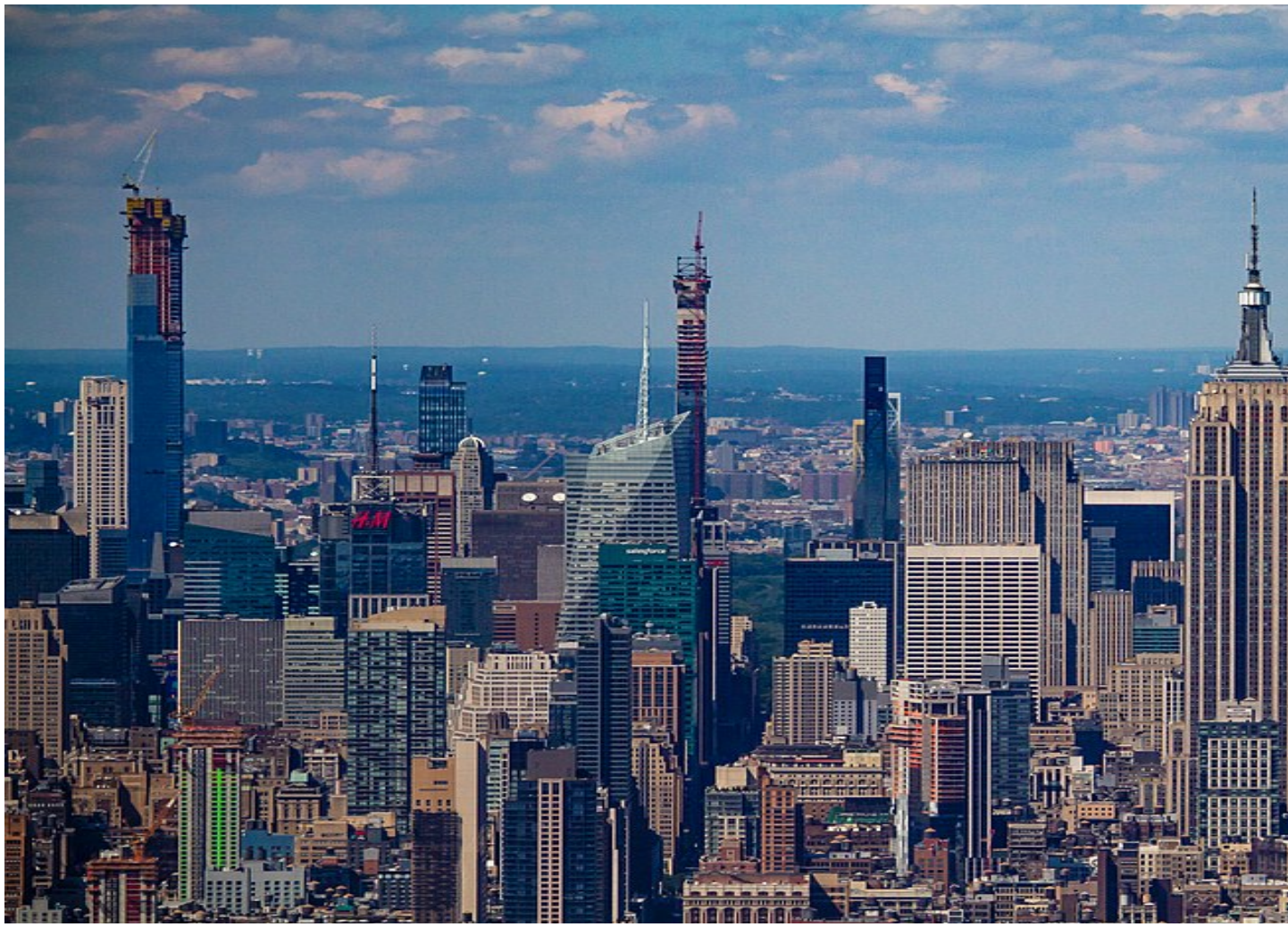
The Battle of the Neighborhoods



Introduction and Business Model

New York city is the most populous city in the United States, with an estimated 2019 population over 8 million, new York is also New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

The cuisine of New York City comprises many cuisines belonging to various ethnic groups that have entered the United States through the city. Almost all ethnic cuisines are well represented in New York City, both within and outside the various ethnic neighborhoods.

