Capstone Project: The Battle of Neighborhoods

New Coffee Shop Location Analysis in London, UK

Created by: Nur Agus Permadi Anwar (bagus.permadi@yandex.com)
February 9, 2021

This project will be divided into 5 major segments:

1. Introduction: I will explain the rationale why I chose the specific topic for this project,
2. Methodology: In general, I will describe about the steps performed to work and analyze the data,
3. Discussion: This is the main part where I elaborate further about the analysis findings,
4. Conclusion: I try to summarize all of the main points of the analysis and how it help to solve the problem.
5. References.

1. Introduction

1.1 Overview

The United Kingdom (UK) is traditionally known as a tea-drinking nation. But, in 2018, the had UK became the fourth-largest coffee market in Europe (EU28 countries) after Germany, France, and Italy, based on market size with an annual consumption value worth 1.9 billion euros. This value was translated into a total coffee consumption volume of around 74.600 tonnes (1). In terms of coffee consumption per capita however, the UK's number is only stood at 1.09 kilogram (kg), which was ranked far below Finland, the Europe's country with the highest per capita coffee consumption at 8.54 kg. This fact tells us about the UK coffee market relative upside potential in comparison with other European markets. The increasing coffee consumption as part of modern urban lifestyle, especially in major cities has been contributed heavily to this trend.

London is the capital and largest metropolis city of England and the UK. The city is divided into 33 districts which consist of 32 boroughs and the City of London Corporation (2). Currently there are more than 9 million people who live in London. The city has a population density of approximately 5,900 people per square km (3). London is also an attractive city for businesses and investors to make investment in a consumer-oriented food and beverages service businesses, such as coffee shop. Why so? Because first of all, London is one of the top financial centres and business hubs in Europe and the UK that housed many headquarters of major global corporations. Due to these characteristics, there are a large number of people working in the city that can be accessed as potential customers for business. Second, the city also has an investor friendly regulatory framework and tax regime, in addition to business property market that is geared to the investor. All of which provides the reasons to choose the city as investment destination for business.

I made this analysis primarily with the perspective of investors who are interested to open a coffee shop in London. It is very important to consider the ideal location to establish the future coffee shop. This project will start with simple analysis such as determining which boroughs have the most attractive features and parameters that can help the decision-making process to select the location candidate. The primary target market for the coffee shop will be offices employees and government workers, in addition to the local population and tourists. To help answering these questions, I will perform analysis which include the creation of tables, maps, charts, and cluster analysis with K-nearest neighborhood (including the cluster evaluations).

1.2 Problem Statement

I narrowed down the main problem as follows:

- **"Where is the most potential location to open a coffee shop in London?"**

1.3 Data Description & How It Will Be Used

To support the analysis, I used data from the following sources:

1. I obtained the postal code and the information of London boroughs from Chris Bell's website that is covering UK postcodes, map tools and some bits of code (4),
2. The London Datastore by the Greater London Authority or GLA (3): GLA data are the main source data for this analysis. Data such as information on boroughs, population data, income, workday population, and information on London businesses.
3. I used the geo spatial data in the format of JSON file format to draw the boundaries in the choropleth maps. I downloaded the files from Stuart Grange (5) and Martin Chorley (6) websites.
4. Nominatim API from Python's GeoPy is used to obtain the coordinates of London (7).
5. Foursquare API is used to show the location of venues in London based on the coordinates and to calculate the number of coffee shops in London. The data from the API is used further in the analysis and in the creation of choropleth maps to visualize the data (8).

2. Methodology

2.1 Data Sources & Collection

The analysis used three main datas as follows:

1. "daytime_population.csv" as the second dataset,
2. "business_Data.csv" as the third dataset,
3. "london-postal-codes.csv" as the first dataset,

I downloaded all of the above files separately, and keep them in my local hard-drive. I renamed all them to make it shorter and easier to import to Pandas' data frame. The reason I did it this way was because I was having difficulty to read the data directly from the website using Pandas'read_csv(). The original links to all of the three files are below:

1. daytime_population.csv
2. business_Data.csv
3. london-postal-codes.csv

2.2 Data Cleaning & Feature Selection

After all files are downloaded, each file will be check separately for the encoding method in order to avoid error when loading them to the Pandas' data frame. There is analysis that employed a combination of datasets to create bar charts and to provide information for the choropleth maps.

The data cleaning process began by inspecting the data frame manually and using the **head(), shape**, and **info()** method. The column headers were inspected next. I removed the leading and trailing white spaces from the column headers, replaced the 'in-between' spaces with underscores, and renamed certain headers. The process was continued by selecting only the necessary columns to be used in the analysis, and to remove all the redundant and problematic features (columns).

Each dataset is mostly grouped based on borogh or neighborhood (ward). Columns with numbers are formatted to float or int data types, exception was made when necessary. The index numbers were mostly reset to streamline the alphabetical order of the boroughs, and in certain data frame to to reflect the applied sorting method. This grouping made all the three datasets and the new data frames created from them are all somewhat resemble one another. The processes are done mostly by using **groupby()** method. The steps performed are more or less the same to clean the data, and are replicated all datasets.

Dataset 1

The focus of the work in dataset 1 is to create a data frame that consist of features as follows: borough names, the number of workday populations in each borough, both including and excluding tourist. Since the main target consumers of the coffee shop are office employees in addition to the local population and tourist, the data of workday population for each borough becomes very important. After a data frame that consist of workday population is created, it will be processed further in the initial analysis steps later on to provide information for the construction of bar chart and choropleth map in order to visualize which borough or area that has the most workday population.

	Borough	Workday_population_include_tourists	Workday_population_exclude_tourists
0	City of London	553103	431384
1	Barking and Dagenham	178326	164584
2	Barnet	356003	331094
3	Bexley	211551	194807
4	Brent	293859	274896

Dataset 2

The focus of the work in dataset 2 is to create a data frame that consist of features as follows: the number of active business, and the two-year business survival rates. The data on the number of active business is important, as it supplemented our knowledge about the number of workday populations per borough. The second data frame will give a better picture to understand which area has the most potential in terms of the coffee shop's potential target market.

The two-year business survival rates is included because initially, I wanted to see which areas have the lowest survival rates. However after further inspection, the rates across the boroughs are not significantly different from one another. This particular data however, might change in the future, which can inform us about the changing business in each of the London's borough.

	Borough	Number_of_active_business	Two_year_business_survival_rates
0	City of London	26130	64
1	Barking and Dagenham	6560	73
2	Barnet	26190	74
3	Bexley	9075	74
4	Brent	15745	74

Dataset 3

The focus of the work in dataset 3 pre-processing was to create a data frame that consist of features as follows: borough names, ward (will be rename to 'neighborhood), average annual income per borough (total in GBP), and geo location data in the forms of latitude and longitude data. The process was simply done using Pandas library. The primary reason for this analysis was to see which borough has the most average income. This information when combine with other data such as the number of workday populations, is one of the important decision-making factors which borough the investor should establish the coffee shop.

	Borough	Neighborhood	Avg_Income	Latitude	Longitude
0	Barking and Dagenham	Thames, Gascoigne, Eastbury, Abbey, Longbridge...	42,599.87	51.546453	0.124953
1	Barnet	Oakleigh, High Barnet, East Barnet, Brunswick ...	55,496.79	51.605511	-0.207739
2	Bexley	Crayford, Barnehurst, Sidcup, Blackfen & Lamor...	50,447.16	51.459192	0.136353
3	Brent	Alperton, Sudbury, Wembley Central, Northwick ...	47,918.76	51.551772	-0.257489
4	Bromley	Bromley Town, Shortlands, Bickley, Plaistow an...	57,706.21	51.391808	0.026273

Note: Dataset 3

- I created a separated data frame ('df3_sort') to provide the basis of the first two choropleth maps creation.
- I tried to obtain the unique names of all the neighborhoods and grouped them all based on their own respective borough.
- I added the column 'Average Income' initially to be used in the analysis. But later on I found out that there were other factors that weigh more on decision about location.

