1. Introduction:

The purpose of this Capstone Project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in New York.

Lots of people are migrating to various states of New York and needed lots of research for finding better place. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, etc.

This Capstone Project aim to create an analysis of features for a people migrating to different boroughs in New York to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both fresh and waste water and excrement conveyed in sewers and recreational facilities.

It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life.

2. Data Section

Data Link: "<https://cocl.us/new_york_dataset>" This data set contains the required information and we use this data set to explore various neighborhoods in New York.

Foursquare API Data:

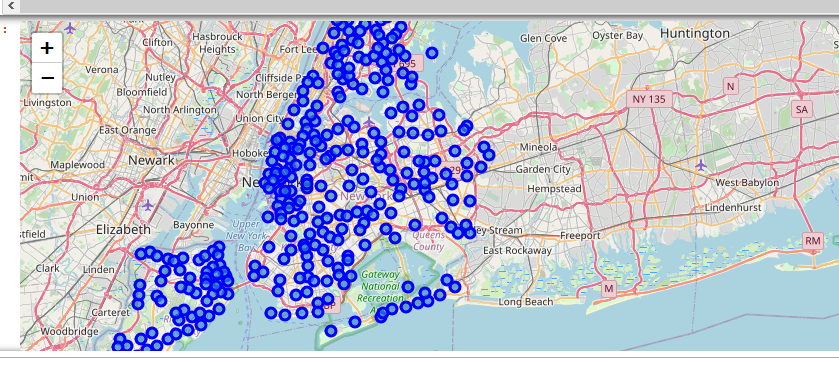
We use Foursquare location information to gain the information about different venues in different neighborhoods of that specific borough. Foursquare is a location data provider with information about all venues and events within an area of interest. Such information includes venues names, locations, photos and so on. So the Foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

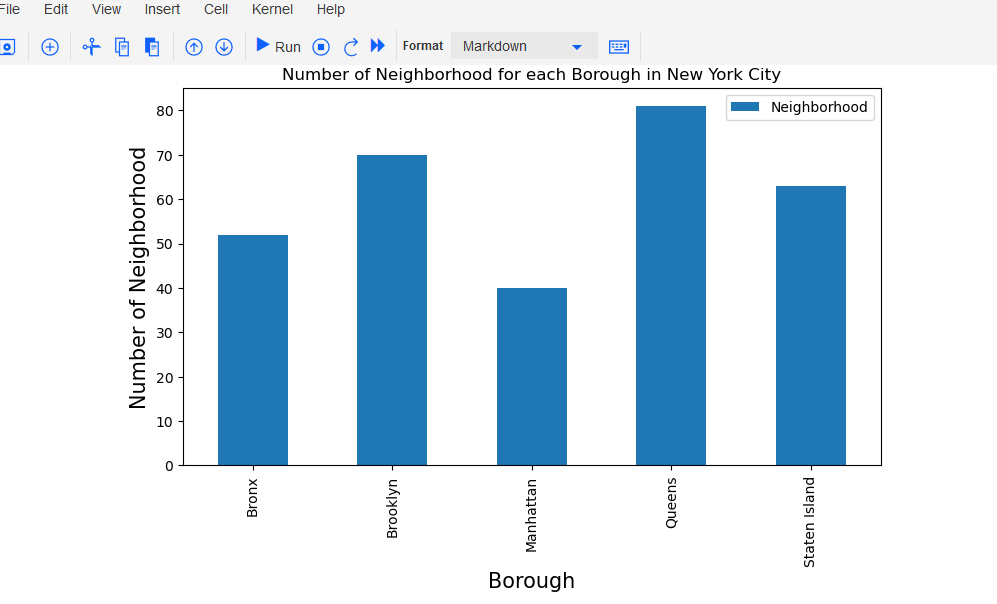
The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

**Map of New York**



**Graph to show different Neighborhoods of New York**



3. Methodology

Clustering Approach:

To compare the similarities of cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

**Using K-Means Clustering Approach** **Most Common Venues**



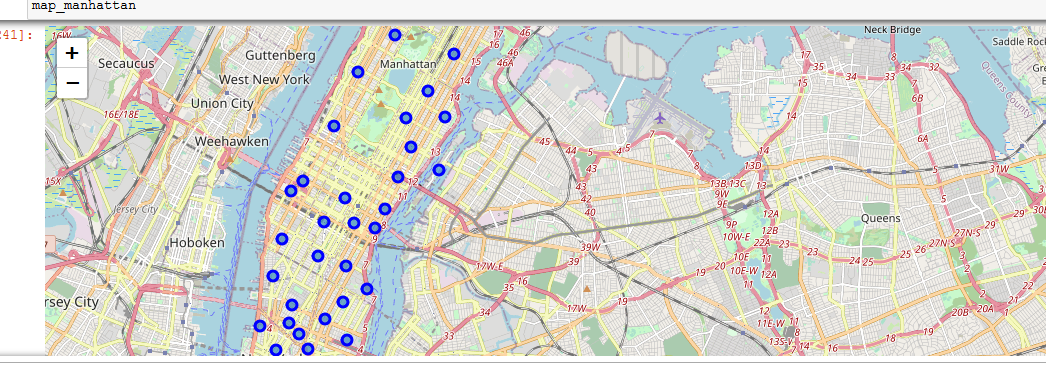
Work Flow:

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

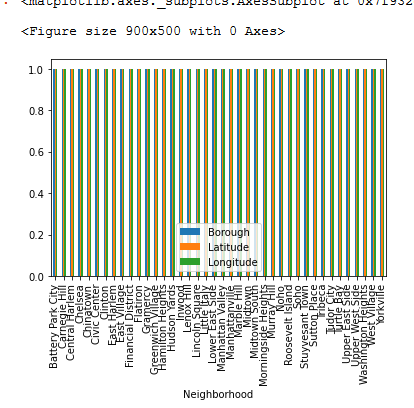
would be set to 500.

4. Results Section

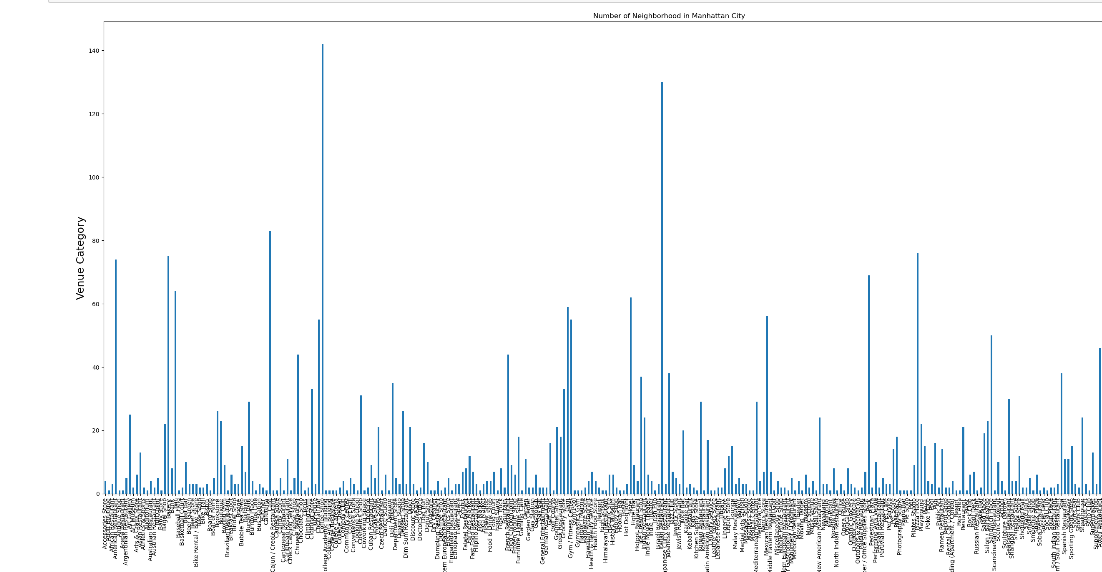
**Map of Clusters in Manhattan**



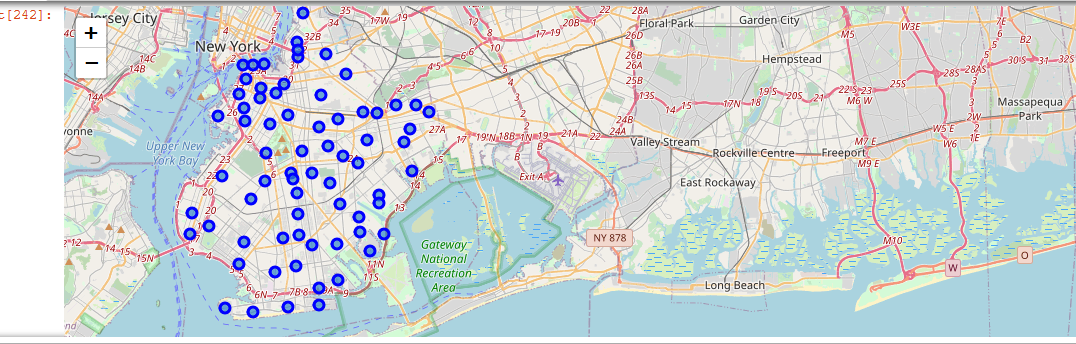
**Graph**



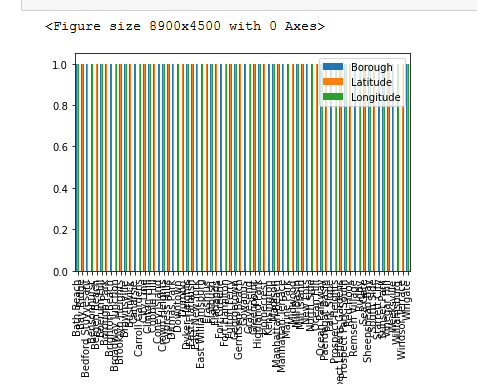
**Graph showing venues**



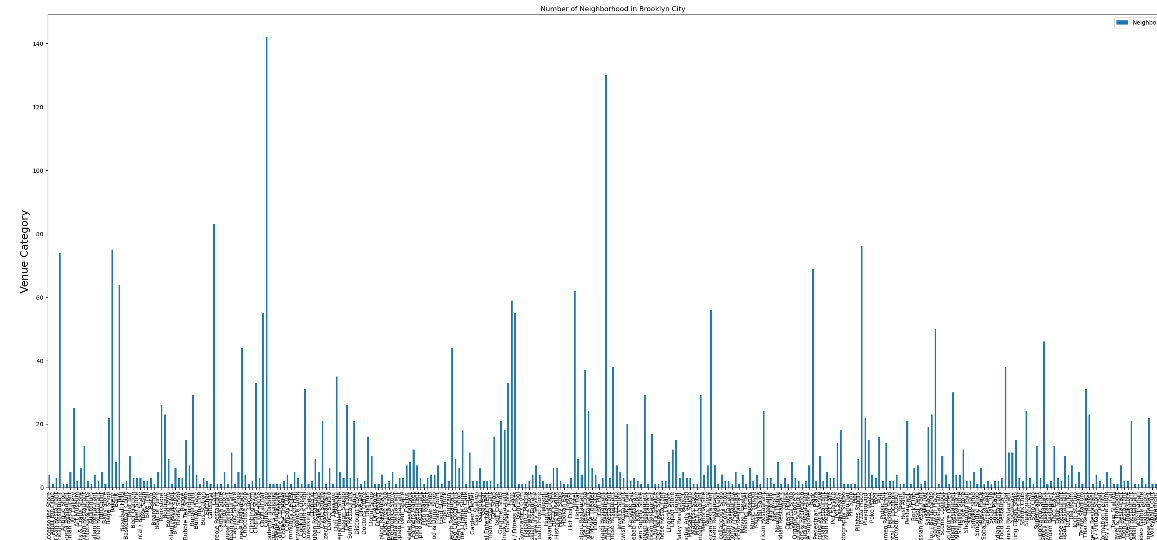
**Brooklyn**

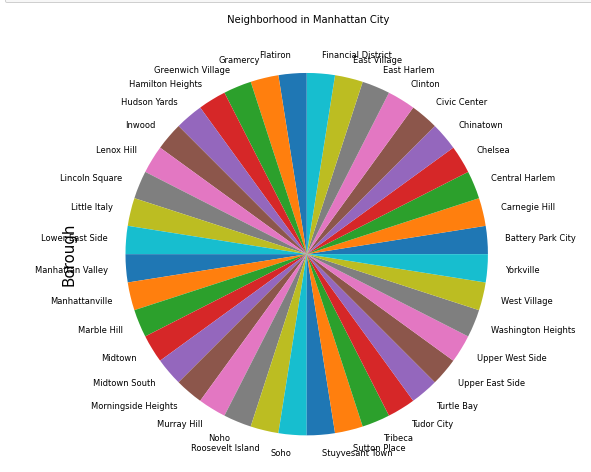


**Graph**

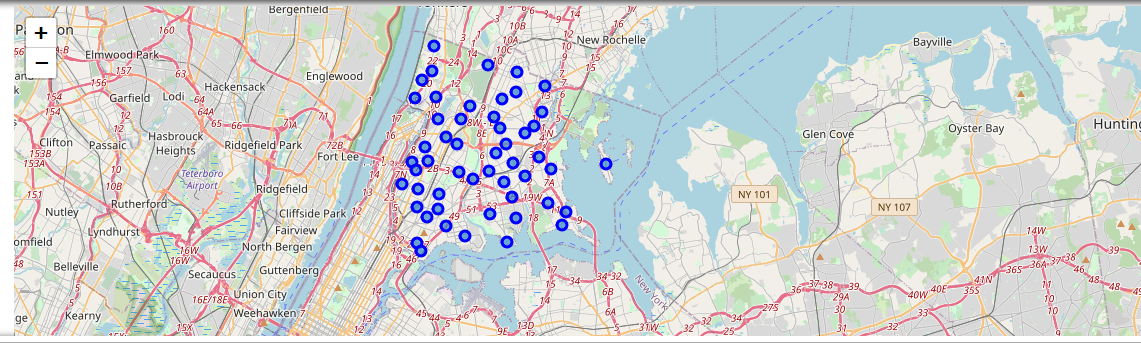


**Graph showing Venues**

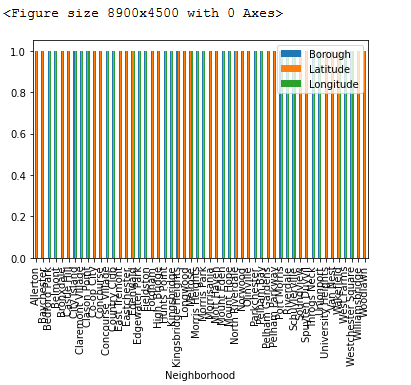


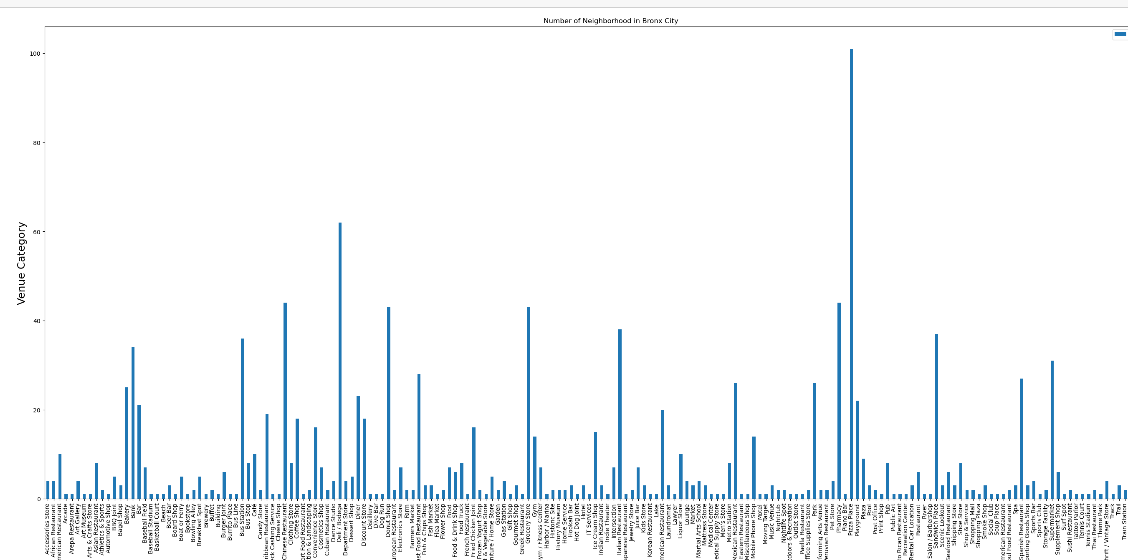
****

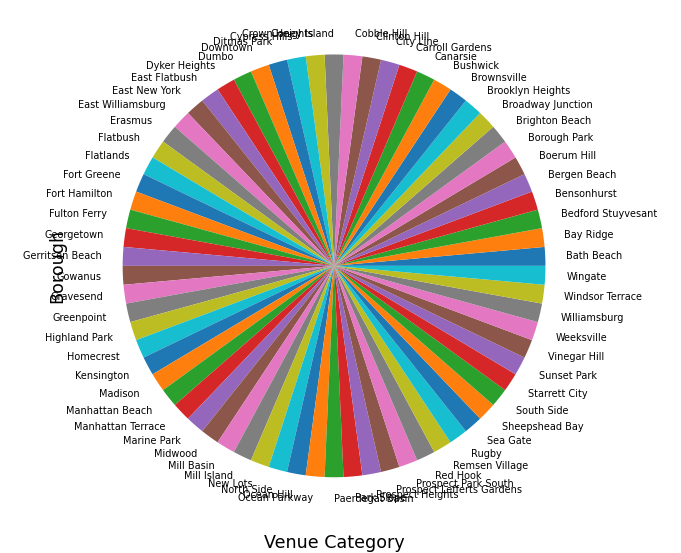
**Bronx**

****

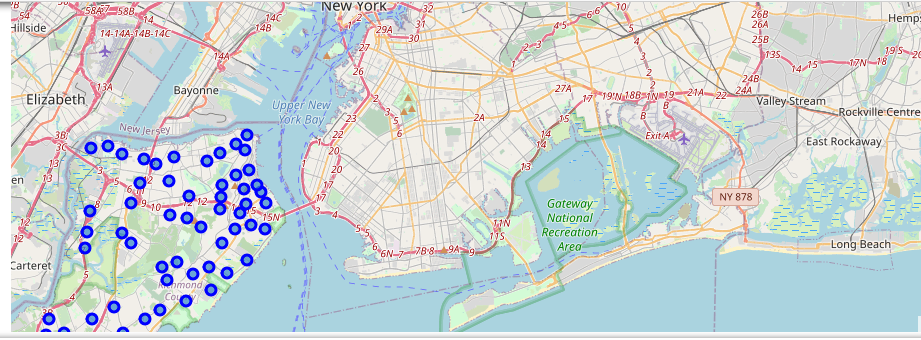
**Graph**

****

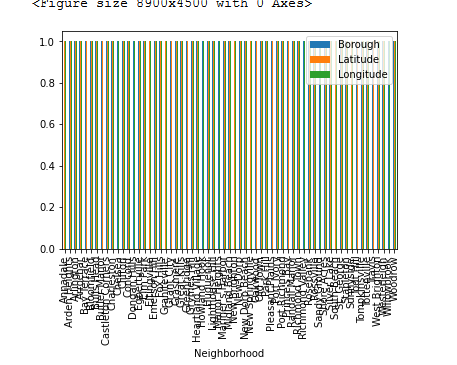
**Graph Showing Venues**

****

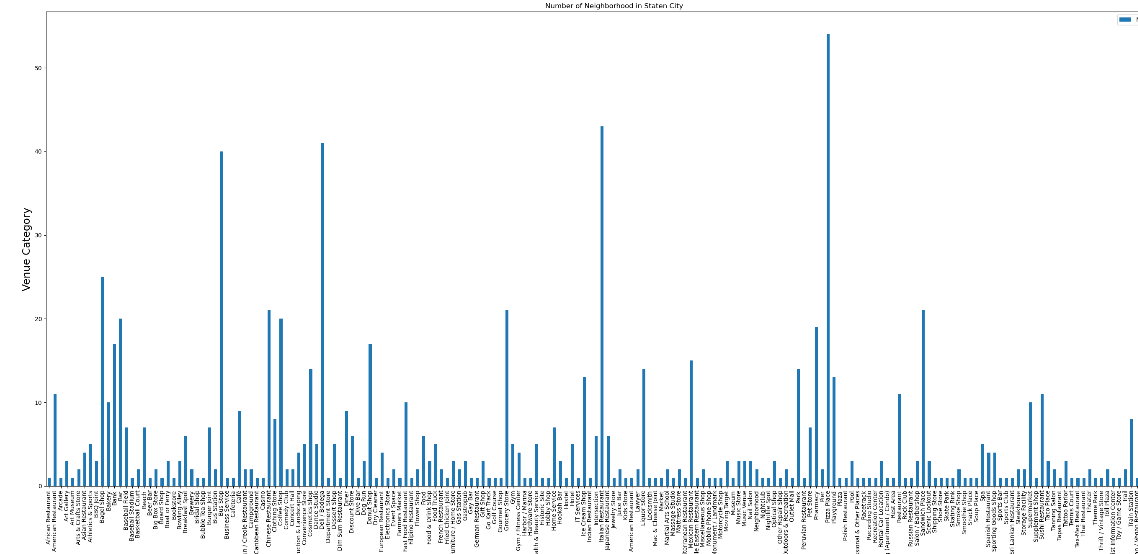
**Staten Island**

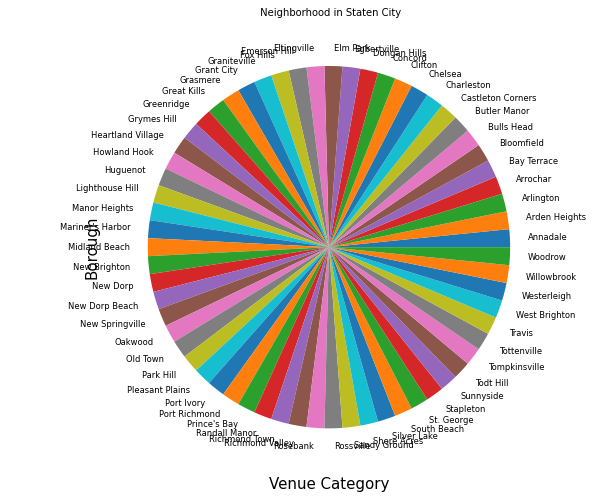
****

**Graph**

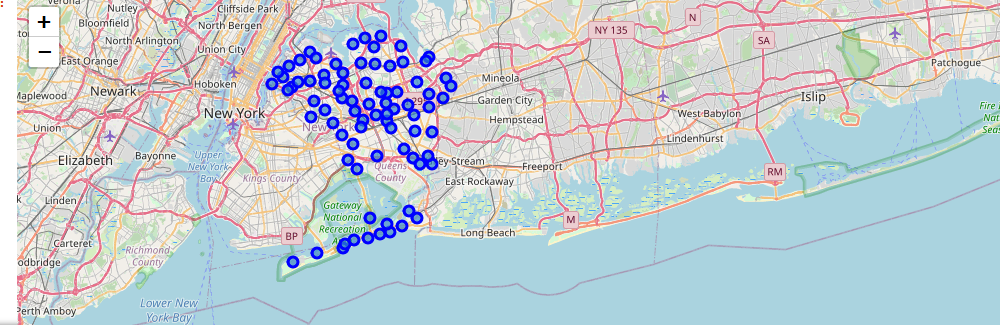
****

**Graph Showing Venues**

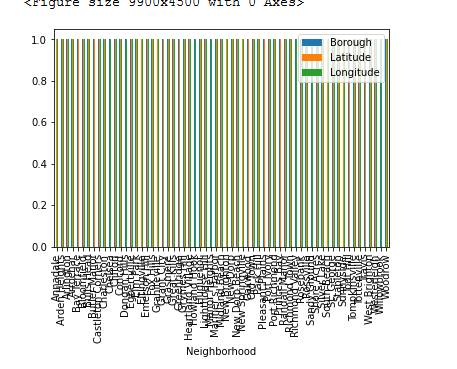
****

****

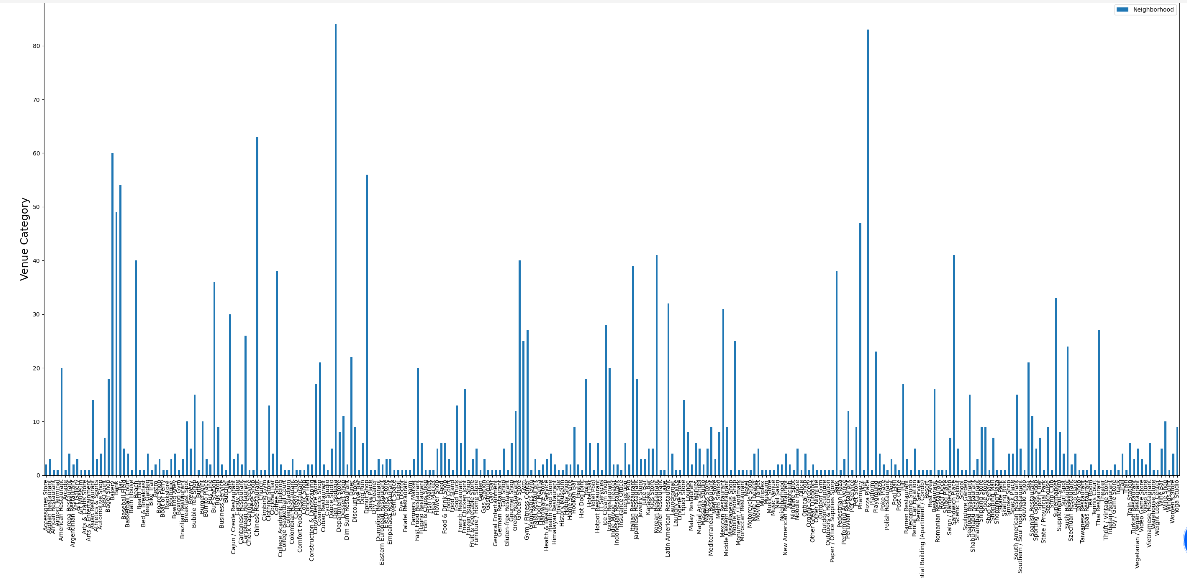
**Queens**

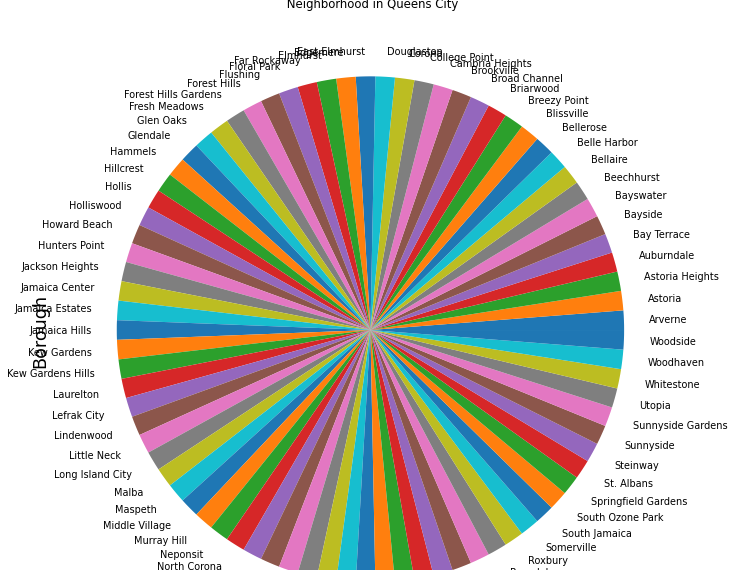
****

**Graph**

****

**Graph Showing Venues**

****



Location:

New York is the most populous city in the United States. It is the most densly populated major city in the United States. Nearly a quarter of New York residents are immigrants, while almost one-fifth of residents are native-born U.S citizens with at least one immigrant parent. Neighbors, business owners, tax payers and workers immigrants are integral