**Location analysis for a coffee chain launch**

Ritwik Pande

March 2020

**1. Introduction**

* 1. **Background**

A supply company that provides trade relationships with local estate coffee growers to coffee chains want to launch their own coffee chain starting with London. Backed by a real estate giant, the organization has sufficient funds to launch a trial and error with openings in London with the main focus on reaching out to a maximum set of customers with their first launch. The rationale provided is the collection of customer data and capture their feedback to understand the likes and tastes of coffee enthusiasts in London. The quantity of data is of utmost importance as this would help in determining not only the changes to be made to the offerings but also to understand the customer demographics and plan for subsequent launches.

* 1. **Problem**

To achieve a high rate of footfall, the problem lies in identifying a vicinity with a relatively higher population density and a favourable competitive landscape. This would require an analysis to identify places where the current competitive landscape is weaker hence providing us a leeway to establish our coffee chain faster and gaining a customer base quicker. The next step would be to gather data sources which provides demographic details on the general population in London across Boroughs. This would further help in segregating the Boroughs into clusters which would be useful to study before launching the next coffee shop.

**2. Data acquisition and cleaning**

**2.1** **Data sources**

Demographic and various other details for London can be found in datasets available in <https://data.london.gov.uk/> which is created and run by the GLA (Greater London Authority) as a step to free London’s data. Datasets need to be explored in the aforementioned websites to find the appropriate data for creating our demographic profiling of Boroughs. The other data source is using the Foursquare API to collect information of the existing coffee stores and their ratings. Basic API calls will help retrieve the competitor coffee shop addresses but the obtaining their ratings require a premium API call which is limited for a free user in a given period of time.

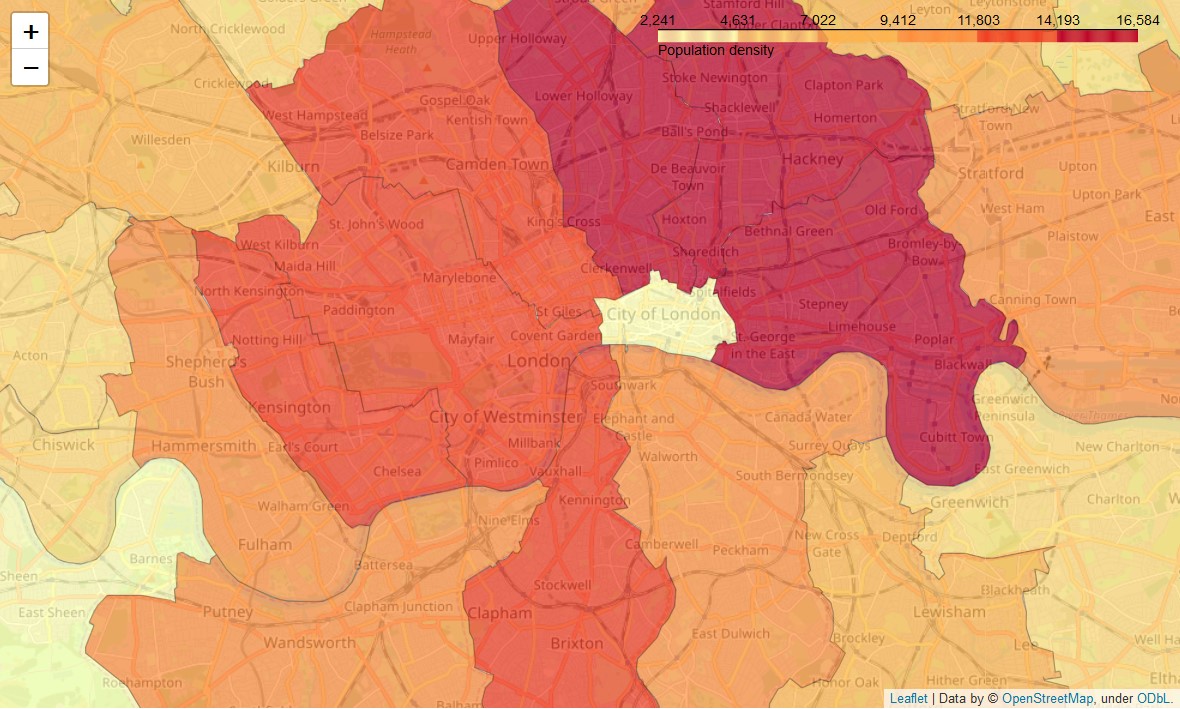
**2.2 Data Cleaning**

Data obtained from multiple datasets from <https://data.london.gov.uk/> were available in xlsx format and did not need any cleaning process. The datasets available were at a Borough level. The dataset for the Borough-level demographic profiling had multiple columns surplus to the requirement for the analysis which were dropped. The entire analysis did not require any substantial level of data cleaning.

