Battle of neighbourhood

London

Introduction

- Business problem:
  - Open a restaurant in London
  - Identify areas in London
  - Identify similar locations where these may be successful
- Success
  - Recommendation with clearly articulated reasons

About london

- Most populous city in UK
- Diverse Culture
- Financial Hub and home to a lot of expats
- Home to people of a number of ethnicities

Deciding on location of restaurant

- Demographics of the location
  - Ethnicity/age etc
- Restaurant coverage
  - How many restaurants and there types
- Are there other venues which may affect desirability of restaurants
- Cuisine

Data Description:

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

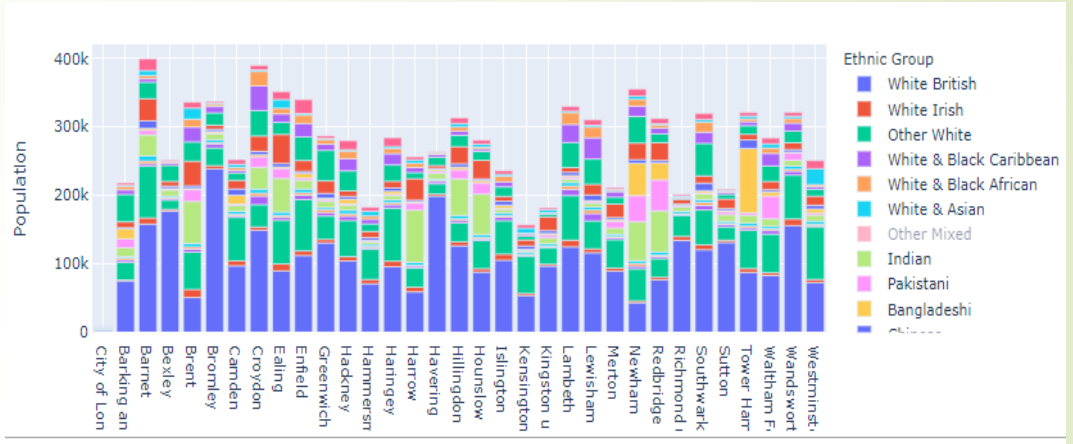
Ethnic Group projections are taken from the following link https://data.london.gov.uk/dataset/ethnic-group-population-projections

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.

Methodology

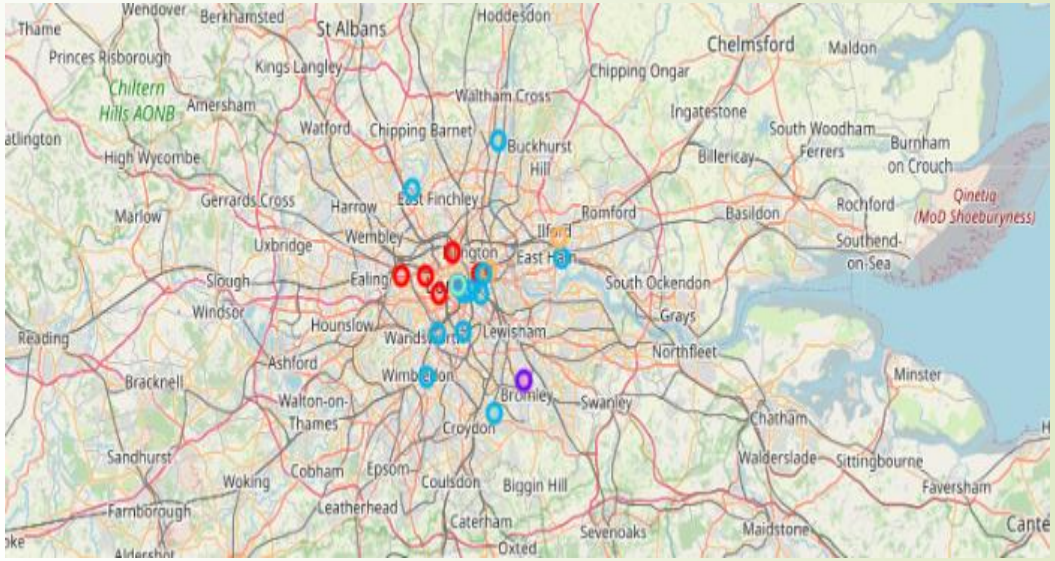
- Demographics of various boroughs were obtained

borough	ethnic_group	2020
City of London	All persons	7452
City of London	White British	2841
City of London	White Irish	141
City of London	Other White	2180
City of London	White & Black Caribbean	35
...	...	...
Greater London	Black Caribbean	349925
Greater London	Other Black	218203
Greater London	Arab	147457
Greater London	Other Ethnic Group	226769
Greater London	BAME	4045840



Methodology

- Clustering was done basis available data to identify clusters with similar demographic mix
- Cluster 5 was selected for further analysis



Methodology

- 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.
- Four Square 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.

Result

- **Clustering results from population demographic:**
  - City of London and west minister have a high no of restaurants likely due to Office hubs
  - 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

Result

- **Venue Analysis**
  - Cluster 1: People seem to prefer coffee shops and pizza place in these locations.
  - Cluster 2: Most promising cluster with just 5 post codes, and general population seems to prefer restaurants as their top 2 choice venues
  - 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.

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

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.

Conclusion

- The analysis performed on limied data with various assumptions
- Best places to open restaurant
  - IG4
  - E12
  - TW4
  - HA8
  - UB1
- Indian restaurant in the above Post Codes basis current preferences

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