# UK Footfall index

This index shows the percentage change in visitors to retail environments or footfall (FF) in the United Kingdom between two different periods of time (for example, consecutive months or days). The FF data is provided by the Local Data Company and covers a period from July 2015 to date and it refers to the number of people walking in front or into a retail unit.

As a global measure, the FF index is capturing major seasonal changes, as the end of the summer or the beginning of the year, where people tends to spend less time in retail areas, compared with the Christmas period for example. It is not designed to pick local changes or particular retail areas characteristics, but to be representative of what is happening in any given town in the UK.

The proposed index is calculated as follows:

(1)

where b = Total weighted footfall at period *b*, a = Total weighted footfall at period *a*, a≠b and *n* is the time elapsed between b and a, measured in any convenient units (moths, days, etc.)

Quantities *b* and *a* depend on the aggregated FF counts at each location, which spatial-temporal distribution exhibits a great heterogeneity in number and composition. To accommodate these variations, a system of weights is applied to each location that allow us to make *b* and *a* statistically comparable to each other. The weighting procedure is explained in Appendix A. Along with this weighted system, the index takes care of possible duplicated counts generated by sensors in close proximity to each other. This is explained in section 3.3.1.

1. Monthly index

First, as in the early months of the project, the number of cities and locations were very limited (less than 10 cities in general and in July 2015 only 2 cities: London and Market Harborough), the index at those months was hardly representative for the whole country. By July 2016 the number of cities reached 52 and the diversity of locations is quite high, so we propose to start the index from that date. Figure 1 shows the FF index taking *b* and *a* as consecutive months. Throughout this period, we can observe different trends according to different seasons. The drop in FF after the Christmas period is evident and rather large while the change at summer time is considerably lower (-2%)

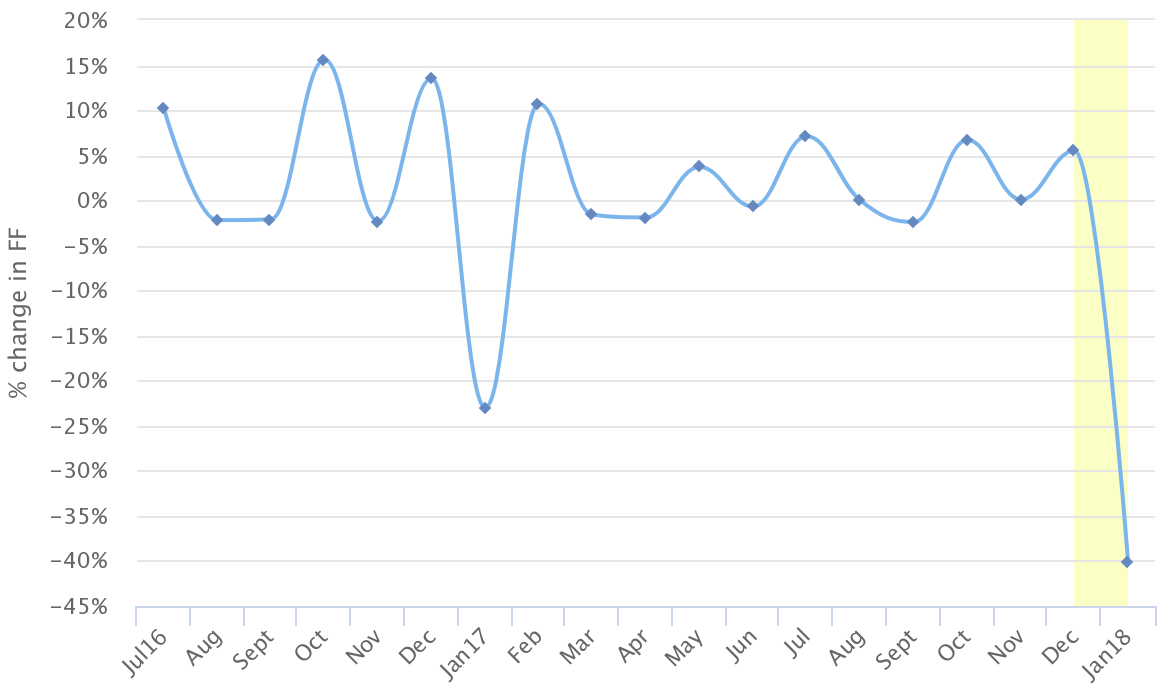


Figure 1. Monthly FF index. Each point represents the change in FF from the previous month. For example, the point Jul16 represents the variation between June 16 and July 16 (10% more). Between the months of Dec-Jan 2017 and 2018, the FF exhibits a quite large drop (-23% and -40% respectively) while in the rest of both years we don’t find changes larger than 15.5%

The FF index can be calculated taking as parameters not only consecutive months but taking a particular month as a base to track the evolution in FF during longer periods. For example, Figure 2 shows the FF index with October 2016 as the baseline. At the first point, we recovered the same value that in the point Nov in Figure 1. The months from March to June 2017 has practically no change in FF compare it with October 2016, while from July to December 2016 the change is only around +1%. There's a significant difference in the change between October 16- December 16 (~6%) and October 16-December 17 (~1%).

