Data

In this project, first, we need to obtain the data set that contains the Boroughs and the Neighborhoods of the New York City. Next, we need to obtain the latitude and longitude of each neighborhood of every borough in the city. Using them, we need to obtain the data of restaurants (along with type of restaurant) inside each Borough. Next, the restaurants related data set will be merged with the initial Borough data to obtain the complete data set. Now, we perform pre-processing of the data to obtain uniformity of the data. Later, we perform exploratory data analysis on each Borough data to obtain some crucial insights(Examples, like, Total number of restaurants in each neighborhood of a Borough, Top 10 most common type of restaurants in a neighborhood etc). Now, we will use K-mean clustering on each Borough data. After that, in each borough, every cluster will be analyzed in order to recommend the type of restaurant that can be opened inside the clusters of each Borough belonging to the New York city.

The data set that contains the details of Boroughs and their neighborhoods of New York city are obtained from https://geo.nyu.edu/catalog/nyu_2451_34572. Using Geocoder of Geopy, we will obtain the latitude and longitude of each neighborhood in the data set. From the FourSquare API location data, we will obtain the data of restaurants and their type using the latitudes and longitudes of neighborhoods present inside a Borough.