Segregating Locations Suitable for Coffee Shop Start-up in London

Er Chun Hsiung

June 21, 2020

1. Introduction

1.1 Background

The people in UK are famous tea drinkers, but UK is also has a fast grow coffee market. The past two decades have seen us switched our affections to the bean with coffee consumption steadily narrowing the gap with the traditional cuppa. According to a recent report by the Allegra Group, we Brits now drink around 55 million cups of coffee per day.

So it's no surprise that the café culture is flourishing in the UK, with 80% of people who visit coffee shops doing so at least once a week – and a sizable 16% visit on a daily basis.

Despite drinking coffee might be a cherish experience; opening a coffee shop might be a different case. Like any form of catering business, make a coffee shop successful takes skills, stamina and hard work. Asides from that, choosing a right location is one of the key point to ensure success for a coffee shop start-up.

1.2 Problem

There are many decisive factors in choosing the right location for coffee shop start-up, and might be very challenging. This project is aimed to let the people to choosing the suitable location for coffee shop start-up from groups of neighbourhoods in London, which has been segregated based on decisive factors such as density of coffee shops, population, rental cost etc.

1.3 Interest

This project is targeted to those who are interested to have their coffee shop start-up at the city of London.

2. Data acquisition and cleaning

2.1 Data sources

In order to complete the analysis, location data for the city of London is required. The list of neighbourhoods and boroughs are extracted from Wikipedia. Besides that, we need the distribution of population for each of the neighbourhoods in London. The location of file is at the database of London city hall, which can be retrieved here.

The data for average rental cost of small shops around neighbourhood in London is available at database of London city hall as well. Here is the location of the file.

2.2 Data pre-processing and cleaning

Data downloaded and combined into an excel file for pre-processing. In order to reduce the size of data, I have taken out the neighbourhood which is at outskirt of London. The method is by including only neighbourhood within London Post Town.

Besides that, there is also some lack of data on the rental value for each neighbourhood. In order to have comprehensive data for analysis, I have fill up the empty rental value for neighbourhood by using the rental value for adjacent neighbourhood.

As for coordinates of the entire neighbourhood in London, I have used Geocoder API to obtain the coordinates by providing address name to the API. A map is plotted to check whether the marker is correctly marked on map as per input of the coordinates.

After getting necessary information such as name of neighbourhood, coordinates, I have used Foursquare API to extract the required data for analysis. This time, I have used explore features in Foursquare API to get the recommended venues for each of the neighbourhoods in London.

Below is the map of neighbourhoods in London shown in blue marker:
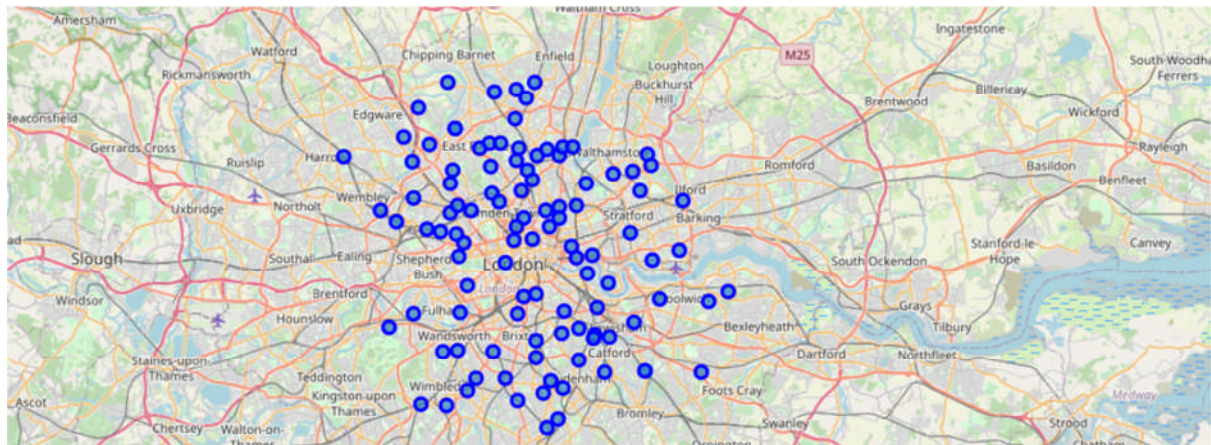
Figure 1: Map for neighbourhood in City of London

2.3 Feature Selection

After retrieving a list of recommended venues in each of neighbourhood in London, I choose to narrow down the data by focusing on coffee shops only. Since my target is only on places which mainly selling coffee, some other places which might also sell coffee like cafe, restaurant, bakery, fast food is omitted from the database.

