**Analysing neighbourhoods of London!**

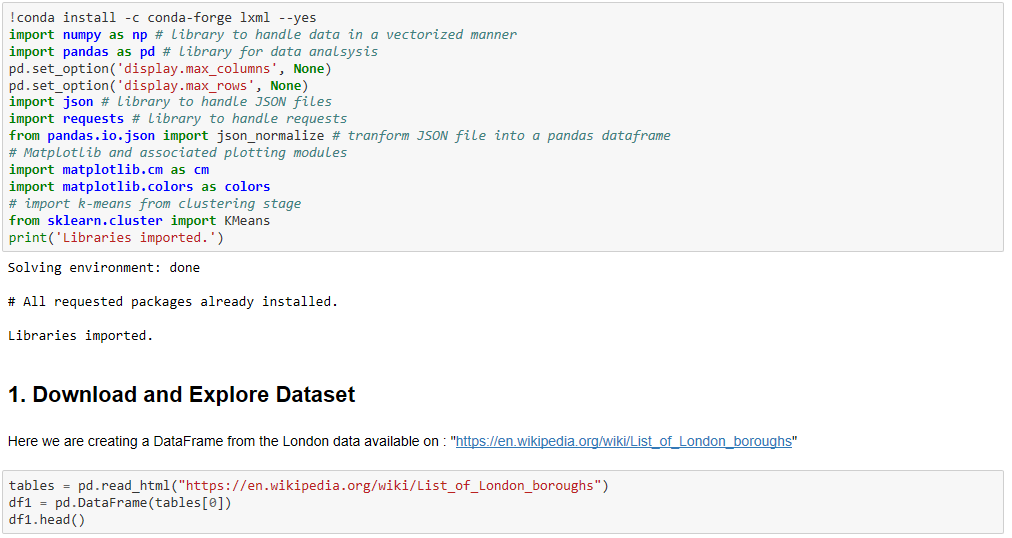
### Introduction:

We intend to analyse the neighbourhoods in City of London and will try to understand and explore neighbourhoods. Our intention is to get the most common venue categories in each neighbourhood, and then use this feature to group the neighbourhoods into clusters.

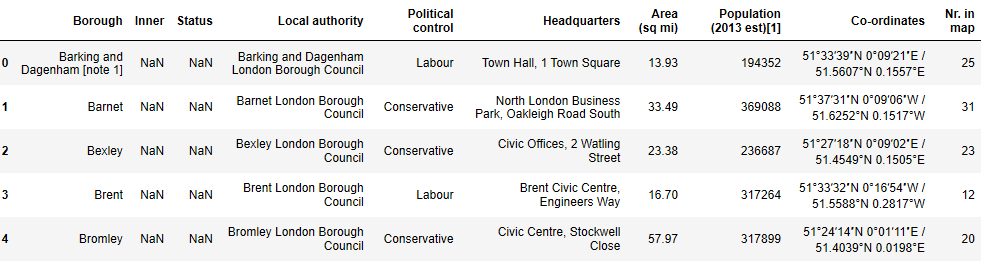
We will use the Foursquare API to explore neighbourhoods and get the relevant data for each neighbourhood.

We will use the k-means clustering algorithm to complete this task. Finally, we will use the Folium library to visualize the neighbourhoods in London City and their emerging clusters. This project will be useful for people coming in the City of London, which will help them with an idea of how similar and diverse different neighbourhoods in the City of London are. It would help them choose/pick the places of their choice easily, for the different activities they would like to do in the City of London.

