-Project-Reportcompressed.pdf, Accessed: November 2019

Parker, C., Ntounis, N., Quin, S. and Millington, S., (2016b). Identifying factors that influence vitality and viability. *Institute of Place Management*, *Manchester Metropolitan University.*

Pavlis, M., Dolega, L. and Singleton, A. (2017). *CDRC Retail Centres and Catchments,* Shape geospatial data, Consumer Data Research Centre, Accessed: February 2019

Porta, S., Strano, E., Iacoviello, V., Messora, R., Latora, V., Cardillo, A., … Scellato, S. (2009). *Street Centrality and Densities of Retail and Services in Bologna, Italy*. Environment and Planning B: Planning and Design, 36(3), 450-465. Avaliable at: https://doi.org/10.1068/b34098. Accessed: October 2020.

Portas, M., (2011). The Portas Review. An independent review into the future of our high streets. Online Report. Gov.UK. Available at: https://www.gov.uk/government/publications/the-portas-review-the-future-of-our-high-streets. Accessed November 2019

Raford, N. and Ragland, D., (2006). Pedestrian volume modeling for traffic safety and exposure analysis: The case of Boston, Massachusetts. Research Report. *University of California Berkeley.*

Ravenscroft, N., Reeves, J. and Rowley, M., (2000). Leisure, property, and the viability of town centres. *Environment and Planning A*, 32(8), 1359-1374.

Scheurer, J. and Porta, S., (2006). Centrality and connectivity in public transport networks and their significance for transport sustainability in cities. Conference Paper. In *World Planning Schools Congress, Global Planning Association Education Network,*. https://strathprints.strath.ac.uk/id/eprint/18466. Accessed January 2020.

Soundararaj, B., Cheshire, J. and Longley, P., (2020). Estimating real-time high-street footfall from Wi-Fi probe requests. *International Journal of Geographical Information Science*, 34(2), 325-343.

Spielman, S.E. and Singleton, A., (2015). Studying neighborhoods using uncertain data from the American community survey: a contextual approach. *Annals of the Association of American Geographers*, 105(5), 1003-1025

Swinney, P. and Sivaev, D., (2013). Beyond the High Street: Why our city centres really matter. Online Report. *Centre for Cities*. https://www.centreforcities.org/wp-content/uploads/2014/08/13-09-10-Beyond-the-High-Street-Why-our-city-centres-really-matter.pdf. Accessed November 2019

Teller, C. and Alexander, A., (2014). Store managers–the seismographs in shopping centres. *European Journal of Marketing,* 48(11/12), 2127-2152.

Tyler, S., Semper, G., Guest, P. and Fieldhouse, B., (2012). The relevance of parking in the success of urban centres, A review for London Councils. *The Means*.

Unwin, J., Symonds, P. and Laike, T., (2017). 'Does lighting affect pedestrian flows? A pilot study in Lund, Market Harborough and Dublin'. *Proceedings of the Lux Europa 2017 European Lighting Conference-Lighting for Modern Society*.

Üsküplü, T., Terzi, F. and Kartal, H., (2020). Discovering Activity Patterns in the City by Social Media Network Data: a Case Study of Istanbul. *Applied Spatial Analysis and Policy*, https://doi.org/10.1007/s12061-020-09336-5.

Vickers, D. and Rees, P., 2007. Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*(2), pp.379-403.

Warnaby, G. and Yip, K.M., (2005). Promotional planning in UK regional shopping centres: an exploratory study. *Marketing Intelligence & Planning.* 23(1), 43-57

Williamson, B., Hargreaves, J., Bond, J. and Lay, H., (2006). The economic costs and benefits of easing Sunday shopping restrictions on large stores in England and Wales. *A report for the Department of Trade and Industry*.

Wrigley, N. and Dolega, L., (2011). Resilience, fragility, and adaptation: new evidence on the performance of UK high streets during global economic crisis and its policy implications. *Environment and Planning A,* 43(10*)*, 2337-2363.
Wrigley, N., Lambiri, D., Astbury, G., Dolega, L., Hart, C., Reeves, C., et al. (2015). British high streets: From crisis to recovery? A comprehensive review of the evidence. *Southampton: University of Southampton/Economic and Social Research Council (ESRC).*
Wrigley, N., Lambiri, D. and Cudworth, K., (2009). Linked trips and town centre viability. *Town & Country Planning*, 78(10).433-438.
Yuo, T.T., Crosby, N., Lizieri, C.M. and McCann, P., (2003). The management of positive inter-store externalities in shopping centres: some empirical evidence. *Paper Presented at the European Real Estate Society 2003 Conference Helsinki, Finland*

Appendix A

Distribution of sensor sample across UK towns and cities

Town/City	Number of sensors (n= 640)	% of sample
Birmingham	5	0.8%
Blackpool	5	0.8%
Boston	2	0.3%
Bradford	2	0.3%
Brighton	19	3.0%
Bristol	14	2.2%
Bromley	11	1.7%
Cambridge	11	1.7%
Cardiff	8	1.3%
Chelmsford	3	0.5%
Chester	18	2.8%
Croydon	6	0.9%
Dorchester	7	1.1%
Gateshead	1	0.2%
Gloucester	14	2.2%
Hove	3	0.5%
Hull	5	0.8%
Kingston Upon Thames	20	3.1%
Leamington Spa	8	1.3%
Leeds	13	2.0%
Leicester	6	0.9%
Liverpool	16	2.5%
London	291	45.5%
Manchester	18	2.8%
Market Harborough	13	2.0%
Newcastle Upon Tyne	7	1.1%
Norwich	14	2.2%
Nottingham	17	2.7%
Orpington	6	0.9%
Oxford	11	1.7%
Plymouth	8	1.3%
Reading	17	2.7%
Salisbury	11	1.7%

Sheffield	8	1.3%
Solihull	2	0.3%
Southampton	8	1.3%
Taunton	6	0.9%
Watford	3	0.5%
Windsor	1	0.2%
York	2	0.3%

Appendix B

Correlation coefficients and significance of the micro-locational footfall influencers used in the clustering

