Capstone Project - The Battle of Neighbourhoods
London New Business Analysis
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 Foursquare 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)

---

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

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 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!
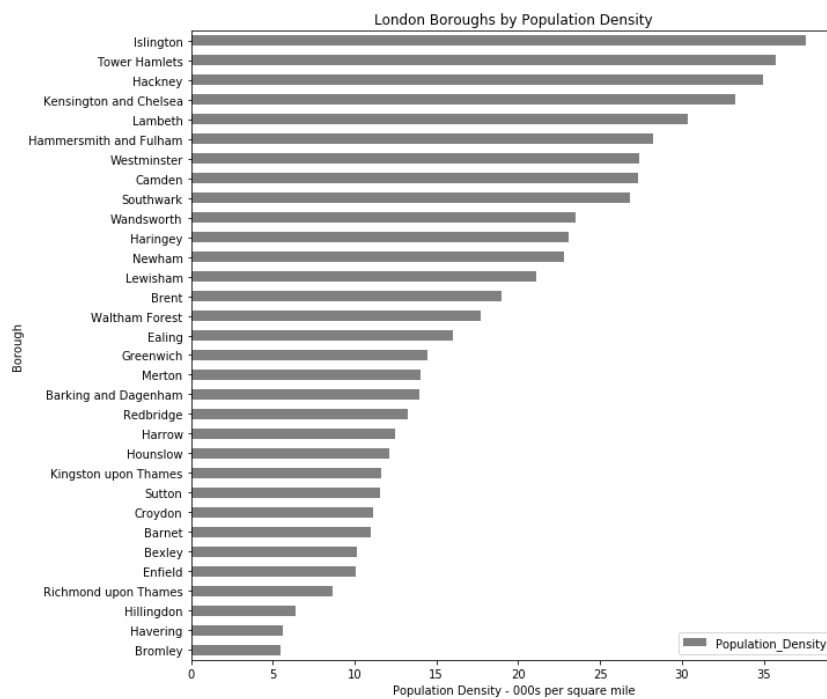
Methodology

1. Import London Boroughs and London Earnings datasets into Panda dataframe

2. Examine Population and Population Density for each Borough using visualisation

3. Examine Earnings in each Borough using visualisation

4. Use Foursquare API to examine businesses located in each Borough.

5. Cluster business data using K-means clustering

6. Plot clusters on map and examine results

7. Attempt to find a borough with highest population density, high hourly wage and a low clustering of coffee shops or pubs.

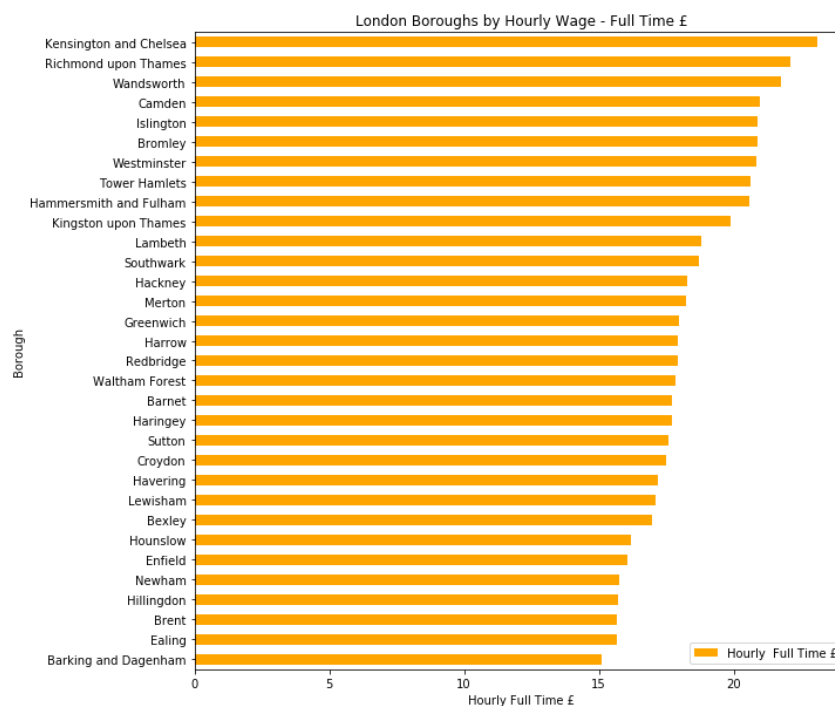For more details refer to Notebook section.