2.3 Initial Analysis

Purpose of the Initial Analysis:

The purpose of the initial analysis is to find the best location to establish a coffee shop in London. I will use two main criteria to determine which location has the most potential. Those data are 'the number of workday populations', and 'the number of active businesses'. These two data I believe will be representative enough to gauge the number of potential customers for the coffee shop. As previously mentioned, the coffee shop is planned to target employees, government workers, as well as the local population and tourists.

I began the initial analysis by importing the necessary Python libraries for geo data, creating API connection, image display, and map visualization. Then, I proceed to obtain London's geographical coordinates (latitude and longitude data) using geopy. These coordinates will be used to display the map of London.

Once the location candidate is chosen, I will then use the Foursquare API to see the competitive landscape in that particular location (neighborhood/ward). I will pay attention to the number of coffee shops in the area, as well as the number of cafe. Cafe is a close substitute for coffee shop, but cafes are mostly focusing more on their food menus, while coffee shop focuses more on their beverage menus (particularly coffee!).

2.3.1 Find Borough with the Highest Number of Workday Population¶

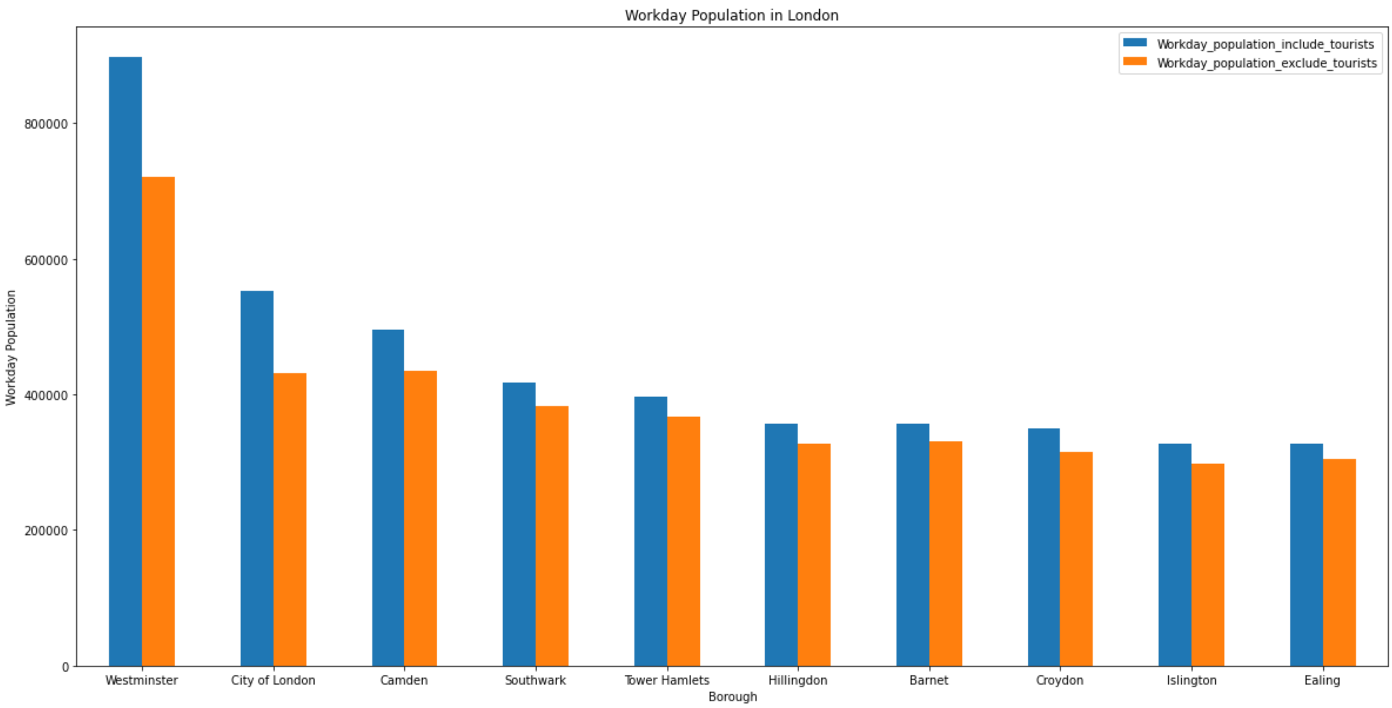
I sorted the dataset 1 ('df1') using sort_values() in descending order based on the workday population number (include tourists) in order to see which borough has the greatest number of potential customers for the coffee shop. The difference in the result is significant. We can see that even in the top 10 rank, Westminster topped other boroughs in both numbers (with or without tourists) with quite considerable gaps.

If we had to make decision based on this data alone, we can confidently say that Westminster is the perfect location for the future coffee shop to access the large number of potential customers. However, we need to check for other data. Next, I created a bar chart using the matplot library plot() function to visualize the number of workday populations in the top 10 boroughs. We can see how the bar that represents Westminster stood tall compares to others..

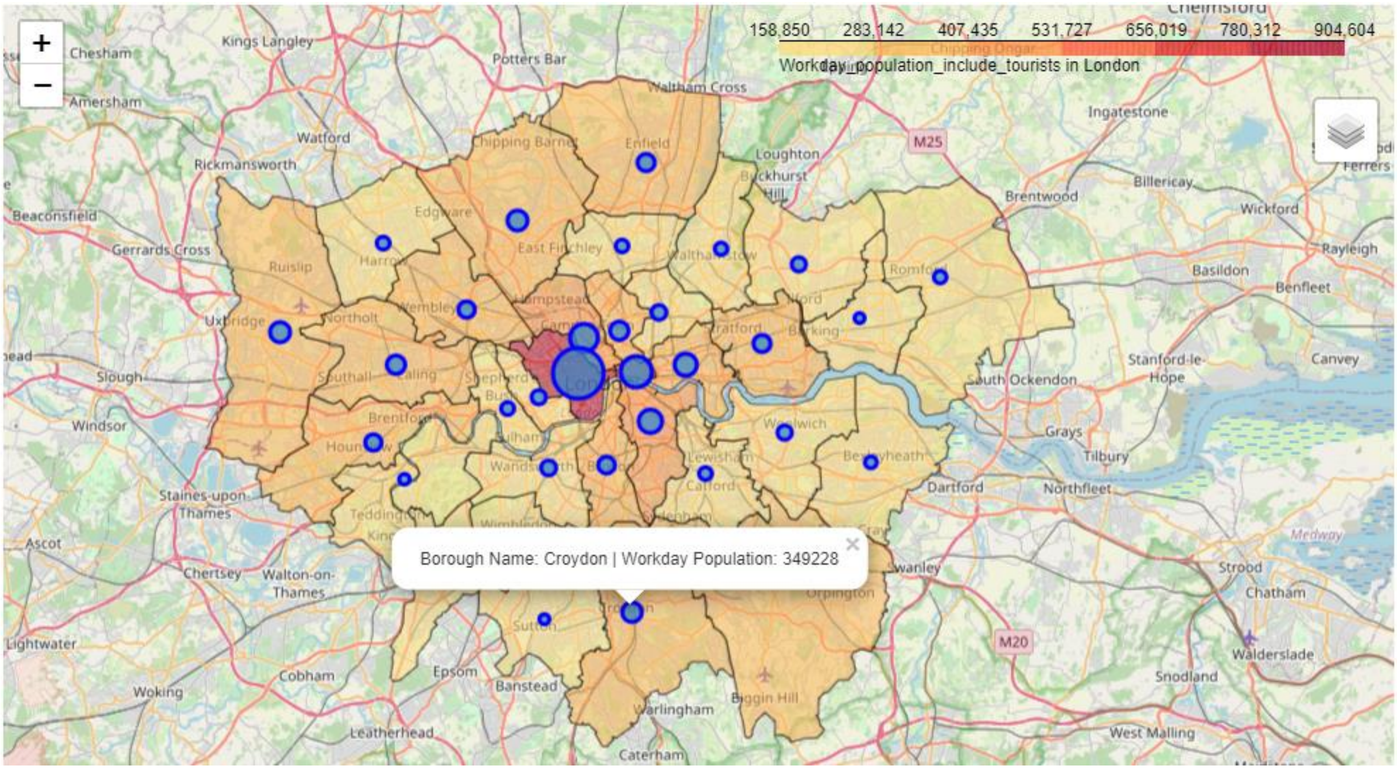
Table 1: Top 10 Boroughs based on Workday Population

	Borough	Workday_population_include_tourists	Workday_population_exclude_tourists
0	Westminster	897293	721351
1	City of London	553103	431384
2	Camden	495332	434279
3	Southwark	417029	382582
4	Tower Hamlets	396939	367677
5	Hillingdon	357295	327032
6	Barnet	356003	331094
7	Croydon	349228	314819
8	Islington	328050	297460
9	Ealing	327625	305316

Bar Chart 1: Top 10 Boroughs based on Workday Population



Choropleth Map 1: London Borough, based on Workday Population (include Tourists)



I created a choropleth map using Folium library that shows the color intensity in London borough based on the number of workday population (include tourists), I also superimposed the clusters on the map to provide information on the name of each borough, In this choropleth map, we can see the color of Westminster is in dark red. It signifies the biggest number of workday population relative to other boroughs. It confirms what the chart shows, but for all boroughs in London.

