Analyzing boroughs in London for Starting a Restaurant

Applied Data Science Capstone Project

By: Boris Gorodilov
Date: May 4, 2021

Table of Contents

Introduction

London is the capital and largest city of England and the United Kingdom. It is one of the world's most important financial, commerce and educational centers. London has a diverse range of people and cultures, and more than 300 languages are spoken in the region. Its estimated population is roughly 9 million, which made it the third-most populous city in Europe. If we are looking to open a new restaurant, this is one of the best cities to consider possible locations. This project can be useful for business owners and entrepreneurs who are looking to invest in a restaurant. The main objective of this project is to carefully analyze appropriate data and find recommendations for the stakeholders.

Data Collection

The data required for this project has been collected from multiple sources. A summary of the data required for this project is given below.

Borough geo coordinates data

The data of the boroughs in London was scraped from https://en.wikipedia.org/wiki/List_of_London_boroughs.

The top 10 rows of the dataframe are shown in Figure 1.

	Borough	Area_sq_mi	Latitude	Longitude
0	Barking and Dagenham	13.93	51.5607	0.1557
1	Barnet	33.49	51.6252	-0.1510
2	Bexley	23.38	51.4549	0.1505
3	Brent	16.70	51.5588	-0.2810
4	Bromley	57.97	51.4039	0.0198
5	Camden	8.40	51.5290	-0.1250
6	Croydon	33.41	51.3714	-0.0970
7	Ealing	21.44	51.5130	-0.3080
8	Enfield	31.74	51.6538	-0.0790
9	Greenwich	18.28	51.4892	0.0648

Figure 1: Top 10 rows of London boroughs data scraped from Wikipedia.

Borough earnings data

Information on the income of the population of the borough is collected on the basis of two sources: data on the income of taxpayers living in the borough https://data.london.gov.uk/dataset/average-income-tax-payers-borough, and data on the income of people working in the borough https://data.london.gov.uk/dataset/earnings-workplace-borough.
The top 10 rows of the dataframe are shown in Figure 2.

	Borough	Tax_payers	Workplace
0	Barking and Dagenham	23900	28553
1	Barnet	28700	32143
2	Bexley	26900	30733
3	Brent	24700	30134
4	Bromley	32000	29819
5	Camden	37300	38147
6	Croydon	27500	32109
7	Ealing	26700	30259
8	Enfield	26300	29134
9	Greenwich	27600	32635

Figure 2: Top 10 rows of collected London boroughs earnings data.

Venue Data

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all boroughs in London and is used to study the popular venues of different boroughs.

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham	51.5607	0.1557	Central Park	51.559560	0.161981	Park
1	Barking and Dagenham	51.5607	0.1557	Lara Grill	51.562445	0.147178	Turkish Restaurant
2	Barking and Dagenham	51.5607	0.1557	The Eva Hart (Wetherspoon)	51.570460	0.130342	Pub
3	Barking and Dagenham	51.5607	0.1557	Costa Coffee	51.576890	0.179497	Coffee Shop
4	Barking and Dagenham	51.5607	0.1557	Harrow Lodge Park	51.555648	0.197926	Park
5	Barking and Dagenham	51.5607	0.1557	The Range	51.575550	0.180254	Furniture / Home Store
6	Barking and Dagenham	51.5607	0.1557	Hoo Hing	51.567561	0.135999	Grocery Store
7	Barking and Dagenham	51.5607	0.1557	Debenhams	51.579097	0.182720	Department Store
8	Barking and Dagenham	51.5607	0.1557	Ciao Bella	51.576103	0.182819	Italian Restaurant
9	Barking and Dagenham	51.5607	0.1557	Pets at Home	51.569605	0.183878	Pet Store

Figure 3: Top 10 rows of data obtained from Foursquare API.

Methodology and Discussion

This section provides details for the methodology used in the project and discuss noted observations and recommendations.

Data Exploring

Since the sizes of boroughs vary significantly, it is necessary to determine the correct radius for eachone before searching venues. Suppose that the shape of the districts tends to be a circle, then it becomes possible to estimate the correct search radius. Results are shown in Figure 4.

