

# IBM Applied Data Science Capstone Project

## The battle of Neighbourhood - Part 2

Shailendra Kumar Gupta

26 July 2020

Topic - To Find out the most suitable areas in London where we can open an Asian restaurant.



# Introduction

London is the capital and largest city of England and the United Kingdom. London is considered to be one of the world's most important global cities and has been called the world's most powerful, most desirable, most influential, most visited, most expensive, sustainable, most investment-friendly, and most-popular-for-work city. It exerts a considerable impact upon the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism and transportation. London ranks 26th out of 300 major cities for economic performance.

It has a large population of people from different foreign countries from Asia, Australia, America, Middle east. Where these people live, work and visit this place as tourist. One of biggest things people want when they visit foreign country, they want the places where they can find the food of their choice. So, in this project we try we use data science of find solution related with this problem. To find best places to open restaurant in London, analyse Ethnic Groups by Borough and number of restaurants in a borough.

## Business Problem

In this project we tried to find out most suitable place to open a Asian restaurant. Depending upon Asian population living in London, number of restaurant and local rent in each borough.

The main purpose this this business problem is to target the people belonging to Asian community these peoples may be

1. Asian student living in London for study.
2. Working professionals from Asian countries.
3. Tourist visiting from Asian countries who mostly prefer to eat their native food.

4. And finally, the local people who love to eat Asian food.

As there will be previously have some Asian or other types of restaurants, so for opening a restaurant we must figure out these places in order to find out most suitable places where we can open an Asian restaurant along with the most favourite venues in London.

## Data

To figure out the business problem the data we are using is the ethnic group data of boroughs which mainly consist of population of Asian people in different boroughs of London, which we can found on

<https://data.london.gov.uk/dataset/ethnic-groups-borough>

The second data we want is the data of number of restaurants in different borough which we will get with the help of Foursquare APIs.

And by using these data sets we try to figure the best suitable place where we can open an Asian restaurant.

# Initial Dataset Obtained:

Neighbourhood	White	Asian	Black	Mix	Total	Latitudes	Longitudes
Barking and Dagenham	109000	54000	36000	15000	215000	51.5541171	0.150504343
Barnet	250000	57000	30000	54000	390000	51.65309	-0.2002261
Bexley	195000	17000	21000	15000	248000	51.4416793	0.150488
Brent	102000	107000	62000	56000	328000	51.5638258	-0.275759656
Bromley	267000	15000	21000	28000	330000	51.4028046	0.0148142
Camden	158000	39000	19000	38000	255000	51.5423045	-0.1395604
Croydon	227000	70000	64000	24000	386000	51.3713049	-0.101957
Ealing	174000	96000	23000	50000	343000	51.5126553	-0.3051952
Enfield	206000	37000	74000	20000	337000	51.6520851	-0.0810175
Greenwich	178000	39000	48000	19000	284000	51.4820845	-0.0045417
Hackney	144000	32000	63000	42000	281000	51.5432402	-0.0493621
Hammersmith and Fulham	119000	20000	23000	19000	180000	51.4920377	-0.2236401
Haringey	180000	18000	34000	44000	276000	51.6014893	-0.1116958
Harrow	110000	98000	13000	28000	249000	51.5967688	-0.337275155
Havering	216000	13000	14000	15000	258000	51.5443687	-0.14430318
Hillingdon	154000	100000	24000	28000	305000	51.5425193	-0.448334931
Hounslow	130000	86000	25000	29000	269000	51.4686132	-0.3613471
Islington	165000	17000	25000	28000	235000	51.5384287	-0.0999051
Kensington and Chelsea	100000	18000	15000	22000	154000	51.4989948	-0.1991229
Kingston upon Thames	123000	30000	3000	21000	177000	51.4096275	-0.3062621
Lambeth	172000	28000	76000	49000	324000	51.5013012	-0.117287
Lewisham	193000	23000	58000	30000	305000	51.4624325	-0.0101331
Merton	135000	35000	16000	21000	207000	51.4108028	-0.188098506
Newham	111000	166000	47000	27000	351000	51.5299996	0.02931796
Redbridge	143000	126000	20000	18000	307000	51.5763203	0.0454097
Richmond upon Thames	170000	11000	-	14000	196000	51.4403723	-0.30571955
Southwark	200000	17000	64000	35000	315000	51.5029222	-0.103458
Sutton	147000	36000	11000	10000	204000	51.3575114	-0.173640123
Tower Hamlets	145000	128000	22000	19000	314000	51.5256294	-0.0335853
Waltham Forest	142000	44000	43000	48000	277000	51.5981694	-0.017836675
Wandsworth	242000	19000	39000	23000	323000	51.4570271	-0.1932607
Westminster	140000	33000	17000	51000	241000	51.5004439	-0.1265398