Note: Create a data frame to support the choropleth map 1

The main difference of this new data frame compare to the previous one used to construct the bar chart is this data frame is merged with the dataset 3 in order to show the coordinates (latitude and longitude) of each borough. The purpose is to enable the choropleth to show the color intensity of each borough based on the number of workday population. The meaning of each different gradient of colors are explained in the legend on the top right corner of the map. The markers and that are added to the map also shows different marker size for each burough based on the number of workday population. The label marker displays the actual number of workday population.

2.3.2 Find Borough with The Highest Number of Active Business

I sorted the dataset 2 ('df2') using sort_values() in descending order based on the number of active businesses in order to gauge the area potential to generate potential customers, that is the office employees as well as government workers for the coffee shop.

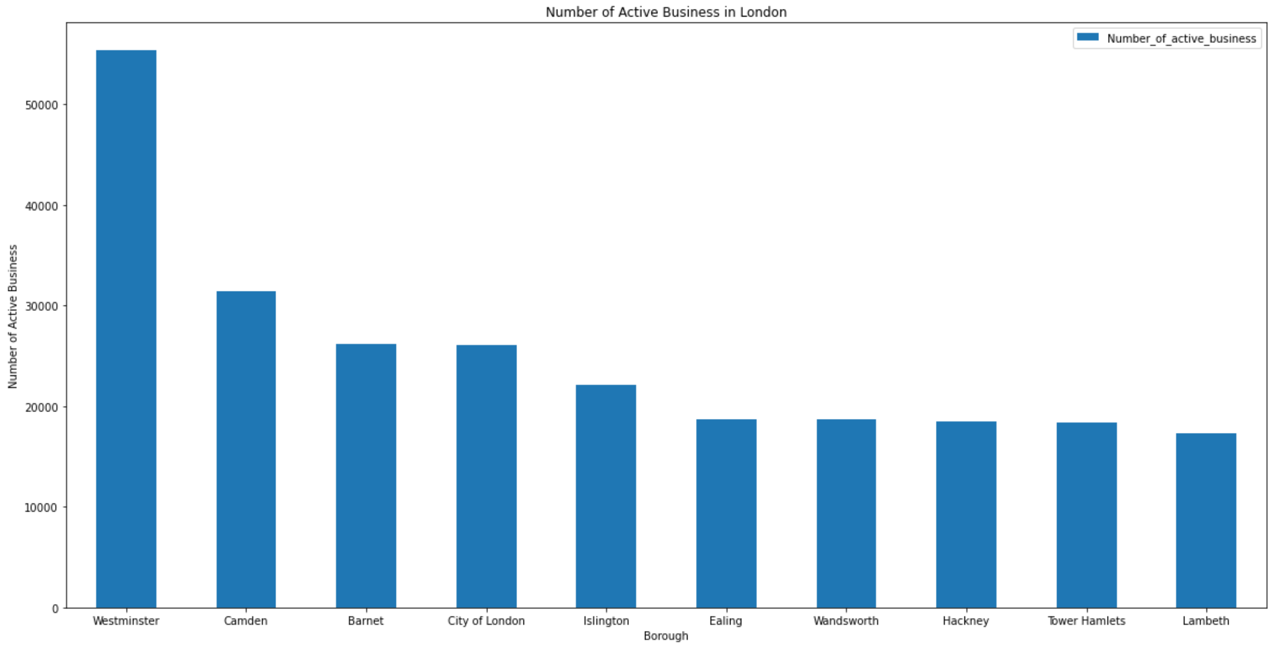
The result shows the same outcome to the previous analysis. Westminster remains at the top. Westminster has a considerably larger number of active businesses compare to any other boroughs, which means this area is presumably have a higher number of workers and employees. Similar to the previous analysis, if we had to make decision based on this data alone, Westminster is the clear winner.

On to the next step, I created two bar charts using the matplot library plot() function to visualize the number of active businesses in the top 10 boroughs and the two-year business survival rates. The first bar chart confirms the findings in the table 2.

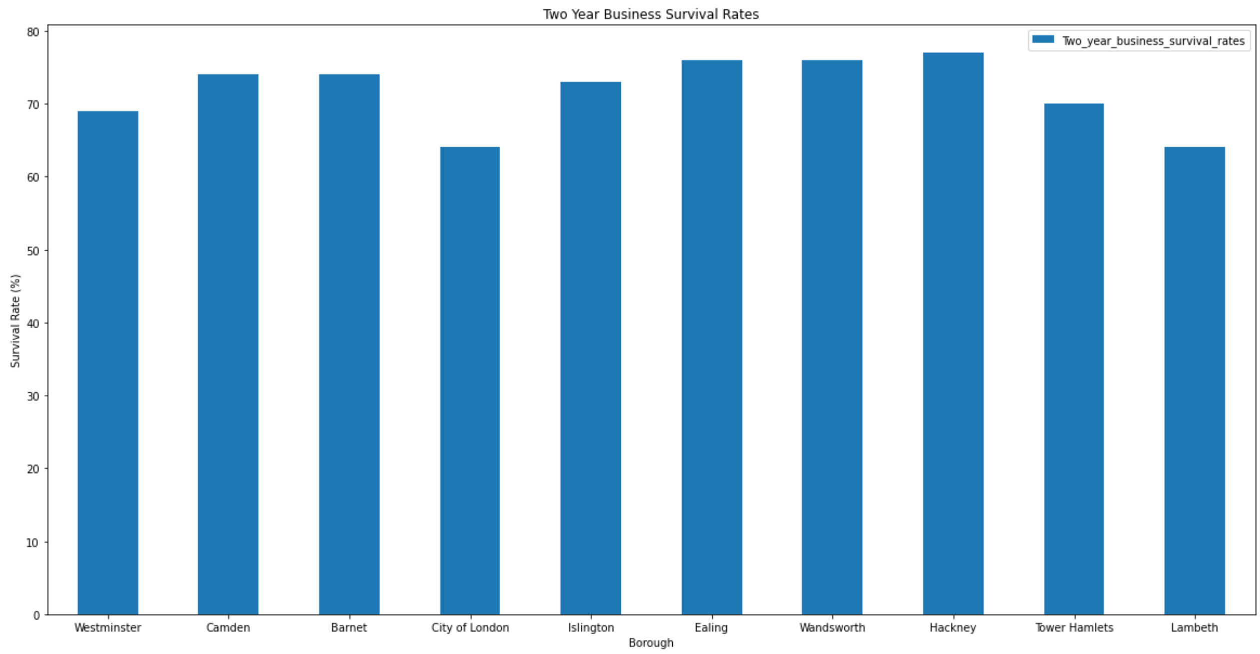
Table 2: Top 10 Boroughs based on Active Businesses