	Borough	Area_sq_mi	Latitude	Longitude	Radius
0	Barking and Dagenham	13.93	51.5607	0.1557	3388.829082
1	Barnet	33.49	51.6252	-0.1510	5254.503444
2	Bexley	23.38	51.4549	0.1505	4390.321866
3	Brent	16.70	51.5588	-0.2810	3710.499205
4	Bromley	57.97	51.4039	0.0198	6913.146454
5	Camden	8.40	51.5290	-0.1250	2631.562871
6	Croydon	33.41	51.3714	-0.0970	5248.223785
7	Ealing	21.44	51.5130	-0.3080	4204.230299
8	Enfield	31.74	51.6538	-0.0790	5115.376082
9	Greenwich	18.28	51.4892	0.0648	3882.059638
10	Hackney	7.36	51.5450	-0.0550	2463.275654

Figure 4: Top 10 rows of borough data with calculated radius of search.

Using folium, a map was plotted to show how the different boroughs are spread all across London. This is shown in Figure 5.

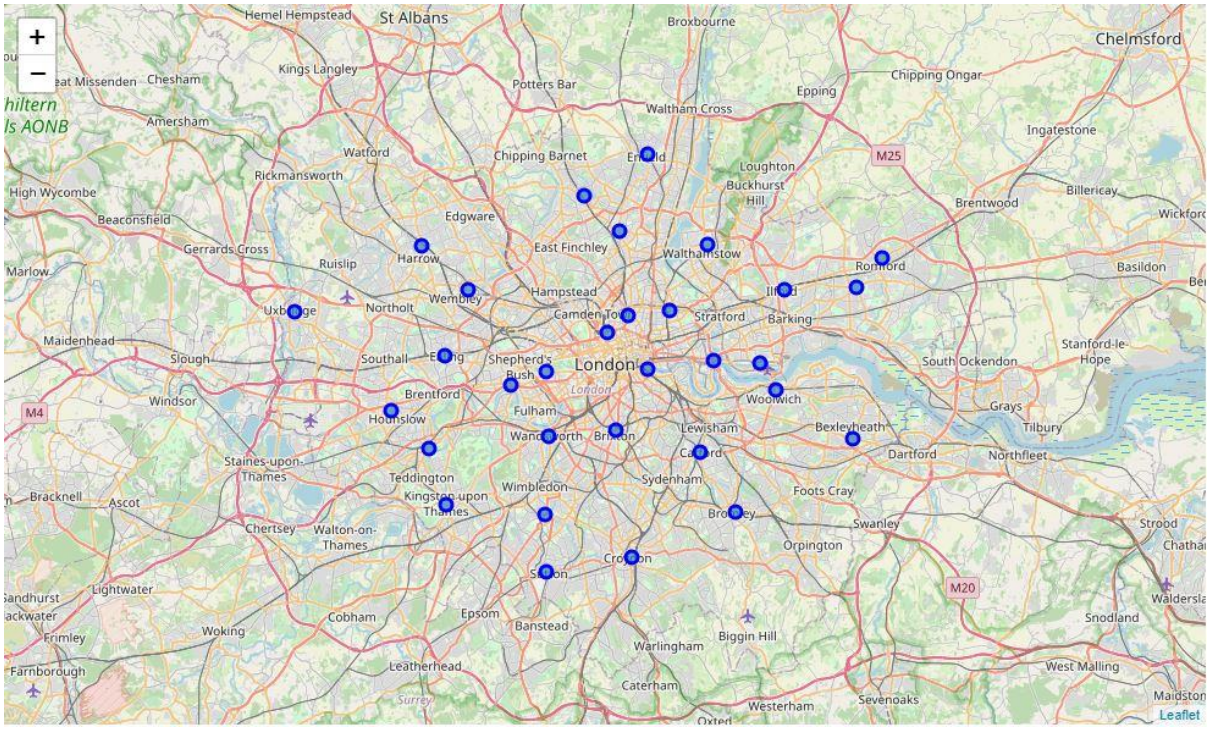

Figure 5: Depicting the boroughs spread across London.

