

City University
MSc Data Science

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

Author: Ludmila Sabelnikova

Academic Supervisor: Professor Jason Dykes

Industry Supervisors: Anne Griffith, Aidan Mcinerney and Steve Gwynne

Company Sponsor: Movement Strategies

Project Introduction

Aim

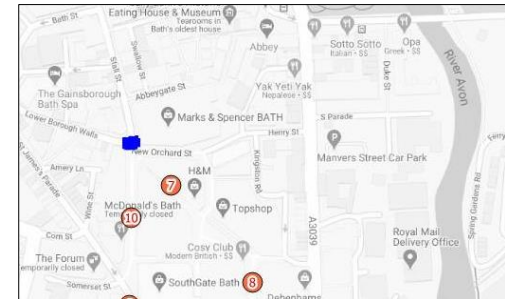
Evaluate the extent to which footfall across different groups can be predicted for event and non-event days using mobile and event listings data.

Objectives

1. Exploratory data analysis to identify which seasonal patterns are important when creating the models
2. Exploratory data analysis to identify the relationship between events and footfall composition in the city centre of Bath
3. Train and optimise two models to predict the footfall for the first few months of 2020
4. Compare and evaluate the two models against each other and the conditions under which the models perform well.

Dataset Characteristics

- Data Source: O2 Mobile Censor in Bath town centre on high street intersection
- Attributes: Daily Level footfall broken down by demographics, with results focused on age/gender and the following 10 age groups: 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64 and 65+
- Date Range: 25th January 2019 – 22nd March 2020



Location of Sensor In Bath

Exploratory Data Analysis (EDA) – Seasonality and Events

EDA – Seasonality

Total footfall

- Seasonal variation, with a peak around the Christmas market, as well as peaks around Bank Holidays.
- Weekday trend, with much higher footfall during Saturdays with the average almost double the footfall of weekday average.

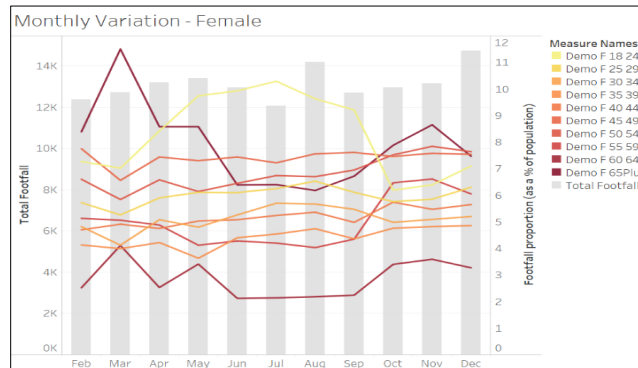
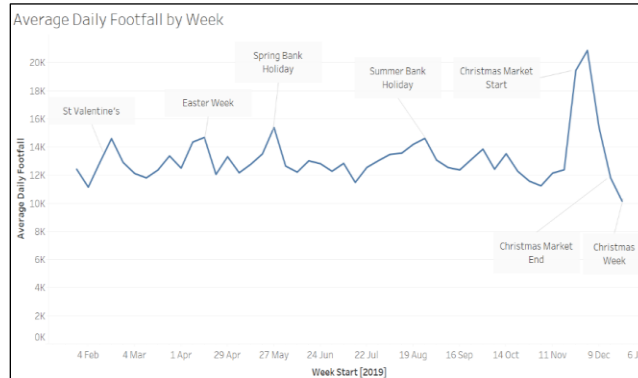
Footfall broken down by age/gender

- 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 Friday – Sunday is the 18-24 age group.

The seasonal analysis showed the importance of seasonal benchmarks for event comparison, as well as including temporal features in the model building.