	Borough	Number_of_active_business	Two_year_business_survival_rates
0	Westminster	55385	69
1	Camden	31385	74
2	Barnet	26190	74
3	City of London	26130	64
4	Islington	22110	73
5	Ealing	18700	76
6	Wandsworth	18695	76
7	Hackney	18510	77
8	Tower Hamlets	18390	70
9	Lambeth	17280	64

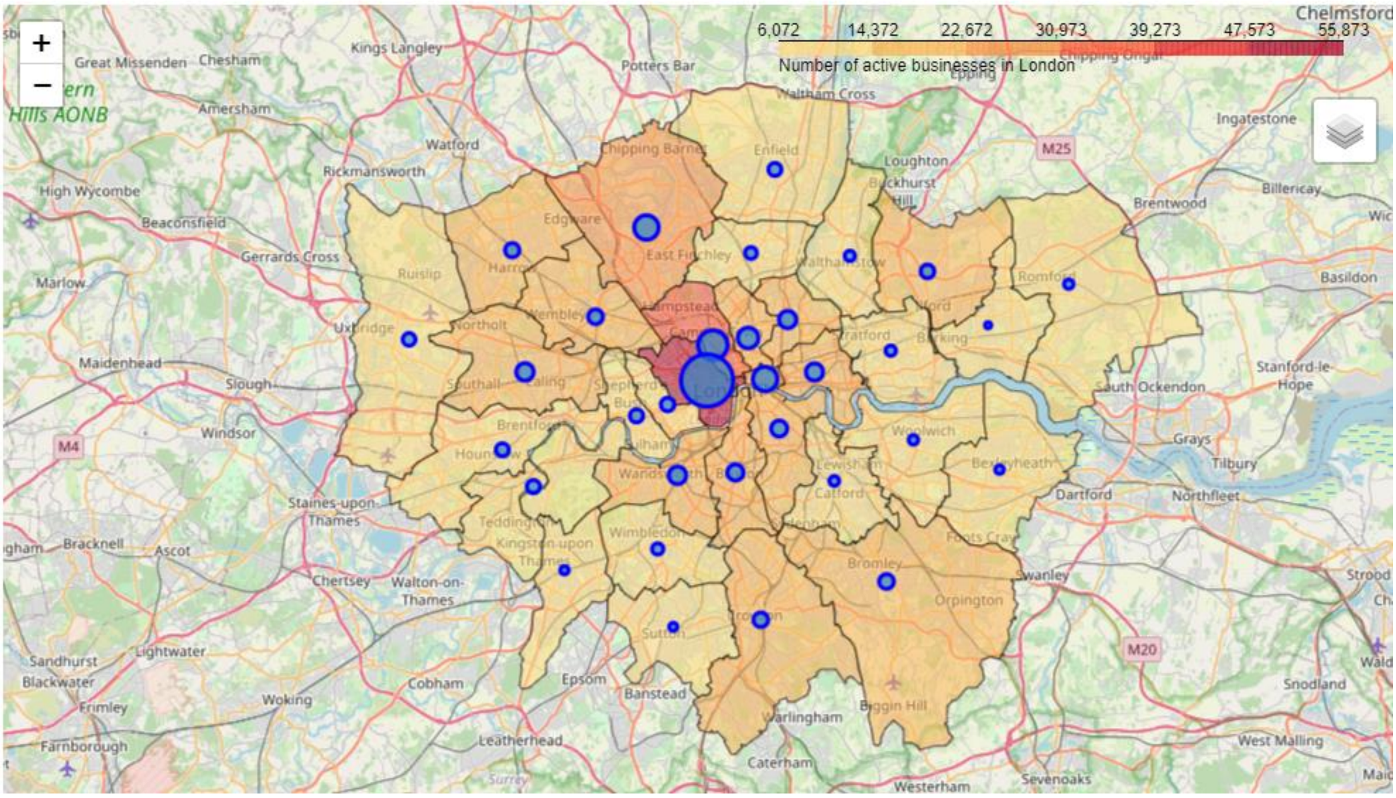
Bar Chart 2: Top 10 Boroughs based Active Business



Bar Chart 3: Top 10 Boroughs based on Business Survival Rate



Choropleth Map 2: London Borough based on Number of Active Businesses



Similar with the first choropleth map, I used Folium library to create a map that shows the color intensity in each borough based on the number of active businesses, I superimposed the clusters on the map to provide information on the name of each borough,

Once again, the color of Westminster is in dark red because it has the highest number of active businesses compares to any other boroughs.

Note: Create a new data frame to support the choropleth map 2

As mentioned before in the 'Note: Create a data frame to support the choropleth map 1', this new data frame for choropleth map 2 is created with relatively the same processes. The difference is in the data that are used to create it. I merged the Table 2 data with the dataset 3 data frame to make the choropleth show the color intensity of each borough based on the number of active businesses.

2.3.3 Summary: Initial Analysis

The result from the two initial analysis have indicated that Westminster is the ideal location to establish the coffee shop. The significant difference with other boroughs in terms of both workday population and number of active businesses becomes a very decisive factor. On the other hand, the level of average income and the two year survival rate were not chosen because they don't really seem to have the similar impact compare to the first two factors. With this result, I decided to choose Westminster as the area candidate to establish the coffee shop. Next, I will explore the the neighborhoods in Westminster further and perform cluster analysis to determine the attractiveness of each neighborhood.

For the next step, I will create a new data frame specifically for Westminster data. I will group the data based on Westminster unique name of neighbourhood data, to avoid having to process more than 30,000 rows on location data. After the unique neighborhood data has beed generated, I drop the borough name. I did this to focus to the neighborhood data, and achieve the level of granularity that matched with Foursquare API data. The other reason was there is a limit to the number of data that can be called from Foursquare API with my free account.

2.4 Foursquare API: Create Connection & Data Retrieval

To get the latitude and longitude data of Westminster, London, I used the geopy library, specifically the Nominatim API to convert an address into latitude and longitude values. The geo coordinates data then used to create a map of Westminster. I used python folium library to visualize geographic details of Westminster.

```
#Foursquare use the term 'Neighborhood' instead of ward!
address = 'Westminster, UK'

geolocator = Nominatim(user_agent="westminster_explorer")
location = geolocator.geocode(address)
lat = location.latitude
lng = location.longitude
print('The geograpical coordinate of Westminster are {}, {}.'.format(lat, lng))
```

```
The geograpical coordinate of Westminster are 51.5004439, -0.1265398.
```

3. Discussion

3.1 Transforming Foursquare API data to a Data Frame

The Foursquare API in the previous segment 2.4 is used to explore all the venues available in the neighborhoods of Westminster. The API connection has a limit of 100 venues per neighborhood. I define the 'search' radius to 500 meters for each neighborhood. The Foursquare API will provide the information of each venues in the neighborhood including their geo coordinates data (latitudes and longitudes).