**3. Location Analysis**

**3.1 Population density**

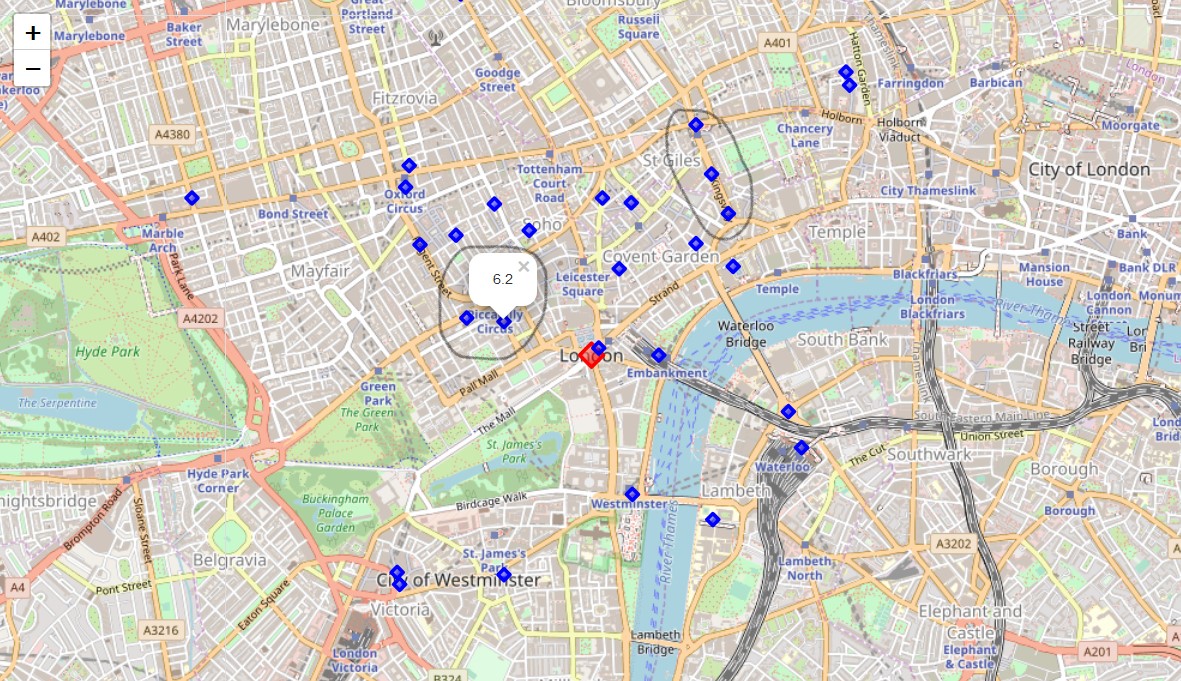
The primary focus was to find out the boroughs with maximum population density. The metric that chosen to serve the purpose was - *population\_per\_square\_kilometre*. The map generated using the folium library to depict the population density across boroughs:



The map is zoomed in to *City of London* Borough and the Boroughs surrounding it. The Boroughs Westminster and Camden stand out in terms of Population Density.

**3.2 Competitive landscape**

The Foursquare search function results will be used to gather a list of current coffee shops in the Boroughs of Camden and Westminster to compare and explore the competitive landscape. A basic API call was made to Foursquare to retrieve the Venues with the rest of the information in JSON normalized into a pandas dataframe. To procure the ratings, a premium call was made to the Foursquare search results. The ratings were then consolidated into the initial dataframe. The latitude and longitude values of the coffee shops from the dataframe were used with folium library to create a map visualization as show below:



The blue markers indicate the coffee shops. The ratings can be seen on clicking the blue markers.

Couple of findings from the above visualization:

* Towards north-east at Hatton Garden, there are 2 highest rated restaurants
* Oxford Circus has the maximum density of coffee shops giving customers plenty of options with some high rated
* There are 3 coffee shops at Kingsway near St.Giles that have below average ratings; the same goes for Picadilly circus

The stores in the vicinity of Picadilly Circus and Kingsway have below average ratings and hence is a favourable competitive landscape. Both of them come under the Borough of Camden which is among the top Boroughs with maximum population density.

**4. Conclusions**

It should be noted that this analysis can prove to be a starting point to explore areas keeping in mind the objectives of the business. The completion of the analysis depends on the availability of data as is the case with cities such as San Francisco and London where the government actively provides the data for free downloads. Business can use the power of these data sources in conjunction with the aforementioned analysis to get a head-start in understanding the landscape of demographics of customers as well as competitors.

**5. Future Directions**

There are more factors that can definitely be incorporated into the analysis including a clustering mechanism that helps cluster the Boroughs based on their demographic profiles. This further analysis is done to help segment the Boroughs into their corresponding demographic profiles. A demonstrative piece of the K-mean clustering technique used to segment the Borough has been included in the final presentation to show the possibility of a further analysis that can be undertaken.

The below illustrative shows all Boroughs similar to Camden based on the selected demographic factors as seen by the column names.

Illustrative:

