**UK Footfall Index**

This document introduces a index showing the percentage change in visitors – or footfall (FF) – to retail environments in the United Kingdom between two different periods of time (months, weeks, days, etc.). The FF index captures major seasonal changes, such as the end of the summer or the beginning of a new year, where people tend to spend less time in retail areas, especially when compared with the Christmas period for example. The UK FF index is not designed to isolate and observe local changes, or particular retail area characteristics, but rather it is designed to be representative of what is happening more generally in any given town in the UK.

Introduction

* Context: incl. what is FF.

The accurate measurement and estimation of human activity is one of the first steps towards understanding the structure of the urban environment. Human activities are highly granular and dynamic in both the spatial and temporal dimensions and estimating them with confidence is crucial for decision-making in numerous applications such as urban management, retail, transport planning and emergency management. Of particularly interest are the people visiting or passing by retail environments, also known as Footfall (FF), due … comprehensive research into the patterns of retail activity in UK high streets

In this context, the SmartStreetSensor project is one of the most comprehensive study carried out on consumer volume and characteristics in retail areas across UK. The project has been organised as a collaboration between Local Data Company (LDC) and Consumer Data Research Centre, University College London (CDRC, UCL). The data for the study is generated independently within the project through sensors installed at around 900 locations across UK and contain details of passive Wi-Fi signal probing from a sensor installed at each of these locations. These data are used as a proxy for estimating FF at retail locations. The potentially identifiable information collected on the mobile devices is hashed at sensor level and the data is sent to the central server via an encrypted channel for storage.

The dataset comprises daily five-minute aggregated footfall counts. The role out of sensors and data collection began in July 2015 and there are now 960 (Dec 2017) sensors in operation distributed in 94 cities across Great Britain:

* Aims and objectives
* About the dataset

# **Scale and Extent**

|  |  |
| --- | --- |
| **Field** | **Value** |
| Data Provider | Local Data Company/CDRC |
| Analytical Units | Postcode |
| Data Format | Postgres DB |
| Temporal Extent | July 2015 - present |
| Geographical Extent | Great Britain |
| Variables | 26 |
| Observations | 4.9 billion records (December 2017) |

# **Citation Information**

The following statement should be included when citing the use of this dataset:

“The data for this research has been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID CDRC [Project Number], ES/L011840/1; ES/L011891/1”

# **Data Classification and Access Summary**

The hashed data are classified as Controlled data and held within the Secure Service. A cleaned CDRC data product available per sensor is available through the Safeguarded Service. Both are available only upon approved application. To make an initial application, please visit: <https://www.cdrc.ac.uk/data-services/using-our-data/>

# **Content**

Field level metadata is provided in the table overleaf.Data are provided in respect of:

* 960 locations identified at address (building number, street, unit postcode)
* General information about the technical and physical restrictions of each sensor and their locations
* Anonymised hashed MAC addresses of detected devices captured by 960 sensors (Controlled data only)
* Signal strength and number of packets of detected devices (Controlled data only)
* All figures reported are verified up to Dec 2017

**Novelty**

The data for the study is generated independently and it is the first unique and comprehensive research into national footfall patterns. These data are one of the most comprehensive sets comprising consumer volume and characteristics in retail areas across Great Britain and it can provide value to researchers, occupiers, landlords, local authorities, investors and consumers within the retail industry.

These data can be extended not only to detect retail activity but to measure all the activity around the sensors and can be linked to transport, work zones and demographic data, etc., to produce, for example, novel functional areas classifications.

**Quality**

The quality of the data are affected by a series of technical limitations relating to the WiFi acquisition process:

1. The range of the sensor: Since the strength of the signal from a mobile device to the WiFi access point varies depending on a variety of factors, the sensors do not have a standard signal range. In other words, the exact delineation of the signal range is different for each sensor.
2. Probe request frequency: This feature varies widely based on the manufacturer, operating system, state of the mobile device and the number of access points already known to the device.
3. MAC address collisions: There are few instances (∼ 0.01%) of MAC address collisions reported where a mobile device known to be in some place is reported elsewhere. This might be due to rogue MAC randomisation by certain mobile devices and the hashing procedure being carried out at two different stages.
4. Human error: These devices are installed at retail points and they may be disconnected from the main power from time to time, resulting in missing data periods.
5. Postprocessing: The process to transform probe request into actual footfall requires a series of assumptions that can potentially lead to over/under counts in the results. The methodology followed for the data delivered is explained in the next section.