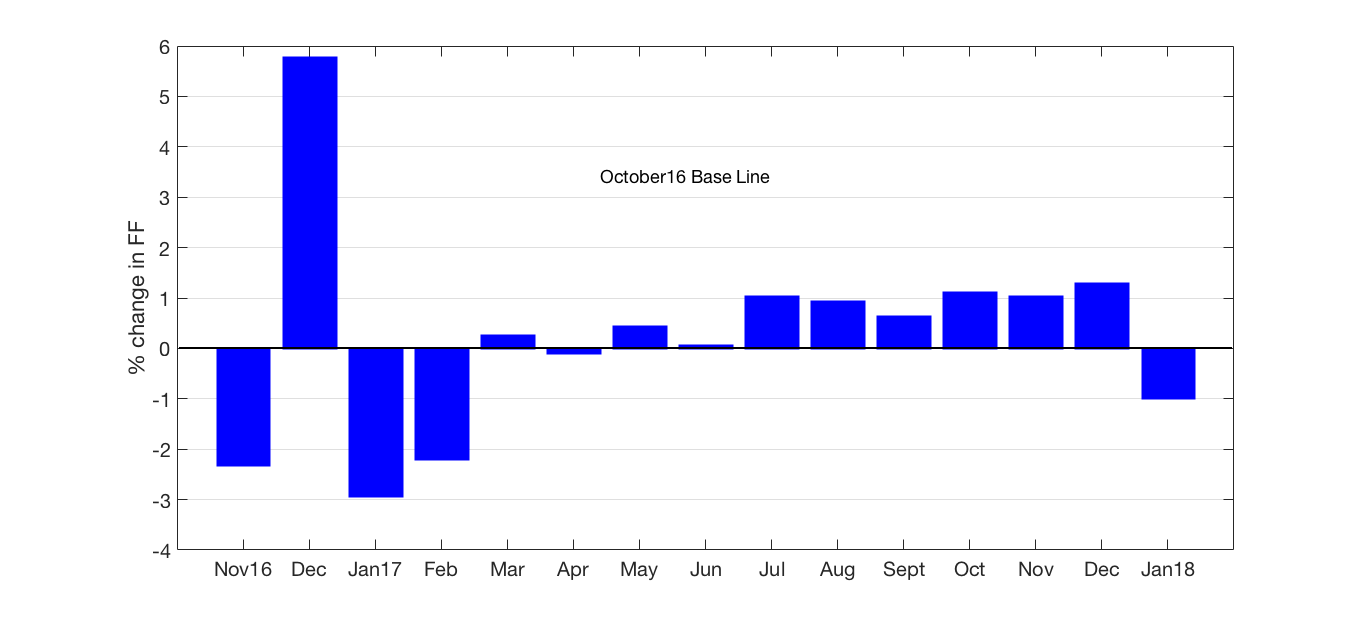


Figure 2. Monthly FF change index. Each bar represents the change between October 2016 and that particular month. As each measure refers to the same baseline, we can compare the area under each bar with the rest.

1. Daily index

Figure 3 shows the FF index on a daily scale, measuring, for example, the difference between any giving Sunday-Monday. The index is picking up the large change in FF between 25th and 26th of December when stores across the UK reopen after Christmas day. Interesting enough, the turnover of people in those days was larger in 2016 than in 2017. With this daily scale, we can detect the expected circadian rhythms found in urban areas, where during weekdays the change in FF is relatively stable (most of the points in Figure 3 around 0-25%) and the positive/negative changes corresponding to the changes between Saturday-Sunday-Monday (points oscillating between ±40%).

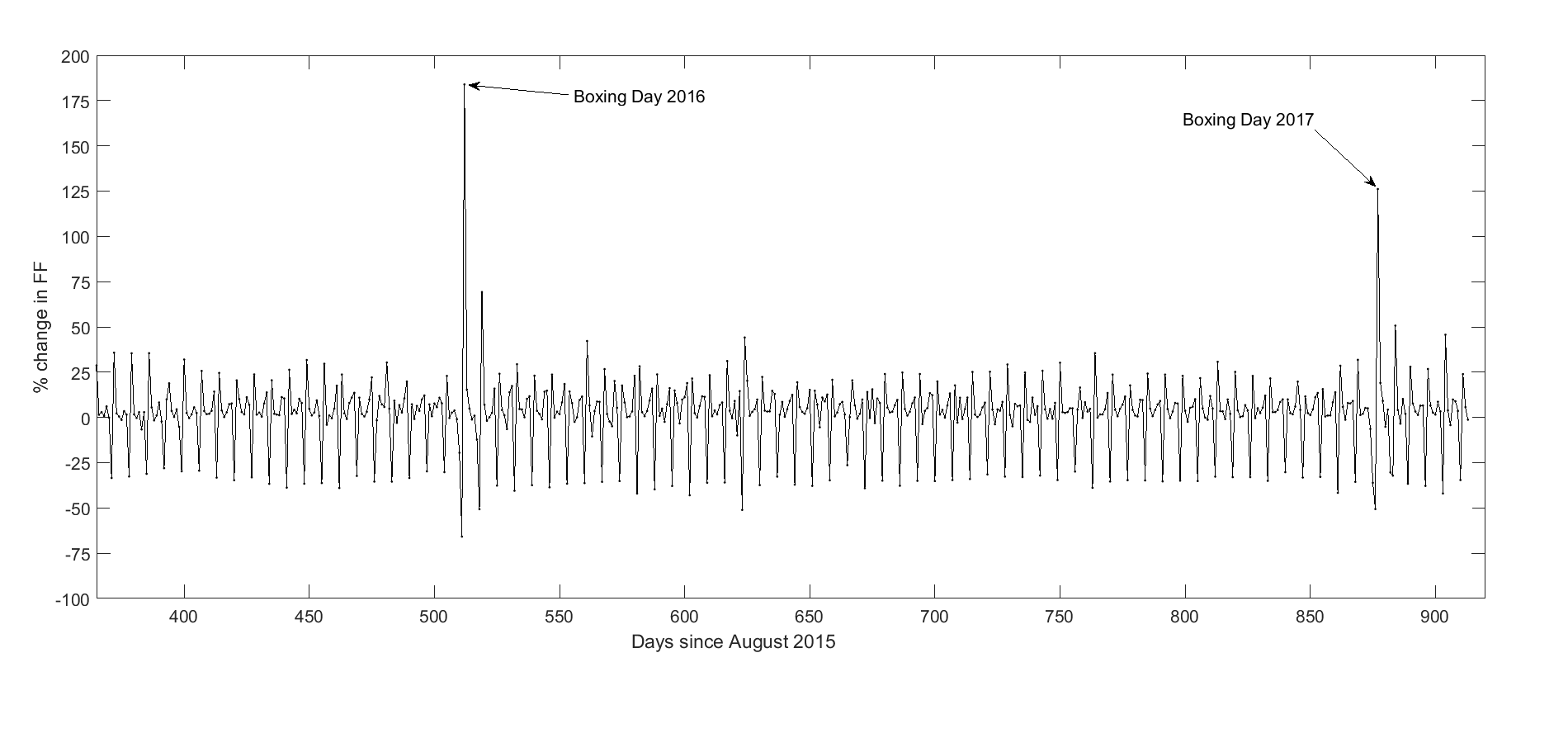
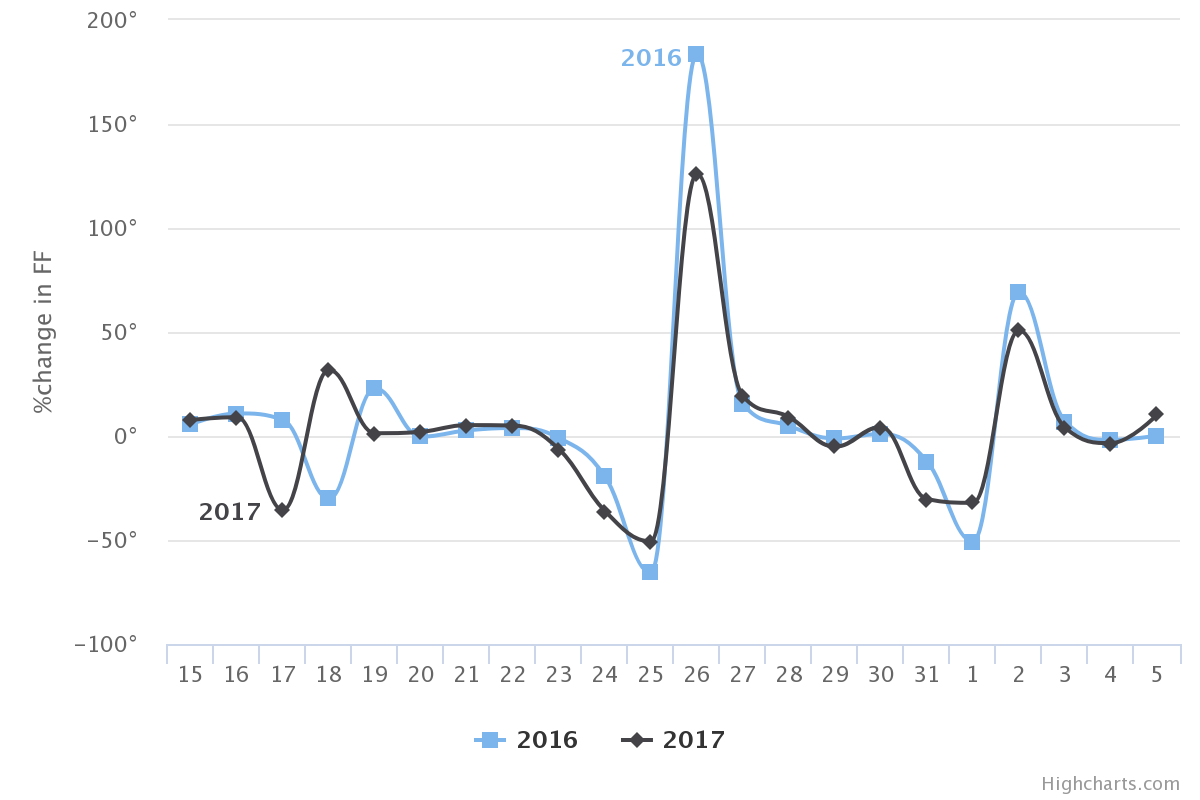


