Contents

The Battle of the Neighbourhoods - London

1. Introduction & Business Problem:

a) Problem Background:

The City of London is the one of the most populous cities in Europe and has very diverse culture. It provides lot of business opportunities and business friendly environment. London has a diverse range of people and cultures, and more than 300 languages are spoken in the region. Its estimated mid-2018 municipal population (corresponding to Greater London) was roughly 9 million, which made it the third-most populous city in Europe. London accounts for 13.4% of the U.K. population. Greater London Built-up Area is the fourth-most populous in Europe, after Istanbul, Moscow, and Paris, with 9,787,426 inhabitants at the 2011 census. The London metropolitan area is the third-most populous in Europe, after Istanbul and the Moscow Metropolitan Area, with 14,040,163 inhabitants in 2016

This also means that the market is highly competitive. As it is such a mature city cost of doing business is also one of the highest probably only below New York. Thus, any new Business would require careful analysis to identify potential opportunities. It would help increase chance of success and reduce chance of failures. The developed nature of city also provides opportunity to source and analyse diverse datasets.

b) Problem Description:

London offers a great variety of cuisine as a result of its ethnically diverse population. Gastronomic centres include the Bangladeshi restaurants of Brick Lane and the Chinese restaurants of Chinatown.

To succeed in such an environment, it is imperative to carefully analyse the available information to analyse where the opportunities may be there. Various factors need to be studied in order to decide on the Location such as: 1) Demographics [Population/Ethnicity - may give an indication of cuisine preferences 2) Existing restaurant data to gauge interest 3) Potential information in restaurant ratings to identify potential gaps 4) Where are existing restaurants located are there hotspots in the city which can be identified, are there are other venues nearby which indicate hot spots.

The Client X need to choose the correct location to start their restaurant and potentially expand a later date.

c) Target Audience:

To recommend range of locations and potentially cuisine, X Company Ltd. The objective is to locate and recommend to client potential range of recommendations where they can potentially launch a restaurant business.

This would interest anyone who wants to start a food serving business in London.

d) Success Criteria:

Range of recommendations and the rationale behind the recommendations.

2. Data:

Data 1: General Information from Wikipedia There are 32 local district authorities. Twelve are designated as Inner London and the Rest are Outer London. More information about these can be found from https://en.wikipedia.org/wiki/London_boroughs

Post Code information for London is available from https://en.wikipedia.org/wiki/London_postal_district.

Data 2: Data from Hygiene Rating https://ratings.food.gov.uk/open-data/en-GB. The implied assumption is every available restaurant in London would be included in the list. The data provides a list of restaurants in London and their Hygiene ratings. It would be useful in determining restaurant densities for various district authorities and hygiene standards expected from boroughs. The data also provides location information in form of Post Codes and geographical co-ordinates.

Data 3: Census data to determine the ethnic mix of the population It can be useful to cluster various boroughs together basis their population mix and then we can analyse the restaurant makeup across boroughs

Ethnic Group projections are taken from the following link https://data.london.gov.uk/dataset/ethnic-group-population-projections
***Data 4:*** London city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood will use the Foursquare API to explore neighbourhoods in New York City.

3. Methodology:

Following data is obtained from Census projections.

The data was loaded from the excel file. The date was cleaned up and only year 2020 was taken since we need a representative population and not time series across years.

Below is the demographic distribution of the data.
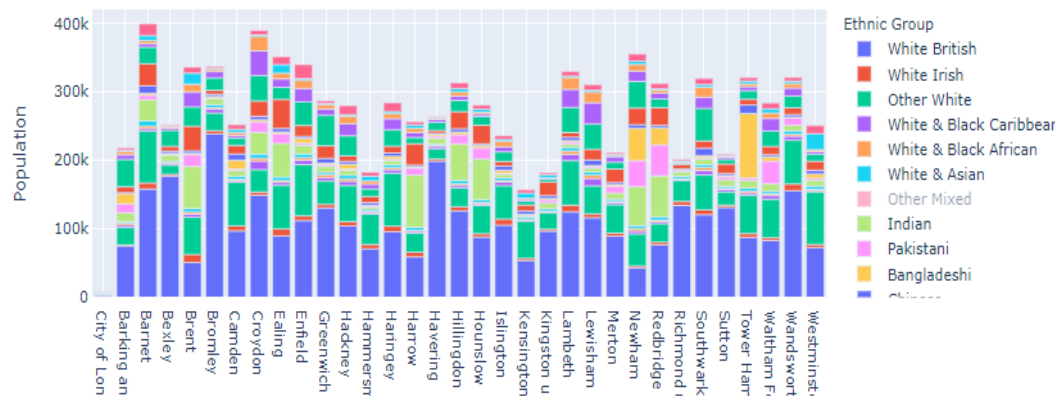
Data indicates there is not too much distribution of population amongst boroughs but individual boroughs have markedly different concentrations of various ethnic groups.

This data was normalised to computed percentage of population of each ethnic group and then clustered into 5 categories using k-means.