Results

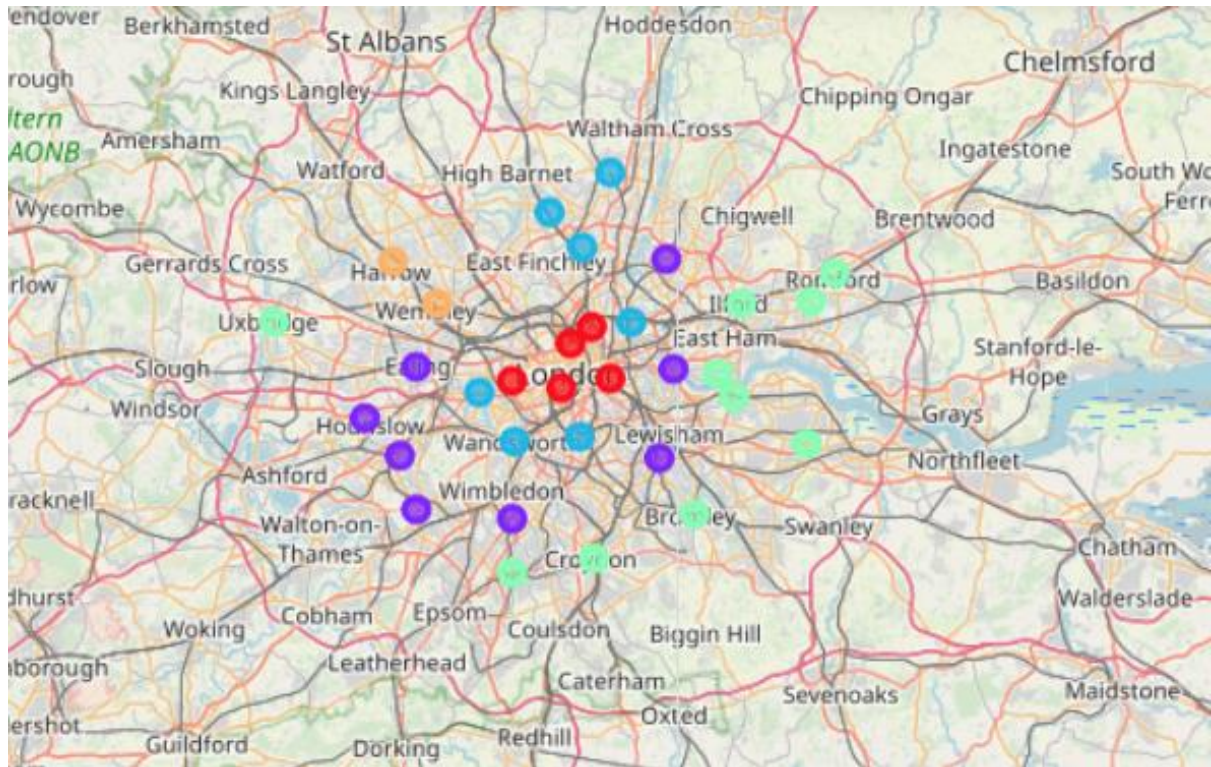The visualisations below were produced in the analysis

1. London Borough by Population Density

2. London Borough by Hourly Wage

3. Clustering Map

Clusters Results Set

Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Camden	0	Hotel	Bookstore	Steakhouse	Pizza Place	Theater	Cocktail Bar	Coffee Shop	Art Gallery	Indian Restaurant	Ice Cream Shop
Islington	0	Hotel	Cocktail Bar	Coffee Shop	Pub	Pizza Place	Market	Steakhouse	Sushi Restaurant	Flea Market	Bakery
Kensington and Chelsea	0	Hotel	Café	Garden	Juice Bar	Park	Art Gallery	Japanese Restaurant	Indian Restaurant	Department Store	Hotel Bar
Southwark	0	Hotel	Theater	Cocktail Bar	Coffee Shop	Scenic Lookout	Street Food Gathering	Sushi Restaurant	Garden	Pub	Pizza Place
Westminster	0	Hotel	Cocktail Bar	Lounge	Art Gallery	Bookstore	Steakhouse	Theater	Park	Art Museum	Plaza
Ealing	1	Pub	Coffee Shop	Park	Café	Hotel	Botanical Garden	Middle Eastern Restaurant	Pizza Place	Bakery	Sushi Restaurant
Hounslow	1	Pub	Park	Coffee Shop	Indian Restaurant	Portuguese Restaurant	Supermarket	Café	Rugby Stadium	Hotel	Gym
Kingston upon Thames	1	Pub	Café	Park	Coffee Shop	Garden	Supermarket	Burger Joint	Gastropub	German Restaurant	Gift Shop
Lewisham	1	Pub	Park	Coffee Shop	Café	Indian Restaurant	Beer Store	Bar	Italian Restaurant	Garden Center	Fish & Chips Shop
Merton	1	Pub	Park	Coffee Shop	Café	Bar	Sushi Restaurant	Italian Restaurant	Tennis Stadium	Supermarket	Burger Joint
Richmond Thames upon	1	Pub	Park	Garden	Coffee Shop	Café	Italian Restaurant	Rugby Stadium	Hotel	Bakery	Scenic Lookout
Tower Hamlets	1	Pub	Park	Coffee Shop	Café	Bar	Garden	History Museum	Farm	Turkish Restaurant	Beer Bar
Waltham Forest	1	Pub	Coffee Shop	Park	Pizza Place	Café	Brewery	Nature Preserve	Italian Restaurant	Restaurant	Gym / Fitness Center
Barnet	2	Café	Coffee Shop	Park	Turkish Restaurant	Supermarket	Mediterranean Restaurant	Greek Restaurant	Bakery	Gym / Fitness Center	Pub
Enfield	2	Pub	Coffee Shop	Turkish Restaurant	Gym / Fitness Center	Park	Café	Garden Center	Supermarket	Grocery Store	Greek Restaurant
Hackney	2	Pub	Coffee Shop	Café	Bakery	Cocktail Bar	Park	Pizza Place	Bookstore	Beer Bar	Brewery
Hammersmith and Fulham	2	Café	Coffee Shop	Pub	Gym / Fitness Center	Pizza Place	Gastropub	Park	Fish & Chips Shop	Bakery	Indian Restaurant
Haringey	2	Turkish Restaurant	Coffee Shop	Café	Pub	Park	Mediterranean Restaurant	Bakery	Trail	Pizza Place	Organic Grocery
Lambeth	2	Park	Coffee Shop	Café	Pizza Place	Italian Restaurant	Theater	Market	Beer Store	Bar	Pub