Figure 3. Daily FF index. The evident increase in FF at the 26th December 2017/2018 is a confirmation that with these index we can detect major shifts in FF throughout Great Britain.

The decrees in FF at the end of 2017 compared with 2016, can be seen in more detail in the next Figure:



Days

Figure 4. Comparative of the daily change in FF from December 15th to January 5th (2016-2017). We can see the differences between a Monday and a Sunday (18th) and how the 25th exhibits the same drop in both years, regardless of being a Monday in 2017.

If we compare December 2016 with December 2017, we found that in total there’s a -23% in FF change. Making this comparison day by day (Figure 4), we observed that in weekdays, the FF is basically the same (we can observe the change in 16-17 (Saturday-Sunday in 2016) and 17-18 (Sunday-Monday in 2017).

1. Representatively

The fundamental question about this index is how it is representative for any location in the UK and its possible bias against different factors, like oversampling areas or types of retail?

* 1. Locations

From July 2016, 20 cities account for 81% of the total FF at every month. From those, London in constantly contributing with ~27% of the total. However, in this index, the bias towards these cities is compensated through the weighted system explained in Appendix A. For example, a typical distribution of each sensor contribution to the total FF is shown in Figure 5. The bulk of the distribution is between 0 and 0.25, where around the 80% of locations are represented. The lognormal-like nature of this distribution confirms that the FF index is not capturing the flow of people around any particular location.