Before opening a restaurant, an important element is to study the competition in the borough. To do this, we will conduct a quick assessment of the competition based on the number of venues in the "restaurant" category.

Borough	Venue	Borough	Venue
Merton	32	Islington	19
Kensington and Chelsea	31	Hillingdon	18
Harrow	31	Tower Hamlets	17
Haringey	27	Barking and Dagenham	16
Hounslow	25	Lambeth	14
Hammersmith and Fulham	25	Richmond upon Thames	13
Redbridge	25	Sutton	13
Croydon	25	Havering	12
Barnet	24	Ealing	12
Brent	22	Camden	12
Southwark	21	Bexley	12
Enfield	20	Newham	10
Bromley	20	Greenwich	10
Kingston upon Thames	19	Hackney	8
Lewisham	19	Wandsworth	8
Waltham Forest	19		

Figure 6: Number of restaurants grouped by borough.

In this regard, it would be logical to shorten the list of boroughs for further analysis. To do this, exclude the districts with the largest (top 25%) and the smallest (bottom 25%) number of venues in the "restaurant" category. Result dataframe is shown in Figure 7.

	Borough	Area_sq_mi	Latitude	Longitude	Radius
0	Barking and Dagenham	13.93	51.5607	0.1557	3388.829082
1	Barnet	33.49	51.6252	-0.1510	5254.503444
3	Brent	16.70	51.5588	-0.2810	3710.499205
4	Bromley	57.97	51.4039	0.0198	6913.146454
8	Enfield	31.74	51.6538	-0.0790	5115.376082
15	Hillingdon	44.67	51.5441	-0.4760	6068.510162
17	Islington	5.74	51.5416	-0.1020	2175.354565
19	Kingston upon Thames	14.38	51.4085	-0.3060	3443.131030
20	Lambeth	10.36	51.4607	-0.1160	2922.496401
21	Lewisham	13.57	51.4452	-0.0200	3344.752840
25	Richmond upon Thames	22.17	51.4479	-0.3260	4275.205093
26	Southwark	11.14	51.5035	-0.0800	3030.516847
27	Sutton	16.93	51.3618	-0.1940	3735.963171
28	Tower Hamlets	7.63	51.5099	-0.0050	2508.051074
29	Waltham Forest	14.99	51.5908	-0.0130	3515.401419

Figure 7: List of boroughs for further exploring.

Feature Extraction

Feature extraction was carried out to obtain features from the Foursquare API data which was used for building the unsupervised learning model. In order to achieve this, the “Venue Category” column had to be converted to some form of numeric value to be used for building the model. This was achieved by the One-hot Encoding method which takes all the unique categories and creates a column for each category. Then, if a borough venue belongs to that category, it would get a value of 1 for that row in that specific category column and if a borough venue does not belong to the particular category, the value would be 0. This process was repeated for all venues in all borough and the result was a sparse matrix containing the borough name and all unique category columns with either 1 or 0 based on whether the borough venue belonged to that category or not. This borough was then grouped by the borough name and the average value was taken for all categories. The result is shown in Figure 8 which shows only the top 10 rows.