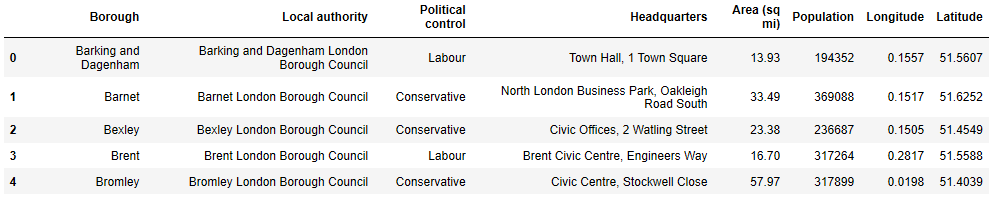
Before we get the data and start exploring it, we download all the dependencies that we will need:



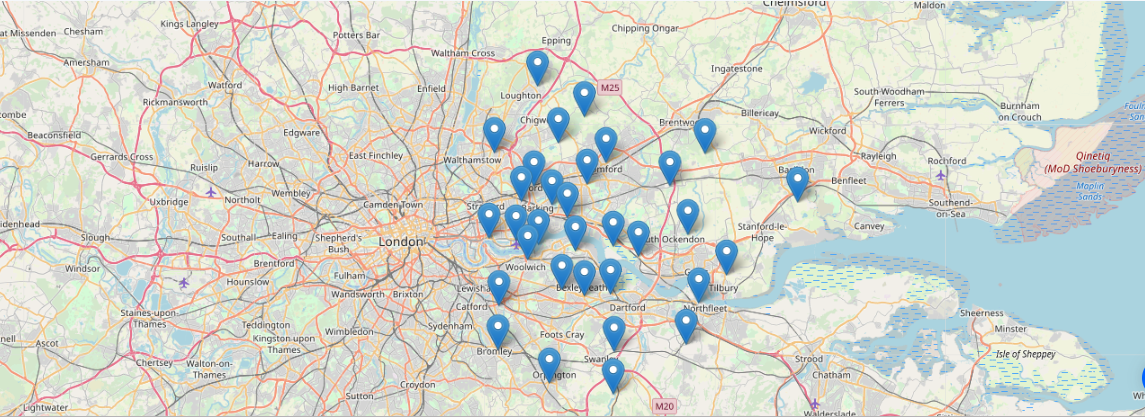
Here is now a sample of Data Frame that we just created:



Now we have (done all kind of Data wrangling and manipulations) and cleaned all relevant data in Data Frame .Now we have our final desired Data Frame on which we will perform our further analysis.

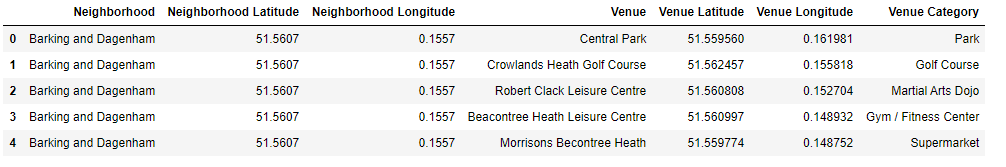


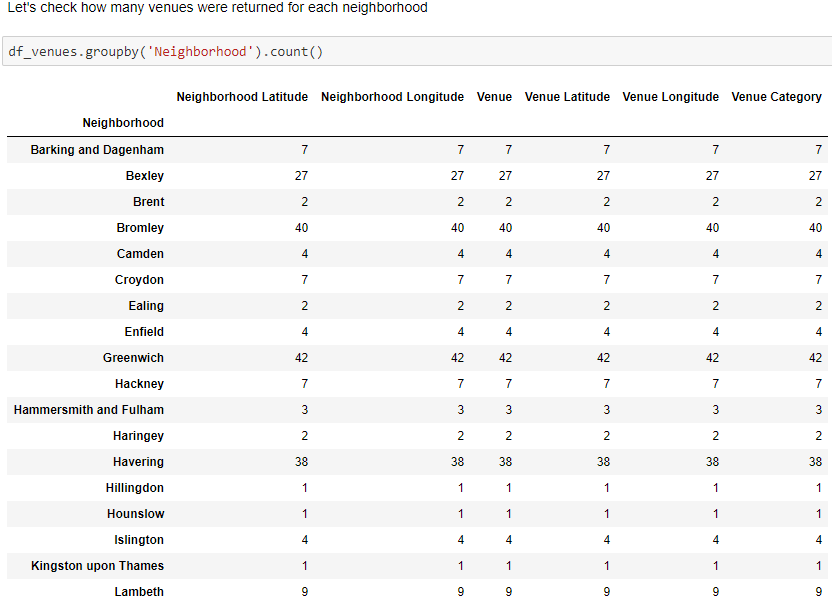
### Create a map of London using Folium with neighbourhoods superimposed on top:



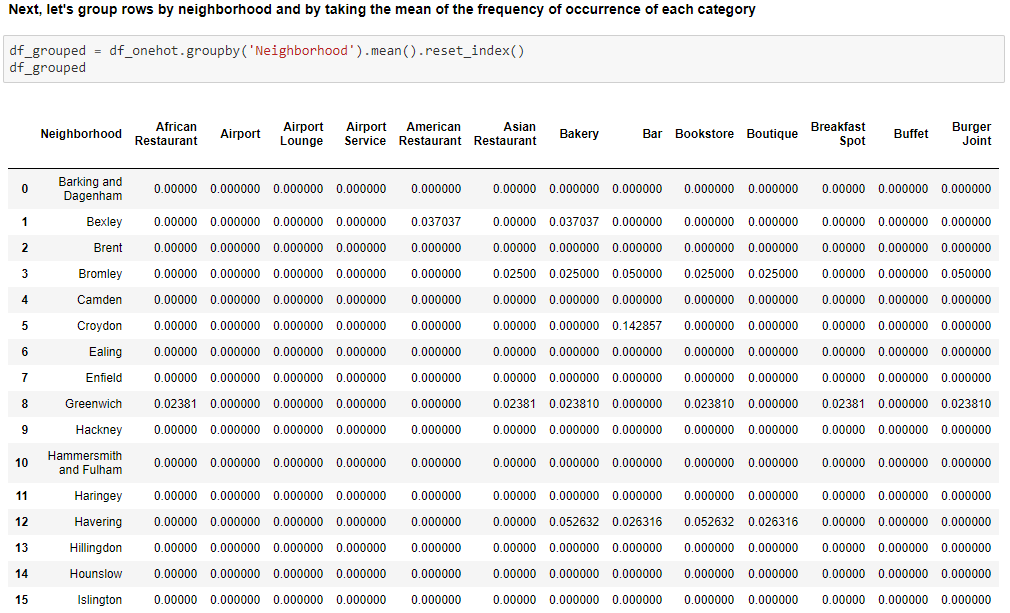
### Explore Neighbourhoods in London:

Next, we are going to start utilizing the Foursquare API to explore the neighbourhoods and segment them.



