The Battle of Neighbourhoods Final

Vishal Chavan *March 29,2020*

1. **Introduction**:

We are well versed on development in london and how they have emerged has a most fine country in the world. London is known for pubs, restaurants, night life, industrial sector. However, we will concentrate on restaurant section which are increasing with varieties. The real deal is that as much as there are many fine restaurants in London – Asian, Middle Eastern, Latin and American restaurants, its very difficult to find good place to dine in the finest of Indian cuisine that has combination of Malvani, konkani, punjabi style.

Tadka, a successful restaurant chain in India is looking to expand operation into Europe through London. They want to create a high-end restaurant that comes with organic mix and healthy. Their target is not only Indians, but all communities which would like to taste something different. Since the London demography is so big, my client needs deeper insight from available data in other to decide where to establish the first Europe “palace” restaurant. This company spends a lot on research and provides customers with data insight into the ingredients used at restaurants.

Target:

Considering the diversity in London where most people are multireligion based. London is a place where different shades live. Most people reside in london are migrants where they come for job or education from India. Definitely, by looking at the population we can determine there is highly shortage of Indian restaurants in London.

2. **Data acquisition and cleaning**:

2.1 Data Acquisition:

This project will rely on public data from Wikipedia and Foursquare.

In this project, London will be used as synonymous to the "Greater London Area" in this project. Within the Greater London Area, there are areas that are within the London Area Postcode. The focus of this project will be the nieghbourhoods are that are within the London Post Code area. The London Area consists of 32 Boroughs and the "City of London".

Our data will be from the link - Greater London Area

https://en.wikipedia.org/wiki/List_of_areas_of_London

Data set contains following columns:

- Location – Location of London
- London Borough – common name
- Post town – Post town of the London

- Post code – Post code of the London areas
- Dial Code – Area dial code

Second set of data was demography of London:

https://en.wikipedia.org/wiki/Demography_of_London

- Local authority – Areas within London
- Demographic subgroup – category of demography

2.2 Data Cleaning:

Data downloaded from multiple sources were structured into one table. There were lot issues like brackets around some data which were cleared, spaces, duplication of data. Also focused on the demography of London where there are predominantly more multicultural groups. According to the proportion of races by London borough as seen in Demography of London where we have sorted white, black, mixed, Asian to check how much count in each location in London.

Latitude and Longitude are obtained and join with location which will help us to identify the distance. To get nearby venues, we have obtain the data from Foursquare so that we can get name and categories of venues like in this case under categories different restaurants such as Italian, African, Indian etc. We have explored (Multiple) Neighbourhoods in the South East London area.

Output will look like below after cleaning 1st data set (Area of london)

	Location	Borough	Post-town	Dial-code	OSGridRef	Postcode
0	Abbey Wood	Bexley, Greenwich	LONDON	020	TQ465785	SE2
1	Acton	Ealing, Hammersmith and Fulham	LONDON	020	TQ205805	W3
1	Acton	Ealing, Hammersmith and Fulham	LONDON	020	TQ205805	W4
10	Angel	Islington	LONDON	020	TQ345665	EC1
10	Angel	Islington	LONDON	020	TQ345665	N1

Output will look like below after cleaning 2nd data set (Demography of London):

	Local authority	White	Mixed	Asian	Black	Other
22	Lewisham	53.5	7.4	9.3	27.2	2.6
27	Southwark	54.3	6.2	9.4	26.9	3.3
21	Lambeth	57.1	7.6	6.9	25.9	2.4
11	Hackney	54.7	6.4	10.5	23.1	5.3
7	Croydon	55.1	6.6	16.4	20.2	1.8

The Coordinates of neighbourhood has been obtain using Google Map API geocoding to get final data set;

	Location	Borough	Postcode	Latitude	Longitude
0	Crofton Park	Lewisham	SE4	51.46268	-0.03558
1	Denmark Hill	Southwark	SE5	51.47480	-0.09313
2	Deptford	Lewisham	SE8	51.48114	-0.02467
3	Dulwich	Southwark	SE21	51.44100	-0.08897
4	East Dulwich	Southwark	SE22	51.45256	-0.07076

The New Data set is used to generate the 10 most common venues for each neighbourhood using foursquare API, finally using k means clustering algorithm to cluster similar neighbour together.

3. Methodology:

3.1 Data Exploration:

The Describe function in python is used to get statistic of the London areas, this return the mean, standard deviation, minimum, maximum, 1st Quantile (25%), 2nd Quantile (50%), 3rd Quantile (75%) for each of the major categories.

