

Segmenting and Clustering Neighborhoods in London

IBM Data Science Professional Certificate Course - Capstone Project

Opening a new Japanese Restaurant in London, UK

Rajesh A C



Introduction

Widely regarded as one of the most diverse cities in the world, London's culinary scene offers an eclectic mixture of classic British and modern, multicultural cuisine.

As the UK's capital, London acts as a representative of everything British to a wide array of visitors from around the world. London gastronomy is teeming with a smorgasbord of different cultural influences and tastes that offer locals and visitors a fantastic culinary experience.

The city has a rich history when it comes to food, and many eateries reflect this in their representation of the nation's favourite cuisines and tastes.



Business Problem

The client is unable to decide which neighbourhood of London would be best suited for the new chain of their Japanese restaurant. The objective of this Capstone project is to analyse the various Neighborhoods of London, UK and select the best location for opening a new restaurant. In other words, If someone is looking to open an Japanese restaurant in the city of London, where would you recommend that they open it?

Data Requirements and Extraction

To solve this problem, we will need to complete the following steps:

- List of Neighborhoods in London

The [Wikipedia](#) page of Boroughs and neighborhoods of London list all the neighborhoods of the city. Using Web Scraping techniques, we will extract the required data, with the help of Python Requests and BeautifulSoup packages.

- Latitude and Longitude coordinates of all the Neighborhoods

The Python Geocoder package would be used to find the location data of these neighborhoods.

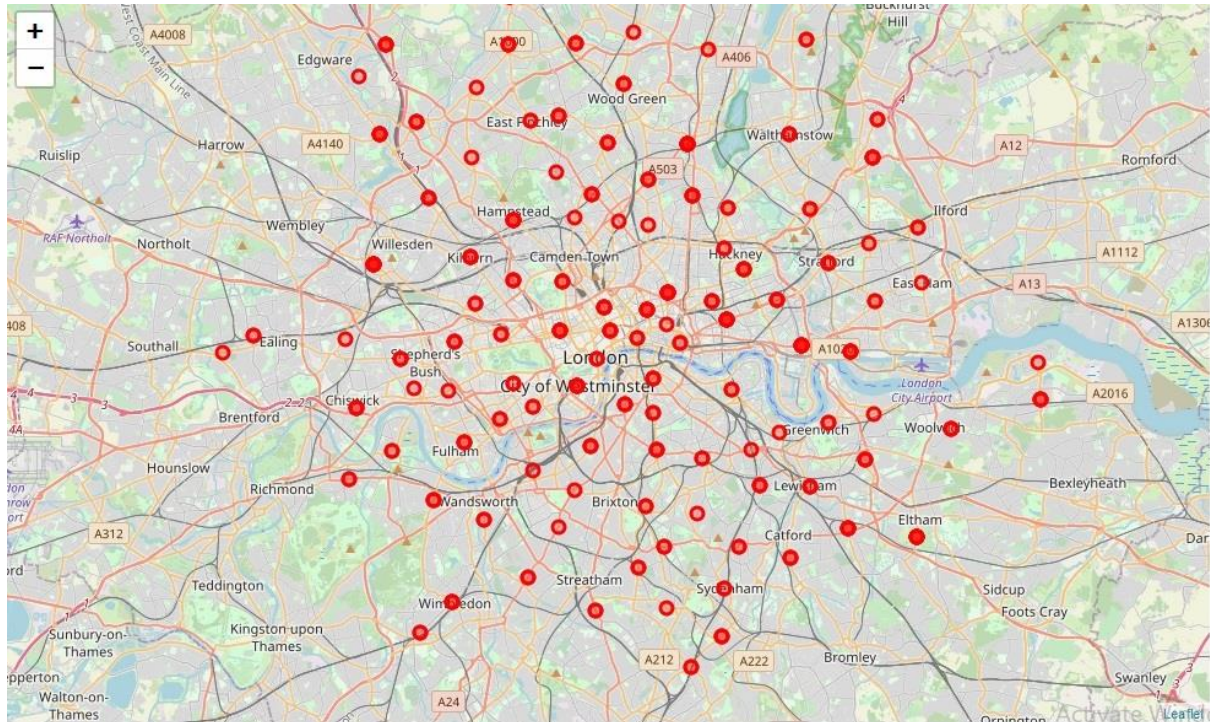
- Venue data (powered by Foursquare), particularly related to Japanese Restaurant

At last, the Foursquare API would be used to extract the data regarding existing restaurants in the city.