**Representation and Bias**

|  |  |
| --- | --- |
| City | Number of locations |
| London | 315 |
| Edinburgh | 39 |
| Wakefield | 29 |
| Manchester | 26 |
| Glasgow | 24 |
| Nottingham | 24 |
| Brighton | 22 |
| Leeds | 21 |
| Kingston Upon Thames | 21 |
| Aberdeen | 20 |
| Liverpool | 20 |
| Gloucester | 18 |
| Reading | 18 |
| Bristol | 17 |
| Cardiff | 14 |
| Chester | 14 |
| Norwich | 14 |
| Market Harborough | 12 |
| Birmingham | 12 |
| Oxford | 12 |
| Southampton | 12 |
| Bromley | 11 |
| Cambridge | 11 |
| Salisbury | 11 |
| Newcastle Upon Tyne | 10 |
| Sheffield | 10 |
| Leicester | 8 |
| Peterhead | 8 |
| Plymouth | 8 |
| Blackpool | 7 |
| Croydon | 7 |
| Leamington Spa | 7 |
| Dorchester | 6 |
| Durham | 6 |
| Gainsborough | 6 |
| Hull | 6 |
| Northallerton | 6 |
| Orpington | 6 |
| Taunton | 6 |
| York | 6 |
| Beckenham | 5 |
| South Queensferry | 5 |
| Bradford | 4 |
| Chelmsford | 4 |
| Coventry | 4 |
| Devizes | 4 |
| Lutterworth | 4 |
| Lymington | 4 |
| Stirling | 4 |
| Watford | 4 |
| Bedale | 3 |
| Gateshead | 3 |
| Hove | 3 |
| Market Rasen | 3 |
| Solihull | 3 |
| Thirsk | 3 |
| Bolton | 2 |
| Boston | 2 |
| Dover | 2 |
| Harrow | 2 |
| Hertford | 2 |
| Ilford | 2 |
| Oldham | 2 |
| Romford | 2 |
| Sale | 2 |
| Twickenham | 2 |
| Windsor | 2 |
| Basingstoke | 1 |
| Bexleyheath | 1 |
| Blackburn | 1 |
| Bognor Regis | 1 |
| Brentford | 1 |
| Burton Upon Trent | 1 |
| Bury | 1 |
| Canvey Island | 1 |
| Clacton-On-Sea | 1 |
| Dartford | 1 |
| Derby | 1 |
| Exeter | 1 |
| Grays | 1 |
| Greenhithe | 1 |
| Ipswich | 1 |
| Knutsford | 1 |
| Maidenhead | 1 |
| Mansfield | 1 |
| Mold | 1 |
| Otley | 1 |
| Richmond Upon Thames | 1 |
| Smethwick | 1 |
| Southend-On-Sea | 1 |
| Staines | 1 |
| Stevenage | 1 |

The date of installation, number of active sensors at any given time and their location can be supplied on request for a particular area.

A third of the locations are in Greater London, particularly in London’s central area, which makes any national aggregated count biased towards this region.

Before July 2016, the number of sensors was limited (no more than 200) and most of these were located in London.

The sensor cannot distinguish between a mobile device and any other WiFi enabled device so the controlled data includes printers and routers for example.

*Production of cleaned footfall estimates*

The probe request detected from a device does not have a one-to-one correspondence with an individual so the initial MAC address detected at each location must go through a cleaning and validation procedure as detailed below.

a. Input: Hashed data. These are the number of probe requests (packets) per MAC address, per sensor during a five-minute period.

b. Mac lookups: Separate probe requests into private and public addresses using MAC lookup table comprising the current complete list from IEEE of the most recognised phone manufacturers.

c. Exclusion of long dwellers.

1. Probe requests from the same mobile device detected during a consecutive time period of 10 minutes are counted only once. For example, a printer or the mobile devices owned by store employees are included only once in a day.
2. We limit the number of times a probe request can appear in a day to 4 times. For example, if a device is detected in a single location for a 5-minutes period at six different points in time in a day, we only count it 4 times.

d. Imputation: Where there are missing data, we fill the gaps using the following methods:

1. Linear interpolation of gaps of maximum 5 minute length. The reference values for interpolation are taken as an average of up to 3 values (15 minutes) on each side of the gap to smooth out any irregularities on the five minute level of the data.
2. Historical imputation, taking a single value from the closest preceding week with data present for the corresponding period.
3. For all remaining gaps, of maximum 1 hour length perform linear interpolation. The reference values for interpolation are taken as an average of up to 9 values (45 minutes) on each side of the gap to smooth out any irregularities on the five minute level of the data).

e. Finally, we aggregate the remaining probes in five-minute packages.

*Manual validation*

For each sensor, a manual verification of the counts has been conducted for specific day/time periods to calibrate the counts against observed pedestrians passing by the premises of the sensor. The ratio manual counts / sensor counts produces an adjustment factor for each sensor. This factor is reported along with the unadjusted counts.

*Potential bias in the footfall counts*

1. MAC address look-up. We do not separate probe requests into private and public addresses.
2. We have not conducted any type of iOS9 or iOS11 modelling to account for randomised IP addresses. A methodology to do this is being investigated by LDC with CDRC and may be available in future updates.

Findings

Footfall index

The FF index is calculated as follows:

(1)

where b = Total footfall at period *b*, a = Total footfall at period *a*, *a≠b,* and *n* is an integer representing the distance between b and a, for example, if a=July 2016 and b=July 2017, n=12.

Quantities *b* and *a* depend on the aggregated FF counts at each location. There is a great heterogeneity in spatial-temporal distribution: we began with 9 locations in July 2015, a figure which rose to 791 by the end of January 2018. To accommodate these variations, a system of weights is applied to each location which enables making *b* and *a* statistically comparable to each other. The weighting procedure is explained in Appendix A. Along with the weighted system, the FF index also accounts for possible duplicated counts which may be generated by sensors in close proximity to each other. This is explained in more detail in section 3.3.1.

1. Monthly index

In the early months of the project, the number of cities and locations of the sensors were very limited. We initially only had sensors in 2 urban locations (London and Market Harborough), so the FF index recorded during this period cannot be considered to be representative of the whole country. However, by July 2016 the number of urban locations with sensors had reached 52. As this numerical increase was coupled with an increase in the diversity of locations, we have chosen to begin our analysis of the FF index from this point. Therefore, Figure 1 shows the FF index on a monthly aggregated basis for the period of July 2016 to January 2018, and in it we can clearly observe different trends across the year. For instance, the large drop in FF after the Christmas period is very apparent; conversely, the variance in FF over the summer months (July-August) is quite small (-2%).