After preparing the initial dataset the second thing have not be done is obtain total number of restaurants and Asian restaurant

in each neighbourhood with the help of Foursquare APIs and Onehot encoding. As shown below

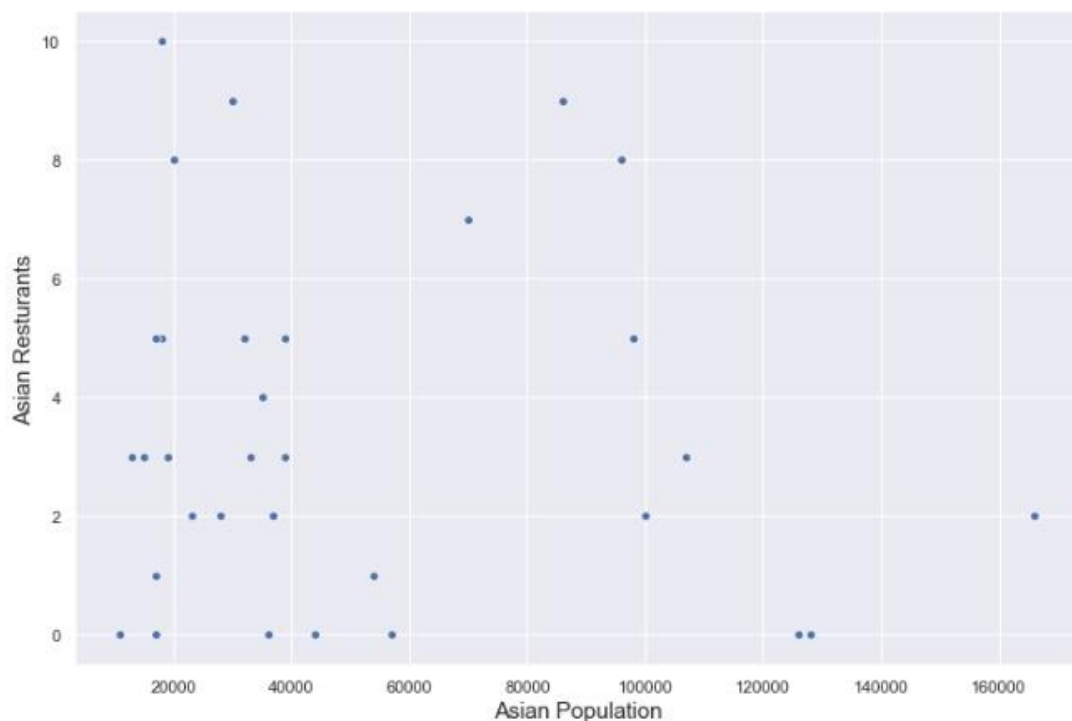
```
In [46]: Neigh_res.head()
```

```
Out[46]:
```

Sushi Restaurant	Japanese Restaurant	...	Chinese Restaurant	Mexican Restaurant	Seafood Restaurant	Taiwanese Restaurant	South Indian Restaurant	Middle Eastern Restaurant	Total_Res	Asian_Res	Asian_Population	Total_Population
0	0	...	1	0	0	0	0	0	1	1	54000	215000.0
0	0	...	0	0	0	0	0	0	3	0	57000	390000.0
0	0	...	0	0	0	0	0	0	4	0	17000	248000.0
0	0	...	0	0	1	0	0	0	12	3	107000	328000.0
1	0	...	0	1	0	0	0	0	9	3	15000	330000.0