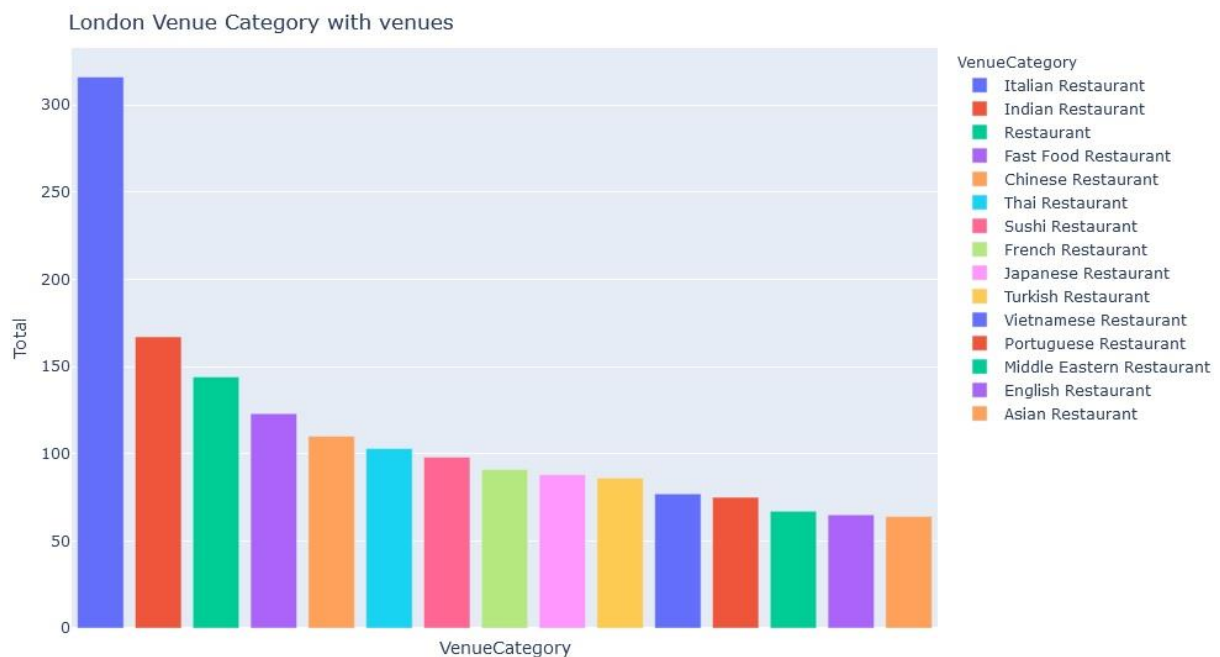
In general, this project would be encompassing a series of Data Science techniques, including, but not limited to, Web Scraping (using BeautifulSoup and Requests), Data Cleaning, Data Wrangling and Machine Learning (K-Means clustering algorithm)

Methodology

To get a list of neighborhoods in the city of London is extracted from the Wikipedia page titled, 'List of areas of London' (https://en.wikipedia.org/wiki/List_of_areas_of_London). Using the BeautifulSoup and Requests packages of Python, the required data is scraped from the webpage. Duplicate neighbourhood name were corrected during the pre-processing stage.



In order to use the Foursquare API, we fetch the location data of all these neighbourhoods from the Python Geocoder package. Next, the Foursquare API is used to get the venues of each neighborhoods.