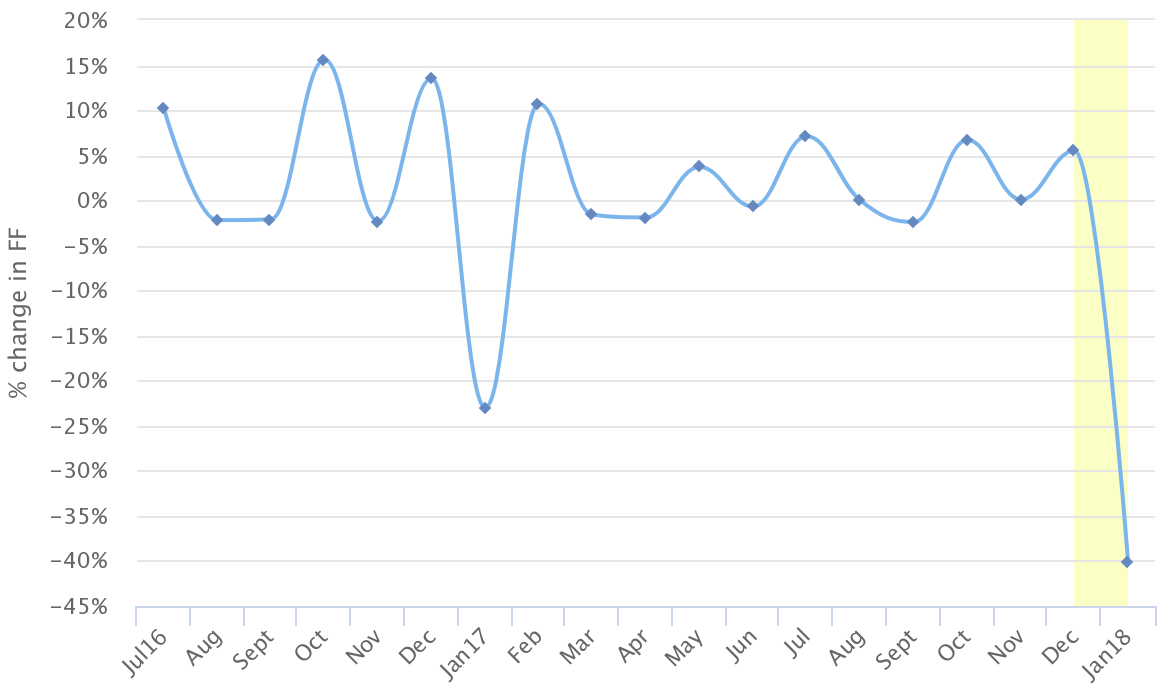
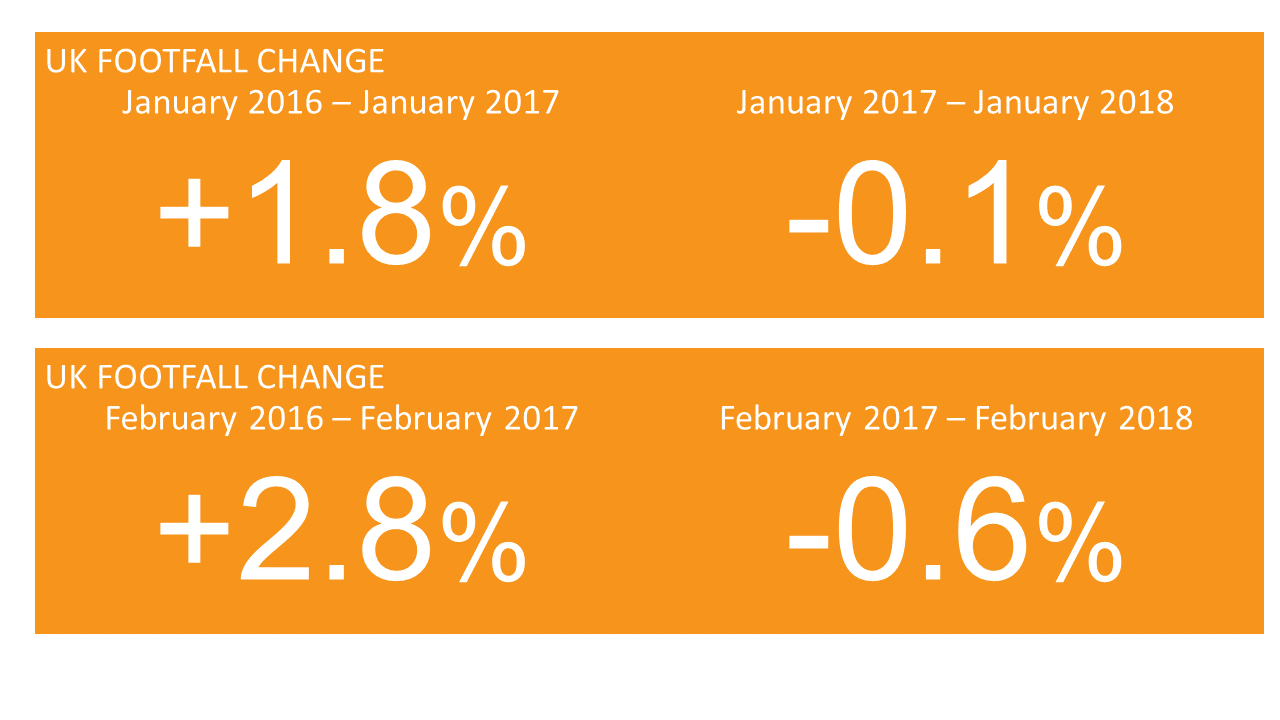


Figure 1. Monthly FF index. We can observe that there is a considerable reduction in footfall in the periods of Dec-Jan 2017 and Dec-Jan 2018 (-23% and -40% respectively), while the remaining months do not demonstrate changes larger than 15.5%.