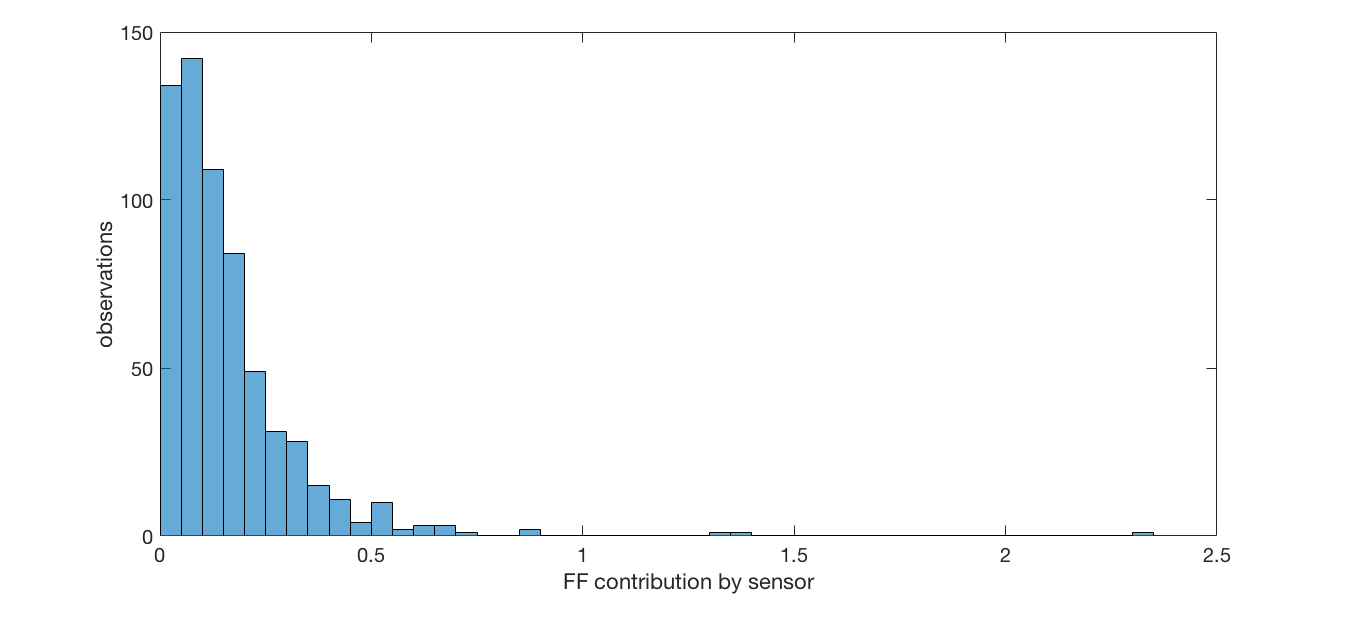


Figure 5. The probability distribution of the contribution to the total FF from each of the 631 locations operating in September 2017. As observed, only three sensors are contributing with a bit more than 1% to the total.

* 1. Type of street

As in the previous point, we can ask the same question in terms of the location’s position inside the street network, i.e., are these sensors predominant located in High streets, retail parks, shopping centres, etc. Using their full addresses, we associate each location with its street type, derived from the OpenStreetMap highway tag definition (https://wiki.openstreetmap.org/wiki/Key:highway). We found that these sensors are located in seven different types (see Figure 6): pedestrian/residential (55% of the locations), primary (16%), secondary (5%), tertiary (7%), service (9%), trunk (3%) and unclassified (5%).

Now, not all pedestrian/residential streets are equal. Particularly, in this classification, sensors installed at shopping centres or retail parks are classified as located in a pedestrian street, as people’s movements at these locations happened in a constraint environment with none or limited motor vehicles circulating. By January 2017, from the 60% of pedestrian locations, 20% (around 90) are actually located in different shopping centres, contributing only with 11% towards the total FF in January 2017.

In short, our index is fundamentally measuring the FF on pedestrian/residential streets, particularly, people walking on suburban high streets and strictly pedestrian ones (not shopping centres).

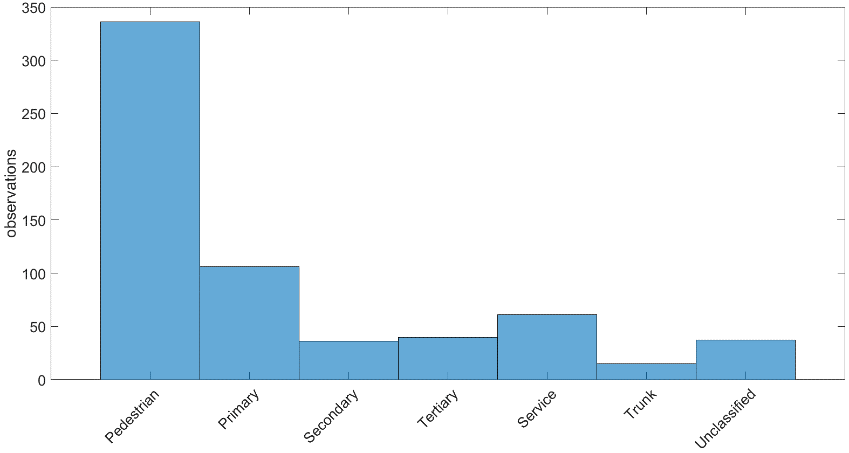


Figure 6. September 2017 probability distribution by type of street. The strong bias towards the pedestrian/residential type is evident, as is contributing to contributes to 55%, while the trunk type only 1%.

An interesting finding from this street analysis is that each type of street has its own characteristic FF signal (FF as a function of time) and this signal doesn’t depend on the type of business, in fact, shops of the same type can generate a whole variety of signals and values. This led us to believe that these sensors are capturing not the FF generated by a particular business, but the FF produced by the local characteristics at the area where the sensors are installed. This hypothesis is explored in Appendix B.

* 1. Type of shop

Finally, we explored if our index is biased towards a particular type of business. In total, from July 2017 to January 2018 there’s has been 105 different shop types within this network, but not all types are present at each month. In fact, only 45% of these types are across these 19 months, and 77% from March 2017. Consequently, the contribution of each type to the overall index varies from month to month, as it depends on the number of shops of that particular type (as all these counts are weighted as explained before). But before showing the actual FF per type of shop, we first need to address a bias in the counts generated by sensors installed close to each other.

* + 1. Nearest Neighbours sensors (NN)

We noticed that where two or more sensors are in close proximity to one another (40 meters or less) they are measuring essentially the same local FF. In our current network, there are 264 pairs of sensors in this situation, distributed in 52 different cities, being London the place with more NN sensors (84). A special case is the Ridings Shopping Centre (Wakefield) where 20 of the 33 sensors installed are in this situation.

For these 264 pairs, we averaged the hourly counts from each one to obtain a single measure (we refer the interested reader to Appendix C for the details about this process). For example, sensors 547 and 276 are located in Bridge St, Chester, and they are 22.3m apart from each other. The daily mean FF (Figure 7) is positive correlated and have quite similar values, even when each shop is in a completely different business category. We assigned a new id to the averaged signal and gave it the shop category “Merged”. We deleted from the index calculation the original FF counts from both merged sensors.

Figure 8 shows the probability distribution for each type of shop in terms of its particular contribution towards the FF index at 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 the 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 30-10% 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 and are representing a vast majority of business.

