**The Battle of Neighbourhoods: Coursera Capstone Project**

# Opening a new coffee shop in Queens, NY

### Introduction of the Business problem

#### Intro

I will explore the best location for a new coffee shop in Queens, NY. Queens is an up-and-coming borough with an emerging coffee culture.

#### Target group

Entrepreneurs looking to invest to a coffee house. Data scientists looking for common applications of foursquare data.

### *Data section*

For this project we need the following data:

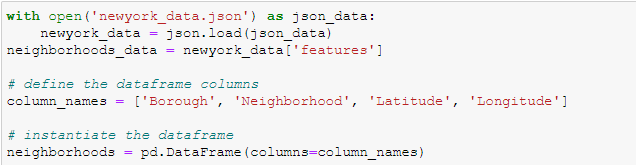
1. New York City data that contains Borough, Neighborhoods along with there latitudes and longitudes Data Source: <https://cocl.us/new_york_dataset> Description: This data set contains the required information. And we will use this data set to explore various neighborhoods of new york city.
2. Coffee shops in Queens neighborhood of new york city. Data Source: Foursquare API Description: By using this API we will get all the venues in the Queens neighborhood. We can filter these venues to get only coffee shops.

### *Approach*

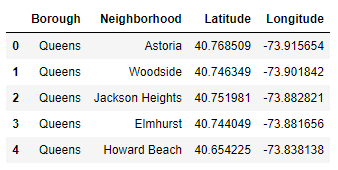
Collect the new york city data from <https://cocl.us/new_york_dataset>.  
Using Foursquare API we will get all venues for each neighborhood.  
Filter out all venues which are coffee shops.  
Data Visualization and some statistical analysis.  
Analyzing using Clustering (especially K-Means): Find the best value of K and Visualize the neighborhood with a number of coffee shops.  
Compare the Neighborhoods to Find the Best Place for Starting up a cafe.  
Inference From these Results and related Conclusions

Data Preparation

We get the data of the New York boroughs and neighbourhoods together with their coordinates.

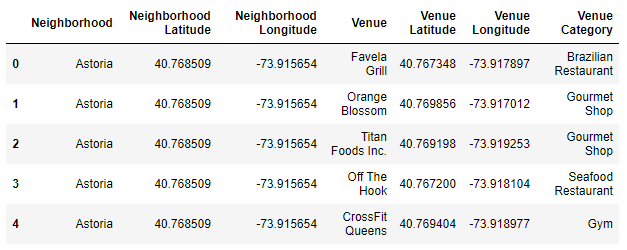


And after filtering for Queens, we get the data frame:



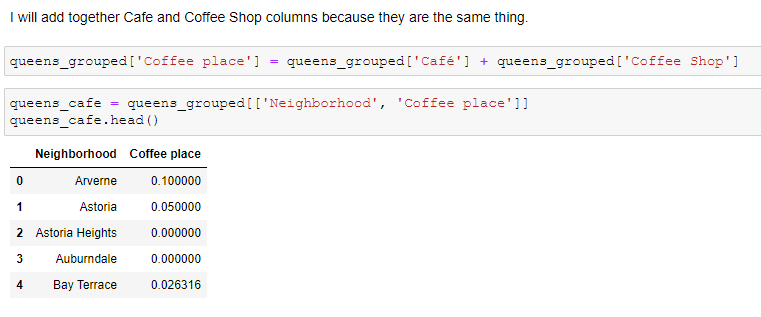
### *Using Foursquare Location Data*

For this business problem I have used, as a part of the assignment, the Foursquare API to retrieve information about the Venue, Venue category with their longitudes and latitudes.



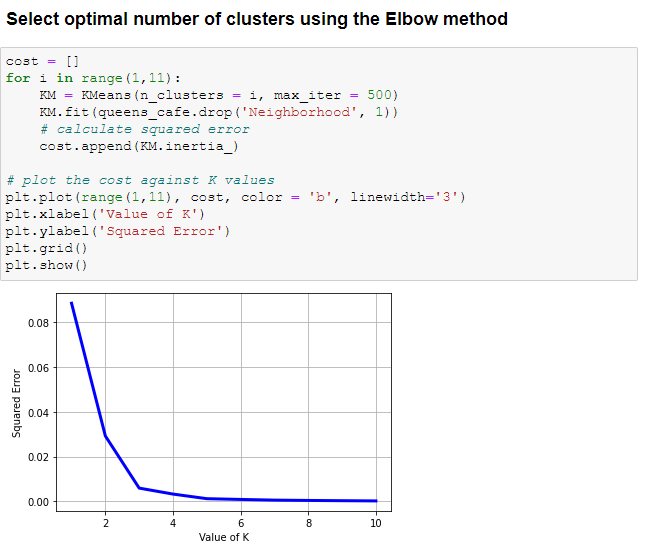
# Exploratory Data Analysis

There are 271 unique categories in which Coffee shop and Café are two of them. We will do one hot encoding for getting dummies of the venue category. Then we will calculate the mean of all venue groups by their neighbourhoods, which is interpreted as the relative frequency of this category.

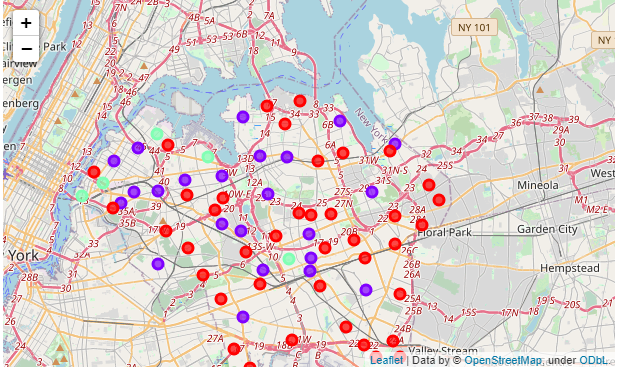


# Clustering the Neighbourhoods

We will extract Coffee place data from the above table and fit this into the code for finding the best value of K.



We split in 3 clusters and visualise using Folium:



The red cluster has the lowest frequency of coffee places, the purple has medium density and the light green the highest one.

### Discussion and Conclusion

Cluster 0 has the lowest frequency of coffee places. I will pick Queensbridge because it is the one closest to Manhattan.