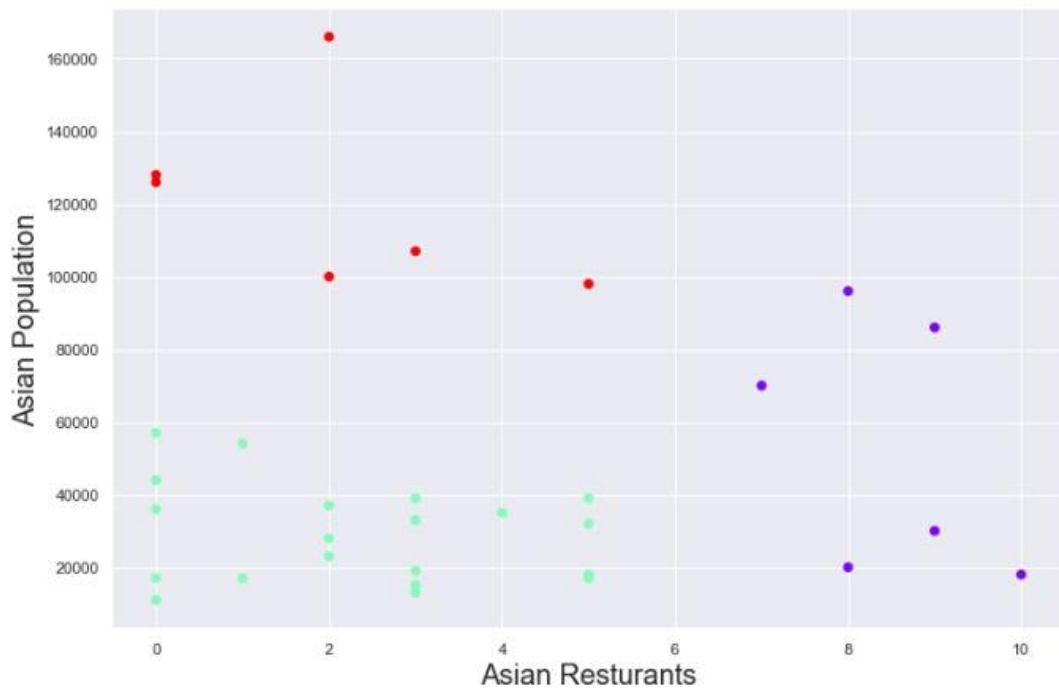
## Methodology:

As we found the required features of the data set, then now it's the time to do some exploratory data analysis, plotting scatterplot of above data using seaborn library.



In above scatterplot we can the distribution of Asian population vs Asian restaurants in different neighbourhood.

As we can see the above distribution, we will use Machine Learning techniques like clustering to segment the data in different segments. And try to find out how we will label the clustered segments.