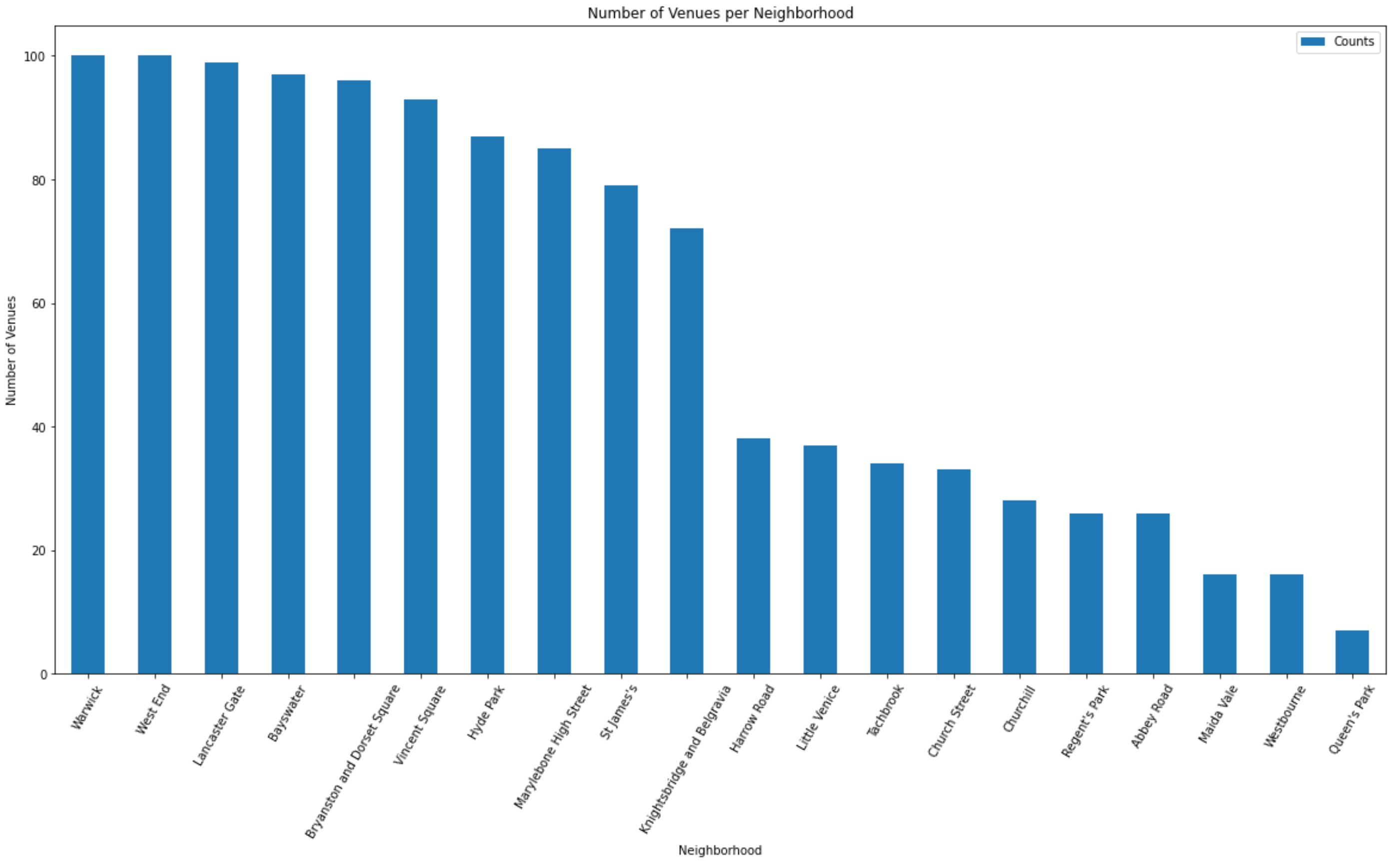
A function called 'getNearbyVenues' is utilized to extract information about venues name and venues geo location from the json file returned from the API. I will then proceed to create a new data frame from the combination of the venues name and geo location with information on Westminster's neighborhood names and its respective geo coordinates (stored in another data frame 'df3_west'). I will name the newly created data frame as 'westminster_venues'.

	Neighborhood	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
0	Abbey Road	Pub	Café	Coffee Shop	Bakery	French Restaurant	Cricket Ground	Gym / Fitness Center	Deli / Bodega	Grocery Store	Thai Restaurant
1	Bayswater	Pub	Café	Persian Restaurant	Gym / Fitness Center	Greek Restaurant	Coffee Shop	Bakery	Garden	Pizza Place	Restaurant
2	Bryanston and Dorset Square	Hotel	Coffee Shop	Sandwich Place	Argentinian Restaurant	Middle Eastern Restaurant	Pub	Restaurant	Juice Bar	Garden	Bakery
3	Church Street	Café	Pub	Coffee Shop	Middle Eastern Restaurant	Garden	Japanese Restaurant	Lebanese Restaurant	Hotel	Fast Food Restaurant	Movie Theater
4	Churchill	Hotel	Italian Restaurant	Deli / Bodega	Coffee Shop	Café	Theater	Sandwich Place	Bakery	Plaza	Gastropub

3.2 Sort the neighborhoods based on the total number of venues & plot into a bar chart

Neighborhood	Counts	Venue
Warwick	100	Pimlico Fresh, FRAME, Artist Residence, Leon, ...
West End	100	Carnaby Street, Goodman Steakhouse, Dishoom, L...
Lancaster Gate	99	Boots, Halepi, Gold Mine, Amorino, Santorini T...
Bayswater	97	JOE & THE JUICE, LOKKANTA, GAIL's Bakery, Plan...
Bryanston and Dorset Square	96	Briciole, Dinings, The Z Hotel Gloucester Plac...
Vincent Square	93	Iris & June, Curzon Victoria, Strutton Ground ...
Hyde Park	87	Paramount Lebanese Kitchen, M&S Simply Food, F...
Marylebone High Street	85	Daunt Books, La Fromagerie, Le Relais de Venis...
St James's	79	National Gallery, Trafalgar Square, East Trafa...
Knightsbridge and Belgravia	72	Chanel Boutique, L'ETO Caffè, Jimmy Choo, Otto...
Harrow Road	38	Mosob, George's Fish Bar, Tsiakkos & Charcoal,...
Little Venice	37	Gogi, LV Lounge, Clifton Nurseries, Real Ale, ...
Tachbrook	34	CASK Pub and Kitchen, Khallouk & Taylor, Deliz...
Church Street	33	Roof Top Kitchen, Indaba, Church Street Market...
Churchill	28	Artist Residence, The Locals Cafe, Omar's Plac...
Regent's Park	26	St John's Wood Church Gardens, Lord's Cricket ...
Abbey Road	26	Abbey Road Studios, Abbey Road Crossing, Heart...
Maida Vale	16	Paddington Recreation Ground, The Elgin, GAIL'...
Westbourne	16	Harbour Club Notting Hill, Tsiakkos & Charcoal...
Queen's Park	7	Yogaloft - Beethoven Street, Fierce Grace West...

Bar Chart 4: Number of Venues per Neighborhood



3.3 Sub-Analysis: Coffee Shops and Cafes in Westminster

We can see that the new 'westminster_venues' data frame has 1169 venues that are categorized into unique 193 categories. The data frame also contains the geo location data for each of the venue. Before I proceed further with clustering analysis to all of this data, I will create a subset of data from the 'westminster_venues' data frame in order to find out specifically about the number of coffee shops and the number of cafes in each neighborhood of Westminster. The purpose is to visualize the competitive landscape in each neighborhood. This sub analysis can provide useful insights about the decision-making process at a later-stages, especially in the appearance of new other data.