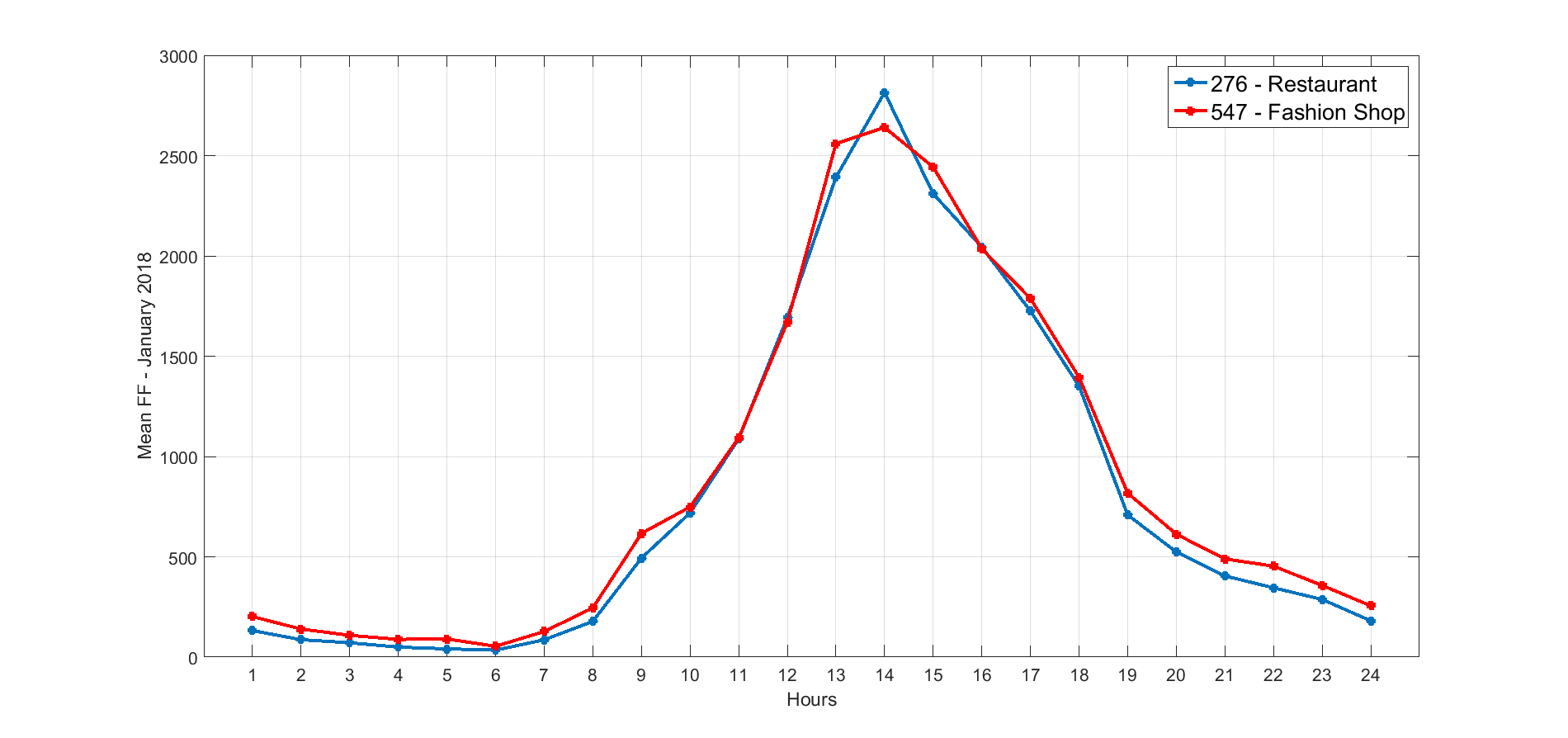


Figure 7. FF signal for a Restaurant and a Fashion shop, both located in a pedestrian street. The correlation matrix between these two variables has correlation coefficients of 0.99 reinforcing the evident interaction between these locations.

The prevalence of the merged type in 13 of the 19 months in Figure 8 gives another clue about how the FF measured is not tied to a particular type of shop but to a cluster of type and/or the particular characteristics.

Just to round up this analysis, if we break up the merged type into its original types, we still have the same type of distribution, but the position of its elements changes (Figure 9). Evidently, there’s no more merged type, and the type Shoe shops appears in the last three months, where originally was only Merged. The reasons for this and the precise evolution of the top contributors from month to month could help us to understand some particular variations at the vicinities of each location associated to something more than merely fewer sensors of a particular. This analysis is beyond the scope of this present work.

