Finding best London borough for Indian Restaurant

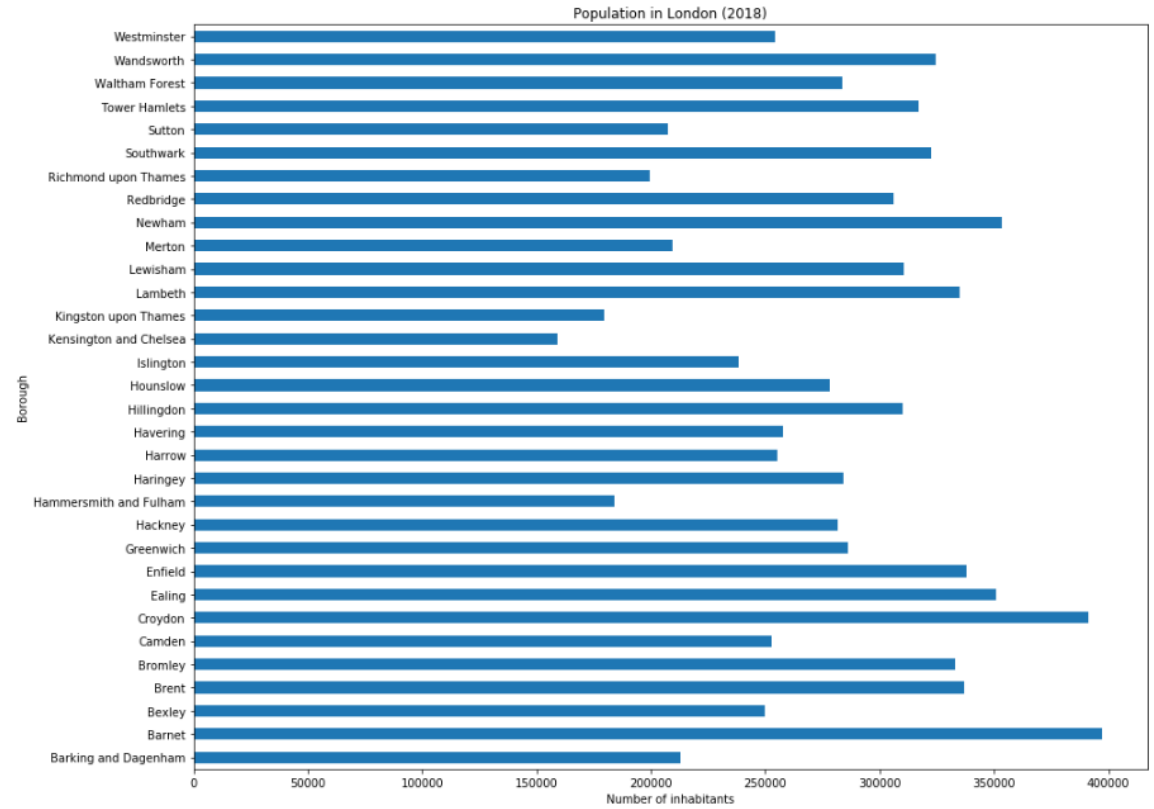
Introduction

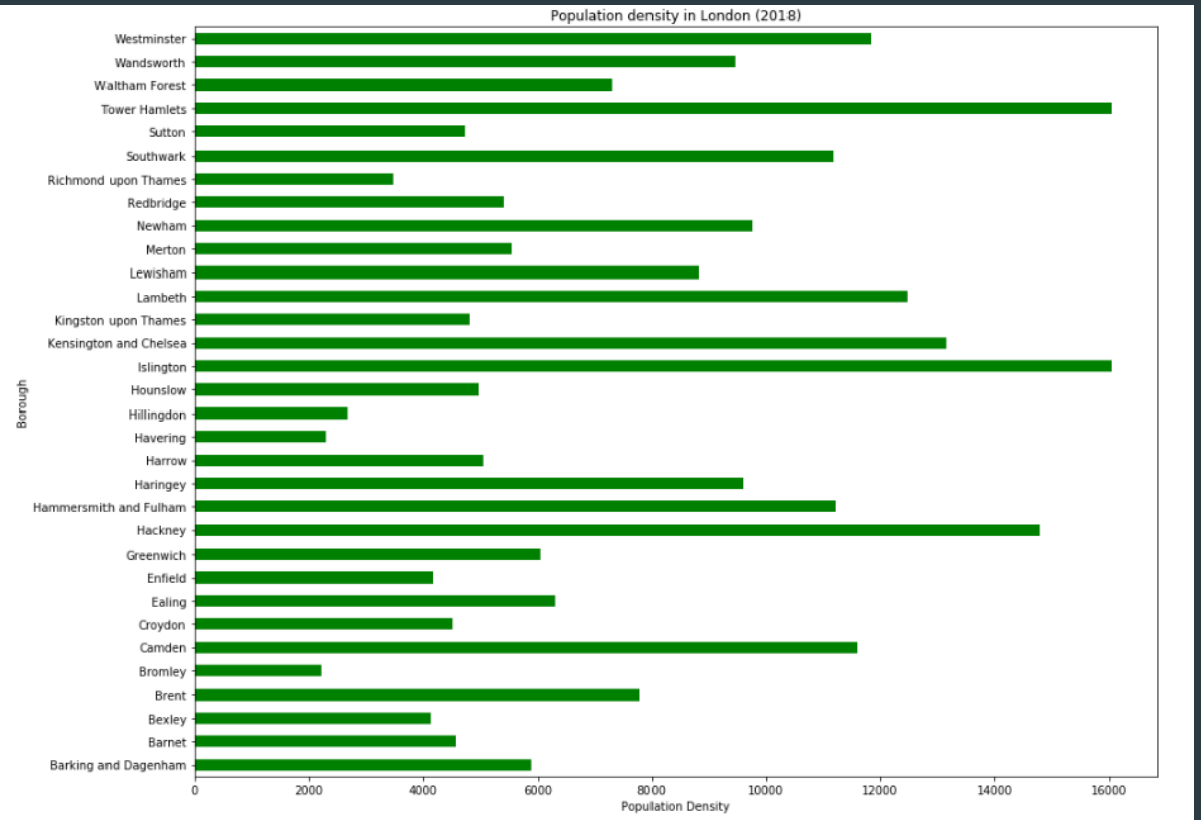
- London has huge variety of restaurant for every taste and thus to start restaurant business may not be an easy task.
- Our stakeholder would like to open an Indian restaurant with middle-high level price.
- There are many criteria which affect the decision to select the location for restaurant to maximize revenue.
- In this project we perform basic analysis to chose most optimum borough in London for opening an Indian restaurant.
- The analysis is purely based on the population, density of current restaurants and income per person in the borough.

Data acquisition and cleaning

- Data related Net income per person, population, population density and Asian community population for each borough was sourced from London datastore https://data.london.gov.uk/dataset/
- Coordinates of all the borough in London was sourced from Wikipedia https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough
- Number of restaurants and Indian restaurants within certain radius of each borough was sourced using Foresquare API.
- All the relevant data from each source was combined in one dataframe, which required cleaning, formatting and harmonizing from each source.
- Cleaned data has 32 rows (for each borough) and 10 columns.

Exploratory Data analysis - population





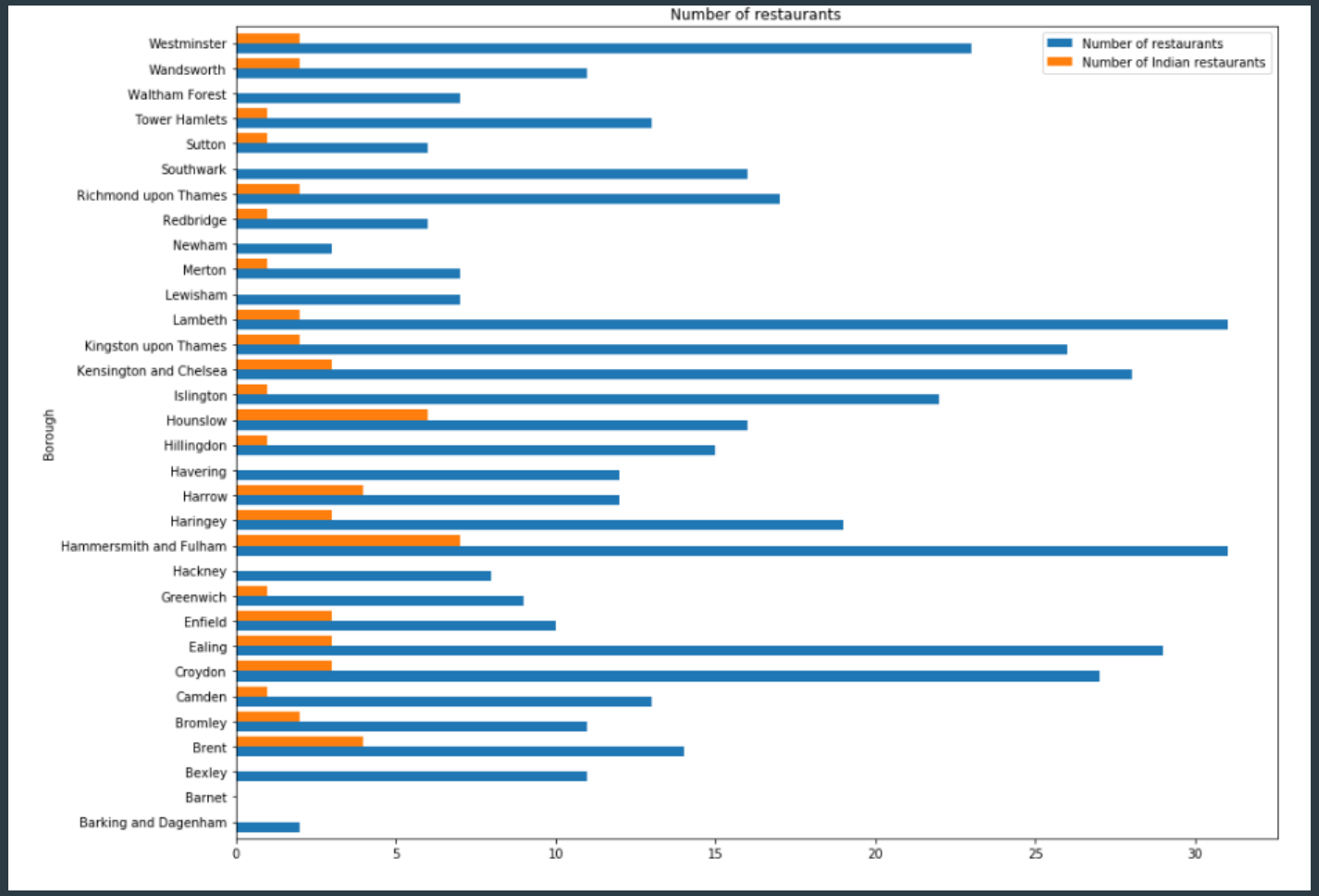
Simple bar chart of population data reveals that even if the most populated borough is Barnet the population density in that borough is only average as the area of this borough is really large. The most densely populated borough on the other hand is Tower Hamlets and Islington.

Exploratory data analysis - Restaurants

Simple bar chart analysis on number of restaurant gives some idea how these are distributed over different boroughs.

Hammersmith and Fulham clearly have large number of restaurants as well as good proportion of that are Indian restaurant.

On the other hand Westminster doesn't have enough proportion of Indian restaurant



Cluster Analysis

- Two cluster analysis was performed
  - First using total population
  - Second using Asian population only

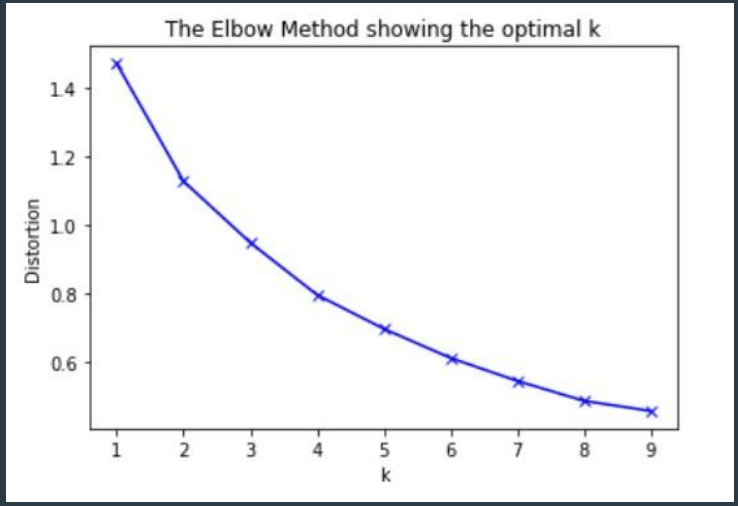
- To identify groups (clusters) with similar characteristics, the unsupervised learning method to our data, namely K-Means algorithm, was applied to our data.

- The strategy was to compare the result and the common set from optimum clusters of two analysis might be the most optimum borough.

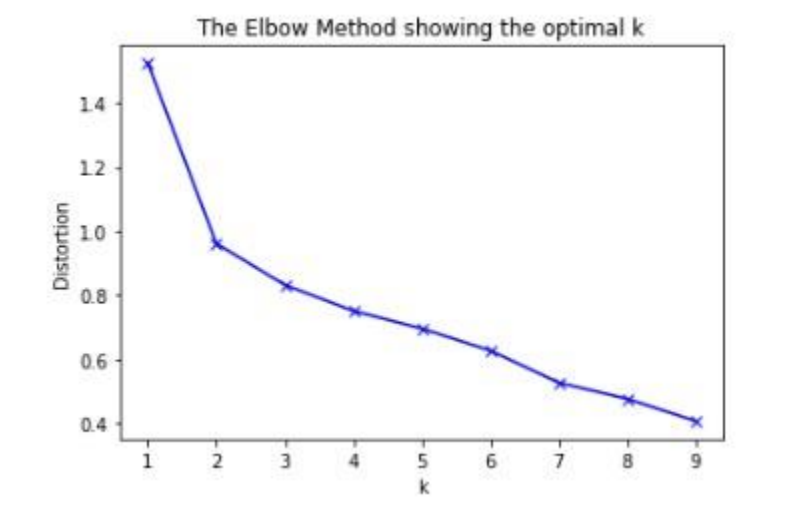
Cluster analysis – No of clusters

- To identify the optimum number of clusters the elbow method was used

- When doing analysis using total population 3 clusters were used whereas only 2 clusters were used when doing analysis using Asian population only



Using total population



Using Asian population only

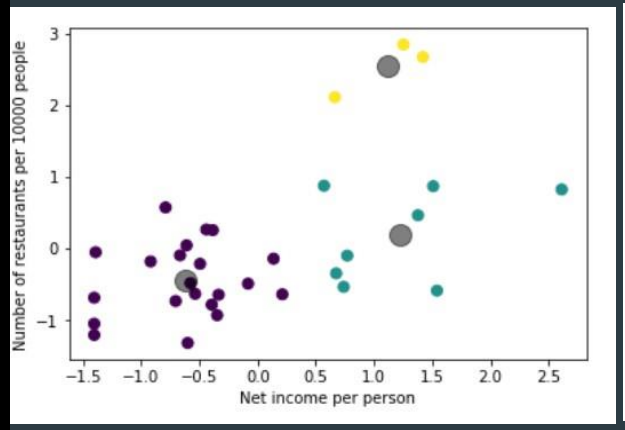
Custer analysis - Result

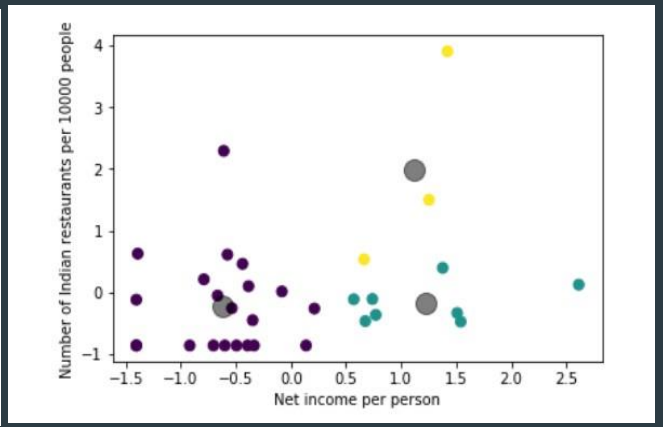
- Scatter plot has been shown. Grey color shows the center of cluster.
- Using total population one cluster is clearly an outlier which has high number of restaurants. Out of these one also has high number of Indian restaurants.
- Both analysis shows there is good number of boroughs in a cluster where income is high but restaurant density is not so high, especially Indian restaurant. Our target should be these clusters.

Restaurant density vs Net income

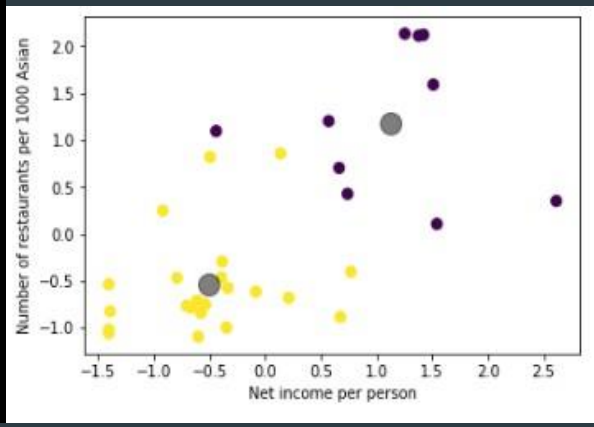
Indian Restaurant density vs Net income

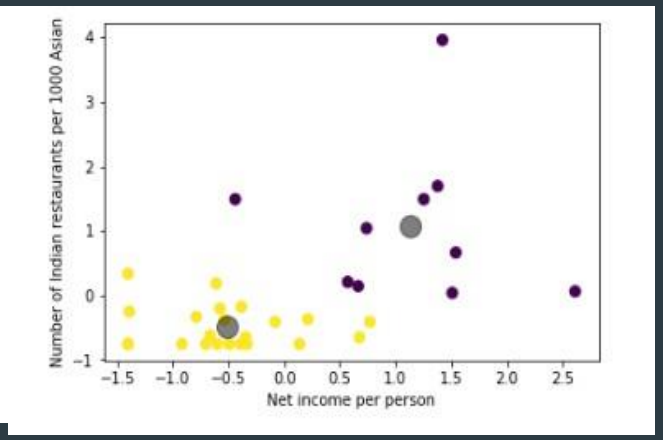
Using total population





Using Asian population only





Cluster analysis – Optimum cluster

Using total population

	Cluster Labels	Borough	Latitude	Longitude	Population	Population Density	Asian Population	Asian Density	Net income per person	Number of restaurants	Number of Indian restaurants
4	1	Bromley	51.4039	0.0198	332733	2216	15000	99	632.5	11.0	2.0
5	1	Camden	51.5290	-0.1255	252637	11594	39000	1789	634.7	13.0	1.0
17	1	Islington	51.5416	-0.1022	238267	16037	17000	1144	687.6	22.0	1.0
20	1	Lambeth	51.4607	-0.1163	334724	12485	28000	1044	620.4	31.0	2.0
25	1	Richmond upon Thames	51.4479	-0.3260	199419	3473	11000	191	678.2	17.0	2.0
28	1	Tower Hamlets	51.5099	-0.0059	317203	16035	128000	6470	627.9	13.0	1.0
30	1	Wandsworth	51.4567	-0.1910	324400	9467	19000	554	689.9	11.0	2.0
31	1	Westminster	51.4973	-0.1372	254375	11838	33000	1535	766.6	23.0	2.0

Using Asian population only

	Cluster Labels	Borough	Latitude	Longitude	Population	Population Density	Asian Population	Asian Density	Net income per person	Number of restaurants	Number of Indian restaurants
4	0	Bromley	51.4039	0.0198	332733	2216	15000	99	632.5	11.0	2.0
11	0	Hammersmith and Fulham	51.4927	-0.2339	184050	11224	20000	1219	681.3	31.0	7.0
12	0	Haringey	51.6000	-0.1119	284288	9604	18000	608	548.1	19.0	3.0
17	0	Islington	51.5416	-0.1022	238267	16037	17000	1144	687.6	22.0	1.0
18	0	Kensington and Chelsea	51.5020	-0.1947	159301	13139	18000	1484	669.3	28.0	3.0
19	0	Kingston upon Thames	51.4085	-0.3064	179581	4819	30000	805	627.1	26.0	2.0
20	0	Lambeth	51.4607	-0.1163	334724	12485	28000	1044	620.4	31.0	2.0
25	0	Richmond upon Thames	51.4479	-0.3260	199419	3473	11000	191	678.2	17.0	2.0
30	0	Wandsworth	51.4567	-0.1910	324400	9467	19000	554	689.9	11.0	2.0
31	0	Westminster	51.4973	-0.1372	254375	11838	33000	1535	766.6	23.0	2.0

Common set of boroughs from the two analysis are Bromley, Islington, Richmond upon thames, Wandsworth and Westminster, out of this Wandsworth and Westmisnter can be most attractive due to high value of income per person.

Conclusion and further direction

- Based on the analysis Westminster & Wandsworth were chosen the most attractive option.
- Further analysis can also be done on other Boroughs in the common set defined from two analysis based on other factors such as distance to city center etc.
- Once borough is selected deeper analysis can be done to find the best exact restaurant location in that borough taking into account factors such as number of parking places in the vicinity etc.
- Foursquare doesn’t represent the full picture as many venues are not listed, so another map such as Google map or Openstreetmap may provide more insights.
- Also Borough have complex geometry and doing analysis based on radius from center might have missed some key venues and prone to error. This can definitely be further improved.