	Borough	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant
0	Barking and Dagenham	0.00	0.00	0.021505	0.00	0.00	0.00	0.00	0.00	0.010753	...	0.00	0.00000	0.00
1	Barnet	0.00	0.00	0.000000	0.00	0.00	0.00	0.00	0.00	0.000000	...	0.00	0.00000	0.00
2	Brent	0.00	0.00	0.010000	0.00	0.00	0.00	0.00	0.00	0.010000	...	0.00	0.00000	0.00
3	Bromley	0.00	0.00	0.010000	0.00	0.00	0.00	0.00	0.00	0.010000	...	0.00	0.00000	0.00
4	Enfield	0.00	0.00	0.000000	0.00	0.00	0.00	0.00	0.00	0.000000	...	0.00	0.00000	0.00
5	Hillingdon	0.00	0.00	0.000000	0.00	0.00	0.00	0.00	0.00	0.020619	...	0.00	0.00000	0.00
6	Islington	0.01	0.00	0.000000	0.01	0.00	0.01	0.00	0.01	0.000000	...	0.00	0.00000	0.00
7	Kingston upon Thames	0.00	0.00	0.000000	0.00	0.00	0.00	0.00	0.00	0.000000	...	0.00	0.00000	0.00
8	Lambeth	0.00	0.01	0.000000	0.00	0.00	0.01	0.00	0.00	0.000000	...	0.01	0.00000	0.00
9	Lewisham	0.00	0.00	0.000000	0.00	0.01	0.00	0.00	0.00	0.000000	...	0.01	0.00000	0.00
10	Richmond upon Thames	0.00	0.00	0.000000	0.00	0.01	0.01	0.00	0.00	0.010000	...	0.00	0.00000	0.00
11	Southwark	0.00	0.00	0.000000	0.00	0.00	0.01	0.02	0.00	0.010000	...	0.00	0.00000	0.01
12	Sutton	0.00	0.00	0.010870	0.00	0.00	0.00	0.00	0.00	0.010870	...	0.00	0.01087	0.00
13	Tower Hamlets	0.00	0.00	0.000000	0.00	0.02	0.01	0.00	0.00	0.010000	...	0.00	0.00000	0.00
14	Waltham Forest	0.00	0.00	0.000000	0.00	0.00	0.02	0.00	0.01	0.000000	...	0.02	0.00000	0.00

Figure 8: One-hot Encoding resulting dataframe.

Notice that most of the values are 0 since there were a large number of unique categories and not all boroughs had venues belonging to each category. This data was used for the unsupervised learning model with the borough name dropped. The unsupervised learning model is explained in the next section.

A dataframe was also created which contained the top 10 most common venues of all boroughs. Though this is not a part of Feature Extraction, it is important to provide a glimpse into what this dataframe looks like as it will be used later to combine the results from the unsupervised learning model. The top 10 rows of this dataframe are shown in Figure 9.

	Borough	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	Barking and Dagenham	Grocery Store	Supermarket	Pub	Coffee Shop	Park	Café	Hotel	Furniture / Home Store	Fast Food Restaurant	Shopping Mall
1	Barnet	Café	Turkish Restaurant	Coffee Shop	Pub	Park	Bakery	Bar	Grocery Store	Greek Restaurant	French Restaurant
2	Brent	Indian Restaurant	Coffee Shop	Hotel	Clothing Store	Park	Gym / Fitness Center	Hookah Bar	Café	Pub	Pizza Place
3	Bromley	Pub	Park	Coffee Shop	Gym / Fitness Center	Pizza Place	Indian Restaurant	Italian Restaurant	Historic Site	Garden Center	Fast Food Restaurant
4	Enfield	Pub	Coffee Shop	Park	Turkish Restaurant	Café	Garden Center	Gym / Fitness Center	Supermarket	Grocery Store	Greek Restaurant
5	Hillingdon	Pub	Indian Restaurant	Coffee Shop	Supermarket	Hotel	Gym / Fitness Center	Park	Pharmacy	Thai Restaurant	Golf Course
6	Islington	Pub	Café	Coffee Shop	Park	Gastropub	Theater	French Restaurant	Gym / Fitness Center	Mediterranean Restaurant	Pizza Place
7	Kingston upon Thames	Pub	Café	Park	Coffee Shop	Garden	Gym / Fitness Center	Italian Restaurant	Gastropub	Thai Restaurant	Hotel
8	Lambeth	Coffee Shop	Pub	Park	Café	Brewery	Market	Beer Bar	Gastropub	Pizza Place	Farmers Market
9	Lewisham	Pub	Park	Coffee Shop	Gastropub	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Café	Farmers Market	Beer Store
10	Richmond upon Thames	Pub	Park	Café	Coffee Shop	Garden	Rugby Stadium	Bakery	Italian Restaurant	Restaurant	Hotel
11	Southwark	Coffee Shop	Hotel	Scenic Lookout	Theater	Restaurant	Pub	Beer Bar	Seafood Restaurant	Pizza Place	Grocery Store
12	Sutton	Pub	Park	Grocery Store	Coffee Shop	Supermarket	Pharmacy	Café	Hotel	Italian Restaurant	Train Station
13	Tower Hamlets	Coffee Shop	Hotel	Bar	Pub	Italian Restaurant	Plaza	Burger Joint	Gym / Fitness Center	Gym	Park
14	Waltham Forest	Pub	Coffee Shop	Park	Brewery	Café	Restaurant	Pizza Place	Supermarket	Bakery	Mediterranean Restaurant