Cluster 5 was selected for further analysis.

Restaurant data

Xml files were downloaded for all the restaurant list for London. These were converted to excel file outside the python notebook. The excel was then read into notebook using pandas data frame. The data was cleaned up and the boroughs were matched with census data.

During exploratory analysis it was found that 2 boroughs (City of London, Westminster) have unusually large restaurant density most likely due to office space and hence large floating population.

The data was used to get different Post Codes in each borough through which further venue analysis could be done.

FourSquare API:

Venue details are obtained from Four Square API for the Post Codes identified in previous exercise and geographical data available from the restaurant data file. These venues are then sorted to get most preferred venue for each of the different Post codes. The resultant data is clustered into 3 clusters through k-means to obtain various clusters.

4. Results:

Clustering results from population demographic:

From the various data we gathered and processed, we focussed on restaurant business and identified 2 clusters using k-means. The fist cluster identified was basis demographics, which gave us 5 clusters of populations. We decided to focus on Cluster 5 (no 4 in data table) since it generally had the least restaurant density.

Cluster 1: Generally, around Central London with very high restaurant density. No exploration further.

Cluster 2: Few boroughs with average restaurant density

Cluster 3: Mis of clusters with a few clusters with very high restaurant density and others with low. It seems the difference may be in affluency of the residents.

Cluster 4: Just one cluster Tower Hamlets with most diverse population mix. This is in line with empirical observation with Tower Hamlets being home to canary wharf where in a lot of expatriates stay.

Cluster 5:

Boroughs on outskirts of inner London with largely South East Asian population (Indian, Pakistani, Bangladesh).

Cluster Labels	borough	rest_density
0	City of London	1123.188406
0	Westminster	104.486809
0	Kensington and Chelsea	54.981746
0	Camden	53.654017
0	Hammersmith and Fulham	39.737017
2	Islington	39.651889
2	Hackney	36.216143
3	Tower Hamlets	30.305912
1	Havering	28.588001
2	Southwark	27.895957
2	Merton	25.279633
2	Wandsworth	23.538810
1	Richmond upon Thames	22.511476
2	Lambeth	21.409745
4	Ealing	21.204363
2	Croydon	20.787985
2	Lewisham	19.292685
2	Kingston upon Thames	18.872618
0	Haringey	18.618456
2	Waltham Forest	17.842811
4	Hounslow	17.074379
2	Hillingdon	16.777311
2	Greenwich	15.698587
4	Brent	15.217949
4	Newham	15.107872
1	Bromley	14.939670
1	Sutton	14.626505
4	Harrow	12.841462
1	Bexley	12.640333
2	Barnet	12.418939
2	Enfield	12.418182
4	Redbridge	11.921996
2	Barking and Dagenham	8.005753

*rest_density refers to restaurants per 10000 people

Withing the various Post Codes in the Cluster 5, k-mean analysis was performed to identify cluster and their preferences with respect to venues.

Clustering results from Venue Analysis:

With in the various clusters identified through k-means following are the results:

1) Cluster 1: People seem to prefer coffee shops and pizza place in these locations.
2) Cluster 2: Most promising cluster with just 5 post codes, and general population seems to prefer restaurants as their top 2 choice venues
3) Cluster 3: Mixed results with a lot of people preferring Grocery store. Most likely these are areas close to housing neighbourhoods, and restaurants are located farther away than these areas.

Results of Cluster 2:

borough	Post	Longitude	Latitude	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
Redbridge	IG4	0.057859	51.581348	1	English Restaurant	Hotel	Thai Restaurant	Yoga Studio	Discount Store	Event Service	Electronics Store
Newham	E12	0.049225	51.545979	1	Indian Restaurant	Hotel	Restaurant	Train Station	Gym / Fitness Center	Grocery Store	Sports Club
Hounslow	TW4	-0.387928	51.473233	1	Indian Restaurant	Hotel	Grocery Store	Pharmacy	Platform	Park	Asian Restaurant
Harrow	HA8	-0.278564	51.597663	1	Food & Drink Shop	Indian Restaurant	Yoga Studio	Discount Store	English Restaurant	Electronics Store	Eastern European Restaurant
Ealing	UB1	-0.371499	51.510821	1	Indian Restaurant	Coffee Shop	Park	Sandwich Place	Food	Supermarket	Gastropub

5. Discussion:

- There is untapped potential across boroughs of Newham, Ealing, Brent, Harrow, Redbridge to open more restaurants given lower restaurant density compared to rest of London.
- The areas generally have south Asian population. Within these 5 identified postcodes indicate people prefer restaurants in these areas.
- Further analysis may be done using hygiene rating etc.

6. Conclusion:

The analysis is performed on limited data abd a number of assumptions are made which may need to be validated.

The analysis indicates the best places to open a restaurant would be IG4/E12/TW4/HA8 and UB1 withing Cluster 5 boroughs with primarily south east Asian population. It also indicated the opportunity in these areas would be to open a Indian restaurant followed by others.

7. Acknowledgements:

Coursera team for the sample exercises