Capstone Project - The Battle of Neighbourhoods
March 2020

Problem Statement:

Where is best area of Greater London to open a new business (either a pub or a coffee shop) in London in 2020.

Description of the problem and background

Greater London is a large and diverse area in the UK. It is made up of 32 boroughs and covers over 600 square miles[1]. Each borough has its own characteristics from busy areas of central London to the relatively quieter residential districts.

Anyone setting up a new business in "London" has a difficult decision to make when choosing a location, so I am seeking to demonstrate how Data Science can help with that decision in terms of analysing characteristics of each borough and using the Fourssquare API to cluster different businesses in each borough.

Data Sources

The majority of the data on London Boroughs will be taken from publicly available dataset on Wikipedia (https://en.wikipedia.org/wiki/List_of_London_boroughs). This data will require a level of manual cleaning to ensure it is fit for analysis.

The Foursquare API (https://developer.foursquare.com) will be used to retrieve information about existing businesses in these Boroughs.

Levels of Weekly pay by London borough are sourced from Office for National Statistics. (http://www.statistics.gov.uk/statBase/product.asp?vlnk=13101)

Using the data I will seek to examine population density, earning levels and similar businesses for each Borough and try to estimate where a new business (pub or coffee shop) is likely to have the best chance of succeeding in Greater London.

Potential limitations

Anecdotally Foursquare is considered less popular in the UK than perhaps North America – so the accuracy of the business data may limit the success of this analysis.

In addition, each Borough in London covers a wide area (compared to previous case studies used in the course so far). To illustrate this the largest Borough, Bromley, covers also 60 square miles. To mitigate for this I will seek to use a wide radius (5000 metres) in the Foursquare API to cover a wide area.

Without paying for other data sources, there is only limited data on London Boroughs available in the public domain. If this was a professional analysis – further analysis could be

---

[1] Source: https://en.wikipedia.org/wiki/Greater_London

considered by paying for additional datasets. There is also a risk that data from Wikipedia could be wrong or could change during the model exercise!

London Boroughs by Population Size

London Boroughs by Population Density