Neighborhood	Avg_Income	Latitude	Longitude	Coffee_Shops_Number	Coffee_Shops_Names	Cafes_Number	Cafes_Names
Abbey Road	58,576.86	51.534412	-0.178185	6.0	Starbucks, Beatles Coffee Shop, Caffè Nero, Co...	8.0	The Peppermint, Boulevard, Raoul's, Fego Caffe...
Bayswater	56,673.68	51.516727	-0.191849	3.0	Arro Coffee - the Temple of Coffee, Amoret Spe...	7.0	TAB x TAB, Haminados, Jusu Brothers, Churreria...
Bryanston and Dorset Square	62,578.10	51.518843	-0.161807	6.0	Caffè Nero, Starbucks, Costa Coffee, Arro Coff...	2.0	The Monocle Café, Daisy Green Food
Church Street	39,800.00	51.523680	-0.169851	5.0	d1, Caffè Nero, Starbucks, Mihbaj Cafe, Costa ...	9.0	Roof Top Kitchen, Darcie & May Green, Beany Gr...
Churchill	50,717.59	51.488757	-0.147685	4.0	The Roasting, Ole & Steen, The Roasting Party ...	4.0	The Locals Cafe, Victoria Cafe, Pimlico Fresh,...
Harrow Road	47,554.90	51.526298	-0.197197	1.0	Starbucks	3.0	The Peppermint, Toast Cafe, Formosina
Hyde Park	56,991.49	51.516236	-0.173652	7.0	Mihbaj Cafe, Caffè Nero, Starbucks, Les Filles...	6.0	Darcie & May Green, Sandro Sandwich Bar, Itali...
Knightsbridge and Belgravia	56,530.63	51.499373	-0.158983	3.0	Starbucks, 39 Steps Coffee Haus, The Roasting ...	7.0	L'ETO Caffè, EL&N, Carpo, Chapati & Karak
Lancaster Gate	60,712.74	51.512922	-0.185634	6.0	Arro Coffee - the Temple of Coffee, Urban Bari...	7.0	Leclerc & Laurent, TAB x TAB, Italian Gardens ...
Little Venice	57,785.78	51.523944	-0.180025	4.0	d1, Mihbaj Cafe, Starbucks, Petit Café	5.0	The Quince Tree, Darcie & May Green, Beany Gre...
Maida Vale	62,475.12	51.529773	-0.189132	5.0	Starbucks, The Coffee Tree, Petit Café, Caffè ...	9.0	The Peppermint, The Quince Tree, Raoul's, Dani...
Marylebone High Street	71,426.32	51.518887	-0.150548	5.0	Workshop Coffee Co., Starbucks, Kiss The Hippo...	1.0	The Monocle Café
Queen's Park	44,816.80	51.529627	-0.208504	8.0	Dark Habit, 333 HotShoe, Cable co., The Coffee...	10.0	Milk Beach, Lowry & Baker, Vicki's of London, ...
Regent's Park	56,618.73	51.529469	-0.168708	3.0	d1, Starbucks, Costa Coffee	12.0	Roof Top Kitchen, Boulevard, The Regent's Bar ...
St James's	58,656.98	51.507343	-0.128691	3.0	Hagen, Monmouth Coffee Company, Rosie & Joe	1.0	WA Cafe
Tachbrook	53,169.82	51.488551	-0.135382	6.0	The Roasting, District, The Black Cab Coffee C...	7.0	Pimlico Fresh, Regency Cafe, Tate Britain Memb...
Vincent Square	56,452.34	51.495303	-0.136006	8.0	Iris & June, Flat Cap Coffee Co, Formative, Th...	6.0	Regency Cafe, Pimlico Fresh, Sapori Café & Res...
Warwick	57,149.91	51.492354	-0.143646	7.0	The Roasting, Ole & Steen, Coffee Addict, Blac...	5.0	Pimlico Fresh, Patisserie Valerie, Regency Caf...
West End	56,552.78	51.513544	-0.140733	0.0	0	0.0	0
Westbourne	47,576.14	51.521871	-0.194398	1.0	Arro Coffee - the Temple of Coffee	6.0	Haminados, Maida Hill Cafe Gallery, The Quince...

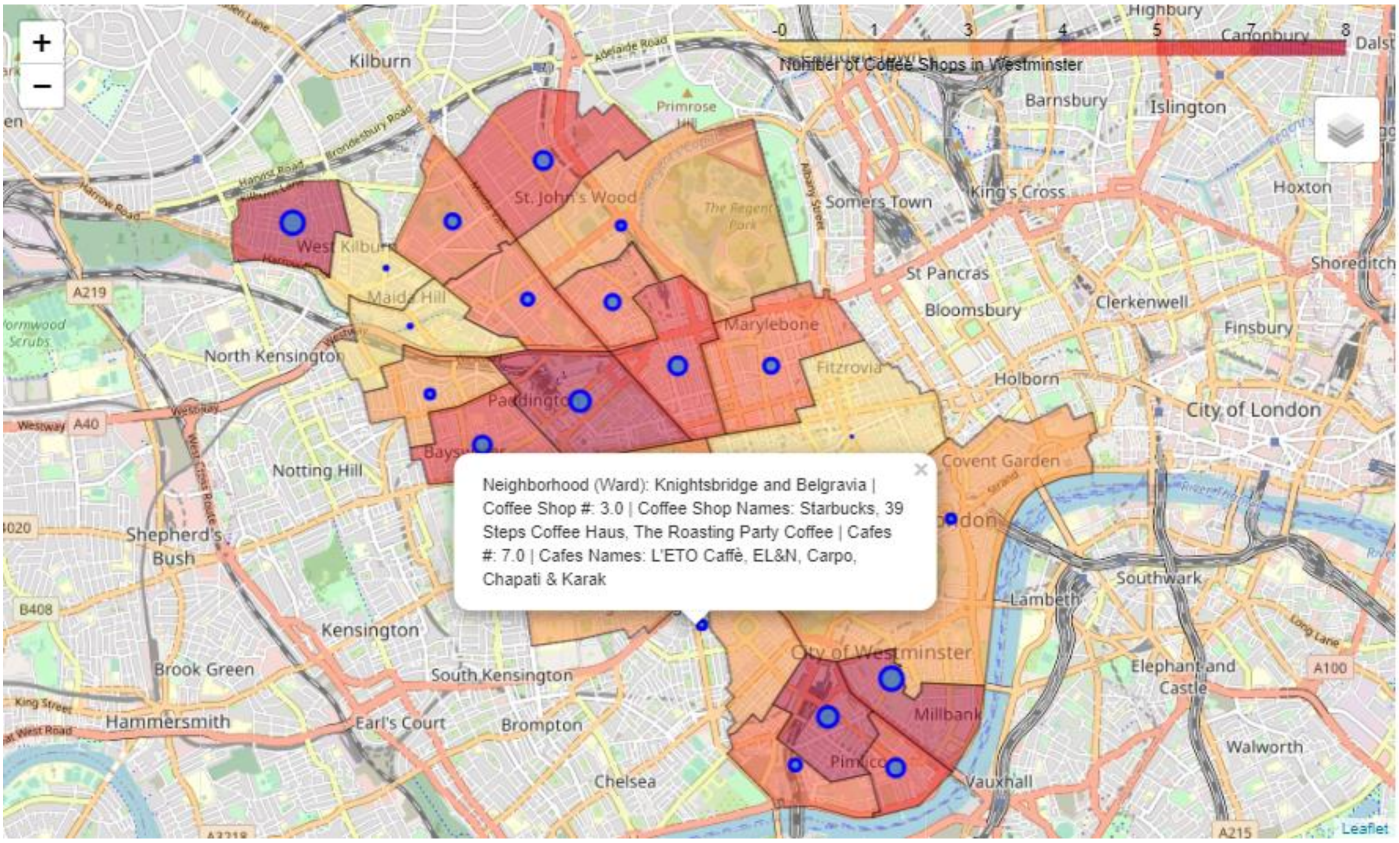
3.2.1 Result of Sub-Analysis

As I mentioned before, I will also pay attention to Cafe despite I am planning to analyze location to establish a coffee shop. The reason is because cafe is a close substitute for coffee shop. However, cafes are mostly focusing more on their food menus, while coffee shop focuses more on their beverage menus (coffee primarily). The sub-analysis has generated several key insights about the competitive landscape as follows:

1. There are a total number of 212 coffee shops and cafes in all Westminster's neighborhoods,
2. There are 97 Coffee shops that make up to 45.75% from the total number,
3. There are 115 Cafes that make up to 54.25% from the total number,
4. Both venue categories (coffee shops and cafes) make up to 18.14% of the total number of venues in Westminster,

5. The top three neighborhoods with the largest number of coffee shops are Vincent's Square and Queens' Park with 8 coffee shops in each neighborhood, followed by Warwick with 7 coffee shops.
6. The bottom three neighborhoods with the smallest number of coffee shops are Harrow Road and Westbourne with 1 cafe in each neighborhood, and followed by Knightsbridge and Belgravia with 3 coffee shops.
7. The bottom three neighborhoods with the largest number of cafes are Regent's Park with 12 coffee shops, followed by Queens' Park with 10 coffee shops, and Maida Value with 9 coffee shops,
8. The bottom three neighborhoods with the smallest number of cafes are Marylebone High Street and St. James' with 1 cafe in each neighborhood, and followed by Bryanston and Dorset Square with 2 cafes.

Choropleth Map 3: Coffee Shops and Cafes in Westminster's Neighborhoods



3.3 Discussion: The Cluster Analysis

Result Overview:

Okay, so based on the result of the previous two analysis (number of workday population and number of active businesses), I have already decided to pick Westminster as the potential area to establish the future coffee shop. In the next step, I increase the granular level of the analysis and ask a question such as: "which neighborhood in Westminster that is ideal to open a coffee shop?".