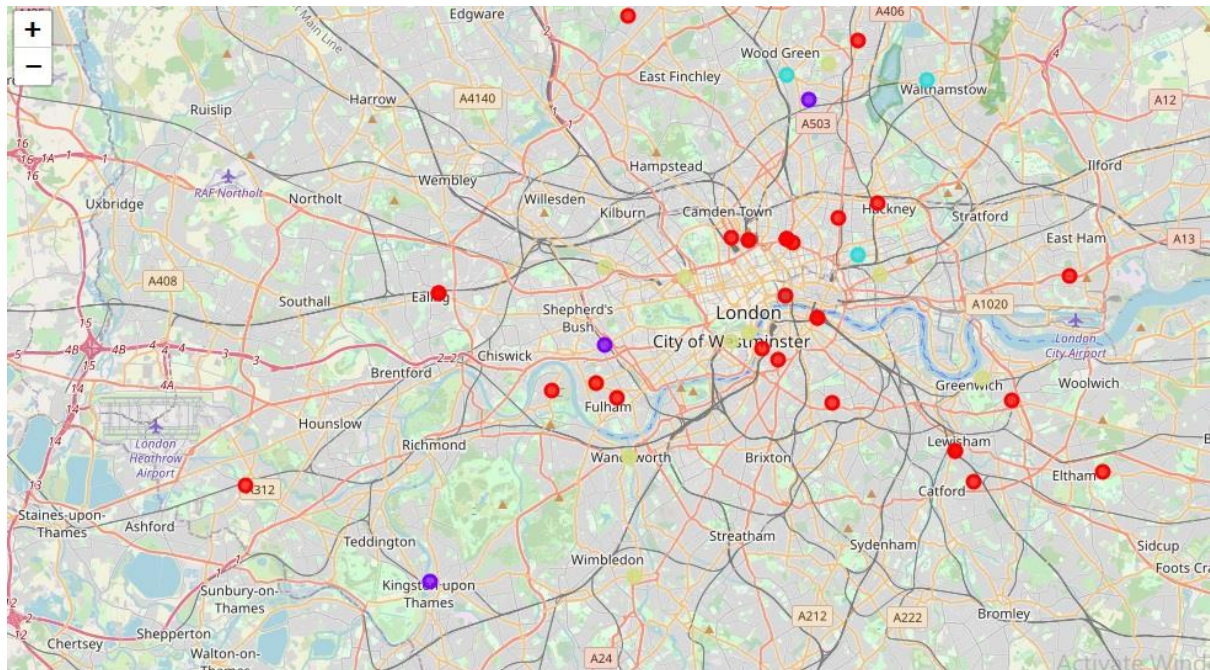


To prepare the data for K-means clustering, we group the data frame by neighborhoods. Lastly, K-means clustering is performed on this data set to return 4 clusters, or categories of neighborhoods in terms of number of Japanese Restaurants.

Results The output from the clustering algorithm shows the division of all the neighborhoods into 4 groups in terms of number of Japanese restaurants present in that area.

- Cluster 0: Neighborhoods with very low number of Japanese restaurants
- Cluster 1: Neighborhoods with low number of Japanese restaurants
- Cluster 2: Neighborhoods with a significant number of Japanese restaurants
- Cluster 3: Neighborhoods with heavy competition of Japanese restaurants

The results are visualised as follows:



Here,

Cluster 0 = Red

Cluster 1 = Purple

Cluster 2 = Light Blue

Cluster 3 = Yellow

Discussion and Conclusion

- Category 3 having intense competition which is not recommended.
- Category 0 also non-existence of competition, hence, it is not recommended
- Japanese restaurant with variety of food, cheap prices etc. can go for Category 1
- But Category 2 is recommended more than Category 1 due to a significantly higher expected target audience.

Note : Appendix section have the complete list of neighborhoods based on these categories.

Appendix

Category 1: Neighborhoods with very low number of restaurants

SL No.	Japanese Restaurant
1	Waltham Forest
2	Southwark
3	Richmond upon Thames
4	Redbridge
5	Newham
6	Lewisham, Southwark
7	Lewisham, Bromley
8	Lewisham
9	Lambeth, Southwark
10	Lambeth
11	Islington, City
12	Islington, Camden
13	Islington
14	Hounslow, Ealing, Hammersmith and Fulham
15	Hounslow
16	Harrow, Brent
17	Haringey, Islington
18	Hackney
19	Greenwich, Lewisham
20	Enfield
21	Ealing, Hammersmith and Fulham
22	Ealing
23	Camden, Islington
24	Bromley
25	Brent, Ealing
26	Brent, Camden
27	Brent
28	Bexley, Greenwich
29	Bexley, Greenwich
30	Bexley
31	Barnet, Brent, Camden

Category 2: Neighborhoods with low number of restaurants

SL No.	Japanese Restaurant
1	Kingston upon Thames
2	Kensington and Chelsea, Hammersmith and Fulham
3	Haringey, Barnet

Category 3: Neighborhoods with a significant number of restaurants

SL No.	Japanese Restaurant
1	Redbridge, Waltham Forest
2	Haringey
3	Camden
4	Brent, Harrow
5	Barnet

Category 4: Neighborhoods crowded with restaurants

SL No.	Japanese Restaurant
1	Westminster
2	Wandsworth
3	Tower Hamlets
4	Merton
5	Lambeth, Wandsworth
6	Kensington and Chelsea
7	Hammersmith and Fulham
8	Greenwich
9	Croydon
10	City, Westminster
11	City