After feature selection, I finally end up with 175 coffee shops at different location in neighbourhoods of London.

3. Exploratory Data Analysis

3.1 Understanding the pattern of coffee shop distribution for all neighbourhood in London

One of the key pattern in coffee shops are the distribution of coffee shops around neighbourhoods of London. A bar chart has been plotted to understand the pattern.
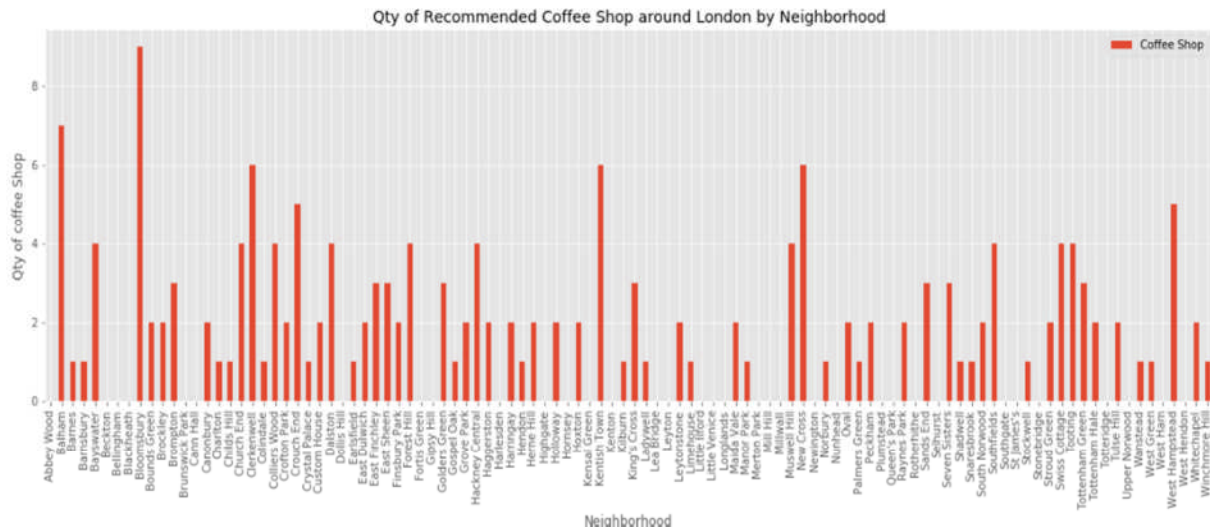
Figure 2: Quantity of recommended coffee chops in London neighbourhood

From this bar chart, we can understand the quantity of coffee shops is roughly at 1 to 2 at each neighbourhood of London, with some exception until 9 coffee shops in single neighbourhood.

### 3.2 Understanding the competition of coffee shops in all neighbourhood in London

Another key point we need to study is to know the competition among the coffee shops in each neighbourhood. In this project, I have decided to focus on competition from the famous coffee shop franchise (such as Costa Coffee, Starbucks etc) which has higher advantage in terms of manpower, funding, supply chain, branding etc.

To understand this, we need to search which coffee shops has the most outlets in London. I have gathered a list showing coffee shops with the highest quantity of outlets as below:

	Count
Venue	
Costa Coffee	32
Starbucks	16
Caffè Nero	11
Brickwood Coffee & Bread	3
Others with less than 3 outlets	113