	Count
count	195.000000
mean	17.600000
std	38.354881
min	1.000000
25%	3.000000
50%	6.000000
75%	17.000000
max	347.000000

3.2 Modelling:

Using Final data set containing the neighbourhood in London with latitude and longitude, we can find all venues within a 500meter radius of each neighbourhood by connecting to the Foursquare API. This returns a json file containing all venues in each neighbourhood which is converted into panda data frame. This data frame contain all venues along with co ordinates and category :

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Crofton Park	51.46268	-0.03558	The Orchard	51.463678	-0.035699	Gastropub
1	Crofton Park	51.46268	-0.03558	Brockley's Rock	51.459457	-0.033868	Fish & Chips Shop
2	Crofton Park	51.46268	-0.03558	Browns Of Brockley	51.464513	-0.037346	Coffee Shop
3	Crofton Park	51.46268	-0.03558	Waterintobeer	51.463712	-0.038826	Beer Store
4	Crofton Park	51.46268	-0.03558	Saka Maka	51.464826	-0.036437	Indian Restaurant

One hot encoding is done on the venues data. The venues data is then grouped by neighbourhood and the mean of the venues are calculated, finally 10 common venues are calculated for each neighbourhoods.

To help people find similar neighbourhoods in the borough we will be clustering similar neighbourhoods using K –means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use cluster size of 5 for this project. The reason to conduct a K –means clustering is to cluster

neighbourhood with similar venues together so that people can short list the area of the interests based on venues/amenities around each neighbourhood.

4. Results:

After running the K-means clustering we can access each cluster created to see which neighbourhoods were assigned to each of the five cluster.

	Neighbourhood	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
0	Bankside	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Seafood Restaurant	Gym / Fitness Center	Art Museum	Garden
1	Bellingham	Grocery Store	Park	Supermarket	Café	Pub	Italian Restaurant	Coffee Shop	Fast Food Restaurant	Train Station
2	Bermondsey	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Seafood Restaurant	Gym / Fitness Center	Art Museum	Garden
3	Blackheath	Pub	Grocery Store	Park	Coffee Shop	Garden	Bakery	Italian Restaurant	Supermarket	Clothing Store
4	Brockley	Coffee Shop	Pub	Café	Park	Gastropub	Bar	Indian Restaurant	Italian Restaurant	Cocktail Bar

To find the optimal value of the number of clusters, k, the number of clusters is iterated corresponding Silhouette Coefficientis calculated for each of the k-values used. The highest Silhouette Coefficient gives the best match to its own cluster:

Cluster 1 :

	Borough	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 Mos Commc Venue
9	Lewisham	0	Pub	Grocery Store	Park	Café	Supermarket	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Train Station
15	Lewisham	0	Pub	Grocery Store	Park	Café	Supermarket	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Train Station
20	Lewisham	0	Grocery Store	Park	Supermarket	Café	Pub	Italian Restaurant	Coffee Shop	Fast Food Restaurant	Train Station
26	Lewisham	0	Grocery Store	Park	Supermarket	Café	Pub	Italian Restaurant	Coffee Shop	Fast Food Restaurant	Train Station
36	Lewisham	0	Grocery Store	Park	Supermarket	Café	Pub	Italian Restaurant	Coffee Shop	Fast Food Restaurant	Train Station
37	Lewisham	0	Pub	Grocery Store	Park	Café	Supermarket	Gym / Fitness Center	Italian Restaurant	Indian Restaurant	Train Station

Cluster 2:

	Borough	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
0	Lewisham	1	Coffee Shop	Pub	Café	Park	Gastropub	Bar	Indian Restaurant	Italian Restaurant	Cocktail Bar
2	Lewisham	1	Pub	Café	Coffee Shop	Bar	Park	Brewery	Trail	Indie Movie Theater	Italian Restaurant
4	Southwark	1	Pub	Café	Pizza Place	Italian Restaurant	Park	Coffee Shop	Gastropub	Burger Joint	Garden Center
11	Lewisham	1	Pub	Café	Gastropub	Park	Garden	Food Truck	Fish & Chips Shop	Coffee Shop	Restaurant
13	Lewisham	1	Coffee Shop	Pub	Café	Park	Gastropub	Bar	Indian Restaurant	Italian Restaurant	Cocktail Bar
16	Lewisham	1	Pub	Café	Gastropub	Park	Garden	Food Truck	Fish & Chips Shop	Coffee Shop	Restaurant
17	Lewisham	1	Pub	Coffee Shop	Café	Bar	Italian Restaurant	Park	Indian Restaurant	Gastropub	Brewery
19	Southwark	1	Pub	Café	Coffee Shop	Pizza Place	Park	Bar	Gastropub	Burger Joint	Cocktail Bar
21	Southwark	1	Pub	Café	Coffee Shop	Pizza Place	Park	Bar	Gastropub	Burger Joint	Cocktail Bar
22	Southwark	1	Pub	Brewery	Coffee Shop	Park	Bar	Café	Food Truck	Gym / Fitness Center	Vietnamese Restaurant