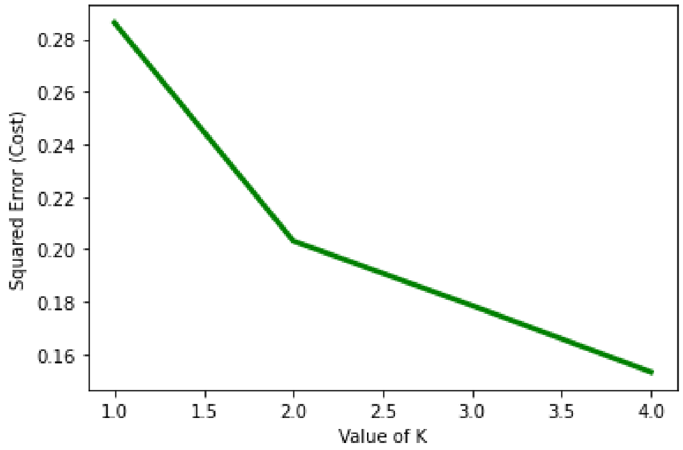
There are __20 neighborhoods (wards)__ in the borough of Westminster. Before we choose a particular neighborhood, it is a very good idea to determine the attractiveness of each neighborhood first. We can perform this type of analysis by processing the incoming data from the Foursquare API to check what types or venues that are available in each neighborhood.

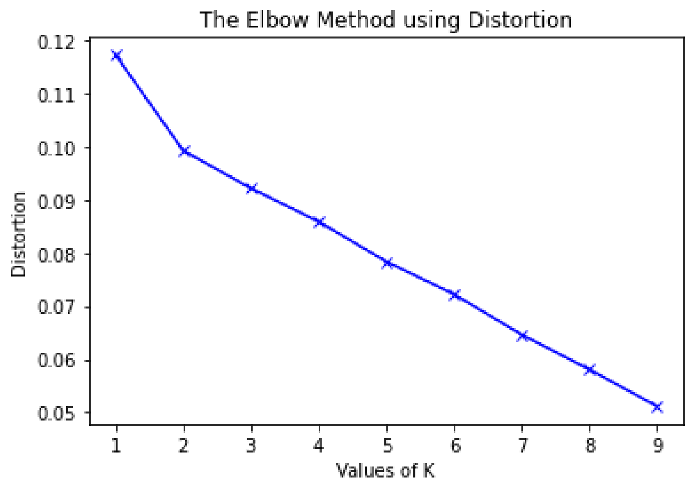
Considering the limit of the Foursquare API, I can only retrieve a maximum of 100 venues per neighborhood. From the total of 1169 venues throughout Westminster, we can see that there are two neighborhood that reached the limit of venues, they are __Warwick__ and __West End__. Lancaster Gate closely behind with 99 venues. On the flip side, the data shows that Queens' Park has the lowest number of venues. In my own opinion, this is caused by the neighborhood area is mainly consist of residentials.

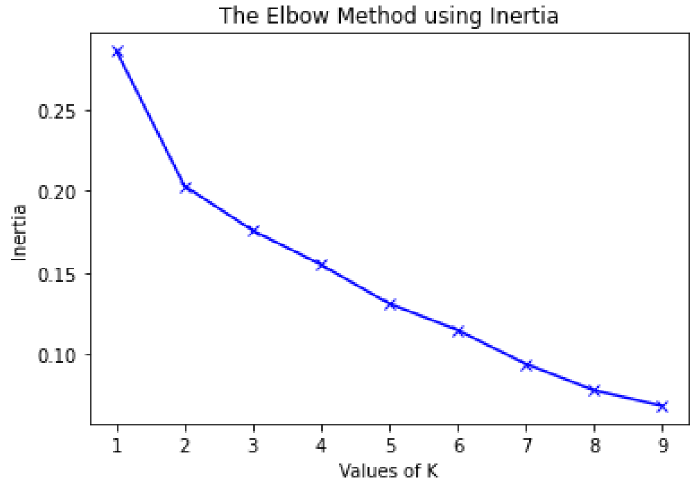
Cluster Analysis:

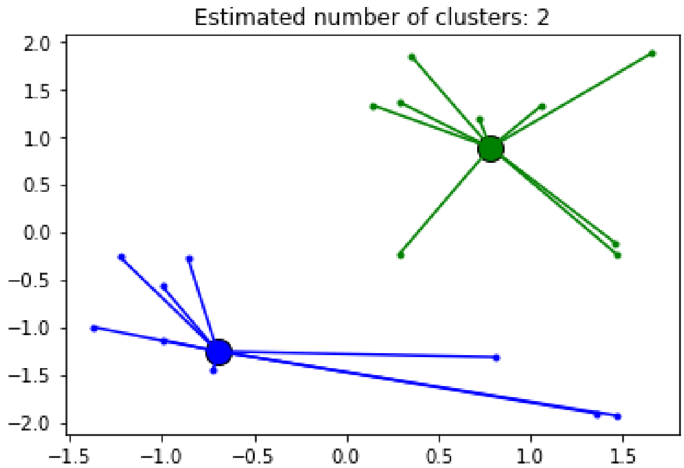
The the type of machine learning algorithm used in this analysis is the 'unsupervised learning' algorithm. Specifically, I used the k-nearest neighbors algorithm (k-NN), which is a non-parametric classification, to classify the neighborhoods in Westminster into cluster areas. It is a predictive classification problems, so k-NN is appropriate for this task. I started with 'k' value = 3 to classify the neighborhood into 3 clusters. After I obtain the result, I ran it into two cluster evaluation methods:

1. The Elbow Method, using 'distortion' and 'inertia', and
2. The Affinity Propagation Method.









The elbow method used that the optimal value for k (#clusters) is 2. The affinity propagation method also shows the same result with optimum k value equals to 2. Afterwards, I ran the k-NN once again, but this time I use the value of k = 2. I plot the result in a choropleth map (Choropleth Map 4: 'The Clusters of Westminster Neighborhoods') that shows the color intensity based on the number of coffee shops in each neighborhood. The clusters of neighborhoods are superimposed on the map a layer above. The cluster members are as follows:

- Cluster 1: Abbey Road, Bayswater, Church Street, Harrow Road, Lancaster Gate, Little Venice, Maida Vale, Queens' Park, Regent's Park, and Westbourne.
- Cluster 2: Bryanston and Dorset Square, Churchill, Hyde Park, Knightsbridge and Belgravia, Marylebone High Street, St James's, Tachbrook, Vincent Square, Warwick, and West End.

Cluster 1

	Neighborhood	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
0	Abbey Road	Pub	Café	Coffee Shop	Bakery	French Restaurant	Cricket Ground	Gym / Fitness Center	Deli / Bodega	Grocery Store	Thai Restaurant
1	Bayswater	Pub	Café	Persian Restaurant	Gym / Fitness Center	Greek Restaurant	Coffee Shop	Bakery	Garden	Pizza Place	Restaurant
3	Church Street	Café	Pub	Coffee Shop	Middle Eastern Restaurant	Garden	Japanese Restaurant	Lebanese Restaurant	Hotel	Fast Food Restaurant	Movie Theater
5	Harrow Road	Pub	Café	Thai Restaurant	Bakery	Taxi Stand	Coffee Shop	Fish & Chips Shop	Restaurant	Food & Drink Shop	French Restaurant
8	Lancaster Gate	Pub	Café	Coffee Shop	Hotel	Garden	Chinese Restaurant	Greek Restaurant	Persian Restaurant	Bookstore	French Restaurant
9	Little Venice	Café	Coffee Shop	Grocery Store	Canal	Pub	Gym / Fitness Center	Bar	Italian Restaurant	Beer Bar	Gym
10	Maida Vale	Pub	Café	Bakery	Coffee Shop	Grocery Store	Pizza Place	Thai Restaurant	Middle Eastern Restaurant	French Restaurant	Deli / Bodega
12	Queen's Park	Café	Pub	Pizza Place	Coffee Shop	Grocery Store	Gym	Yoga Studio	Middle Eastern Restaurant	Bakery	Garden
13	Regent's Park	Café	Pub	Cricket Ground	Coffee Shop	Grocery Store	Gym / Fitness Center	Middle Eastern Restaurant	Bakery	French Restaurant	Restaurant
19	Westbourne	Pub	Café	Persian Restaurant	Gym / Fitness Center	Fish & Chips Shop	Pizza Place	Italian Restaurant	Garden	Restaurant	Bakery

