

Understanding and Predicting Consumer Behaviour in Music Festivals with Machine Learning

Luis Francisco Mejia Garcia¹, Guy Lansley¹ and Ben Calnan²

¹University College London, ²Movement Strategies

Project Background

Music festivals have become very popular social events across the globe and make a large proportion of their revenue from the sale of consumables within the festival sites. However, as these purchases are typically concentrated over a couple of days in temporary locations, relatively little insight can be achieved about consumers relative to the more sophisticated data and modelling approaches available to longstanding retailers with an established network of stores. Therefore, new technologies and innovative techniques could be useful to estimating temporal patterns in footfall across a festival site in order to model patronage at pop-up catering facilities.

Data and Methods

This study presents an exploratory analysis of newly available GPS data collected from the FYF music festival in the United States in order to estimate consumer behaviour inside the event. The data was originally collected from a mobile phone app made available to festival visitors. The festival app used the mobile phones positioning systems to record users locations at various different time intervals. In total the data included a 100 million records of user locations at various times, social media information on some users and the festival schedule. 6 features were engineered to represent the factors that might be influencing the users in the decision when they go to the bar areas. Machine learning algorithms such as Random Forest and Artificial Neural Networks were subsequently trained using these features to identify which are the most influential factors for estimating visits to the bar areas across the festival.

Key Findings

The influence of the 6 variables over time on Sunday have been shown in figure 1. The 'total time spent by the users in the festival' was shown to be the most influential factor on the consumer followed by 'time since their last visit to the bar'. The 'distance from the closest bar' feature didn't prove to be a significant influencing factor. The 'artist's popularity' seems to have more influence in the case of upcoming shows than in the past shows. 'Gender' feature is the less influential factor according to the results.

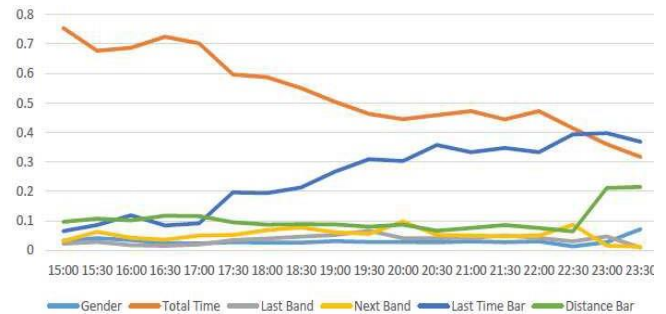


Figure 1. Feature Importances on Sunday

A prediction model was also devised and used on a sample of data from Sunday which were withheld from our initial analysis. The prediction model based on artificial neural networks presented an accuracy of 75% when compared to the actual results – despite the festival data only pertaining to two days of data.

It was also proved that the modification of organisational factors could lead to an increment in the occupancy of the bars, for example, with a decrement of 30% of the distance between the users and the bar areas the occupancy could be incremented by 7% on average.

Value of the Research

The results of this study demonstrate the importance of a good feature engineering process as the behaviour of people can be deducted from a relatively simple location dataset. The Technology used for the collection of data in this project proved to be a good source of information useful for modelling the occupancy in bar areas.

The classification and predictive models analysed in this work could represent an opportunity for companies to implement a similar process in other scenarios. Occupancy of public spaces or the analysis of consumer behaviours across retail sites could be analysed with the correct feature selection and the appropriate machine learning models if similar data is made available.