1. Daily index

Figure 2 shows the FF index on a daily scale, measuring, for example, the difference between any given Sunday-Monday. The index clearly demonstrates the large change in FF between 25th and 26th of December when stores across the UK reopen on Boxing Day. Interestingly enough, the turnover of people on those days was larger in 2016 than in 2017. With this daily scale, we can detect the expected circadian rhythms found in urban areas: during weekdays the change in FF is relatively stable (most of the points in Figure 2 are around 0-25%), whereas the larger positive/negative changes correspond to Saturday-Sunday-Monday (the points oscillate between ±40%).



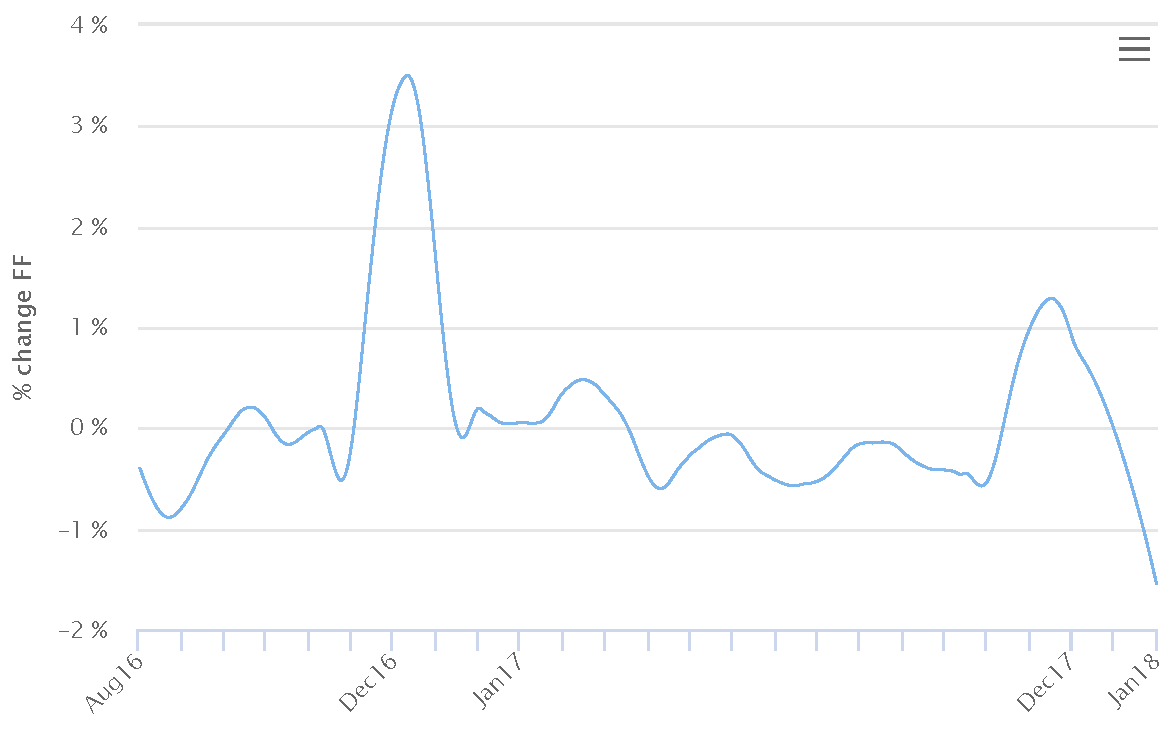


Figure 3. Smooth representation of Figure 2.

With the above smooth representation, we can more clearly see the large variance in FF between December and the rest of the year, as well as the consistency in FF between January–December 2017. This index also captures the decrease in FF in December 2016/2017 when compared with December 2015/2016.

The decrease in FF at the end of 2017 when compared with 2016 can be seen in more detail in Figure 4:

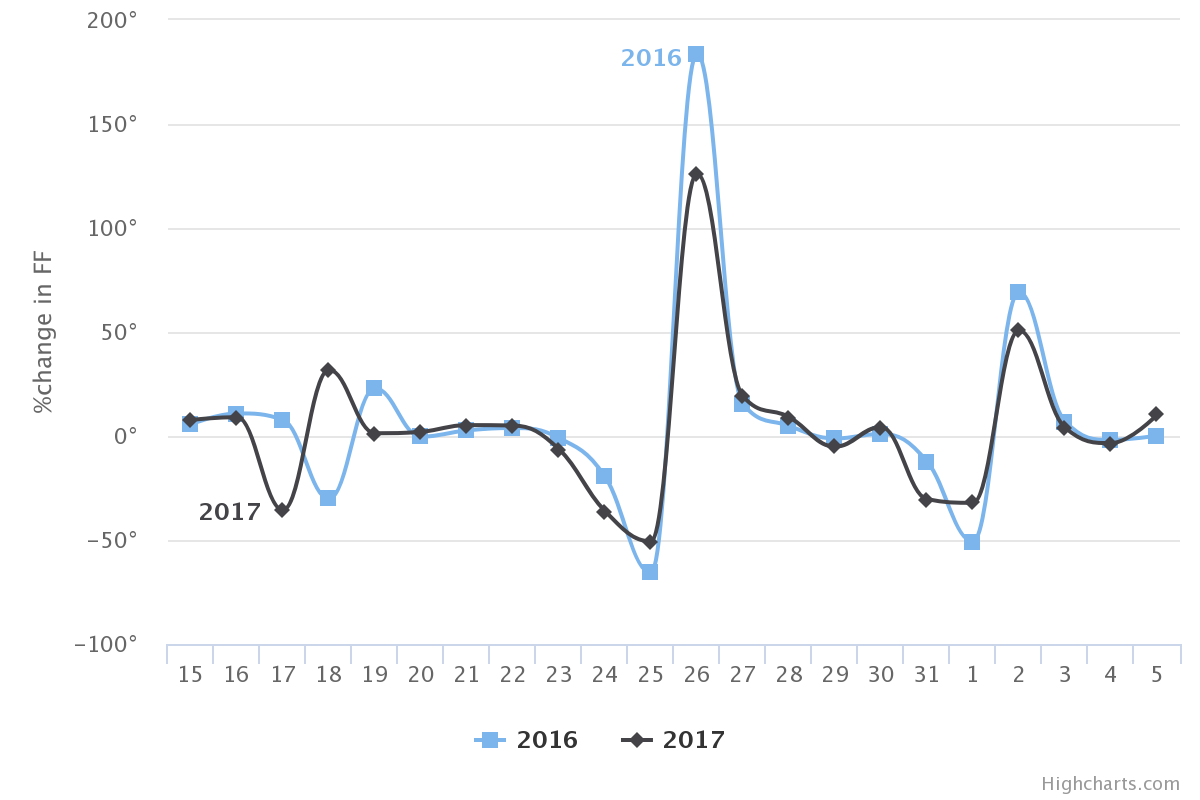


Figure 4. Comparison of the daily change in FF from 15th December 2016 to 5th January for the years 2016 and 2017. The 18th illustrates the difference between a Monday (2017) and a Sunday (2016). The 25th exhibits a similar drop in FF in both years, regardless of the fact that it was a Monday in 2017.

When comparing December 2016 with December 2017, we found a total FF change of -23%.

1. Representativeness

The fundamental question about this index is how representative it is for any retail centre in the UK and whether it contains any biases towards locations, types of retail areas, and other factors? We will address these questions below.

* 1. Locations

Since July 2016, 20 urban locations have accounted for 81% of the total FF in any given month, with London constantly contributing ~27% of that total. However, in our index, the bias towards these urban locations is compensated for by the weighted system explained in Appendix A. For example, a typical distribution of each sensor’s contribution to the total FF is shown in Figure 5. The bulk of the distribution – 80% of locations – is between 0-0.25%, which confirms that the FF index is not capturing the flow of people around any particular location.

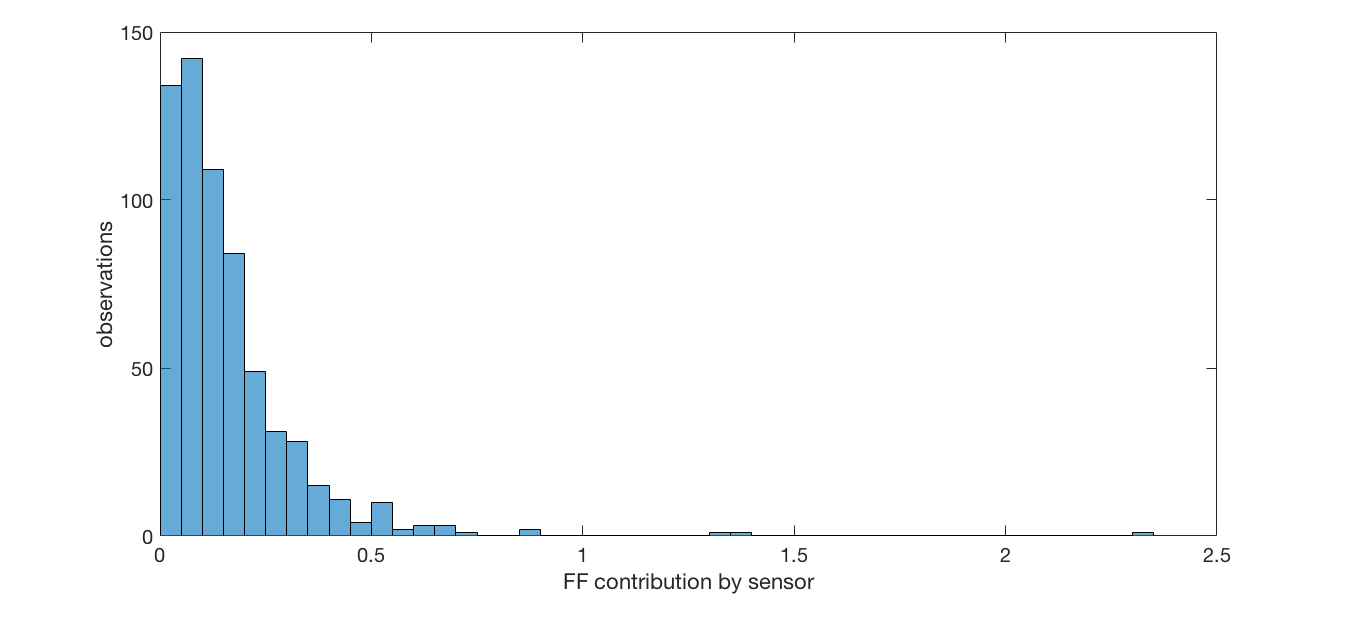


Figure 5. The probability distribution of the contribution to the total FF from each of the 631 locations operating in September 2017. As can be observed, only three sensors are contributing slightly more than 1% of the total.

* 1. Type of street

Once we established that the distribution of sensor locations was not causing undue bias, we then queried their position in respect of the type of street they are positioned on: i.e. are these sensors predominantly located in high streets, retail parks, shopping centres, etc.? Using the full address of the location of the sensors, we classified each location with its street type as derived from the OpenStreetMap highway tag definition.[[1]](#footnote-1) We found that the sensors are located in seven different types of areas: pedestrian/residential (55% of the locations), primary (16%), secondary (5%), tertiary (7%), service (9%), trunk (3%), and unclassified (5%). Consequently, the FF index is dominated by the pedestrian/residential street type (see Figure 6).

That said, not all pedestrian/residential streets are equal. For instance, sensors installed at shopping centres and retail parks are also classified as being located in pedestrian/residential streets, this is because people’s movements in shopping centres occur in a constrained environment with limited or no motor vehicles circulating. From January 2017 to date, 60% of the sensors are in pedestrian/residential locations. Within that figure, 20% of the sensors (approximately 90) are located in various shopping centres, and contribute just 11% of the total FF counted in January 2017. In short, our index is fundamentally measuring the FF on pedestrian/residential streets as opposed to shopping centres. In particular, our FF index is a measure of people walking on suburban and pedestrianised high streets.

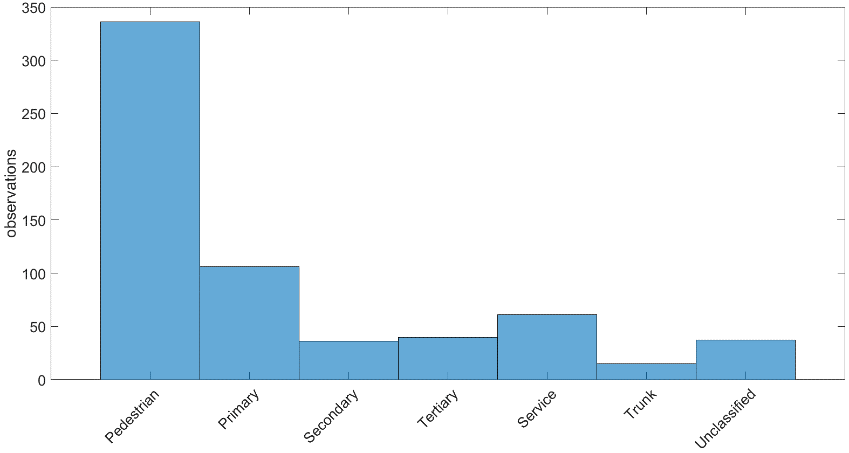


Figure 6. September 2017 probability distribution by type of street. The strong bias towards the pedestrian/residential street type is evident: 55% of sensors are located here, whereas the trunk type of location only contains 1% of sensors.

An interesting finding from this street type analysis is that each type of street has its own characteristic FF signal (FF as a function of time) and this signal does not depend on the type of business. In fact, shops of the same type can generate a whole variety of signals and values. From this, we inferred that the sensors are not capturing the FF generated by any particular business, but instead are capturing the FF produced more broadly by the characteristics of the areas in which the sensors are installed.

* 1. Type of shop

Finally, we explored if our index is biased towards a particular type of business. From July 2017 to January 2018, 105 shop types have been included in the sensor network, although not all shop types are present at each month. In fact, only 45% of shop types are included in the total 19 months, rising to 77% of shop types from March 2017. Consequently, the contribution of each shop type to the overall index varies from month to month, as it depends on the number of shops of that particular type (these counts are weighted as explained above). Before showing the actual FF per type of shop, we first need to address a bias in the counts generated.

* + 1. Nearest Neighbours sensors (NN)

We observed that when two (or more) sensors are in close proximity to one another (40 meters or less) they essentially measure the same local FF. In our current network, there are 264 pairs of sensors (representing 239 unique sensors) in this situation, distributed over 52 different urban locations with London having the greatest number of NN sensors (84). A special case is the Ridings Shopping Centre (Wakefield), where 20 of the 33 sensors installed are in close proximity to one another.

For these 264 pairs, we averaged the hourly counts from each sensor to obtain a single measure (we refer the interested reader to Appendix B for the details about this process). For example, sensors 547 and 276 are located in Bridge St, Chester, 22.3m apart from one another. As can be observed from Figure 7, the daily mean FF is positively correlated and has quite similar values, even though each shop is in a completely different retail category. We therefore assigned a new ID to the averaged signal and classified it as the category ‘Merged’, being sure to also remove the original FF counts for the merged sensors.