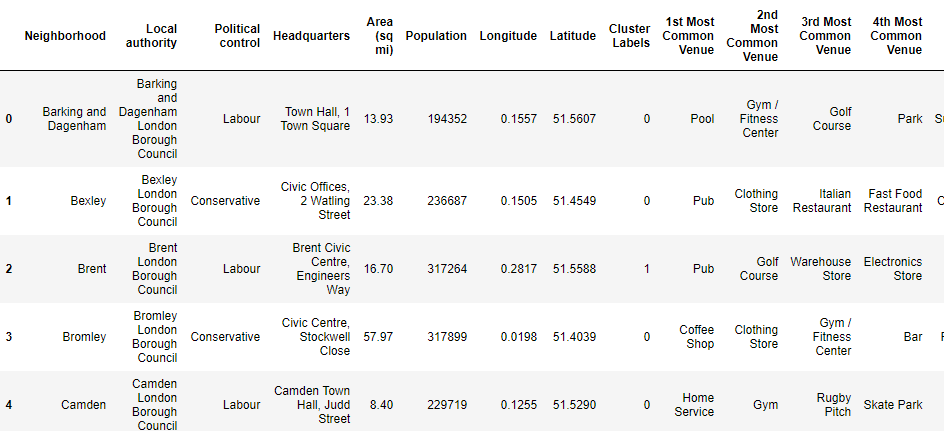


## Analyse Each Neighbourhood: (One-hot Encoding)

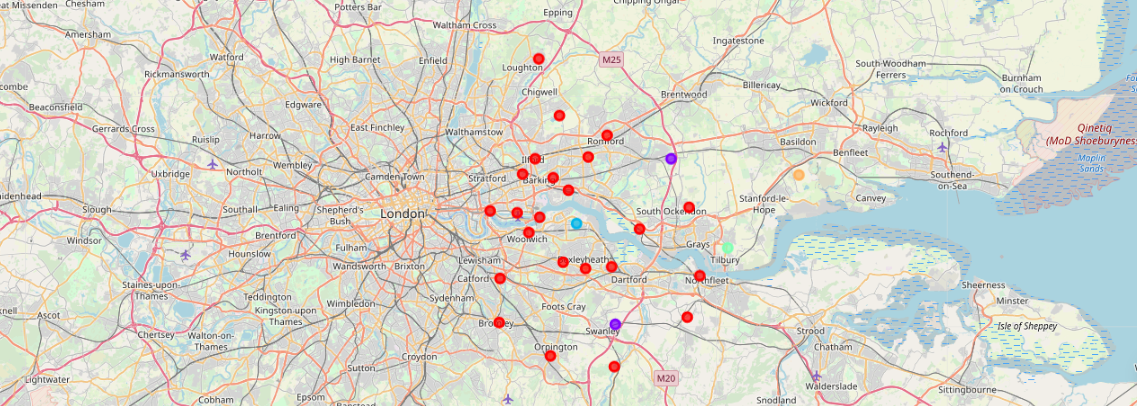


## Cluster Neighbourhoods: (Using K- means clustering)

Run k-means to cluster the neighbourhood into five clusters.



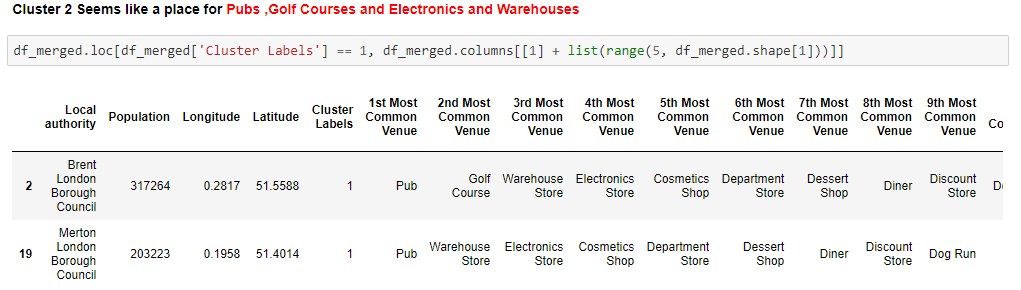
Finally, let us visualize the resulting cluster using Folium map:

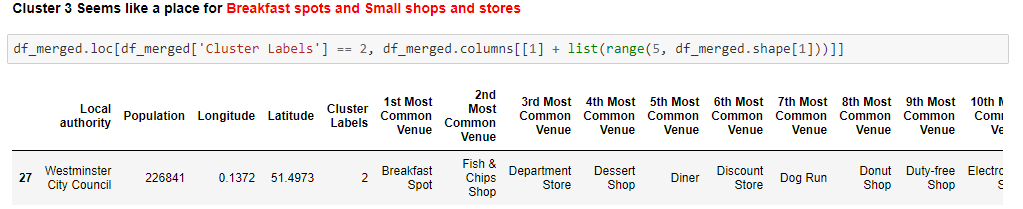


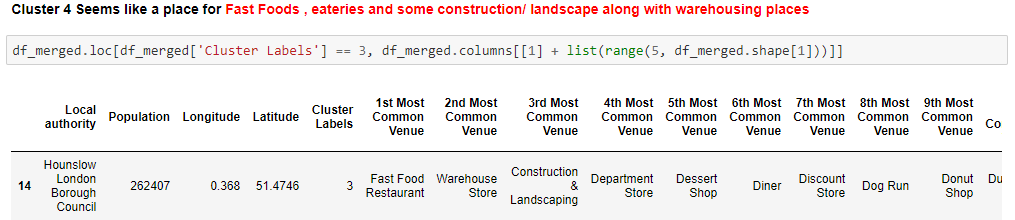
## Examine Clusters:

Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster. Here we can clearly see below how distinct and clearly similar/dissimilar the different clusters are among themselves.





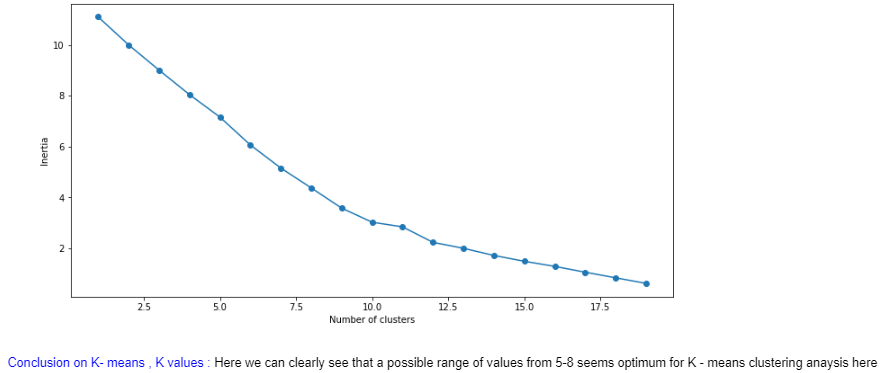






### Possible K values for optimal analysis of clusters {Elbow point Observation}:

We can also go ahead and try to analyse what K values we can choose to define different types of Clusters. There is a popular method known as elbow method, which is used to determine the optimal value of K to perform the K-Means Clustering Algorithm. The basic idea behind this method is that it plots the various values of cost with changing k. As the value of K increases, there will be fewer elements in the cluster. So average distortion will decrease. The lesser number of elements means closer to the centroid. Therefore, the point where this distortion declines the most is the elbow point.



### Results: Conclusion on K- means, K values:

Here we can clearly see that, a possible range of values, from 5-8 seems optimum for K - means clustering analysis used here.

### Discussion:

Our conclusion is two folds here:

We have identified that we can go for K=8 for optimum results on K-means clustering. We can then easily identify the clusters and their different categories for (Cafe , Pubs , Pools and Supermarkets) or (Pubs ,Golf Courses and Electronics and Warehouses) or (Breakfast spots and Small shops and stores) or (Fast Foods , eateries and some construction/ landscape along with warehousing places) or (Distinct with Stables , Cosmetics and Departmental store).

Another point is that, we had used above, value of LIMIT = 100 # limit of number of venues returned by Foursquare API. If we increase this number to a larger value, then we would have more data and venues to cluster and which would refine our data and analysis, both. This would result in crisp and much detailed findings.

### Conclusion:

Our conclusion is two folds here. We have identified that we can go for K= 8 for optimum results on K-means clustering. We can then easily identify the clusters and their different categories for (Cafe, Pubs, Pools and Supermarkets) or (Pubs, Golf Courses and Electronics and Warehouses) or (Breakfast spots and Small shops and stores) or (Fast Foods, eateries and some construction/ landscape along with warehousing places) or (Distinct with Stables, Cosmetics and Departmental store)

Another point is that we had used above LIMIT = 100 # limit of number of venues returned by Foursquare API. If we increase this number to a larger value, then we would have more data and venues to cluster, which would refine our data, and analysis both. This would result in crisp and much detailed findings.