Borough	Cluster	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Wandsworth	2	Park	Coffee Shop	Café	Pub	Pizza Place	French Restaurant	Tennis Stadium	Soccer Stadium	Bakery	Yoga Studio
Barking and Dagenham	3	Supermarket	Grocery Store	Coffee Shop	Park	Pub	Italian Restaurant	Café	Fast Food Restaurant	Bakery	Gym / Fitness Center
Bexley	3	Pub	Grocery Store	Supermarket	Coffee Shop	Italian Restaurant	Clothing Store	Furniture / Home Store	Pharmacy	Park	Fast Food Restaurant
Bromley	3	Coffee Shop	Pub	Park	Grocery Store	Pizza Place	Indian Restaurant	Gym / Fitness Center	Café	Mediterranean Restaurant	Gastropub
Croydon	3	Pub	Park	Grocery Store	Coffee Shop	Supermarket	Pizza Place	Mediterranean Restaurant	Clothing Store	Hotel	Gym / Fitness Center
Greenwich	3	Pub	Park	Coffee Shop	Hotel	Gym / Fitness Center	Historic Site	Grocery Store	Clothing Store	Garden	Asian Restaurant
Havering	3	Coffee Shop	Grocery Store	Supermarket	Pub	Café	Italian Restaurant	Park	Fast Food Restaurant	Bakery	Clothing Store
Hillingdon	3	Pub	Coffee Shop	Grocery Store	Pharmacy	Pizza Place	Indian Restaurant	Supermarket	Sandwich Place	Hotel	Gym / Fitness Center
Newham	3	Pub	Hotel	Park	Coffee Shop	Gym / Fitness Center	Café	Turkish Restaurant	Historic Site	Plaza	Burger Joint
Redbridge	3	Park	Pub	Grocery Store	Supermarket	Café	Indian Restaurant	Pizza Place	Train Station	Gym	Coffee Shop
Sutton	3	Park	Coffee Shop	Pub	Grocery Store	Supermarket	Café	Italian Restaurant	Pizza Place	Clothing Store	Hotel
Brent	4	Indian Restaurant	Pub	Supermarket	Coffee Shop	Middle Eastern Restaurant	Park	Gym / Fitness Center	Café	Portuguese Restaurant	Pizza Place
Harrow	4	Indian Restaurant	Coffee Shop	Park	Pub	Café	Gym / Fitness Center	Portuguese Restaurant	Italian Restaurant	Supermarket	Middle Eastern Restaurant

Conclusions

The analysis shows that the both Cluster 0 and Cluster 4 has the lowest number of pubs and coffee shops of all the London Boroughs.

Both of these clusters would present an interesting option for opening either of the above restaurants. However, using the population density and hourly earnings data we can try to differentiate further between these two clusters.

If we examine the boroughs with the highest population density and highest earning, we can see that Islington (ranked 5th for Hourly Wage, and ranked 1st for Population Density) would appear to be the best Borough for a new business.

Although Islington is the second smallest borough it benefits from 215k people in 5.7 square miles and has 5th highest wage at £20.9 per hour in 2019. In addition, coffee shops and pubs are only the 3rd and 4th most common venue of Foursquare for this borough.

Limitations of the Analysis

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.

Borough in London covers a wide area (compared to previous case studies used in the course so far) and a standard radius (5000 metres) does not cover or limit the analysis sufficiently.

Wage data was taken from 2019 and takes not account of earnings of visitors or employees who visit the areas each day.

No consideration of rents in the area which has a big impact on success of new business Analysis of businesses using Foursquare API only considers ranking of businesses in borough and not the number of coffee shops. Just because coffee shops are 3rd most common venue if Foursquare doesn't mean there is shortage of coffee shops in the Borough.