

## Combining footfall from mobile network data and event listings to predict consumer behaviour patterns

Ludmila Sabelnikova<sup>1</sup>, Professor Jason Dykes<sup>1</sup>, Anne Griffith<sup>2</sup>, Aidan Mcinerney<sup>2</sup> and Steve Gwynne<sup>2</sup>  
<sup>1</sup>City University, <sup>2</sup>Movement Strategies

### Background and Motivation

This project firstly explores the relationship between the demographic profile footfall (number of people at a place at a given time) and several events in the town centre of Bath, UK and secondly evaluates the extent to which the demographic profile footfall can be predicted using time series models. Footfall counting has been commonly used in retail for marketing and businesses planning and more recently in town centres uncovering new insights about how town centres are used and the temporal patterns of usage.

### Data and Methods

Using data from an O2 mobile network sensor that measures footfall across age/gender and spend categories on the high street in the town centre of Bath, I explored the relationship between events and demographic composition, as well as the footfall seasonal patterns across 20 demographic groups (10 age groups and 2 genders) and used the findings to inform the development of two footfall predictive models. The date range is 25th January 2019 – 22nd March 2020, with the exploratory data analysis being focused on 2019 data, while 2020 data is used for the model evaluation.

Data visualisation helped identify seasonal trends (month, day of the week). To understand the relationship between event listings and footfall composition, I identified suitable events – Rugby games, Half Marathon, Bath Festival and the Christmas Market, which could be sufficiently large to influence the demographic composition.

Using Python and combining the main dataset with weather data, I developed the Long Short-Term Memory (LSTM) and the Extreme Gradient Boosting (XGBoost) machine learning models, which have been previously shown to have a strong predictive performance with footfall data.

### Key Findings

For overall footfall patterns, there is a peak around the Christmas market, as well as peaks around Bank Holidays. There is also weekday

trend, with much higher footfall on Saturdays. For footfall segmented by age/gender, the most prevalent group during the summer months is the 18-24, while during the rest of the year is 65+. During Monday – Thursday the most prevalent group is 65+, while during Friday – Sunday it is the 18-24. The seasonal analysis showed the importance of seasonal benchmarks for event comparison, as well as including temporal features in the model building.

Whilst event comparison to the annual benchmark showed some variations in the demographic composition, such as population composition being more evenly distributed across all age groups during the Christmas Market, when comparing to seasonal benchmarks the differences are small, indicating footfall composition is not affected by events as much as expected.

With the walk forward method being implemented, where the models are retrained on 2020 data as it becomes available to predict one day in advance at a time, more accurate results were seen for the XGBoost, with the Root Mean Squared Error (RMSE) as the evaluation metric. Further after testing 3 lag windows (1 day, 3 day and 7 day), I compared the 3-day lag XGBoost and the 7-day lag LSTM, with the XGBoost having a better performance. Both models do not predict well after the 15th March, which is the week before lockdown was implemented, suggesting adjustments should be applied to account for behavioural changes.

### Value of the research

With strong seasonal trends, the research can aid local business decision making, with businesses choosing to run seasonal promotions for students during weekends and summer. Some local promotions may run based on event, for example rugby games could attract more people to local pubs. The research may also aid staff management, as higher footfall expected on Saturday and during the Christmas period. The models also showed how previous footfall patterns did not hold during the lead up to the national lockdown, with businesses needing to consider uncertainty.