part of New York's diverse and thriving communities and make extensive contributions to all. In 2018, 4.4 million immigrants (foreign-born individuals) comprised 23 percent of the population.

New York City is composed of five boroughs, each of which is a county of the State of New York. The five boroughs—Brooklyn, Queens, Manhattan, the Bronx, and Staten Island—were consolidated into a single city in 1898.

Foursquare API:

This Capstone project have used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

5. Discussion Section

We know that while searching for a good place to live it will be hard to meet the essential requirements. All of them have their own needs one have to select proper place according to their own interest. Here we will provide all datas of the cities in New York and the resources available in the cities. We are providing a detailed report of the cities so one can select appropriate neighborhood for them to settle. Consider if one person want to settle in New York but he doesn't know anything detailed about the city so he will search online or ask some of his friends or relatives about the places, they will provide information about thier own interest which may not match with his requirements. So here we are providing an elaborate information about the cities so he can select place according to his own demands.

Main work done in this project is providing some information about the cities in New York and detailed datas about the resources offering in this cities. Each boroughs are clearly examined and provides a detailed information about the boroughs and neighborhood

6. Conclusion Section

In this Capstone project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different latitude and longitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented neighborhood based on the availability of different resources.

I feel rewarded with the efforts and believe this course with all the topics covered is well worth of appreciation.  
This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.  
The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

Future Works:

This Capstone project can be continued for making it more precise in terms to find best house in New York. Best means on the basis of all required things(daily needs or things we need to live a better life) around and also in terms of cost effective.

Libraries Which are Used to Develop the Project:

*Pandas: For creating and manipulating data frames.*

*Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.*

*Scikit Learn: For importing k-means clustering.*

*JSON: Library to handle JSON files.*

*XML: To separate data from presentation and XML stores data in plain text format.*

*Geocoder: To retrieve Location Data.*

*Beautiful Soup and Requests: To scrap and library to handle http requests.*

*Matplotlib: Python Plotting Module.*