Problem Background

Doing a business in New York City is very competitive and expensive compare to other major metropolitan cities like Los Angeles, Houston or Chicago. We need to analyze any business

carefully for any future growth or starting a new business. One should expect higher investment in the early stage due to real estate, labor markets, sales etc.
New York City known to support vast ethnic cultures, lifestyles and food etc.

Problem Description

The cuisine of New York City consists of many cuisines belonging to various ethnic groups that have entered. Almost all cuisines are well represented in New York City, both within and outside the various ethnic neighborhoods.
As a Data Scientist, we are given task to find out which neighborhood can support and financially sustainable for good Indian restaurant.

Target Audience

ABC Corp is hiring data scientist to analyze the New York City neighborhoods to find out which borough and neighborhood would be a good location to start a new restaurant.
We pull data from multiple sources to analyze the market based on the population, foot traffic, business locations, supply chain.
We will analyze the market for starting an Indian restaurant chain who wants to start and grow in the region.

Data Acquisition:

For this project we will use the following data sets:

Data 1: Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.
This dataset exists for free on the web: https://geo.nyu.edu/catalog/nyu_2451_34572
Sample screenshot of boroughs and their coordinates can be seen in the following data frame.

[27]:

	Borough	Neighborhood	Latitude	Longitude
286	Staten Island	Sandy Ground	40.541140	-74.217766
287	Staten Island	Egbertville	40.579119	-74.127272
288	Queens	Roxbury	40.567376	-73.892138
289	Brooklyn	Homecrest	40.598525	-73.959185
290	Queens	Middle Village	40.716415	-73.881143
291	Staten Island	Prince's Bay	40.526264	-74.201526

Data 2: Below data sets are from Wikipedia for the analysis.
We will analyze data from Wikipedia about the population demographics, types of cuisines already in New York City and economic data to support for new business opportunity.

https://en.wikipedia.org/wiki/New_York_City

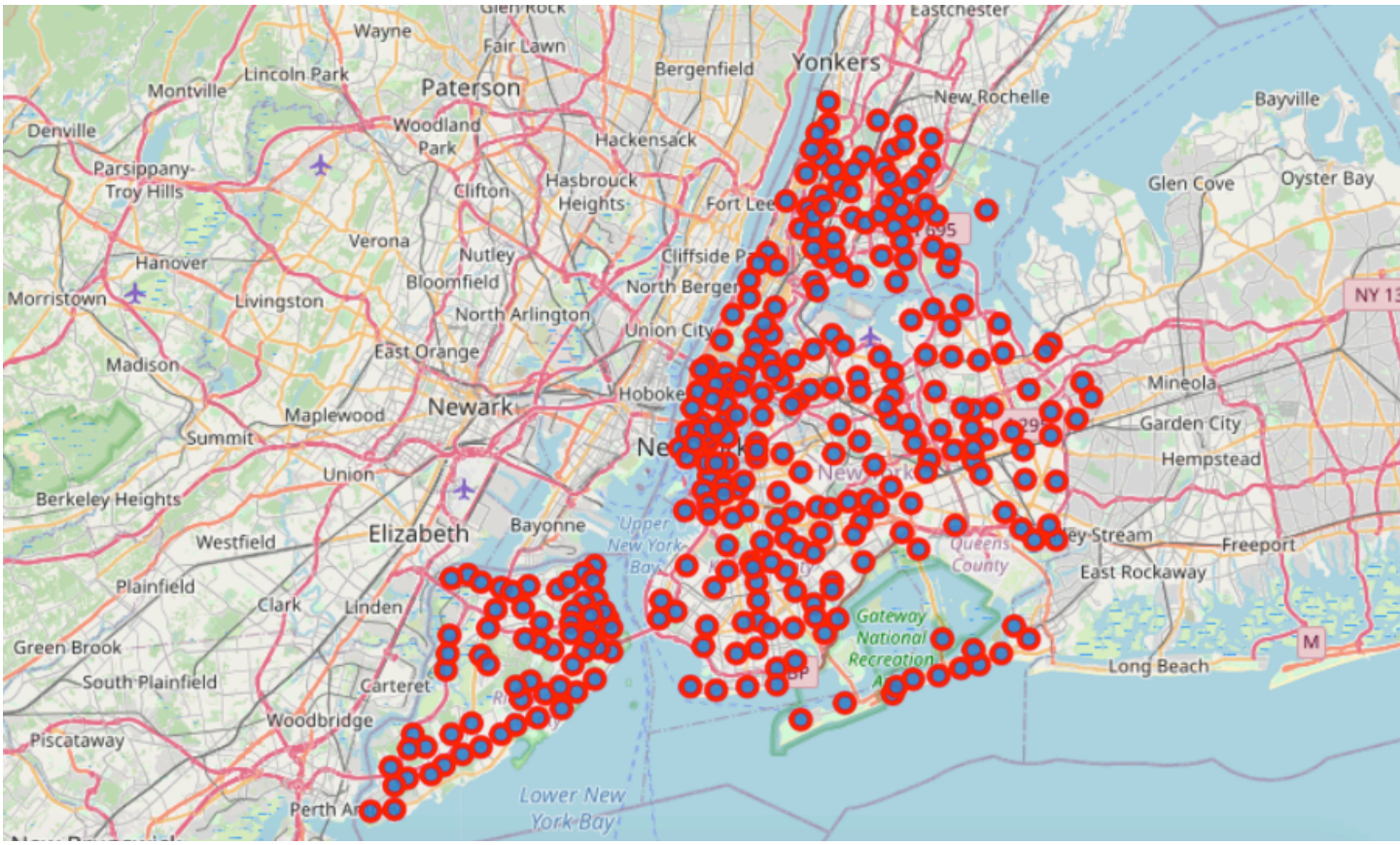
https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

https://en.wikipedia.org/wiki/Economy_of_New_York_City

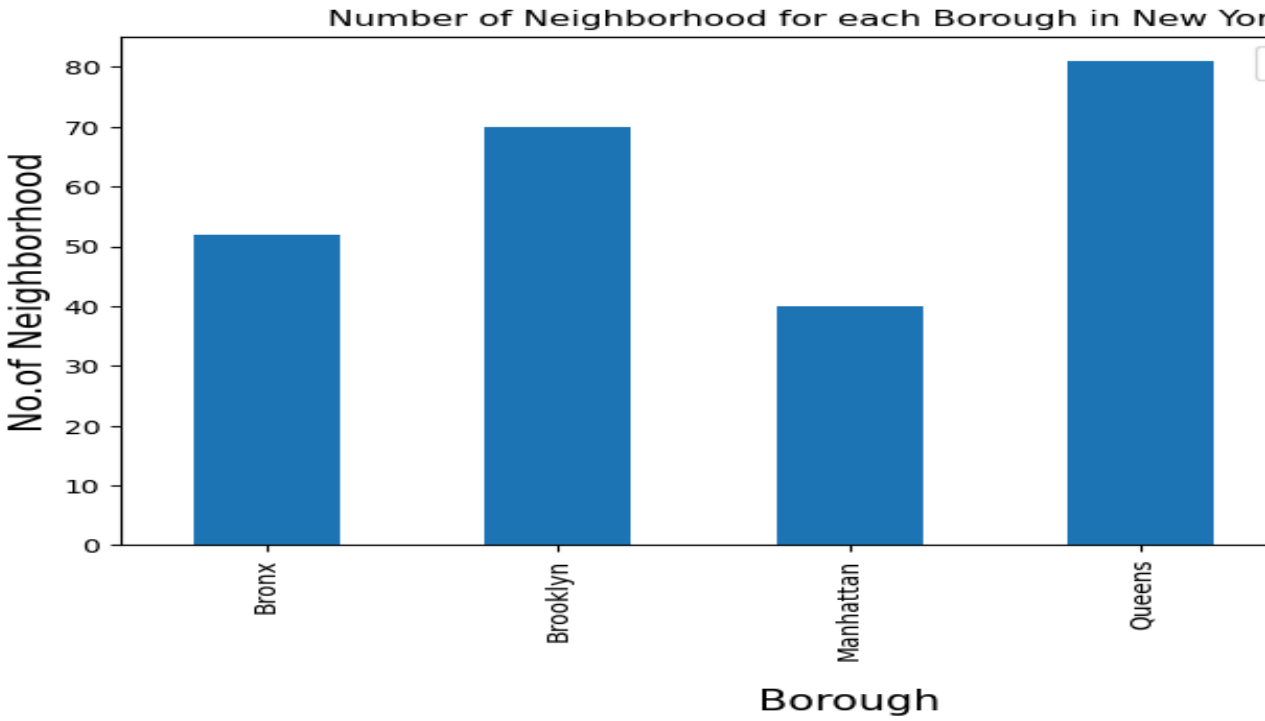
Data 3: In this data, we are going to start utilizing the Foursquare API to explore the neighborhoods from New York City and segment them.
Foursquare data will help to segment the clusters and find out opportunity for new restaurant location in the given neighborhood.

Methodology:
Our goal is to locate the best locations for the clients who would like to open the restaurant in New York City.

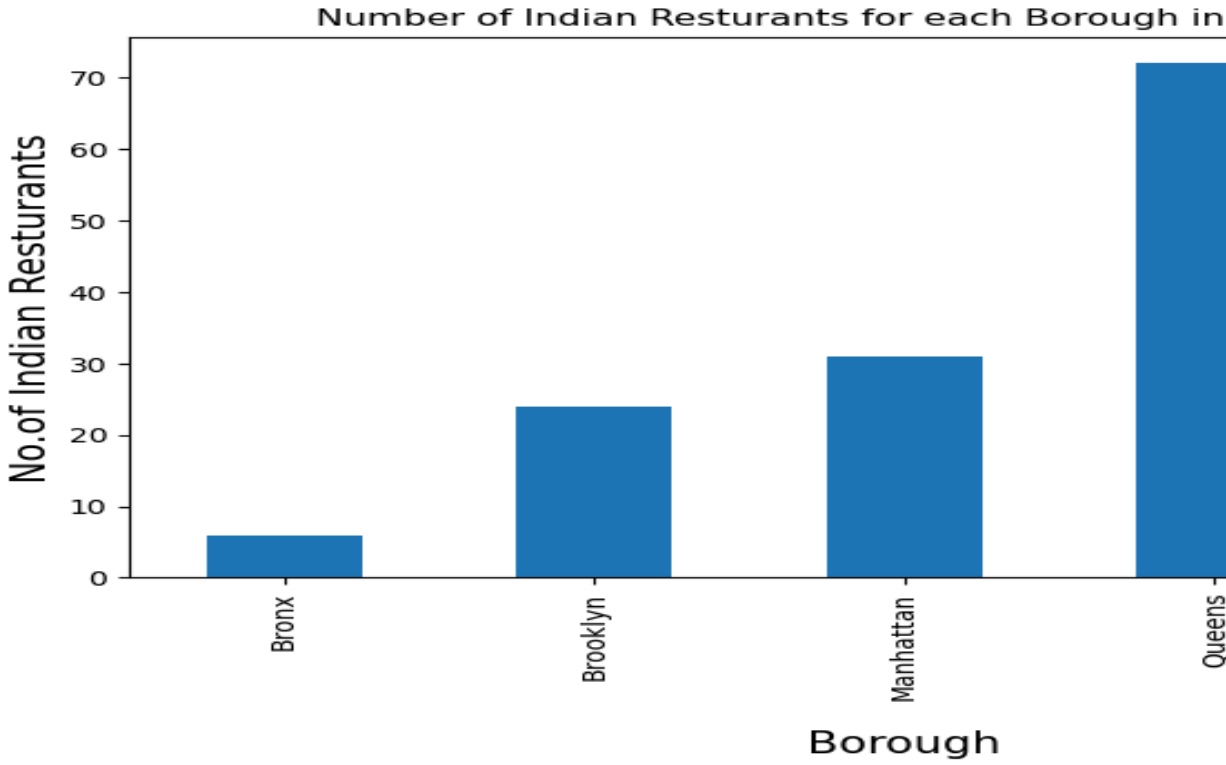
Data preparation: New York data is downloaded from IBM course site for preparation
We can see the New York city using folium map with latitude and longitude with neighborhoods.
We will use this data to explore the neighborhoods, venues using Foursquare data.



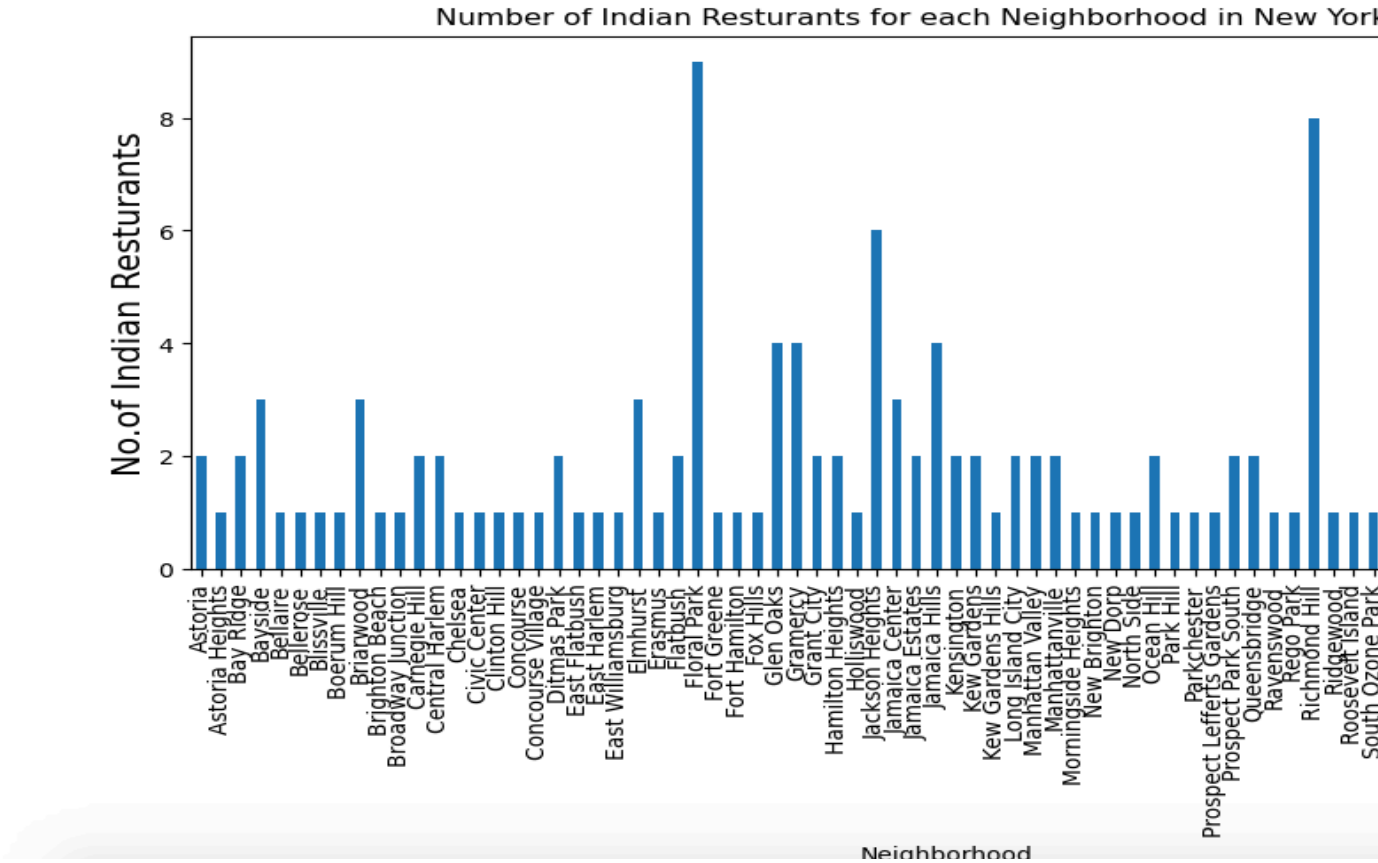
Data Analysis: The following bar chart shows the number of neighborhoods in each Borough, we can see the Queens got the highest number of neighborhood



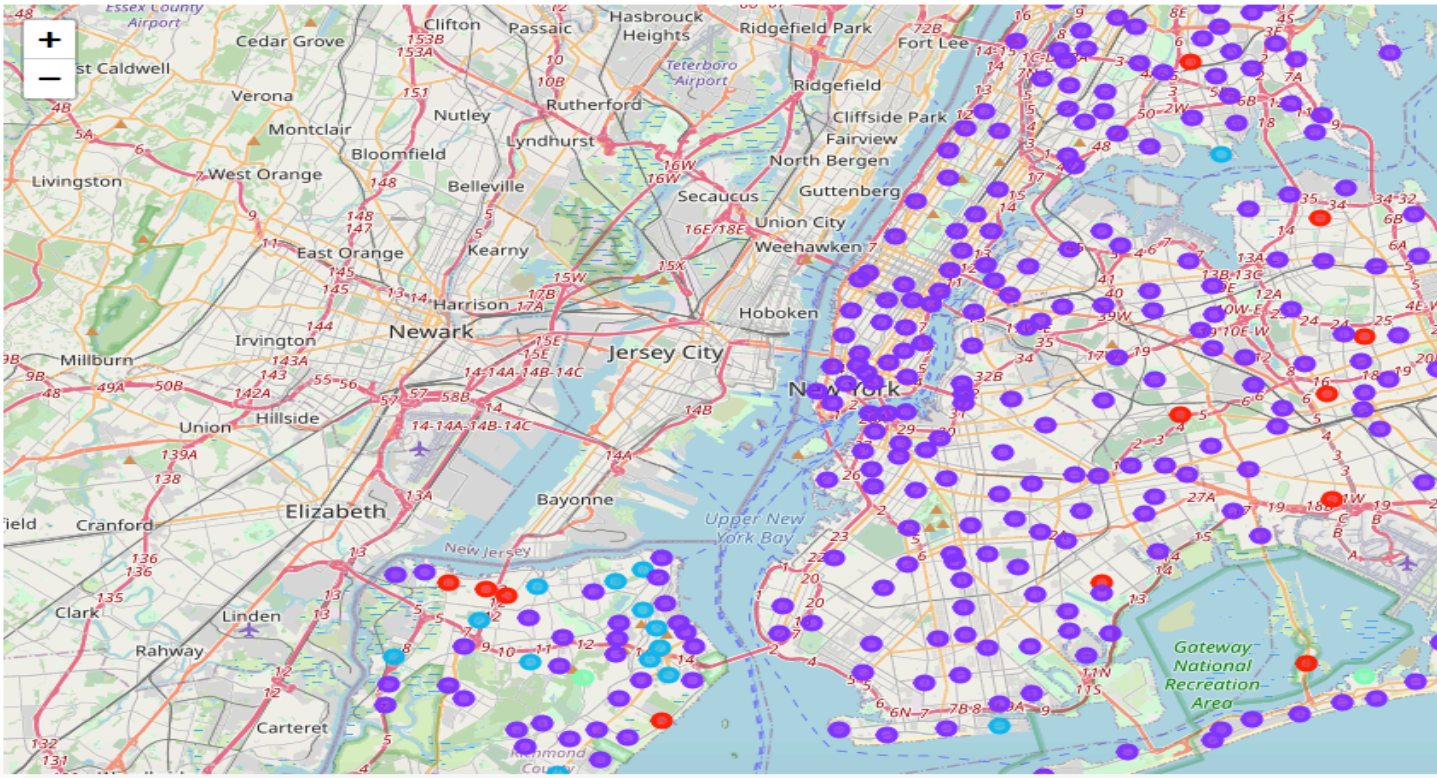
In this chart we see Queens got the highest number of Indian restaurants followed by Manhattan and Brooklyn.



The following chart represents the number of Indian restaurants locations in each neighborhood. Floral part, Richmond Hill and Wood side ranks the top three.



Machine Learning algorithm used: Unsupervised K-Mean clustering algorithm applied to form clusters of different categories of various Burroughs, neighborhoods for a better location for staring a new Indian restaurant.



Results:

After using K-means Algorithm we clustered in to 5 categories:

Cluster-1(red)- In this cluster, we see Staten Island and Queens mainly and 1st most common venues are pizza and Deli.

Cluster-2 (Purple)- This cluster is the biggest of all, we see the number one venue is restaurants in multiple boroughs across followed by another restaurant or family fun place.

Cluster-3(Light Blue)- In this cluster, the common location is bus stops and other commercial locations.

Cluster-4(Light Green)- This cluster presents with park, yoga studios for relaxation/ sightseeing.

Cluster-5(Light Brown)- We see this cluster presents with restaurant as most common venue, with Caribbean.

In conclusion, cluster 2 represents the most common location to start Indian restaurant due to increased Asian population, other attractions which will pull customers for a startup business.

Discussion:

Based on our data acquisition, preparation and analysis we can see Queens has the highest number of restaurants and also has highest number of neighborhoods. We can understand that there is correlation between population and number of venues available for people to choose. Manhattan has the second highest Indian restaurant but lacks with neighborhoods. Even though it has highest number of business population that are mainly there during business hours but got a good potential for new startup.

Conclusion:

With K-means clustering, we can conclude that both Queens and Manhattan have higher potential to have Indian restaurant based on population and other supporting business structures.

Limitations: Our recommendations were limited with Foursquare data availability at the time of this project.

References:

https://en.wikipedia.org/wiki/Cuisine_of_New_York_City
https://en.wikipedia.org/wiki/New_York_City
www.Foursquare.com
https://geo.nyu.edu/catalog/nyu_2451_34572
https://en.wikipedia.org/wiki/New_York_City#/media/File:Midtown_Manhattan_2019.jpg