Table 1: Quantity of coffee chops in London neighbourhood sorted by venue name

A pie chart is plotted to further understand the pattern
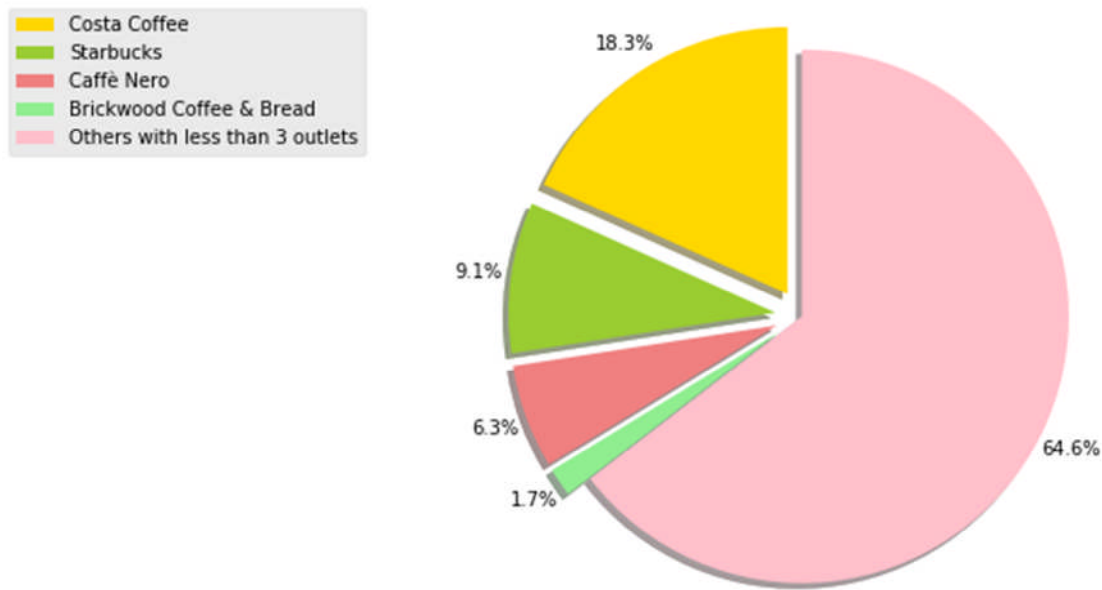
Figure 3: Distribution of coffee shops by name in London neighbourhood

It seems that Costa Coffee constitutes around 18% of total numbers of coffee shops around London, followed by Starbucks (9%), Caffe Nero (6%) and Brickwood Coffee & Bread (1.7%). There are also some other small coffee shops with less than 3 outlets (around 65%).

This data also shows that despite stiff competition from coffee shops franchise, the market is still mainly dominated by non-franchise coffee shops. This shows a sign that the consumers in London are also quite accepting non-franchise coffee shops. From above pattern, it shows that a non-franchise coffee shop start-up in London might sounds feasible.