	Premium Stores	Entertainment Venues	Anchor Stores	Workplace Population	Transport Hubs	Car Parks	Density of units	Value Stores	Independent Stores	Night-time Economy	Service prominence	Vacancy Rate	Centrality
Premium Stores	*1.00*												
Entertainment Venues	*.25* *a*	*1.00*											
Anchor Stores	*.24* *a*	*.41* *a*	*1.00*										
Workplace Population	*-.26* *a*	*-.06* *b*	*-.07* *c*	*1.00*									
Transport Hubs	*.19* *a*	*.26* *a*	*-.02* *c*	*-.19* *a*	*1.00*								
Car Parks	*.07* *c*	*.09* *b*	*.19* *a*	*.11* *a*	*-.12* *a*	*1.00*							
Density of units	*-.23* *a*	*-.20* *a*	*-.28* *a*	*-.06* *b*	*.06* *b*	*-.14* *a*	*1.00*						
Value Stores	*.43* *a*	*.03* *b*	*-.04* *c*	*-.33* *a*	*.18* *a*	*.02* *c*	*-.08* *b*	*1.00*					
Independent Stores	*.24* *a*	*.11* *a*	*.44* *a*	*-.18* *a*	*-.03* *c*	*.07* *c*	*-.16* *a*	*.05* *c*	*1.00*				
Night-time Economy	*-.07* *c*	*-.09* *b*	*.29* *a*	*.34* *a*	*-.18* *a*	*.09* *c*	*-.33* *a*	*-.27* *a*	*.27* *a*	*1.00*			
Service prominence	*.09* *b*	*.11* *a*	*.20* *a*	*-.05* *b*	*-.01* *c*	*.08* *b*	*-.30* *a*	*.01* *c*	*.27* *a*	*.25* *a*	*1.00*		
Vacancy Rate	*-.10* *b*	*-.04* *b*	*-.09* *b*	*-.09* *b*	*.13* *a*	*-.10* *b*	*.09* *b*	*.11* *a*	*-.09* *b*	*-.18* *a*	*.00* *c*	*1.00*	
Centrality	*.40* *a*	*.32* *a*	*.14* *a*	*-.41* *a*	*.29* *a*	*-.03* *c*	*.02* *c*	*.30* *a*	*.00* *c*	*-.30* *a*	*.26* *a*	*.05* *c*	*1.00*

Notes: [a] correlation significant to 0.01; [b] correlation significant to 0.05; [c] correlation not significant

Appendix C

Specifics for the derivation of some of the footfall descriptors, informed by the conditions in Dolega, Pavlis & Singleton (2016)

Value stores

Store Name: Aldi, Lidl, Iceland, Primark, Farmfoods, Poundworld, Poundstretcher, Home Bargains, Savers, B&M Bargains, Pound Bakery

Category: Discount & Surplus Stores, Charity And Secondhand Shops, Pawnbroking And Cheque Cashing

Subcategory: Bookmakers.

Night-time economy locations

Category: Bars Pubs And Clubs, Off Licenses And Restaurants

Subcategory: Fast Food Takeaway, Take Away Food Shops, Fish And Chip Shops, Pizza Takeaway, Chinese Fast Food Takeaway, Indian Takeaway, Fast Food Delivery, Amusement Parks & Arcades, Theatres & Concert Halls, Cinemas, Snooker, Billiards & Pool Halls, Bowling Alleys

Ratio of service to retail

Retail over service where retail is:

Classification: Comparison

Category: Groceries, Supermarkets & Food Shops, Bakers, Confectionery, Tobacco, Newsagents, Off Licenses, Butchers & Fishmongers

And service is:

Classification: Service

Anchor Stores

Store name: Tesco (excluding Express and in store services), Sainsburys (excluding Local and in store services), Waitrose (excluding Little Waitrose), Morrisons (excluding in store services), ASDA (excluding in store services), John Lewis, Debenhams, Marks & Spencer, Harvey Nichols, H&M, Primark, Zara, Boots, Next, B&Q and House of Fraser.

Premium Stores

Store name: Waitrose, John Lewis, Harvey Nichols, Laura Ashley, Ted Baker, Tommy Hilfiger, Fat Face, Superdry, Seasalt, Jack Wills, White Stuff, Crew Clothing, Boss, Cath Kidston, Joules, Swarovski, Lacoste, Diesel, Apple Store, Bose, Hotel Chocolat, Radley, Karen Millan, Michael Kors, The White Company, Reiss, All Saints, Tessuti, Flannels, Ralph Lauren, Kate Spade, Mulberry, Burberry, Armani, Calvin Klein, Coach, Dune, Diesel, Fossil, Fred Perry, French Connection, Guess, Hobbs, Karl Lagerfeld, Kurt Geiger, L'Occitane, Lacoste, Levi, Lindt, Osprey, Swarovski, Timberland and Toms

Entertainment activities

Subcategory: Cinemas, Theatres & Concert Halls, Amusement Parks & Arcades, Museums & Art Galleries, Sports Grounds & Stadiums, Tourist Attractions, Party Venues & Function Rooms, Bingo Halls, Bowling Alleys, Ticket Outlets & Box Offices, Golf Courses, Snooker, Billiards & Pool Halls, Driving Ranges, Ice Rinks, Booking Agents, Paintball & Combat Games and Karting

Appendix D

Values for final cluster centroids, un-standardised for comprehensibility

	CCR micro-locations	**BI micro-locations**	**VOCR micro-locations**
Within cluster sum of squares	2130	2340	1951
Number of observations	343	254	43
Distance to nearest anchor store (m)	85.98	199.93	216.53
Distance to nearest premium store (m)	122.79	234.07	1860.41
Distance to nearest entertainment activity (m)	120.79	157.09	328.91
Mean workplace population	409.03	770.28	94.07
Distance to nearest transport hub (m)	159.36	93.79	249.72
Distance to nearest car park (m)	160.69	227.58	276.44
Density of units (unit per $100^2\pi$ m^2)	79.38	48.86	52.63
Proportion of value stores	5%	2%	12%
Proportion of independent stores	38%	55%	58%
Proportion of night-time economy stores	15%	34%	15%
Ratio of service to retail	0.45	0.87	0.84
Proportion of vacant stores	9%	5%	6%
Centrality of street	0.08	0.05	0.16

The prevalence of each category of store as defined by LDC's survey (2017) for each cluster and for the entire sample (n=222953). The nearest 25 stores to each sensor were considered when calculating a total percentage for the cluster. This information was used alongside the cluster centroids in Appendix C to create the cluster pen portraits.

LDC Categories	**CCR micro-locations (%)**	**BI micro-locations (%)**	**VOCR micro-locations (%)**	**Entire sample (%)**
Accommodation	0.5	3.3	0.0	1.9
Auto & Accessories	0.0	0.0	0.2	0.3
Auto Services	0.0	0.1	0.5	1.5
Bakers	1.2	1.1	1.0	1.0
Banks, Financial Services & Building Societies	4.2	3.0	3.4	1.7
Bars, Pubs & Clubs	3.0	7.8	2.1	4.6
Books, Arts & Crafts, Stationery, Printers	2.8	3.0	2.3	1.9
Butchers & Fishmongers	0.3	0.1	2.4	0.6
Cafes & Fast Food	8.6	15.2	9.7	10.7
Car & Motorbike Showrooms	0.0	0.0	0.0	0.7
Charity & Secondhand Shops	1.8	0.6	5.9	1.6
Chemists, Toiletries & Health	4.1	2.0	3.2	2.6
Confectionery, Tobacco, Newsagents	2.3	1.8	0.9	1.9
Department Stores & Mail Order	0.8	0.3	0.0	0.3
Discount & Surplus Stores	0.5	0.0	2.6	0.6
DIY, Hardware, Builder's Merchants & Household Goods	0.7	0.6	0.7	1.7
Electrical Goods & Home Entertainment	5.4	2.7	5.0	2.6
Employment & Post Offices	1.4	0.8	1.5	1.5
Entertainment	2.5	3.3	4.0	2.5
Estate Agents & Auctioneers	0.9	2.1	3.1	2.9
Fashion & General Clothing	15.1	6.9	5.3	4.6
Florists & Garden	0.2	0.3	0.2	0.5
Footwear	3.0	0.7	1.2	0.7
Furniture, Carpets, Textiles, Bathrooms & Kitchens	1.8	1.2	2.4	3.0
Gifts, China & Leather Goods	1.4	1.5	0.6	0.8
Groceries, Supermarkets & Food Shops	2.5	3.7	6.8	6.9
Hairdressing, Health & Beauty	8.3	7.7	13.8	10.8

Household & Home	0.1	0.1	0.0	0.6
Jewellers, Clocks & Watches	3.1	0.9	1.0	1.1
Launderettes, Dry Cleaners & Other	0.4	1.0	0.8	1.1
Locksmiths, Clothing Alterations & Shoe Repairs	0.6	0.5	0.6	0.4
Medical	0.3	0.4	0.0	0.8
Miscellaneous	0.6	0.8	1.4	1.6
Non-Retail	1.5	2.3	3.3	3.5
Off Licences	0.2	0.6	0.8	0.6
Pawnbroking & Cheque Cashing	0.4	0.2	1.3	0.4
Pet Shops & Pet Supplies	0.0	0.0	0.0	0.3
Petrol Filling Stations	0.0	0.0	0.0	0.7
Restaurants	4.7	14.8	3.7	5.8
Royal Mail Delivery Offices	0.0	0.0	0.0	0.1
Shopping Centres & Markets	0.3	0.1	0.4	0.1
Sports, Toys, Cycle Shops & Hobbies	3.6	1.3	1.3	1.4
Transport	0.5	1.3	0.8	2.3
Travel Agents & Tour Operators	1.5	0.6	0.6	0.6
Vacant	8.9	5.1	5.3	8.3

Appendix 4.2 – Specifics for derivation of micro-locational characteristics

Specifics for the derivation of some of the footfall descriptors, informed by the conditions in Dolega, Pavlis and Singleton (2016).

Some brands included in these specifications have closed down their physical outlets in recent years (e.g. Debenhams, House of Fraser). As the analysis captures 2017 data, they are included as they were still active and open during that time.

Value stores

Store Name:	Aldi, Lidl, Iceland, Primark, Farmfoods, Poundworld, Poundstretcher, Home Bargains, Savers, B&M Bargains, Pound Bakery
Category:	Discount & Surplus Stores, Charity And Secondhand Shops, Pawnbroking And Cheque Cashing
Subcategory:	Bookmakers.

Night-time economy locations

Category:	Bars Pubs And Clubs, Off Licenses And Restaurants
Subcategory:	Fast Food Takeaway, Take Away Food Shops, Fish And Chip Shops, Pizza Takeaway, Chinese Fast Food Takeaway, Indian Takeaway, Fast Food Delivery, Amusement Parks & Arcades, Theatres & Concert Halls, Cinemas, Snooker, Billiards & Pool Halls, Bowling Alleys

Ratio of service to retail

Retail over service where retail is:

Classification:	Comparison

Category:	Groceries, Supermarkets & Food Shops, Bakers, Confectionery, Tobacco, Newsagents, Off Licenses, Butchers & Fishmongers
And service is:	
Classification:	Service

Anchor Stores

Store name:	Tesco (excluding Express and in store services), Sainsburys (excluding Local and in store services), Waitrose (excluding Little Waitrose), Morrisons (excluding in store services), ASDA (excluding in store services), John Lewis, Debenhams, Marks & Spencer, Harvey Nichols, H&M, Primark, Zara, Boots, Next, B&Q and House of Fraser (

Premium Stores

Store name:	Waitrose, John Lewis, Harvey Nichols, Laura Ashley, Ted Baker, Tommy Hilfiger, Fat Face, Superdry, Seasalt, Jack Wills, White Stuff, Crew Clothing, Boss, Cath Kidston, Joules, Swarovski, Lacoste, Diesel, Apple Store, Bose, Hotel Chocolat, Radley, Karen Millan, Michael Kors, The White Company, Reiss, All Saints, Tessuti, Flannels, Ralph Lauren, Kate Spade, Mulberry, Burberry, Armani, Calvin Klein, Coach, Dune, Diesel, Fossil, Fred Perry, French Connection, Guess, Hobbs, Karl Lagerfeld, Kurt Geiger, L'Occitane, Lacoste, Levi, Lindt, Osprey, Swarovski, Timberland and Toms

Entertainment activities

Subcategory:	Cinemas, Theatres & Concert Halls, Amusement Parks & Arcades, Museums & Art Galleries, Sports Grounds & Stadiums, Tourist Attractions, Party Venues & Function Rooms, Bingo Halls, Bowling Alleys, Ticket Outlets & Box Offices, Golf Courses, Snooker, Billiards & Pool Halls, Driving Ranges, Ice Rinks, Booking Agents, Paintball & Combat Games and Karting

Appendix 4.3 – Correlation and summary statistics for micro-locational characteristics

Correlation coefficients and significance of the micro-locational footfall influencers used in the clustering

	Premium Stores	Entertainment Venues	Anchor Stores	Workplace Population	Transport Hubs	Car Parks	Density of units	Value Stores	Independent Stores	Night-time Economy	Service prominence	Vacancy Rate	Centrality
Premium Stores	*1.00*												
Entertainment Venues	*.25* [a]	*1.00*											
Anchor Stores	*.24* [a]	*.41* [a]	*1.00*										
Workplace Population	*-.26* [a]	*-.06* [b]	*-.07* [c]	*1.00*									
Transport Hubs	*.19* [a]	*.26* [a]	*-.02* [c]	*-.19* [a]	*1.00*								
Car Parks	*.07* [c]	*.09* [b]	*.19* [a]	*.11* [a]	*-.12* [a]	*1.00*							
Density of units	*-.23* [a]	*-.20* [a]	*-.28* [a]	*-.06* [b]	*.06* [b]	*-.14* [a]	*1.00*						
Value Stores	*.43* [a]	*.03* [b]	*-.04* [c]	*-.33* [a]	*.18* [a]	*.02* [c]	*-.08* [b]	*1.00*					
Independent Stores	*.24* [a]	*.11* [a]	*.44* [a]	*-.18* [a]	*-.03* [c]	*.07* [c]	*-.16* [a]	*.05* [c]	*1.00*				
Night-time Economy	*-.07* [c]	*-.09* [b]	*.29* [a]	*.34* [a]	*-.18* [a]	*.09* [c]	*-.33* [a]	*-.27* [a]	*.27* [a]	*1.00*			
Service prominence	*.09* [b]	*.11* [a]	*.20* [a]	*-.05* [b]	*-.01* [c]	*.08* [b]	*-.30* [a]	*.01* [c]	*.27* [a]	*.25* [a]	*1.00*		
Vacancy Rate	*-.10* [b]	*-.04* [b]	*-.09* [b]	*-.09* [b]	*.13* [a]	*-.10* [b]	*.09* [b]	*.11* [a]	*-.09* [b]	*-.18* [a]	*.00* [c]	*1.00*	
Centrality	*.40* [a]	*.32* [a]	*.14* [a]	*-.41* [a]	*.29* [a]	*-.03* [c]	*.02* [c]	*.30* [a]	*.00* [c]	*-.30* [a]	*.26* [a]	*.05* [c]	*1.00*