Figure 9: Top 10 most common venues for boroughs.

Unsupervised Learning (venue data)

K-means unsupervised learning technique was used to cluster the boroughs based on the category of venues near the boroughs. Since the number of exploring boroughs is small, the number of clusters for machine learning algorithm would be set to 3.

The clustering model then clusters the boroughs in London and provides a label for each borough which is representative of the cluster it belongs to. The cluster labels were then added to the dataframe in Figure 9 along with the Location, Latitude, and Longitude columns to provide a complete summary of the clustering. The top 10 rows are shown in Figure 10.

Borough	Latitude	Longitude	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
Barking and Dagenham	51.5607	0.1557	0	Grocery Store	Supermarket	Pub	Coffee Shop	Park	Café	Hotel	Furniture / Home Store	Fast Food Restaurant	Shopping Mall
Barnet	51.6252	-0.1510	0	Café	Turkish Restaurant	Coffee Shop	Pub	Park	Bakery	Bar	Grocery Store	Greek Restaurant	French Restaurant
Brent	51.5588	-0.2810	1	Indian Restaurant	Coffee Shop	Hotel	Clothing Store	Park	Gym / Fitness Center	Hookah Bar	Café	Pub	Pizza Place
Bromley	51.4039	0.0198	2	Pub	Park	Coffee Shop	Gym / Fitness Center	Pizza Place	Indian Restaurant	Italian Restaurant	Historic Site	Garden Center	Fast Food Restaurant
Enfield	51.6538	-0.0790	0	Pub	Coffee Shop	Park	Turkish Restaurant	Café	Garden Center	Gym / Fitness Center	Supermarket	Grocery Store	Greek Restaurant
Hillingdon	51.5441	-0.4760	2	Pub	Indian Restaurant	Coffee Shop	Supermarket	Hotel	Gym / Fitness Center	Park	Pharmacy	Thai Restaurant	Golf Course
Islington	51.5416	-0.1020	2	Pub	Café	Coffee Shop	Park	Gastropub	Theater	French Restaurant	Gym / Fitness Center	Mediterranean Restaurant	Pizza Place
Kingston upon Thames	51.4085	-0.3060	2	Pub	Café	Park	Coffee Shop	Garden	Gym / Fitness Center	Italian Restaurant	Gastropub	Thai Restaurant	Hotel
Lambeth	51.4607	-0.1160	2	Coffee Shop	Pub	Park	Café	Brewery	Market	Beer Bar	Gastropub	Pizza Place	Farmers Market
Lewisham	51.4452	-0.0200	2	Pub	Park	Coffee Shop	Gastropub	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Café	Farmers Market	Beer Store

Figure 10: Clustering boroughs in London (venue data).

Furthermore, boroughs in each individual cluster can be extracted using cluster labels and thus the details of specific clusters can be seen. This is done below for all clusters.

	Borough	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	Barking and Dagenham	Grocery Store	Supermarket	Pub	Coffee Shop	Park	Café	Hotel	Furniture / Home Store	Fast Food Restaurant	Shopping Mall
1	Barnet	Café	Turkish Restaurant	Coffee Shop	Pub	Park	Bakery	Bar	Grocery Store	Greek Restaurant	French Restaurant
4	Enfield	Pub	Coffee Shop	Park	Turkish Restaurant	Café	Garden Center	Gym / Fitness Center	Supermarket	Grocery Store	Greek Restaurant
12	Sutton	Pub	Park	Grocery Store	Coffee Shop	Supermarket	Pharmacy	Café	Hotel	Italian Restaurant	Train Station

Figure 11: Cluster 1.