A bar chart shows distribution of Top 4 franchise coffee shops among neighbourhood in London are plotted as below. The bar chart shows most of the neighbourhood has got at least 1 of Top 4 coffee shop outlets.
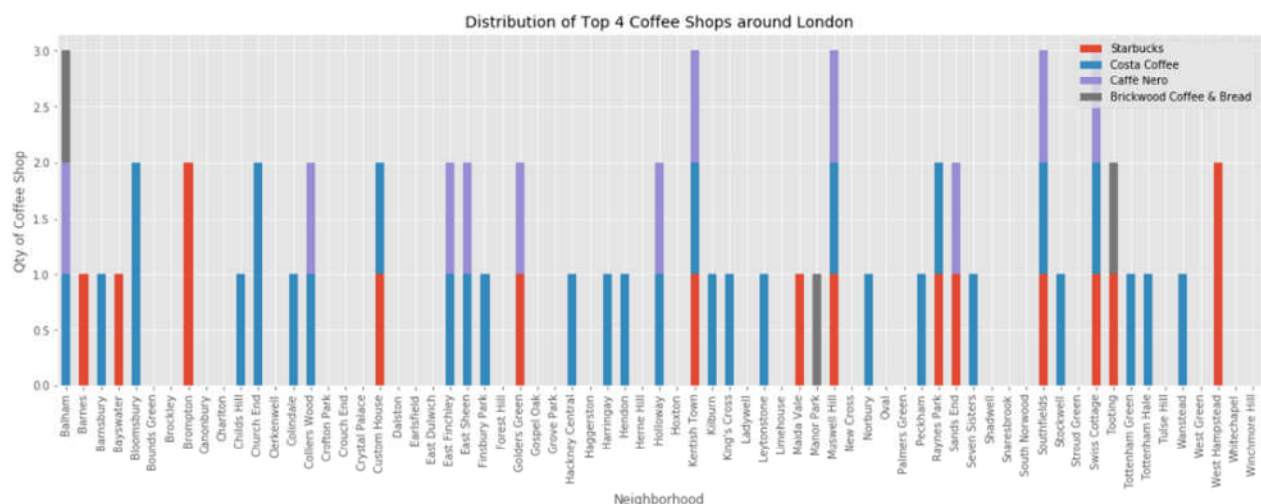
Figure 4: Distribution of Top 4 Coffee Shop Franchise in London Neighbourhood
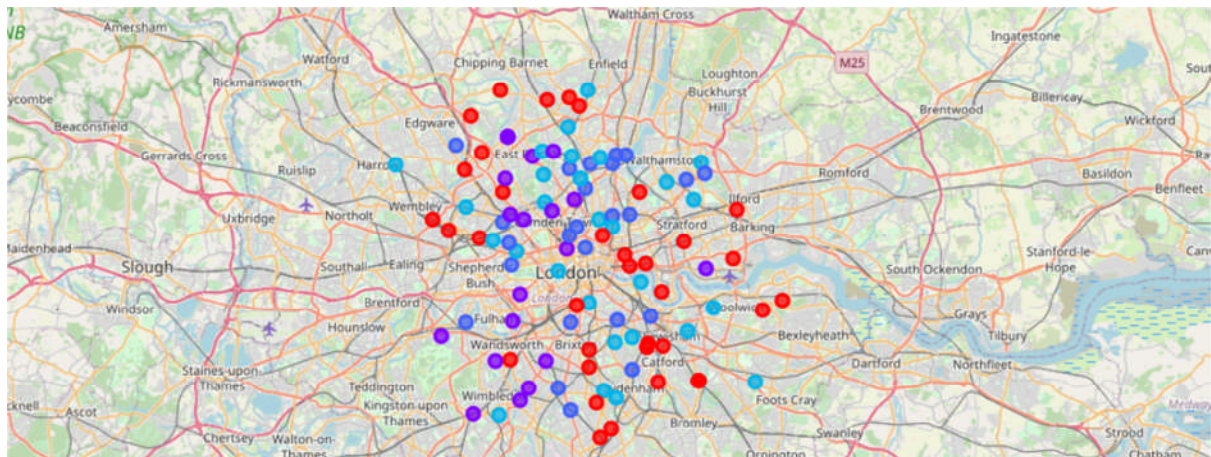
4. Clustering the Neighbourhood of London

4.1 Clustering the Neighbourhood using K-Mean factor

In this project, I had choose to cluster the neighborhood into maximum 4 clusters. The algorithm shall divide the neighbourhoods into 4 clusters according to following priority in decending order.

After running the K-Mean algorithm, a map showing different clusters is plotted.

Figure 5: Map showing 4 clusters of London neighbourhood marked in different colour

Cluster 1 is in red colour, cluster 2 is in purple colour, cluster 3 is in dark blue colour, and cluster 4 is in light blue colour.

4.2 Characteristics of each Cluster the Neighbourhood

Finally, we need to summarize the characteristic of the clusters so that people can choose the location that suits their requirement.

Below is the summary showing the characteristics of each cluster:

Clusters	Coffee Shop Density	Qty of Top 4 Coffee Shop	Population	Rental
Cluster 1	Lowest	Lowest	Middle to High	Low
Cluster 2	Highest	Highest	Low to Middle	Middle
Cluster 3	Middle	Middle	Middle	Middle to High
Cluster 4	Low	Low	Low to Middle	Low to Middle

Table 2: Summary of characteristics for each cluster

A list of neighbourhood for each cluster is also retrieved to let people understand which clusters the neighbourhood belongs to.