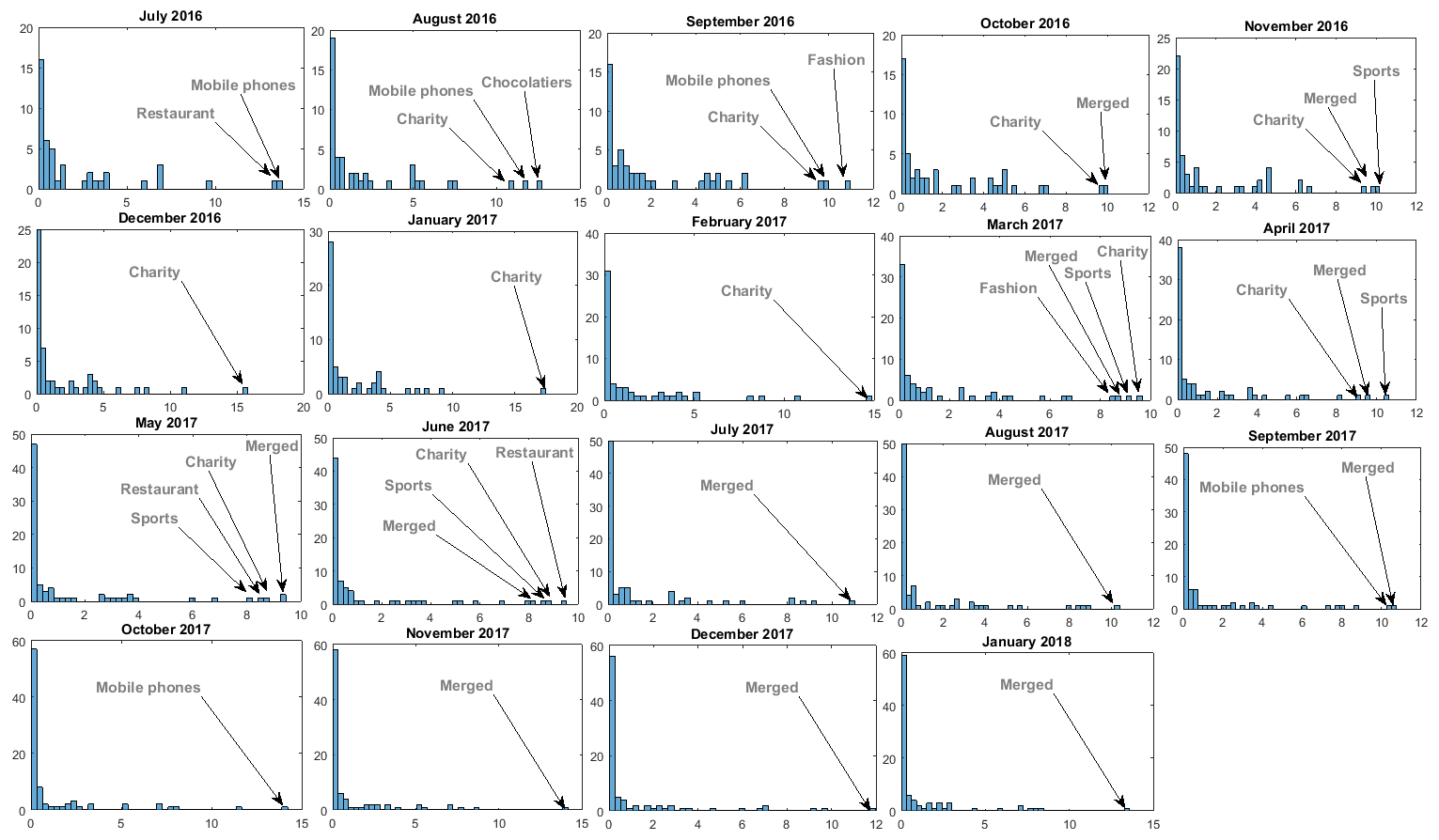


Figure 8. Distribution of FF by Type of Store (NN included).

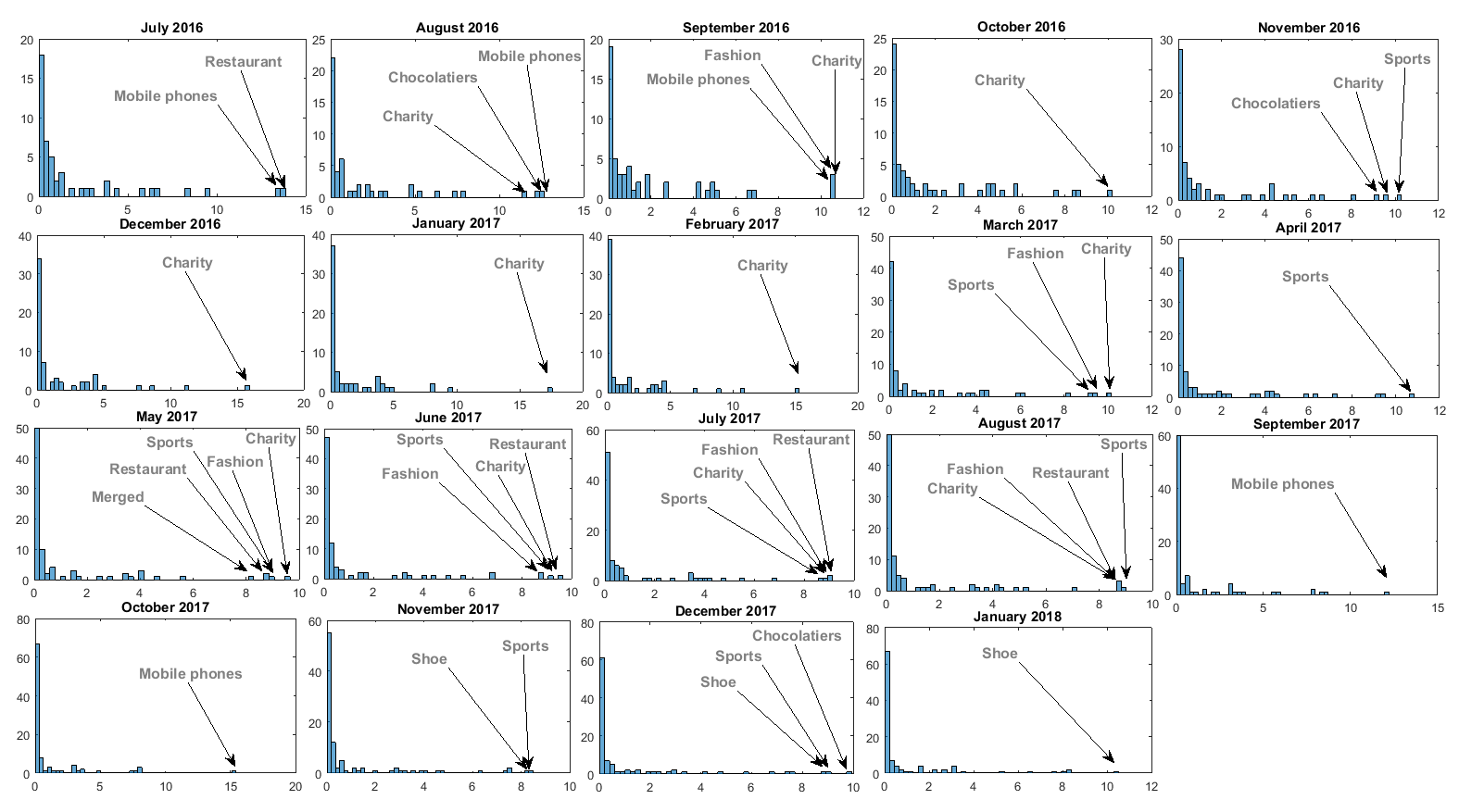


Figure 9. Distribution of FF type of shop (No NN included).

Appendix A

As mentioned, the formulation of the FF index (Eq. 1) is quite simple and only depends on the aggregated counts between the two different periods to be compared. The real challenge is the actual construction of the periods b and a, as the number of sensors between two-time periods is not necessarily the same. In fact, even if the number of sensors is the same, the number of hours measured by the same sensor between two different days could be different. This discrepancy makes, in principle, b and a incomparable as an increase in FF between two periods could come just as a result of having more sensors in one them. As a first step, we can weight each period by the mean number of active sensors at precisely that period. For example, if a = 10,000 (100 sensors) and b=17,000 (130 sensors), the weighted counts will be a=100, b=130. The problem is that, as mentioned, the number of active sensors is not constant in a single month. Let us look at Figure A.1, where the number of active sensors, at every five minutes during December 2017 is plotted. The circadian patterns at 28 of the 30 days show the stability of the system in almost the whole month. And December the 3rd and the 31st show at the same time its eventual weakness. In these two of days, the network suffered almost a complete failure through the whole day.

|  |  |
| --- | --- |
| C:\2017\ldc\indicator\oneHour\figures\fiveMinSensors.png  a) | C:\2017\ldc\indicator\oneHour\figures\fiveMinWeek.png  b) |

Figure A.1 a) Number of active sensors at December 2017(Except the 31st). Each point corresponds to a single minute. B) Active sensors at December the 1st. The maximum number of active sensors form a plateau from 10 to 18 hrs. The raise (decay) in numbers coincide with the early (late) hours ins a single day.

Isolating a single day in this data (Figure A.1 b), we can observe the precise hours where the number of sensors is constant.

The large variation in the number of active sensors in a single day makes our first attempt to weight the FF counts by the mean number of sensors in a given month, not an accurate one. To fully normalize the counts across different periods, we need to weight the aggregated 1-hour counts, by the number of active sensors at each hour as follows:

1. Let be the hourly counts at sensor at hour and let be the number of active sensors at hour during a given period P, so the aggregated counts by hour in a single period P, AH, are defined as

(A1)

Basically, we are creating an array of 24 points, where each point is the weighted sum of each hour in P for all sensors with counts in that particular hour.

1. Finally, the total weighted FF in P is calculated as:

(A2)

A2 is just the sum over all the weighted hours present in a single period.

In essence, what we are doing is normalize the counts by the number of sensors using the number of individual hours with data as a proxy for the number of sensors.

Appendix C

To detect the NN sensors, we created a distance matrix between all sensors in the network and extract all pairs that are within a distance of 40m or less, obtaining 264 pairs. Over these, we performed a correlation analysis by pairs to investigate if the FF signal from both sensors followed the same pattern or not. In different degrees, all pairs exhibit a positive correlation. It is important to notice that this correlation is only in terms of the signal, not in terms of the nominal value at each point in time.

To create an average FF value at each hour for the NN sensors, we proceeded as follows:

1. Case one. Both sensors have FF counts at the same of hours. The merged FF is just the mean FF at each hour from each pair. An example of these type of pairs is shown in Figure C.1, where we can observe how the FF signal of that two sensors is practically the same hour by hour both in shape and value.

