**Clustering Neighbourhoods – Opening a new venue in London**

Valerie van der Spree, October 2020

1. **Introduction**
   1. **Background**

London is one of the most popular and populated cities in the world. The cuisine here and the activities available are some of the most diverse in the world. This makes it both a very attractive and a challenging location to open a new venue. Opening a new venue requires a lot of knowledge about the area, the already existing venues, as well as investment. Therefore, it is advantageous to get a feel for the different neighbourhoods in London and the already existing venues in these areas, to better orient oneself before deciding on a location to open a new pub.

* 1. **Problem**

Data might contribute to better orient oneself in this great city. Data about boroughs, the area, population, average rent and the existing venues will be very helpful as a first step towards deciding where to open a new pub. This will take into account the population living here, part of the investment costs, as well as the competition.

* 1. **Interest**

Entrepreneurs looking to open a pub, café or restaurant would be very interested in the analysis, as this will save them a lot of time and money and it will also help them set up at the best location taking into account their constraints. Others who are interested in venues in different areas of London, such as already existing venues, might also find this analysis interesting.

1. **Data acquisition and cleaning**
   1. **Data sources**

The data used for this report consist of the following:

* Data about the boroughs in London from [Wikipedia](https://en.wikipedia.org/wiki/List_of_London_boroughs)
* Data about average rent in London from the [UK Government website](https://www.gov.uk/government/statistics/private-rental-market-summary-statistics-april-2018-to-march-2019)
* Geographical data: using geopy the coordinates for each borough will be found
* Venue data: Foursquare will be used to collect data about the existing venues in each borough After the data is collected a k-means clustering method will be used to cluster the boroughs and visualise this data on a map.
  1. **Data cleaning**

Data downloaded or scraped from the multiple sources were combined into one table. The table was cleaned:

* The data from Wikipedia was cleaned; removing odd characters, notes etc.
* Columns were renamed.
* It was ensured that the data in the columns were in the right data format (e.g. population in float instead of object).
* Removing the inner boroughs of London as the rent here might be too expensive for an entrepreneur.

There was no problem with outliers as the data was geographic. Most of the cleaning had to do with aesthetics such as column titles and merging tables.

1. **Data Analysis**

**3.1. Exploratory analysis**

Not a lot only exploratory analysis is relevant to this type of data. The boroughs were visualised on a map of London to be able to get a feel for their relative location. Additional exploratory analysis included creating a bar chart of the rents, populations and areas, to get a better feel for the data

* 1. **Cluster Analysis**

All of the venues in each borough were imported from the Foursquare API, they were one-hot encoded, grouped, and the most popular venues merged with the list of boroughs. The boroughs (total of 20) were then clustered in five clusters and merged with the initial data set. These clusters were then mapped (see figure 1). Each of these clusters was then analysed individually.



*Figure 1: Map of London with the chosen Boroughs*

1. **Results**

The individual analysis of the clusters showed that there was a big variety of the type of venues that already existed in each borough.

The first cluster that was analysed was already very much saturated with pubs, cafes and restaurants. Hence, this cluster was not deemed attractive as it would be very hard to compete with already so many existing venues.

The second cluster existed of only one borough, Havering, in which the most popular venues were an airfield and IT services. Hence, this borough is not very attractive.

The third cluster existed of only one borough as well, Brent, and here the most popular venues were fast food restaurants and a convenience store. Depending on what the borough looks like in real life, and the demand here for a pub or café, it might be attractive. However, it was decided that it was not a great place to start out due to this ambiguity.

The fourth cluster also existed of only one borough, Sutton, which did seem promising. The most popular venues included some restaurants of different cuisines and a hotel, suggesting that there might be demand for a pub or café here.

Finally, the fifth cluster included a mix of boroughs out of which Bromley, Croydon, Enfield and Waltham Forest looked the most promising. The first three already have a pub but they are closely followed by a coffee shop, a park and some restaurants, so this might be interesting! Waltham Forest is probably one of the most interesting as it has lots of restaurants but no pub and a sizeable population.

Some overall other findings include:

* The most popular social venues, outside of Inner London boroughs are Pubs and Coffee shops.
* Northern boroughs are more prone to visiting pubs, whereas southern boroughs are most likely to shop and have the social life from home.
* Within the top 5 places of interest in every borough is an ethnic restaurant.
* Rent price does not seem to be much of a factor for going out - the demand is not affected by difference in costs.

1. **Discussion**

Looking at the data Waltham Forest, Bromley, Enfield and Sutton seem to be the best places outside of Central London where a new venue is worth opening. However, a lot of information is not taken into account, and cannot be obtained from Foursquare Developer:

* Bromley and Enfield's rent is slightly higher (approx £200 per month) so this will need to be taken into account. However, demand here might be higher and therefore it might be worth the extra cost. This can only be determined by visiting the boroughs.
* Higher ethnic presence in a given borough can and will influence the popularity of a given cuisine.
* Closer proximity to Inner boroughs and better transport links allows people to travel to the neighbouring borough and impact the measurements
* Many small venues are not registered in Foursquare and are marketed via word-of-mouth, and are not taken into account

Regardless, the analysis provided an insight into what people like and opt for, when it comes to going out in their own neighbourhoods.

1. **Conclusion**

In this study, I analysed the different boroughs in London using cluster analysis, based on their location, most popular venue type and rent. I identified the most promising venues in the Outer London area to open a new venue at based on these factors. This could help entrepreneurs in the first few stages when deciding where to open a new venue.