	Borough	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	Brent	Indian Restaurant	Coffee Shop	Hotel	Clothing Store	Park	Gym / Fitness Center	Hookah Bar	Café	Pub	Pizza Place
11	Southwark	Coffee Shop	Hotel	Scenic Lookout	Theater	Restaurant	Pub	Beer Bar	Seafood Restaurant	Pizza Place	Grocery Store
13	Tower Hamlets	Coffee Shop	Hotel	Bar	Pub	Italian Restaurant	Plaza	Burger Joint	Gym / Fitness Center	Gym	Park

Figure 12: Cluster 2.

	Borough	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
3	Bromley	Pub	Park	Coffee Shop	Gym / Fitness Center	Pizza Place	Indian Restaurant	Italian Restaurant	Historic Site	Garden Center	Fast Food Restaurant
5	Hillingdon	Pub	Indian Restaurant	Coffee Shop	Supermarket	Hotel	Gym / Fitness Center	Park	Pharmacy	Thai Restaurant	Golf Course
6	Islington	Pub	Café	Coffee Shop	Park	Gastropub	Theater	French Restaurant	Gym / Fitness Center	Mediterranean Restaurant	Pizza Place
7	Kingston upon Thames	Pub	Café	Park	Coffee Shop	Garden	Gym / Fitness Center	Italian Restaurant	Gastropub	Thai Restaurant	Hotel
8	Lambeth	Coffee Shop	Pub	Park	Café	Brewery	Market	Beer Bar	Gastropub	Pizza Place	Farmers Market
9	Lewisham	Pub	Park	Coffee Shop	Gastropub	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Café	Farmers Market	Beer Store
10	Richmond upon Thames	Pub	Park	Café	Coffee Shop	Garden	Rugby Stadium	Bakery	Italian Restaurant	Restaurant	Hotel
14	Waltham Forest	Pub	Coffee Shop	Park	Brewery	Café	Restaurant	Pizza Place	Supermarket	Bakery	Mediterranean Restaurant

Figure 13: Cluster 3.

Based on the clusters shown above, the boroughs can once again be plotted on a map of London, however, this time with different color markers to distinguish between different clusters. This is shown in Figure 14.