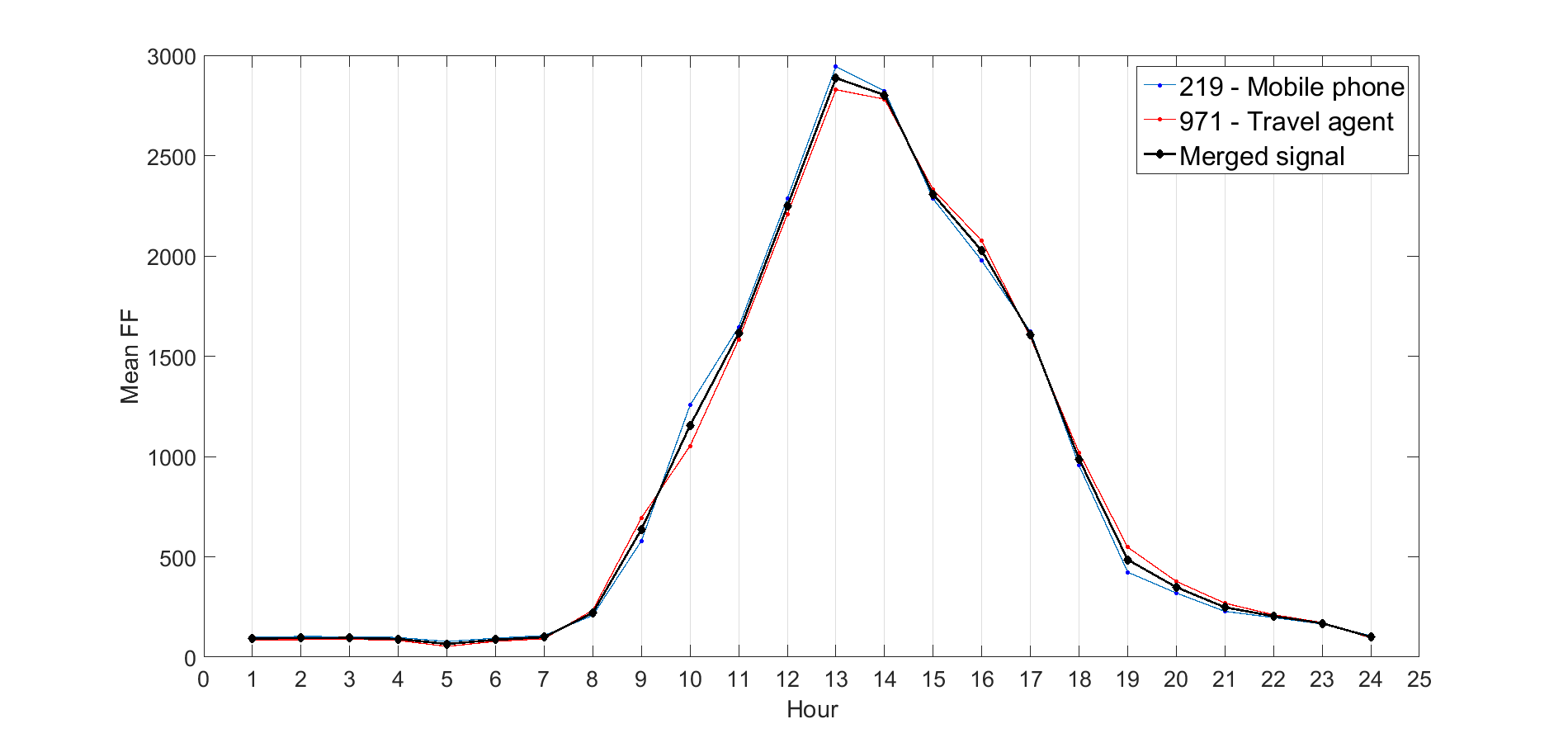


Figure C1. FF signals for two sensors (blue and red lines) at King Edward St, Hull, separate it by 30m. This street is a pedestrian one, and the "Bell shape" obtained is a typical signature of this. The black line is the merged signal.

1. Case two. Both sensors don't have FF counts in the same hours. In this case, we only average between common hours and the rest of the hours are left out of the merging process. For example, sensors 200 and 473, both at the Ridings shopping Centre (Wakefield) and separated by 20m only have counts in common hours from 8 to 19, so, the merged signal is only for these 11 hours. In this case, we delete the FF counts from sensor 200, kept the FF from 1 to 7 from sensor 473 and the mean FF for the merged counts is provided with a new id and assigned to the Merged shop category.

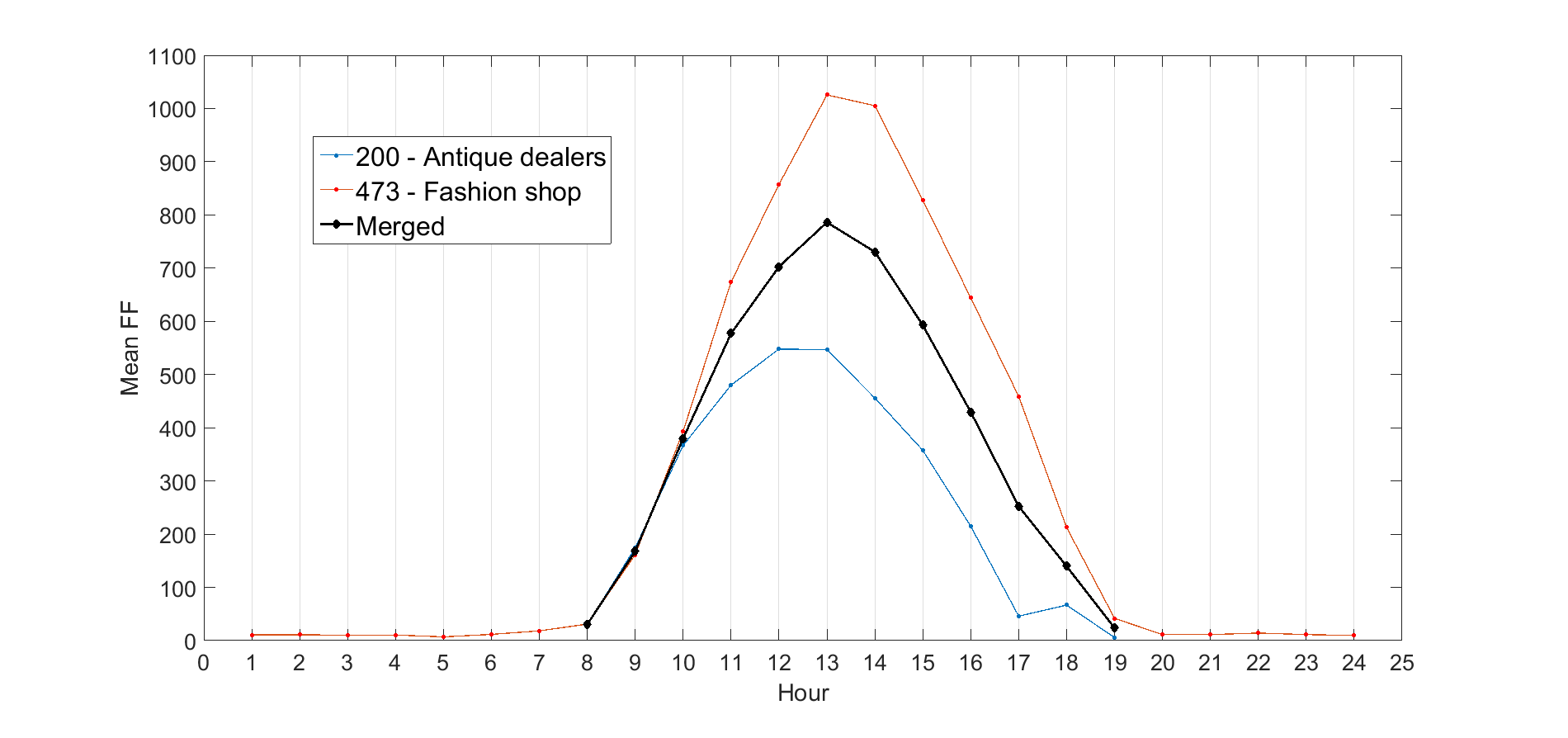


Figure C2. FF signals for two sensors (blue and red) and the mean FF (black) between them hour by hour.

There’s still some relevant questions about the NN sensors that would be significant for measures not as broad as the FF index or for comparing FF between Retail Areas. Some of these are:

* Investigate the wide range of correlation coefficients found among these pairs. This could highlight micro differences across locations.
* Look into the nominal FF values and not only the shape. It has been suggested that this could highlight the differences between shop types: higher values could indicate a long dwellers effect and not actual FF.
* Instead of calculated the mean between the two sensors, the merging process could be done using another centrality measure, like the maximum or the minimum values. The latter, for example, could be used to highlight the actual FF around a location.
* At this moment, we did the mean only between pairs, but there’s actually clusters of NN where three or more sensors are really close to each other. Collapsing these groups into one sensor could lead to more acquired local measures.