Cluster 3:

	Borough	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
1	Southwark	2	Café	Coffee Shop	Park	Pub	Italian Restaurant	Cocktail Bar	Middle Eastern Restaurant	Brewery	Pizza Place
10	Lambeth	2	Coffee Shop	Pub	Café	Brewery	Market	Pizza Place	Cocktail Bar	Burger Joint	Caribbean Restaurant
29	Lambeth	2	Coffee Shop	Pub	Café	Brewery	Market	Pizza Place	Cocktail Bar	Burger Joint	Caribbean Restaurant
35	Southwark	2	Café	Coffee Shop	Park	Pub	Italian Restaurant	Cocktail Bar	Middle Eastern Restaurant	Brewery	Pizza Place

Cluster 4:

	Borough	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
3	Southwark	3	Pub	Café	Park	Coffee Shop	Grocery Store	Bakery	Gym / Fitness Center	Farmers Market	Pizza Place
7	Lewisham	3	Pub	Coffee Shop	Café	Park	Grocery Store	Supermarket	Gym / Fitness Center	Japanese Restaurant	Pizza Place
8	Lambeth	3	Pub	Park	Coffee Shop	Italian Restaurant	Grocery Store	Café	Train Station	Gastropub	Bakery
12	Lewisham	3	Pub	Coffee Shop	Café	Park	Grocery Store	Supermarket	Gym / Fitness Center	Japanese Restaurant	Pizza Place
23	Croydon	3	Pub	Grocery Store	Café	Supermarket	Coffee Shop	Park	Tram Station	African Restaurant	Plaza
25	Croydon	3	Pub	Grocery Store	Café	Supermarket	Coffee Shop	Park	Tram Station	African Restaurant	Plaza
30	Croydon	3	Pub	Park	Coffee Shop	Italian Restaurant	Grocery Store	Café	Train Station	Gastropub	Bakery
32	Lewisham	3	Pub	Grocery Store	Park	Coffee Shop	Garden	Bakery	Italian Restaurant	Supermarket	Clothing Store
33	Lambeth	3	Pub	Grocery Store	Coffee Shop	Café	Bakery	Park	Pizza Place	Gym / Fitness Center	Italian Restaura

Cluster 5:

	Borough	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
5	Southwark	4	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Seafood Restaurant	Gym / Fitness Center	Art Museum	Garden
6	Southwark	4	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Seafood Restaurant	Gym / Fitness Center	Art Museum	Garden
14	Lambeth	4	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Seafood Restaurant	Gym / Fitness Center	Art Museum	Garden
18	Southwark	4	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Seafood Restaurant	Gym / Fitness Center	Art Museum	Garden
24	Southwark	4	Coffee Shop	Hotel	Pub	Italian Restaurant	Theater	Seafood Restaurant	Gym / Fitness Center	Art Museum	Garden

The following are the highlights of the 5 clusters above: 1] Pubs, Grocery Stores are popular in the South East London. 2] As for restaurants, the Italian Restaurants are very popular in the South East London area. Especially in Southwark and Lambeth areas. 3] With the Lewisham area being the most condensed area of Indian Restaurant in the South East Area, it is surprising to see how in the top 10 venues, you can barely see restaurants in the top 5 venues. 4] Although, the Clusters have variations, a very visible presence is the predominance of pubs.

5. Discussion and Conclusion:

Its been visible from cluster 3 and 4 that Indian Restaurant can play vital role in Restaurant competition and very good chance to set up. Their proximity to other amenities and accessibility to station are huge. These 2 clusters do not have top restaurants that could rival their standards if they are created. And the chances to resources needed is quite high as Lewisham and Lambeth. In conclusion, this project would have had better results if there were more data in terms of per capita income data within the area, traffic access, corporates of more venues exploration with the Foursquare. Also, getting the ratings and feedbacks of the current restaurants within the clusters would have helped in providing more insight into the best location.