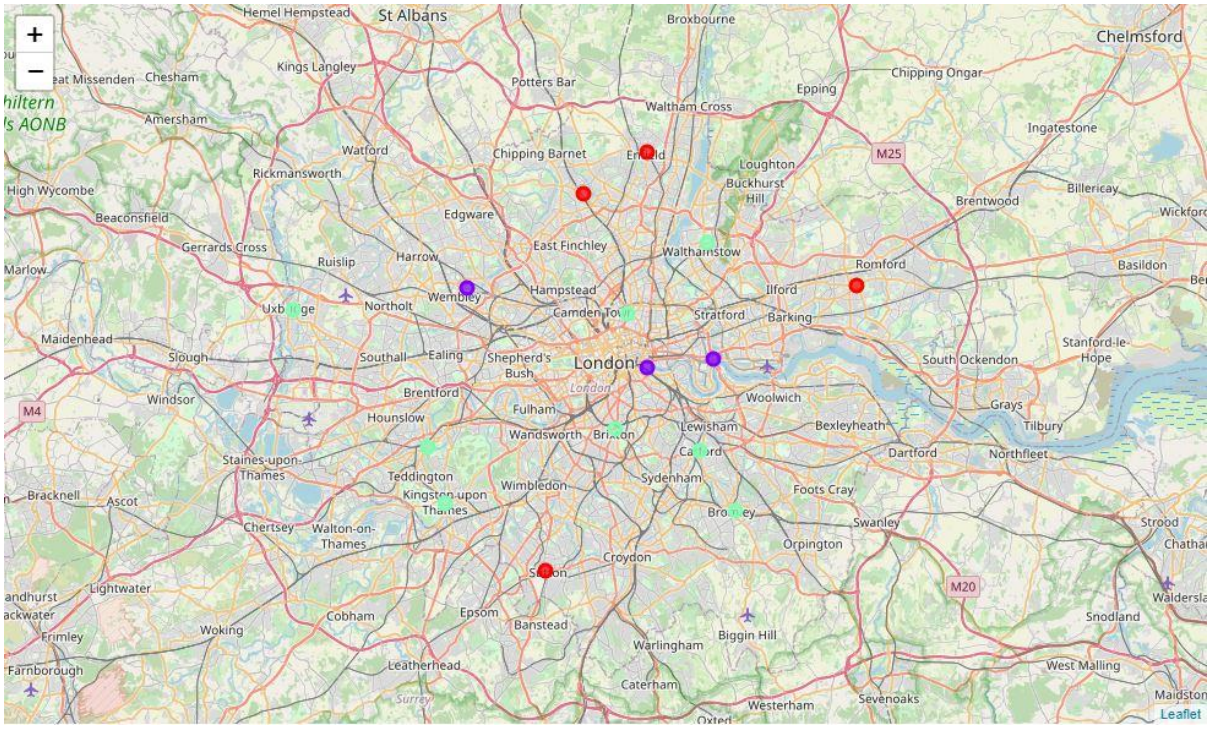

Figure 14: Visualizing the clustering of boroughs in London.

It looks like the boroughs from the second cluster (purple marks) have the greatest potential for opening a new restaurant, because in the top 3 categories of venues there are: no parks (dense development) and there are hotels (tourists and business travelers can significantly increase attendance).
On opposite side top-10 venues of boroughs from the first cluster (red marks) are made up of parks, supermarkets, furniture stores, grocery store and etc. This set of venues characterizes these boroughs as outskirts. Most of the venues are

likely to be used by locals and suburbanites who drive through the area. This makes the opening of a new restaurant in these boroughs the least attractive.

Unsupervised Learning (income data)

Since visiting restaurants is not a matter of basic necessity, when determining the most suitable boroughs, it should also take into account the income level of the population living and working there.

For these needs was used K-means unsupervised learning technique to cluster the boroughs based on income of the working and living population. The clustering model then clusters the boroughs in London and provides a label for each borough which is representative of the cluster it belongs to. The cluster labels were then added to the dataframe in Figure 2. The result is shown in Figure 15.

	Borough	Tax_payers	Workplace	Income Cluster Labels
0	Barking and Dagenham	23900	28553	0
1	Barnet	28700	32143	0
3	Brent	24700	30134	0
4	Bromley	32000	29819	2
8	Enfield	26300	29134	0
15	Hillingdon	27100	33596	0
17	Islington	33400	39348	2
19	Kingston upon Thames	32400	31308	2
20	Lambeth	29900	35036	2
21	Lewisham	27300	33294	0
25	Richmond upon Thames	36100	32141	2
26	Southwark	29400	37601	2
27	Sutton	28200	27945	0
28	Tower Hamlets	30200	49369	1
29	Waltham Forest	25500	28284	0

Figure 15: Clustering boroughs in London (income data).

Results

To summarize the results, a data frame was created with the borough name, geo-coordinates information and the results of both clustering processes. The results of the clustering processes have been combined under a single value for the final letter-based rating ("A +" - most preferred, "C-" - least preferred) and shown in Figure 16.

	Borough	Latitude	Longitude	Cluster Labels	Income Cluster Labels	Final cluster
0	Barking and Dagenham	51.5607	0.1557	C	-	C-
1	Barnet	51.6252	-0.1510	C	-	C-
2	Brent	51.5588	-0.2810	A	-	A-
3	Bromley	51.4039	0.0198	B		B
4	Enfield	51.6538	-0.0790	C	-	C-
5	Hillingdon	51.5441	-0.4760	B	-	B-
6	Islington	51.5416	-0.1020	B		B
7	Kingston upon Thames	51.4085	-0.3060	B		B
8	Lambeth	51.4607	-0.1160	B		B
9	Lewisham	51.4452	-0.0200	B	-	B-
10	Richmond upon Thames	51.4479	-0.3260	B		B
11	Southwark	51.5035	-0.0800	A		A
12	Sutton	51.3618	-0.1940	C	-	C-
13	Tower Hamlets	51.5099	-0.0050	A	+	A+
14	Waltham Forest	51.5908	-0.0130	B	-	B-

Figure 16: Rating of boroughs in London.