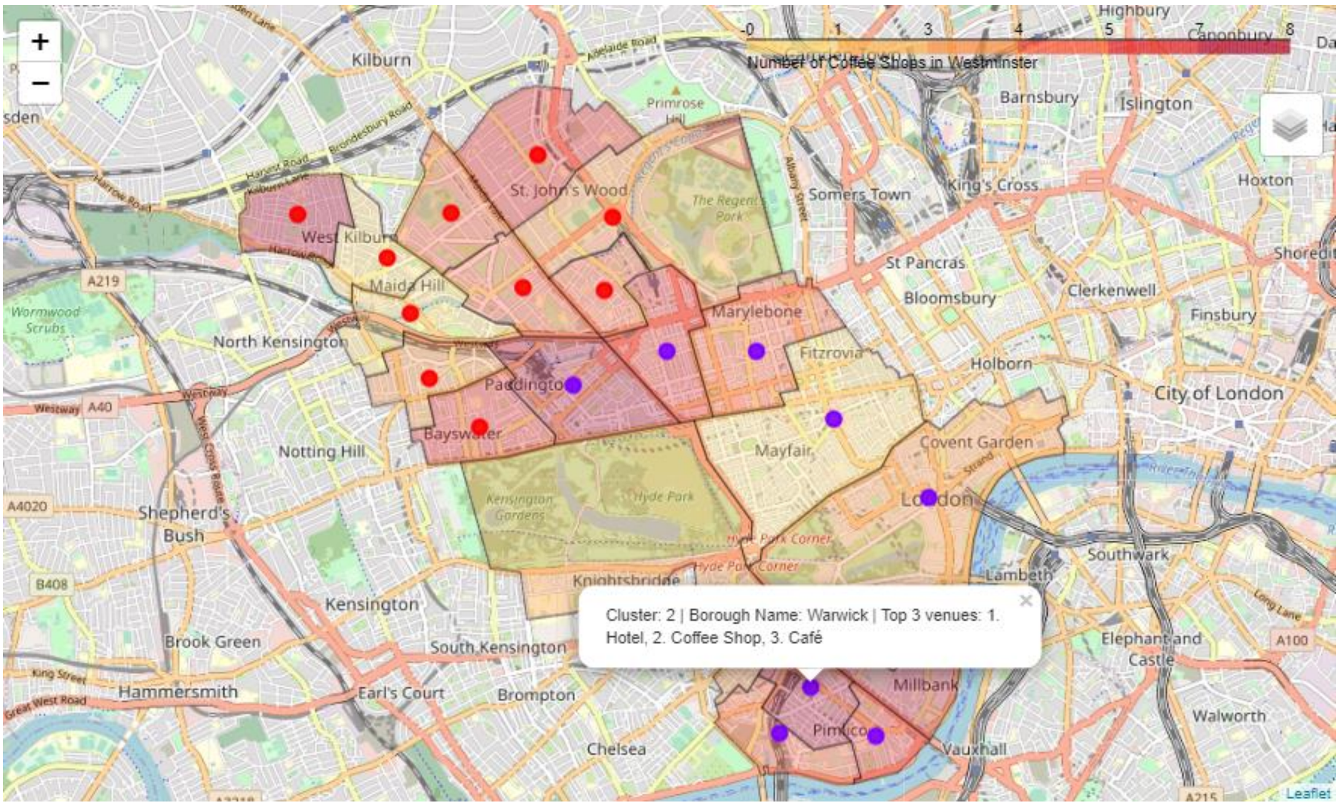
Cluster 2:

	Neighborhood	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
2	Bryanston and Dorset Square	Hotel	Coffee Shop	Sandwich Place	Argentinian Restaurant	Middle Eastern Restaurant	Pub	Restaurant	Juice Bar	Garden	Bakery
4	Churchill	Hotel	Italian Restaurant	Deli / Bodega	Coffee Shop	Café	Theater	Sandwich Place	Bakery	Plaza	Gastropub
6	Hyde Park	Coffee Shop	Hotel	Café	Garden	Pub	Middle Eastern Restaurant	Italian Restaurant	Beer Bar	Indian Restaurant	Greek Restaurant
7	Knightsbridge and Belgravia	Hotel	Café	Boutique	Plaza	Italian Restaurant	Restaurant	Coffee Shop	Japanese Restaurant	Shoe Store	Lebanese Restaurant
11	Marylebone High Street	Hotel	Juice Bar	Coffee Shop	Burger Joint	Clothing Store	Hotel Bar	Movie Theater	Dessert Shop	Food Court	Restaurant
14	St James's	Theater	Plaza	Hotel	Boutique	Indian Restaurant	Lounge	Bookstore	Garden	Coffee Shop	Monument / Landmark
15	Tachbrook	Hotel	Café	Coffee Shop	Sandwich Place	Park	Pub	Italian Restaurant	Mediterranean Restaurant	Tapas Restaurant	Plaza
16	Vincent Square	Hotel	Coffee Shop	Café	Theater	Sushi Restaurant	Park	Sandwich Place	Italian Restaurant	Pub	Sporting Goods Shop
17	Warwick	Hotel	Coffee Shop	Café	Theater	Gym / Fitness Center	Gastropub	Pub	Sandwich Place	Deli / Bodega	Italian Restaurant
18	West End	Coffee Shop	Clothing Store	Lounge	Dessert Shop	Hotel	Boutique	Cocktail Bar	Department Store	French Restaurant	Indian Restaurant

From cluster examination, I can a clear characteristics from each cluster. Cluster 1(0) is dominated by the pubs and cafes that make up the first and second most common venues, while cluster 2 is dominated by the hotels and coffee shops as the first and second most common venues. If we examine the choropleth map 4 further, we can see that the clusters divided the Westminster borough approximately in 'half'.

Our focus of analysis is to find a location that is suited for a coffee shop. In this case we will be focusing on the cluster 2 since tcoffee shops have more presence in this cluster. The cluster marks on the cloropleth map 4 also tells us an interesting fact that coffee shop is situated much closer to the office area within Westminster neighborhood such as in St. James, Vincent Square and Mayfair. Other than office area, we can also see government offices and tourist destinations also mainly located in cluster 2 in Westminster. Hence this result findings confirms our initial target market/customers: office employees and government workers, as well as tourists and the local population.

Choropleth Map 4: The Clusters of Westminster's Neighborhoods



4. Conclusion

Finding the perfect location to set up or invest in a business is a critical step that will contribute to the business future success. In this project, I began with a hypotesis-like statement that the coffee shop will target several types of customers (employees, gov't workers, tourists, locals), this determine what kind of data or parameters that will be used in the analysis. Accordingly, the analysis is performed to find a suitable location based on these specified criteria. The result confirms the hypothesis, coffee shop thrive in areas where office spaces, government offices, and tourists destination are located or intersected. These types of areas have the most potential to generate potential customers in the future.

Based on this result I recommend three neighborhood in Westminster to establish a coffee shop. These neighborhoods are: St. James, Vincent Square, Warwick, and Mayfair.

5. References:

1. European Coffee Federation. European Coffee Report 2018-2019. Brussels : European Coffee Federation, 2019. p. 20, Market Report.
2. Wikipedia. London. San Fransisco : Wikimedia Foundation Inc., 2021.
3. Greater London Authority. Population Density (GLA Estimates/Projections). London : Greater London Authority, 2021.
4. www.doogal.co.uk. London Postcodes.
5. Grange, Stuart. London Boroughs Boundaries. Zurich : Stuart Grange, 2021.
6. https://github.com/martinjc
7. pypi.org. geopy 2.1.0.
8. Foursquare. Developers API. 2021.