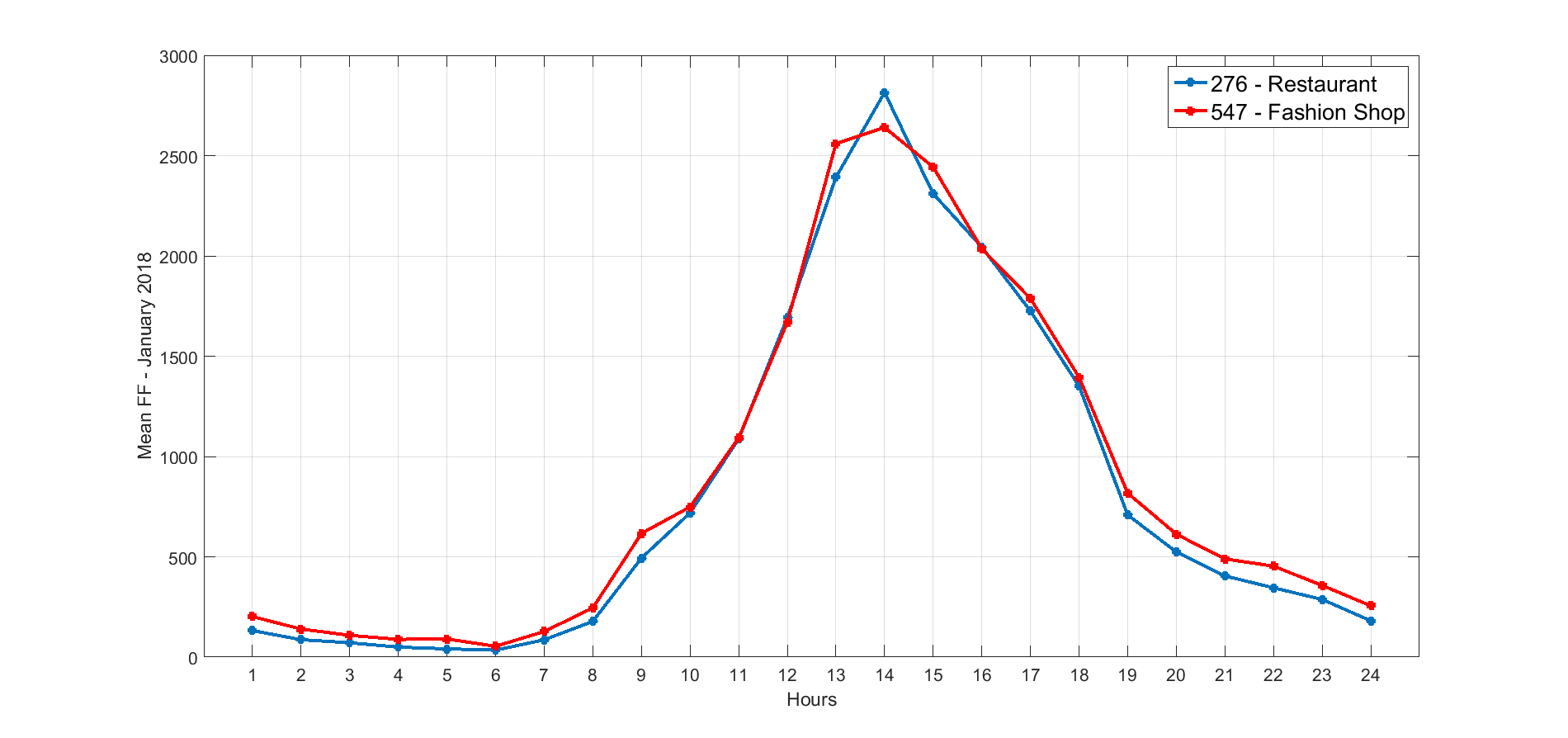


Figure 7. FF signal for a Restaurant and a Fashion shop located on a pedestrian street. The correlation matrix between these two variables has correlation coefficients of 0.99 [ref] reinforcing the evident interaction between these locations.

Figure 8 shows the probability distribution for each type of shop in terms of its particular contribution towards the FF index in each month since July 2016. All months share the same lognormal-like distribution, where the bulk is concentrated in the lower contributors. In general, between 70-90% of the total FF is generated by 95% of the shops, and only 7 types (Merged, Mobile Phones, Restaurant, Charity, Chocolatiers, Fashion, and Sports) account for the 10-30% of the total. The heterogeneity of shop types is an indicator that, as in the case of locations, these sensors are not capturing the FF outside a particular type of shop per se, but rather can be viewed as representative of a vast range of retail businesses.

The prevalence of the Merged category in 12 of the 19 months in Figure 8 provides another indication of how the FF measured is not tied to a particular type of shop, but rather to a cluster of types and/or the particular characteristics of a given retail area. If we break up the Merged category into its original types as in Figure 9, we still have the same type of distribution, but the location of the specific categories differ slightly. For instance, the category of shoe shops appear in the last three months in Figure 9, where originally there was only Merged (Figure 8). Understanding the reasons for this, and the precise evolution of the top contributors from month to month, could help us to more fully understand some of the particular variations of each location over and above the quantity of sensors in the retail location. This analysis is beyond the scope of this present work.