Based on the information shown above, the boroughs can be plotted on a map of London. This is shown in Figure 17.

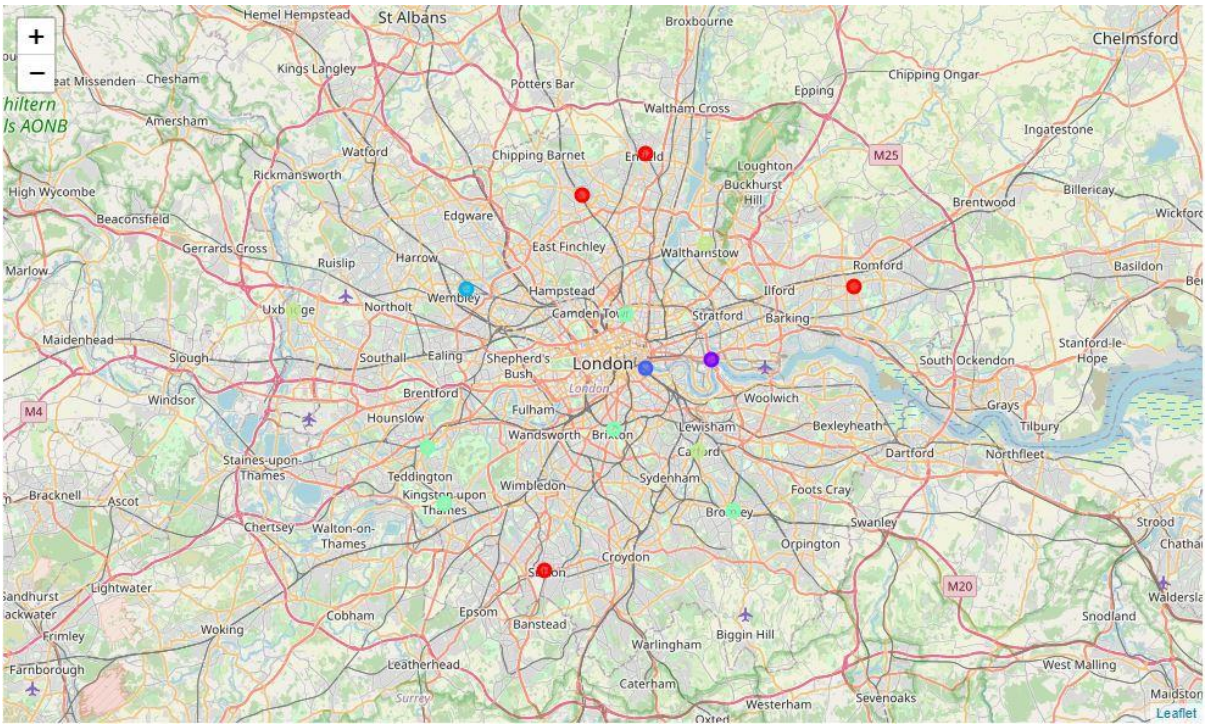

Figure 17: Visualizing the ratings of boroughs in London.

Borough rating	Description
A+ A A-	Boroughs rated "A" are the best option to open a new restaurant, especially those rated "A +".
B B-	Boroughs rated 'B' may be considered for open a new restaurant, but they are not the best options.
C-	Boroughs rated 'C' should not be considered for a new restaurant opening. However, with a more detailed exploring and identification of points of attraction of foot-traffic, they can be a good option.

Conclusion

In this project, the boroughs in London, England have been successfully analyzed for determining which would be the best borough for opening a new restaurant. Based on the analysis carried out, boroughs have been rated based on venue and population income data. The stakeholders and investors can further tune this by considering various other factors like transport, legal requirements, and costs associated which were out of the scope for this project and thus were not considered.