	Neighborhood	Cluster Labels	Coffee Shop	Total Top 4	Population	Rental per m2	latitude	longitude
0	Abbey Wood	0	0	0.0	15704	300	51.4876	0.11405
5	Beckton	0	0	0.0	15141	400	51.5161	0.0594257
6	Bellingham	0	0	0.0	14775	275	51.4311	-0.0245145
10	Brockley	0	2	0.0	17156	538	51.4578	-0.0360868
12	Brunswick Park	0	0	0.0	16394	440	51.6262	-0.14819
16	Childs Hill	0	1	1.0	20049	538	51.563	-0.197249

Table 3: Partial List of neighbourhood for cluster 1

	Neighborhood	Cluster Labels	Coffee Shop	Total Top 4	Population	Rental per m2	latitude	longitude
1	Balham	1	7	3.0	14751	969	51.4456	-0.150364
8	Bloomsbury	1	9	2.0	10892	2045	51.5231	-0.126066
11	Brompton	1	3	2.0	8839	1400	51.4918	-0.178326
17	Church End	1	4	2.0	11516	440	51.6011	-0.19189
18	Church End	1	4	2.0	11516	440	51.6011	-0.19189

Table 4: Partial List of neighbourhood for cluster 2

	Neighborhood	Cluster Labels	Coffee Shop	Total Top 4	Population	Rental per m2	latitude	longitude
2	Barnes	2	1	1.0	10299	700	51.4719	-0.238744
3	Barnsbury	2	1	1.0	12201	2368	51.5389	-0.114735
4	Bayswater	2	4	1.0	10300	950	51.5123	-0.188385
21	Clerkenwell	2	6	0.0	11490	2368	51.5237	-0.105555
22	Colindale	2	1	1.0	17098	1023	51.5954	-0.24993

Table 5: Partial List of neighbourhood for cluster 3

	Neighborhood	Cluster Labels	Coffee Shop	Total Top 4	Population	Rental per m2	latitude	longitude
7	Blackheath	3	0	0.0	14039	700	51.4663	0.0085624
9	Bounds Green	3	2	0.0	13725	335	51.6074	-0.124905
13	Cann Hall	3	0	0.0	13799	300	51.5574	0.0156818
14	Canonbury	3	2	0.0	12072	400	51.5438	-0.0906318
15	Charlton	3	1	0.0	14385	1023	51.4826	0.0365359

Table 6: Partial List of neighbourhood for cluster 4

4.3 Analysis on suitable location for coffee shop start-up

Cluster 1 and 4 is a better choice for those who desired to have lower density of coffee shops and coffee shop franchises in one area, with lower rental cost. But in the other hand, some neighbourhoods in these 2 clusters do not have any coffee shops. So this points to a risk where the people for neighbourhoods in these 2 clusters do not favour much on coffee.

Cluster 2 is the area congested with coffee shops and coffee shop franchises, thus the competition of coffee shops in these areas is at the highest level. But from the other point of view, cluster 2 might have highest number of coffee lover population. Therefore most of coffee shops are located in these neighbourhoods.

Cluster 3 has the characteristics between cluster 1,4 and cluster 2. Since there is already some amount of coffee shops established in the neighbourhoods of cluster 3, means there is already some population in these areas having habits in drinking coffee. Starting a coffee shop within neighbourhoods in cluster 3 might have higher chance of success.

5. Conclusion

From the segregation of neighbourhoods in London based on some important criteria such as density of coffee shops in 1 area, competition analysis based on quantity of coffee shop franchises, population and rental, one can judge which location might have higher chance of success in starting a coffee shop business.

But there are lots more to be done in choosing the correct locations. Some examples like understanding the accessibility of the shop locations, understanding the surroundings of the shops (whether mainly is industrial area, or a residential area, or a corporate office area, a recreational area etc.).

6. Future Directions

In order for this segregation model to be more accurate, a more updated data for rental is required. This might help in having a more accurate clustering process. Besides that, as this model is built based on neighbourhood locations, one might found it difficult to choose from multiple location in the same neighbourhood. Therefore a model which further clustering multiple location in one neighbourhood can improve this situation.

Besides that, further refinement of population data might be helpful in getting more accurate clustering results. In case data for coffee drinker population distribution among neighbourhoods of London is available, bringing these data into model shall create huge improvement on clustering result.