Notes: [a] correlation significant to 0.01; [b] correlation significant to 0.05; [c] correlation not significant

Summary statistics of the micro-locational footfall influencers used in the clustering

Variable	Min	Median	Mean	Max	IQR
Distance to nearest anchor store (m)	0	97	139	2494	128
Distance to nearest premium store (m)	0	121	284	3404	234
Distance to nearest entertainment activity (m)	0	108	149	2684	118
Workplace population	10	397	531	2981	439
Distance to nearest transport hub (m)	3	95	139	2827	113
Distance to nearest car park (m)	30	168	195	1493	117
Density of units (unit per $100^2\pi$ m^2)	7	60	66	189	47
Proportion of value stores	0%	3%	4%	30%	5%
Proportion of independent stores	5%	44%	46%	96%	27%
Proportion of night-time economy stores	2%	19%	22%	68%	19%
Ratio of service to retail	0	0.44	0.64	9.5	0.48
Proportion of vacant stores	0%	6%	7.4%	39%	7%
Centrality of street	0.02	0.07	0.07	0.53	0.05

Appendix 4.4 – Distribution of sensor sample across UK towns and cities

Town/City	Number of sensors (n= 640)	% of sample
Birmingham	5	0.8%
Blackpool	5	0.8%
Boston	2	0.3%
Bradford	2	0.3%
Brighton	19	3.0%
Bristol	14	2.2%
Bromley	11	1.7%
Cambridge	11	1.7%
Cardiff	8	1.3%
Chelmsford	3	0.5%
Chester	18	2.8%
Croydon	6	0.9%
Dorchester	7	1.1%
Gateshead	1	0.2%
Gloucester	14	2.2%
Hove	3	0.5%
Hull	5	0.8%
Kingston Upon Thames	20	3.1%
Leamington Spa	8	1.3%
Leeds	13	2.0%
Leicester	6	0.9%
Liverpool	16	2.5%
London	291	45.5%
Manchester	18	2.8%
Market Harborough	13	2.0%
Newcastle Upon Tyne	7	1.1%
Norwich	14	2.2%
Nottingham	17	2.7%
Orpington	6	0.9%
Oxford	11	1.7%
Plymouth	8	1.3%
Reading	17	2.7%
Salisbury	11	1.7%

Sheffield	8	1.3%
Solihull	2	0.3%
Southampton	8	1.3%
Taunton	6	0.9%
Watford	3	0.5%
Windsor	1	0.2%
York	2	0.3%

Appendix 4.5 – Correlations between micro-locational characteristics and footfall with time

Correlations between the 13 micro-locational characteristics and time of day on weekdays (*** $p < 0.001$ ** $p < 0.01$ * $p < 0.05$).

Weekday	**Night**			**Early Morning**			**Morning**			**Afternoon**			**Early Evening**			**Evening**		
	Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value	
Distance to Premium Store	-0.06	0.186		-0.01	0.874		-0.11	0.019	*	-0.19	0.000	***	-0.12	0.006	**	-0.10	0.030	*
Distance to Entertainment Venue	-0.02	0.721		0.03	0.538		-0.03	0.464		-0.10	0.027	*	-0.05	0.278		-0.02	0.688	
Distance to Anchor Store	0.02	0.639		-0.01	0.811		-0.14	0.002	**	-0.21	0.000	***	-0.11	0.019	*	0.02	0.676	
Workplace Population	0.28	0.000	***	0.26	0.000	***	0.31	0.000	***	0.34	0.000	***	0.42	0.000	***	0.37	0.000	***
Distance to Transport Hub	-0.23	0.000	***	-0.29	0.000	***	-0.25	0.000	***	-0.19	0.000	***	-0.28	0.000	***	-0.28	0.000	***
Distance to Car Park	0.09	0.056		0.17	0.000	***	0.11	0.020	*	0.04	0.331		0.11	0.020	*	0.09	0.052	
Density of Units	-0.09	0.041	*	-0.19	0.000	***	-0.06	0.214		0.12	0.007	*	-0.02	0.724		-0.09	0.051	
Proportion of Value stores	-0.15	0.001	***	-0.07	0.117		-0.13	0.003	**	-0.19	0.000	***	-0.20	0.000	***	-0.18	0.000	***
Proportion of Independent stores	0.04	0.342		-0.02	0.661		-0.23	0.000	***	-0.39	0.000	***	-0.20	0.000	***	0.02	0.726	
Proportion of Night-time economy	0.37	0.000	***	0.26	0.000	***	0.13	0.005	**	0.01	0.785		0.23	0.000	***	0.43	0.000	***
Ratio of service to retail	-0.01	0.879		0.04	0.342		-0.07	0.134		-0.18	0.000	***	-0.10	0.029	*	-0.02	0.719	
Proportion of Vacant stores	-0.20	0.000	***	-0.15	0.001	**	-0.10	0.028	*	-0.04	0.421		-0.15	0.001	**	-0.21	0.000	***
Centrality of Street	-0.23	0.000	***	-0.18	0.000	***	-0.19	0.000	***	-0.19	0.000	***	-0.25	0.000	***	-0.27	0.000	***

Correlations between the 13 micro-locational characteristics and time of day on Saturdays (*** p < 0.001 ** p < 0.01 * p < 0.05).

Saturday	Night			Early Morning			Morning			Afternoon			Early Evening			Evening		
	Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value	
Distance to Premium Store	-0.09	0.044	*	-0.01	0.892		-0.14	0.003	**	-0.22	0.000	***	-0.18	0.000	***	-0.12	0.006	*
Distance to Entertainment Venue	-0.07	0.143		0.01	0.852		-0.10	0.025	*	-0.15	0.001	**	-0.10	0.026	*	-0.07	0.135	
Distance to Anchor Store	0.02	0.609		-0.01	0.750		-0.25	0.000	***	-0.25	0.000	***	-0.14	0.001	**	0.01	0.790	
Workplace Population	0.23	0.000	***	0.21	0.000	***	0.10	0.030	*	0.20	0.000	***	0.34	0.000	***	0.32	0.000	***
Distance to Transport Hub	-0.18	0.000	***	-0.27	0.000	***	-0.15	0.001	**	-0.11	0.012	*	-0.21	0.000	***	-0.23	0.000	***
Distance to Car Park	0.05	0.262		0.14	0.003	**	0.03	0.498		-0.01	0.820		0.03	0.540		0.04	0.431	
Density of Units	-0.03	0.565		-0.14	0.001	**	0.18	0.000	***	0.30	0.000	***	0.16	0.001	***	0.01	0.866	
Proportion of Value stores	-0.14	0.002	**	-0.07	0.129		-0.10	0.030	*	-0.19	0.000	***	-0.22	0.000	***	-0.18	0.000	***
Proportion of Independent stores	0.05	0.248		-0.01	0.909		-0.38	0.000	***	-0.44	0.000	***	-0.26	0.000	***	0.02	0.692	
Proportion of Night-time economy	0.40	0.000	***	0.26	0.000	***	-0.14	0.002	**	-0.12	0.007	**	0.15	0.001	**	0.44	0.000	***
Ratio of service to retail	-0.01	0.836		0.02	0.692		-0.19	0.000	***	-0.26	0.000	***	-0.19	0.000	***	-0.05	0.325	
Proportion of Vacant stores	-0.16	0.000	***	-0.14	0.002	***	0.00	0.916		0.03	0.533		-0.08	0.071		-0.14	0.002	**
Centrality of Street	-0.21	0.000	***	-0.18	0.000	***	-0.05	0.242		-0.11	0.015	*	-0.22	0.000	***	-0.26	0.000	***

Correlations between the 13 micro-locational characteristics and time of day on Sundays (*** p < 0.001 ** p < 0.01 * p < 0.05).

Sunday	Night			Early Morning			Morning			Afternoon			Early Evening			Evening		
	Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value		Cor	P-value	
Distance to Premium Store	-0.09	0.041	*	-0.02	0.646		-0.13	0.003	**	-0.20	0.000	***	-0.14	0.002	**	-0.08	0.088	
Distance to Entertainment Venue	-0.08	0.090		0.01	0.847		-0.06	0.177		-0.11	0.014	*	-0.04	0.354		-0.01	0.810	
Distance to Anchor Store	0.02	0.684		0.00	0.943		-0.18	0.000	***	-0.23	0.000	***	-0.08	0.073		0.02	0.617	
Workplace Population	0.20	0.000	***	0.23	0.000	***	0.17	0.000	***	0.21	0.000	***	0.34	0.000	***	0.26	0.000	***
Distance to Transport Hub	-0.15	0.001	**	-0.25	0.000	***	-0.21	0.000	***	-0.17	0.000	***	-0.27	0.000	***	-0.26	0.000	***
Distance to Car Park	0.03	0.505		0.13	0.006	**	0.07	0.146		0.02	0.713		0.05	0.275		0.06	0.197	
Density of Units	0.00	0.989		-0.14	0.002	***	0.08	0.092		0.24	0.000	***	0.04	0.444		-0.08	0.099	
Proportion of Value stores	-0.13	0.004	**	-0.10	0.032	*	-0.16	0.001	**	-0.20	0.000	***	-0.22	0.000	***	-0.15	0.001	**
Proportion of Independent stores	0.06	0.217		0.01	0.777		-0.28	0.000	***	-0.40	0.000	***	-0.14	0.002	**	0.06	0.225	
Proportion of Night-time economy	0.40	0.000	***	0.32	0.000	***	0.03	0.579		-0.07	0.138		0.26	0.000	***	0.41	0.000	***
Ratio of service to retail	-0.01	0.746		0.01	0.863		-0.15	0.001	**	-0.25	0.000	***	-0.13	0.004	**	-0.01	0.820	
Proportion of Vacant stores	-0.13	0.006	**	-0.15	0.001	**	-0.09	0.054		-0.04	0.373		-0.15	0.001	**	-0.18	0.000	***
Centrality of Street	-0.19	0.000	***	-0.19	0.000	***	-0.11	0.013	*	-0.12	0.007	**	-0.24	0.000	***	-0.23	0.000	***

Appendix 4.6 – Cluster centroids and information used for class derivation

Values for final cluster centroids, un-standardised for comprehensibility

	CCR micro-locations	BI micro-locations	VOCR micro-locations
Within cluster sum of squares	2130	2340	1951
Number of observations	343	254	43
Distance to nearest anchor store (m)	85.98	199.93	216.53
Distance to nearest premium store (m)	122.79	234.07	1860.41
Distance to nearest entertainment activity (m)	120.79	157.09	328.91
Mean workplace population	409.03	770.28	94.07
Distance to nearest transport hub (m)	159.36	93.79	249.72
Distance to nearest car park (m)	160.69	227.58	276.44
Density of units (unit per $100^2\pi$ m^2)	79.38	48.86	52.63
Proportion of value stores	5%	2%	12%
Proportion of independent stores	38%	55%	58%
Proportion of night-time economy stores	15%	34%	15%
Ratio of service to retail	0.45	0.87	0.84
Proportion of vacant stores	9%	5%	6%
Centrality of street	0.08	0.05	0.16

The prevalence of each category of store as defined by LDC's survey (2017) for each cluster and for the entire sample (n=222953). The nearest 25 stores to each sensor were considered when calculating a total percentage for the cluster. This information was used alongside the cluster centroids in to create the cluster pen portraits.

LDC Categories	**CCR micro-locations (%)**	**BI micro-locations (%)**	**VOCR micro-locations (%)**	**Entire sample (%)**
Accommodation	0.5	3.3	0.0	1.9
Auto & Accessories	0.0	0.0	0.2	0.3
Auto Services	0.0	0.1	0.5	1.5
Bakers	1.2	1.1	1.0	1.0
Banks, Financial Services & Building Societies	4.2	3.0	3.4	1.7
Bars, Pubs & Clubs	3.0	7.8	2.1	4.6
Books, Arts & Crafts, Stationery, Printers	2.8	3.0	2.3	1.9
Butchers & Fishmongers	0.3	0.1	2.4	0.6
Cafes & Fast Food	8.6	15.2	9.7	10.7
Car & Motorbike Showrooms	0.0	0.0	0.0	0.7
Charity & Secondhand Shops	1.8	0.6	5.9	1.6
Chemists, Toiletries & Health	4.1	2.0	3.2	2.6
Confectionery, Tobacco, Newsagents	2.3	1.8	0.9	1.9
Department Stores & Mail Order	0.8	0.3	0.0	0.3
Discount & Surplus Stores	0.5	0.0	2.6	0.6
DIY, Hardware, Builder's Merchants & Household Goods	0.7	0.6	0.7	1.7
Electrical Goods & Home Entertainment	5.4	2.7	5.0	2.6
Employment & Post Offices	1.4	0.8	1.5	1.5
Entertainment	2.5	3.3	4.0	2.5
Estate Agents & Auctioneers	0.9	2.1	3.1	2.9
Fashion & General Clothing	15.1	6.9	5.3	4.6
Florists & Garden	0.2	0.3	0.2	0.5

Footwear	3.0	0.7	1.2	0.7
Furniture, Carpets, Textiles, Bathrooms & Kitchens	1.8	1.2	2.4	3.0
Gifts, China & Leather Goods	1.4	1.5	0.6	0.8
Groceries, Supermarkets & Food Shops	2.5	3.7	6.8	6.9
Hairdressing, Health & Beauty	8.3	7.7	13.8	10.8
Household & Home	0.1	0.1	0.0	0.6
Jewellers, Clocks & Watches	3.1	0.9	1.0	1.1
Launderettes, Dry Cleaners & Other	0.4	1.0	0.8	1.1
Locksmiths, Clothing Alterations & Shoe Repairs	0.6	0.5	0.6	0.4
Medical	0.3	0.4	0.0	0.8
Miscellaneous	0.6	0.8	1.4	1.6
Non-Retail	1.5	2.3	3.3	3.5
Off Licences	0.2	0.6	0.8	0.6
Pawnbroking & Cheque Cashing	0.4	0.2	1.3	0.4
Pet Shops & Pet Supplies	0.0	0.0	0.0	0.3
Petrol Filling Stations	0.0	0.0	0.0	0.7
Restaurants	4.7	14.8	3.7	5.8
Royal Mail Delivery Offices	0.0	0.0	0.0	0.1
Shopping Centres & Markets	0.3	0.1	0.4	0.1
Sports, Toys, Cycle Shops & Hobbies	3.6	1.3	1.3	1.4
Transport	0.5	1.3	0.8	2.3
Travel Agents & Tour Operators	1.5	0.6	0.6	0.6
Vacant	8.9	5.1	5.3	8.3

Appendix 5.1 – Deriving retail snapshots for case study micro-locations

Retail Snapshots for each micro-location were derived from the Local Data Company Retail Unit dataset for July 2018. The units were clipped to a buffer of 100m around the sensors used for that micro-location. 100m was used as it is large enough to give a clear idea of the general retail composition of the area without making the sample too large. The Store Types were as defined as 'Category' in the Local Data Company dataset. The order is as follows, Liverpool ONE, Manchester Market Street, Edinburgh Old Town and Edinburgh New Town.

Liverpool ONE micro-location

Store Type	Number of Stores	Percentage
Fashion & General Clothing	49	23%
Cafes & Fast Food	22	10%
Sports, Toys, Cycle Shops & Hobbies	20	9%
Restaurants	17	8%
Chemists, Toiletries & Health	14	7%
Electrical Goods & Home Entertainment	14	7%
Footwear	11	5%
Vacant Property	11	5%
Hairdressing, Health & Beauty	9	4%
Jewellers, Clocks & Watches	8	4%
Books, Arts & Crafts, Stationery, Printers	5	2%
Groceries, Supermarkets & Food Shops	5	2%
Banks, Financial Services & Building Societies	4	2%
Bars, Pubs & Clubs	4	2%
Confectionery, Tobacco, Newsagents	4	2%
Department Stores & Mail Order	3	1%
Entertainment	3	1%
Gifts, China & Leather Goods	3	1%
Discount & Surplus Stores	2	1%
Accommodation	1	0%
Bakers	1	0%
Furniture, Carpets, Textiles, Bathrooms & Kitchens	1	0%
Miscellaneous	1	0%
Transport	1	0%
Travel Agents & Tour Operators	1	0%
Total	214	

Manchester (Market St) micro-location

Store Type	Number of Stores	Percentage
Fashion & General Clothing	31	16%
Vacant Property	24	13%
Cafes & Fast Food	20	11%
Electrical Goods & Home Entertainment	15	8%
Sports, Toys, Cycle Shops & Hobbies	12	6%
Footwear	10	5%
Jewellers, Clocks & Watches	10	5%
Chemists, Toiletries & Health	8	4%
Groceries, Supermarkets & Food Shops	8	4%
Hairdressing, Health & Beauty	8	4%
Confectionery, Tobacco, Newsagents	7	4%
Banks, Financial Services & Building Societies	6	3%
Restaurants	6	3%
Bakers	3	2%
Gifts, China & Leather Goods	3	2%
Bars, Pubs & Clubs	2	1%
Discount & Surplus Stores	2	1%
Entertainment	2	1%
Furniture, Carpets, Textiles, Bathrooms & Kitchens	2	1%
Transport	2	1%
Travel Agents & Tour Operators	2	1%
Books, Arts & Crafts, Stationery, Printers	1	1%
Department Stores & Mail Order	1	1%
DIY, Hardware, Builder's Merchants & Household Goods	1	1%
Employment & Post Offices	1	1%
Household & Home	1	1%
Medical	1	1%
Total	189	

Edinburgh Old Town micro-location

Store Type	Number of Stores	Percentage
Restaurants	22	15%
Fashion & General Clothing	18	13%
Gifts, China & Leather Goods	17	12%
Cafes & Fast Food	15	11%
Bars, Pubs & Clubs	14	10%
Accommodation	10	7%
Travel Agents & Tour Operators	7	5%
Vacant Property	7	5%
Entertainment	5	4%
Hairdressing, Health & Beauty	4	3%
Off Licences	4	3%
Confectionery, Tobacco, Newsagents	3	2%
Groceries, Supermarkets & Food Shops	3	2%
Books, Arts & Crafts, Stationery, Printers	2	1%
Furniture, Carpets, Textiles, Bathrooms & Kitchens	2	1%
Miscellaneous	2	1%
Bakers	1	1%
Banks, Financial Services & Building Societies	1	1%
Charity & Secondhand Shops	1	1%
Chemists, Toiletries & Health	1	1%
Jewellers, Clocks & Watches	1	1%
Shopping Centres & Markets	1	1%
Sports, Toys, Cycle Shops & Hobbies	1	1%
Total	142	

Edinburgh New Town micro-location

Store Type	Number of Stores	Percentage
Fashion & General Clothing	54	21%
Restaurants	31	12%
Hairdressing, Health & Beauty	23	9%
Bars, Pubs & Clubs	21	8%
Cafes & Fast Food	21	8%
Vacant Property	18	7%
Jewellers, Clocks & Watches	14	5%
Accommodation	8	3%
Chemists, Toiletries & Health	7	3%
Employment & Post Offices	7	3%
Banks, Financial Services & Building Societies	6	2%
Footwear	6	2%
Furniture, Carpets, Textiles, Bathrooms & Kitchens	5	2%
Groceries, Supermarkets & Food Shops	5	2%
Sports, Toys, Cycle Shops & Hobbies	5	2%
Gifts, China & Leather Goods	4	2%
Electrical Goods & Home Entertainment	3	1%
Estate Agents & Auctioneers	3	1%
Locksmiths, Clothing Alterations & Shoe Repairs	3	1%
Miscellaneous	3	1%
Bakers	2	1%
Books, Arts & Crafts, Stationery, Printers	2	1%
Entertainment	2	1%
Travel Agents & Tour Operators	2	1%
Charity & Secondhand Shops	1	0%
Confectionery, Tobacco, Newsagents	1	0%
Department Stores & Mail Order	1	0%
DIY, Hardware, Builder's Merchants & Household Goods	1	0%
Off Licences	1	0%
Pawnbroking & Cheque Cashing	1	0%
Transport	1	0%
Total	262	

Appendix 5.2 – Bank Holiday Dates

Name	**2017**	**2018**
New Year's Day	2nd January	1st January
2nd January (Scotland)	3rd January	2nd January
Easter Weekend	14th – 17th April	30th March – 2nd April
Early May Bank Holiday	1st May	7th May
Spring Bank Holiday	29th May	28th May
Summer Bank Holiday	7th August	6th August
Black Friday	24th November	23rd November
St Andrew's Day (Scotland)	30th November	30th November
Super Saturday	23rd December	22nd December
Christmas Day	25th December	25th December
Boxing Day	26th December	26th December

Appendix 5.3 – Full derived rankings and explanations for case study locations

Liverpool ONE micro-location

Date	Rank	Event	Date	Rank	Event
02/12/2017	4.8	Christmas (Sat)	01/12/2018	2.2	Christmas (Sat)
22/12/2017	7.3	Christmas (Fri)	24/11/2018	3.1	Christmas (Sat)
24/11/2017	8.5	Black Friday	23/11/2018	4.9	Black Friday
25/11/2017	9.0	Christmas (Sat)	22/12/2018	6.1	Christmas (Sat)
23/12/2017	10.7	Christmas (Sat)	06/10/2018	6.5	Giants Spectacular
21/12/2017	10.9	Christmas (Thu)	27/12/2018	8.6	Boxing Day Sales
18/11/2017	11.7	Christmas (Sat)	21/12/2018	11.4	Christmas (Fri)
09/12/2017	12.6	Christmas (Sat)	08/12/2018	11.4	Christmas (Sat)
16/12/2017	14.0	Christmas (Sat)	28/12/2018	12.0	Boxing Day Sales
28/12/2017	16.2	Boxing Day Sales	10/11/2018	12.6	Christmas (Sat)
27/12/2017	17.8	Boxing Day Sales	17/11/2018	14.2	Christmas (Sat)
11/11/2017	18.4	Christmas (Sat)	03/11/2018	17.1	N/A
01/12/2017	18.9	Christmas (Fri)	20/12/2018	19.6	Christmas (Thu)
20/12/2017	24.5	Christmas (Wed)	20/10/2018	21.2	Half Term
18/12/2017	24.7	Christmas (Mon)	16/11/2018	23.7	Christmas (Fri)
29/12/2017	27.3	Boxing Day Sales	29/09/2018	24.2	Freshers Week
04/11/2017	28.3	N/A	30/11/2018	28.6	Christmas (Fri)
25/03/2017	28.4	Liverpool Home Game	17/12/2018	29.4	Christmas (Mon)
27/10/2017	29.8	Halloween (Fri) Half Term	14/12/2018	29.4	Christmas (Fri)
19/12/2017	31.7	Christmas (Tue)	15/12/2018	29.4	Christmas (Sat)

Manchester (Market Street) micro-location

Date	Rank	Event	Date	Rank	Event
22/12/2017	2.2	Christmas (Fri)	24/11/2018	2.0	Christmas (Sat)
02/12/2017	3.0	Christmas (Sat)	17/11/2018	2.5	Christmas (Sat)
24/11/2017	5.4	Black Friday	23/11/2018	2.5	Black Friday
09/12/2017	6.3	Christmas (Sat)	01/12/2018	3.9	Christmas (Sat)
25/11/2017	6.6	Christmas (Sat)	10/11/2018	5.1	Christmas (Sat)
01/12/2017	7.3	Christmas (Fri)	27/10/2018	8.6	Halloween Half Term
08/12/2017	7.9	Christmas (Fri)	06/10/2018	9.6	Food & Drink Festival
23/12/2017	8.6	Christmas (Sat)	20/10/2018	12.2	Half Term
21/12/2017	11.5	Christmas (Thu)	21/12/2018	13.1	Christmas (Fri)
27/12/2017	14.8	Boxing Day Sales	05/05/2018	13.3	Early May Bank Holiday
17/11/2017	18.4	Christmas (Fri)	29/09/2018	14.2	Food & Drink Festival
27/10/2017	18.8	Halloween Half Term	30/11/2018	16.4	Christmas (Thu)
06/12/2017	19.0	Christmas (Wed)	08/12/2018	16.4	Christmas (Sat)
11/11/2017	19.9	Christmas (Sat)	22/09/2018	18.3	Freshers Week
25/10/2017	20.3	Halloween Half Term	26/10/2018	19.3	Halloween Half Term
18/11/2017	20.3	Christmas (Sat)	28/04/2018	20.6	N/A
11/12/2017	22.3	Christmas (Mon)	14/07/2018	22.0	England World Cup
16/12/2017	23.0	Christmas (Sat)	13/10/2018	23.3	N/A
30/09/2017	24.8	Food & Drink Festival	14/12/2018	23.7	Christmas (Fri)
28/12/2017	25.2	Boxing Day Sales	26/05/2018	25.5	Late May Bank Holiday

Edinburgh (High Street/Old Town) micro-location

Date	Rank	Event	Date	Rank	Event
13/08/2017	1.5	Fringe Festival	25/08/2018	2.4	Fringe Festival
09/08/2017	4.4	Fringe Festival	17/08/2018	6.1	Fringe Festival
10/08/2017	5.8	Fringe Festival	10/08/2018	11.0	Fringe Festival
12/08/2017	8.8	Fringe Festival	24/08/2018	11.7	Fringe Festival
05/08/2017	10.2	Fringe Festival	09/08/2018	13.3	Fringe Festival
30/12/2017	14.3	New Year's Eve	08/08/2018	16.9	Fringe Festival
19/08/2017	14.6	Fringe Festival	04/08/2018	17.2	Fringe Festival Bank Holiday
24/03/2017	16.2	N/A	20/08/2018	18.2	Fringe Festival
02/12/2017	17.4	Christmas (Sat)	22/08/2018	19.8	Fringe Festival
04/08/2017	17.5	Fringe Festival	23/08/2018	20.2	Fringe Festival
08/08/2017	20.4	Fringe Festival	06/10/2018	21.9	N/A
11/08/2017	21.9	Fringe Festival	03/08/2018	23.1	Fringe Festival
25/08/2017	23.9	Fringe Festival	13/08/2018	23.1	Fringe Festival
15/08/2017	26.8	Fringe Festival	22/09/2018	24.7	N/A
07/08/2017	27.7	Fringe Festival	15/08/2018	26.7	Fringe Festival
25/11/2017	29.0	Christmas (Sat)	31/12/2018	27.2	New Year's Eve
27/10/2017	30.3	Halloween	07/08/2018	28.2	Fringe Festival
31/12/2017	31.0	New Year's Eve	24/11/2018	28.4	Christmas (Sat)
16/12/2017	33.7	Christmas (Sat)	12/08/2018	28.7	Fringe Festival
27/01/2017	34.5	N/A	27/08/2018	29.5	Fringe Festival

Edinburgh (New Town) micro-location

Date	Rank	Event	Date	Rank	Event
09/12/2017	2.9	Christmas (Sat)	08/12/2018	4.0	Christmas (Sat)
02/12/2017	4.0	Christmas (Sat)	01/12/2018	4.4	Christmas (Sat)
16/12/2017	5.7	Christmas (Sat)	22/12/2018	12.3	Christmas (Sat)
31/12/2017	9.0	New Year’s Eve	15/12/2018	14.2	Christmas (Sat)
30/12/2017	10.7	New Year’s Eve	14/12/2018	14.4	Christmas (Fri)
22/12/2017	11.3	Christmas (Fri)	24/11/2018	15.5	Christmas (Sat)
15/12/2017	13.4	Christmas (Fri)	27/12/2018	19.9	Boxing Day Sales
18/11/2017	14.1	Christmas (Sat)	11/08/2018	20.7	Fringe Festival
23/12/2017	15.5	Christmas (Sat)	28/12/2018	23.6	Boxing Day Sales
21/12/2017	18.2	Christmas (Thu)	29/12/2018	24.0	Boxing Day Sales
18/12/2017	22.8	Christmas (Mon)	17/11/2018	24.9	Christmas (Sat)
08/12/2017	23.1	Christmas (Fri)	25/08/2018	29.1	Fringe Festival
28/12/2017	25.2	Boxing Day Sales	21/12/2018	29.1	Christmas (Fri)
20/12/2017	25.6	Christmas (Wed)	16/12/2018	33.3	Christmas (Sun)
24/11/2017	25.9	Black Friday	17/12/2018	34.5	Christmas (Mon)
17/12/2017	28.9	Christmas (Sun)	31/05/2018	41.9	N/A
19/12/2017	30.2	Christmas (Tue)	23/11/2018	42.3	Black Friday
17/11/2017	37.5	Christmas (Fri)	23/12/2018	42.3	Christmas (Sun)
29/12/2017	37.5	Boxing Day Sales	24/02/2018	42.6	Six Nations
19/11/2017	38.8	Christmas (Sun)	30/06/2018	43.7	N/A

Appendix 6.1 – Specifics for derivation of micro-locational characteristics

Specifics for the derivation of some of the footfall descriptors, informed by the conditions in Dolega, Pavlis and Singleton (2016) and adapted from those given in Appendix 4.1

Value stores

Store Name:	Aldi, Lidl, Iceland, Primark, Farmfoods, Poundland, Poundstretcher, Home Bargains, Savers, B&M Bargains, Pound Bakery
Category:	Discount & Surplus Stores, Charity And Secondhand Shops, Pawnbroking And Cheque Cashing
Subcategory:	Bookmakers.

Night-time economy locations

Category:	Bars Pubs And Clubs, Off Licenses And Restaurants
Subcategory:	Fast Food Takeaway, Take Away Food Shops, Fish And Chip Shops, Pizza Takeaway, Chinese Fast Food Takeaway, Indian Takeaway, Fast Food Delivery, Amusement Parks & Arcades, Theatres & Concert Halls, Cinemas, Snooker, Billiards & Pool Halls, Bowling Alleys

Anchor Stores

Store name:	Tesco (excluding Express and in store services), Sainsburys (excluding Local and in store services), Morrisons (excluding in store services), ASDA (excluding in store services), John Lewis, Debenhams, Marks & Spencer, Harvey Nichols, H&M, Primark, Zara, Boots, Next, and B&Q

Premium Stores

Store name:	Waitrose, Little Waitrose, John Lewis, Laura Ashley, Ted Baker, Tommy Hilfiger, Fat Face, Superdry, Seasalt, Jack Wills, White Stuff, Crew Clothing, Boss, Cath Kidston, Joules, Swarovski, Lacoste, Diesel, Apple Store, Bose, Hotel Chocolat, Radley, Karen Millan, Michael Kors, The White Company, Reiss, All Saints, Tessuti, Flannels, Ralph Lauren, Kate Spade, Mulberry, Burberry, Armani, Calvin Klein, Coach, Dune, Diesel, Fossil, Fred Perry, French Connection, Guess, Hobbs, Karl Lagerfeld, Kurt Geiger, L'Occitane, Lacoste, Levi, Lindt, Osprey, Swarovski, Timberland and Toms

Entertainment activities

Subcategory:	Cinemas, Theatres & Concert Halls, Amusement Parks & Arcades, Museums & Art Galleries, Sports Grounds & Stadiums, Tourist Attractions, Party Venues & Function Rooms, Bingo Halls, Bowling Alleys, Ticket Outlets & Box Offices, Golf Courses, Snooker, Billiards & Pool Halls, Driving Ranges, Ice Rinks, Booking Agents, Paintball & Combat Games and Karting

Appendix 6.2 – Specifics for derivation of holiday days

Holidays identified in the footfall dataset

- Bank Holidays
    - Easter, Good Friday, Easter Monday, New Years Day, Christmas Eve, Christmas Day, Boxing Day, May Day, Summer Bank holiday
- Pre-Christmas Footfall
    - Every Saturday from and including Black Friday until Christmas

Appendix 6.3 – Correlation between percentage error by location and spatial characteristics

Variable	Value	P-Value
Transport Hubs	-0.10	0.26
Premium Stores	-0.10	0.24
Entertainment	-0.04	0.65
Anchor Stores	-0.09	0.29
Workplace Population	-0.02	0.85
Betweenness/Centrality	-0.05	0.56
Population	0.10	0.28
Vacant Stores	-0.08	0.34
Independent Stores	-0.04	0.68
Value Stores	-0.06	0.49
Density of Stores	0.19	0.03
Night-Time Economy locations	0.04	0.62
Car Parks	-0.02	0.83