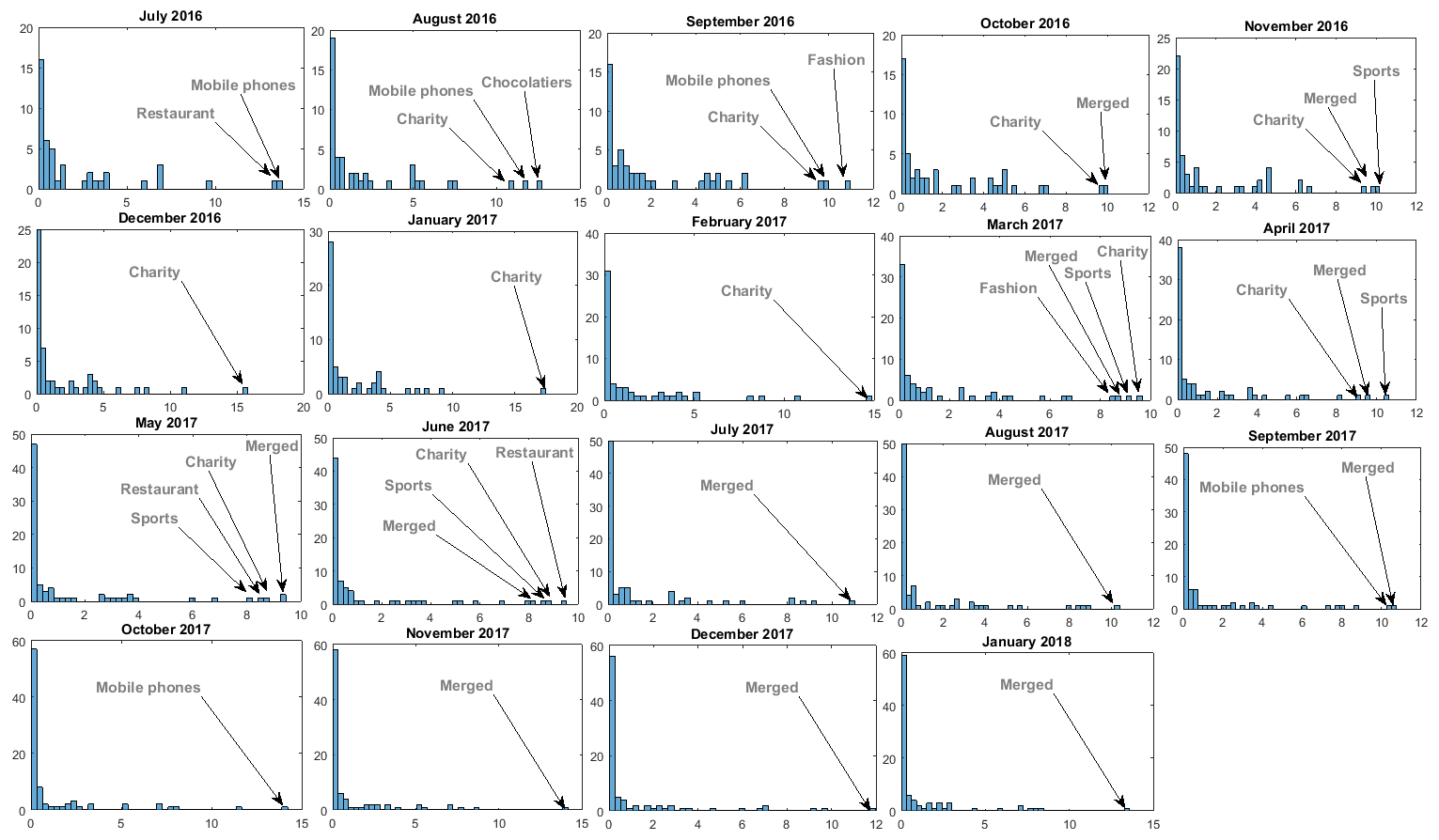


Figure 8. Distribution of FF by type of shop with the ‘Merged’ category.

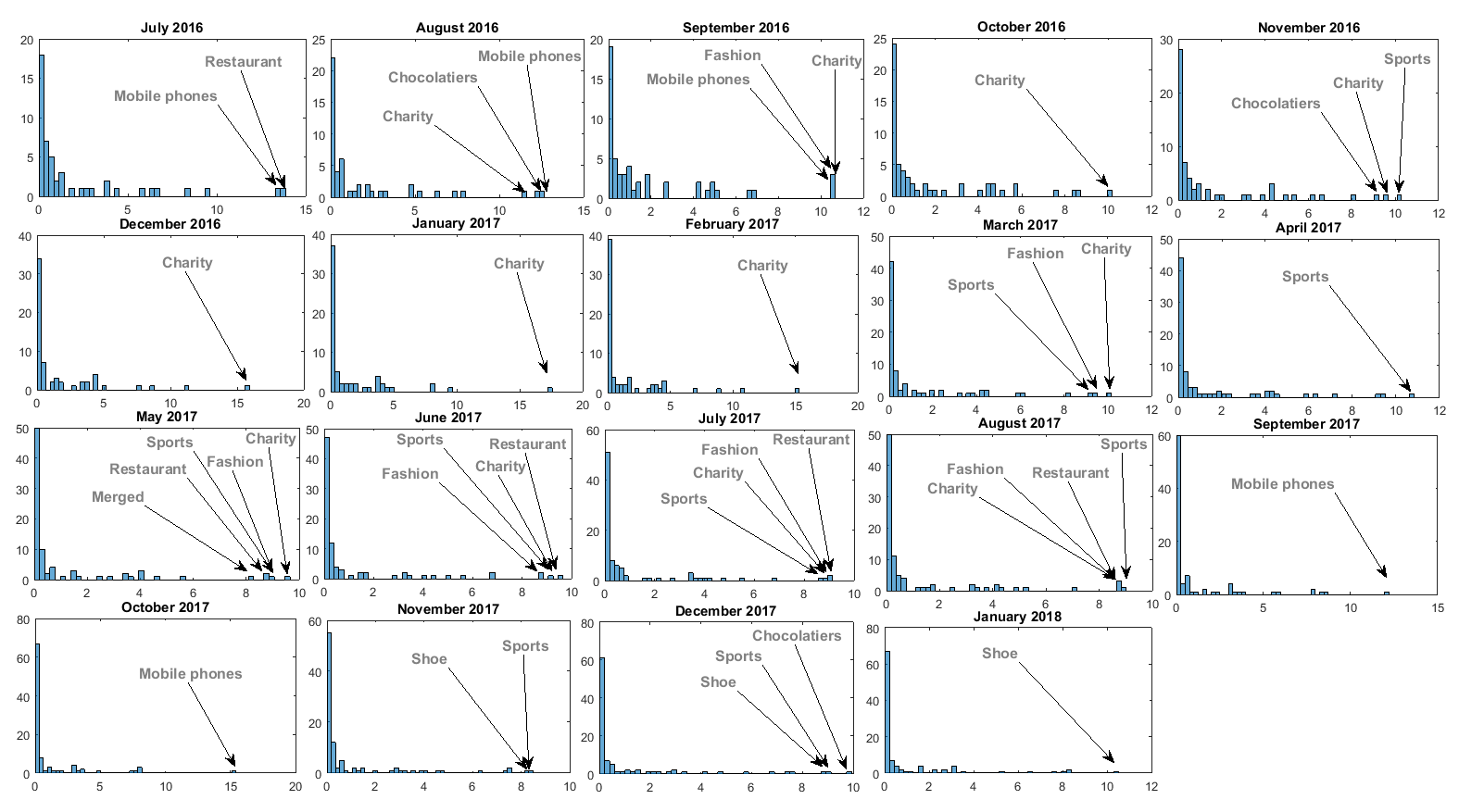


Figure 9. Distribution of FF type of shop without the ‘Merged’ category.

Summary

* Sum up
* Recommendations

Appendices

1. https://wiki.openstreetmap.org/wiki/Key:highway. [↑](#footnote-ref-1)