Insights & Recommendations

- Seasonal Promotions for students during weekends and summer months
- Local promotions based on event – Rugby games could attract more people to local pubs and restaurants
- Staff Management – Higher footfall expected on Saturday and during the Christmas Period



Darker Colour lines – Older Age Groups

EDA – Events

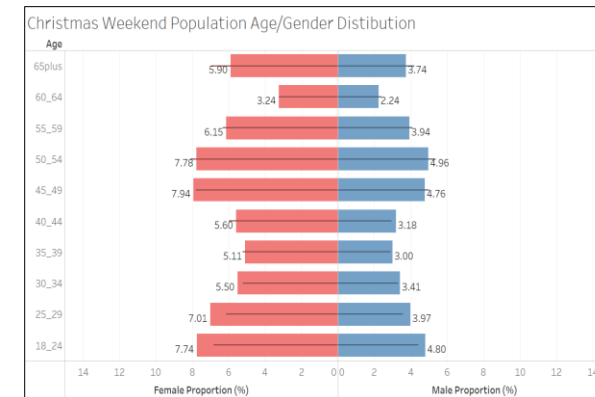
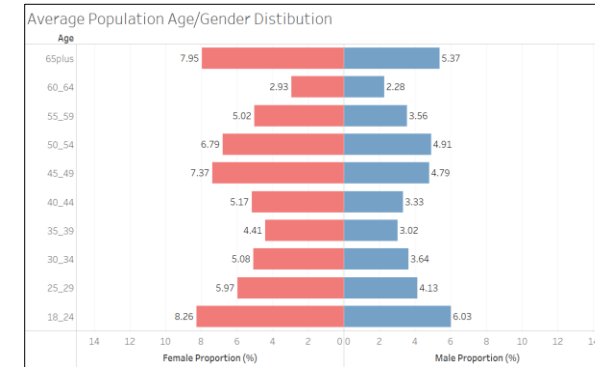
Benchmark (2019 data)

- Higher female proportion at 59% vs 41% males, which could be due to the positioning of the sensor on the high street.
- Two dominant groups – 18-24 and 65+, which is consistent with Bath's census data.

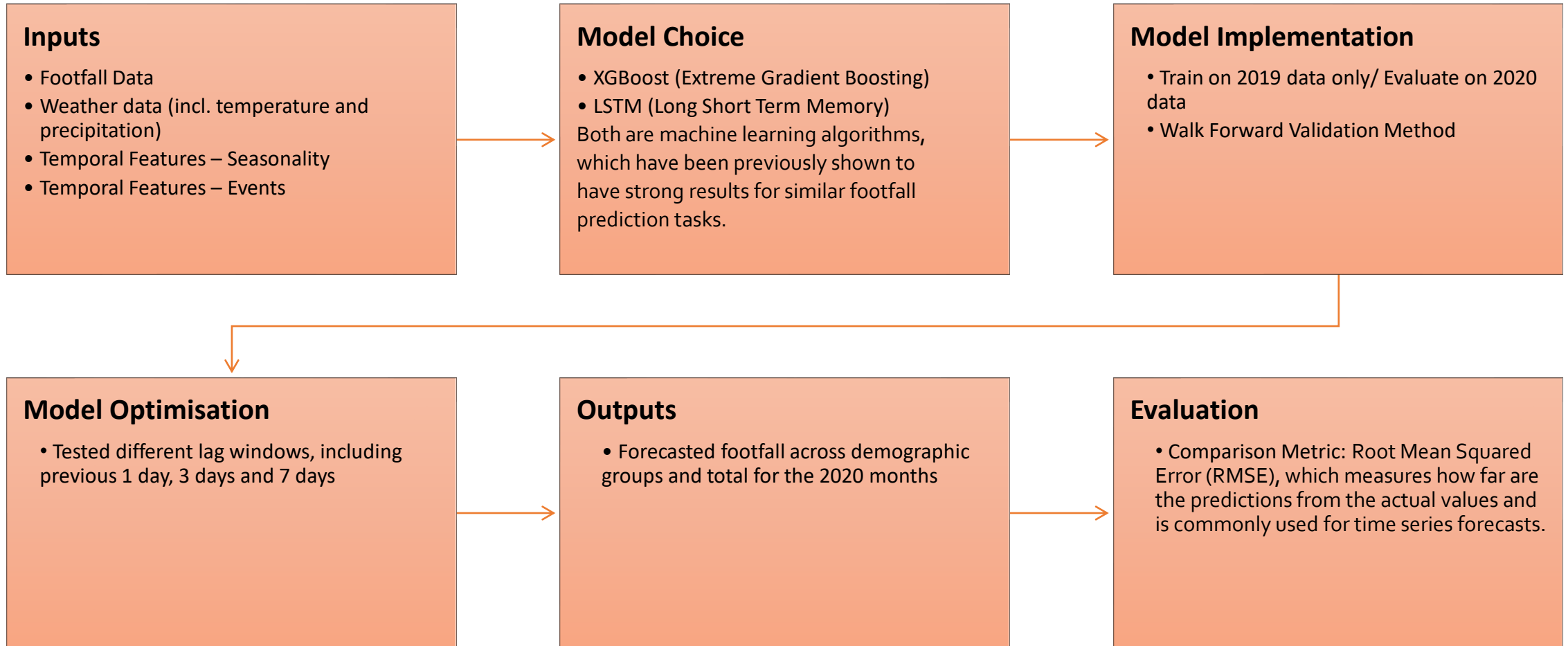
Comparison to annual benchmark:

- Bath Festival Weekend** - Higher proportion of 18-24 females.
- Half Marathon** - Unusually higher proportion of 65+ females.
- Rugby Games** - No significant differences compared to the average.
- Christmas Market weekends** - More distributed across all age groups, with more people in 45-54 age groups than average.

Whilst comparing to the annual benchmark, showed variations in the demographic composition, when including the seasonal benchmarks the differences are small, indicating footfall composition is not affected by events as much as expected.



Model Building



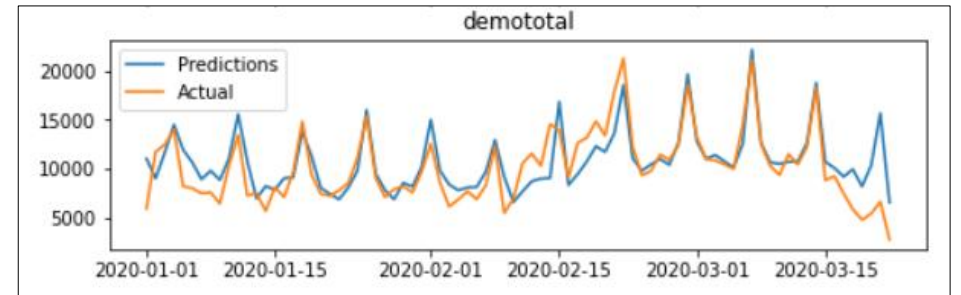
Model Results

Model Results and Comparison

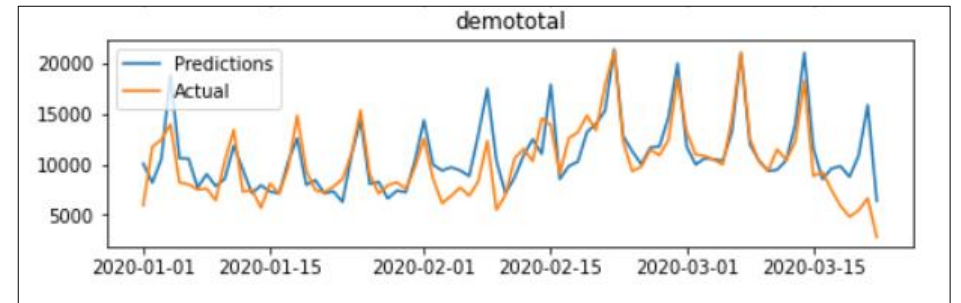
- For the initial models, which were trained on 2019 data only, the LSTM had a better performance.
- With walk forward implementation the XGBoost (3-day lag) had the better performance, as shown in the table. After testing different lag windows the XGBoost (3-day lag) and the LSTM (7-day lag) were the best.
- The XGBoost performed better across all the demographic groups, weekday/weekends and for both events – Half Marathon and Rugby.
- Both models do not predict well after the 15th March, which is the week before lockdown was implemented, as behaviour patterns changed.
- Feature importance, which was generated by the XGBoost model, showed that weather features were the most important indicating seasonality is important for footfall.

Model	RMSE (initial models)	RMSE (walk forward method 1-day lag)	RMSE (walk forward method 3-day lag)	RMSE (walk forward method 7-day lag)
XGBoost	688.20	610.38	542.39	584.70
LSTM	583.76	727.84	853.36	630.71

Comparison of XGBoost and LSTM models



XGBoost Predictions for Total Footfall – 3 day lag



LSTM Predictions for Total Footfall – 7 day lag

Insights & Recommendations

- Considering footfall uncertainty during Covid-19 – Different type of modelling could be more suitable.
- Importance of weather and feature inclusion in models – Footfall is highly seasonal, with weather incorporating the seasonality well.