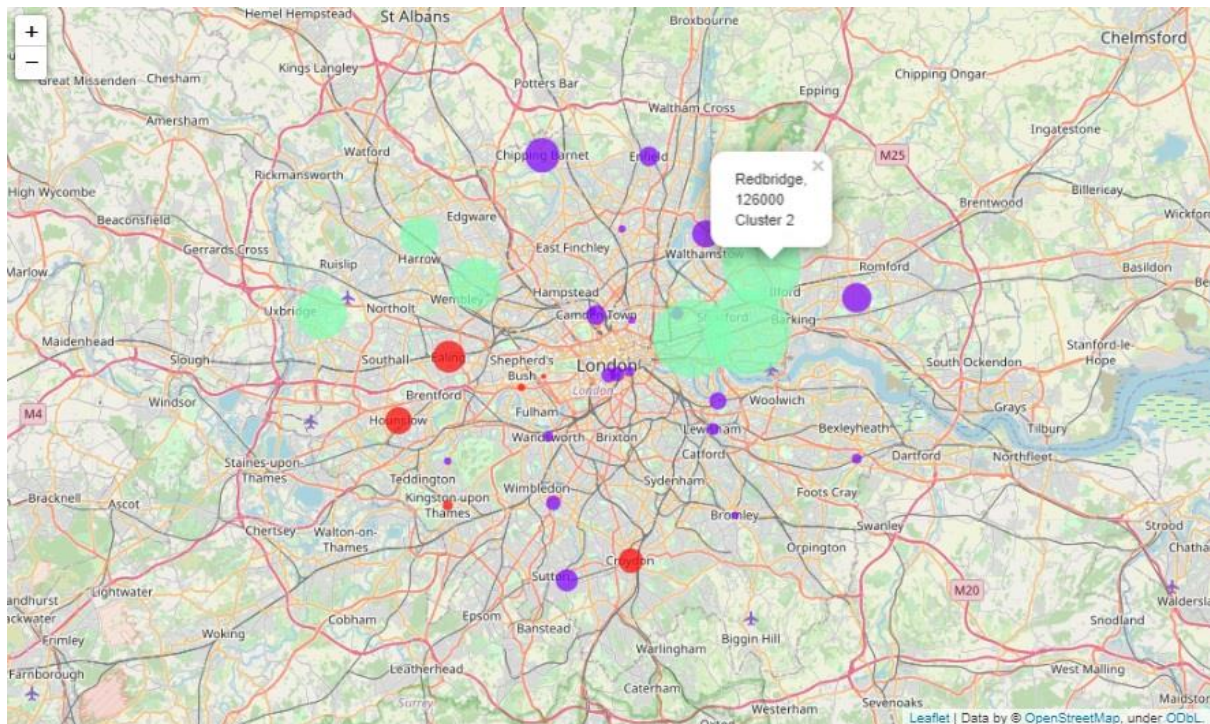
Depending Upon Above Cluster We can define restaurant venues in 3 Categories

1. Low Number of Restaurants, Low Population
2. High Number of Restaurants, Low Population
3. Low Number of Restaurants, High Population

Depending upon above categories We can say 3<sup>rd</sup> type of category is most suitable venue to do Asian restaurant business.



# Geo-spatial plotting of cluster



In this geospatial plotting we can see cluster number 2 in the desired cluster, this plot represents the location of the cluster, type of the cluster and the size of circle radius is corresponding to the more favourable neighbourhood within the same cluster.

## Result

By analysing the observations obtained from all plotting and clusters, we can say within all neighbourhood places the most suitable places are –

```
In [62]: data_with_clusters.head()
```

```
Out[62]:
```

	Neighborhood	Asian_Res	Asian_Population	Cluster	latitudes	longitudes
0	Barking and Dagenham	1	54000	0	51.554117	0.150504
1	Barnet	0	57000	0	51.653090	-0.200226
2	Bexley	0	17000	0	51.441679	0.150488
3	Brent	3	107000	2	51.563826	-0.275760
4	Bromley	3	15000	0	51.402805	0.014814

Also depending upon the other factor like number of people per restaurant in each neighbourhood we can find the most suitable place to maximise the profit.

## Discussion

- In this project the created data can be used for furthermore analysis like topmost common venue of each neighbourhood, or best restaurants in each neighbourhood.
- Inside business understanding section the main target of this business is to target Asian population living in London, but the same analysis we be used to target another group.
- Other clustering techniques like Density based clustering can be used to cluster a particular type of venue within all venues.
- In this analysis by adding more features like rent and average income of each neighbourhood we can explore more to maximise our profit.

## Conclusion

In this project, initially to get the required data web scraping, foursquare APIs, folium, geopy and basic data analysis python libraries have been used to create the initial data. As the initial data get prepared exploratory data analysis, machine learning techniques like clustering have been used to find out the required information we want, like most favourable place to open a particular type restaurant or top venues of any particular area.



## References:

1. <https://github.com/shailendra546/githubg>
2. <https://medium.com/@shailendragupta546/ibm-capstone-project-asian-restaurants-business-in-london-d3dd8ecac184>
3. <https://data.london.gov.uk/dataset/ethnic-groups-borough>
4. <https://www.coursera.org/learn/applied-data-science-capstone>
5. <https://developer.foursquare.com/docs/>