Segmenting and Clustering Neighborhoods in Brooklyn

1. Introduction

New York is a city of neighborhoods. There are 5 boroughs in the New York city: Bronx, Brooklyn, Manhattan, Queens, Staten Island

Brooklyn is the most populous borough of New York City, with an estimated

population of 2.649 million residents based on 2017 census. Today, if each borough were ranked as a city, Brooklyn would rank as the third-most populous in the U.S., after Los Angeles and Chicago.

Brooklyn has experienced a dramatic house price and a decrease in housing affordability. It has evolved into a thriving hub of entrepreneurship and high technology startup firms.

As the city is growing, there is a continuous need for people for product and services such as home, schools, food, entertainment to name a few. Entrepreneurs use different tools, analysis to analyze the area with respect to business such as what is trending and what is in monopoly, what is expensive to people and what is not easily available etc. other business needs such as plain greed to make money by doing any business which generates revenue.

This data science project aims to help the business to see the various cluster in the neighborhood for different types of venues which can help them to launch a new business or move their business to the area.

2. Data

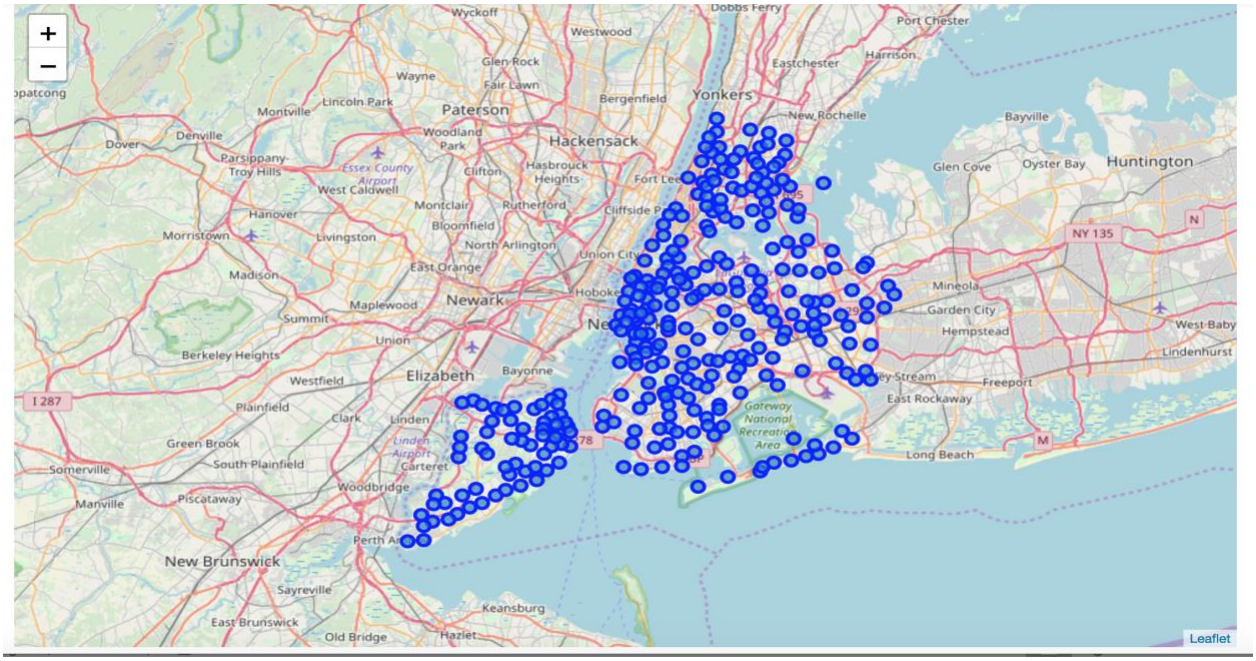
The project would leverage Foursquare – location provider like Google Places, Yelp – API to explore neighborhoods in New York City. In addition, the project has used the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

3. Methodology

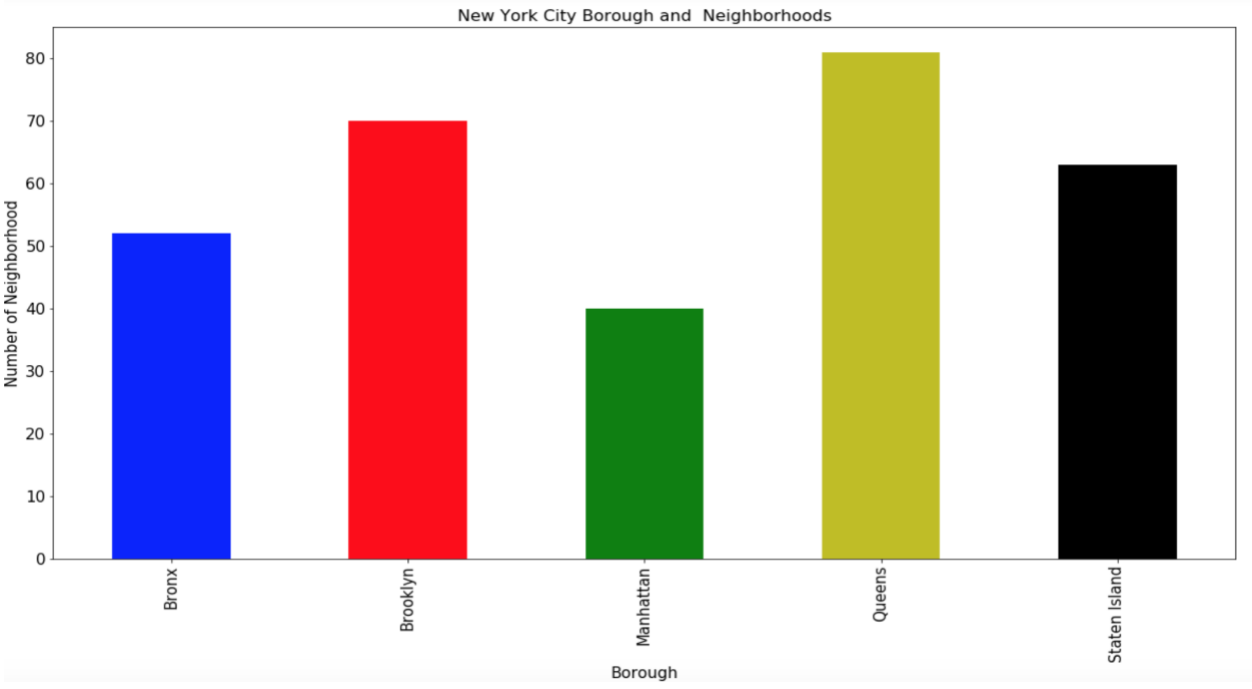
The addresses are converted into their equivalent latitude and longitude values. In addition, I have used the Foursquare API to explore neighborhoods in New York City and Brooklyn. The explore function of Foursquare API is to get the most common venue categories in Brooklyn. The result is used to group the neighborhoods into clusters. I have used k-mean clustering algorithm because the project aims to find the different clusters of neighborhoods. To display the result, I have to Folium library to visualize the neighborhoods and their emerging clusters.

Exploratory analysis

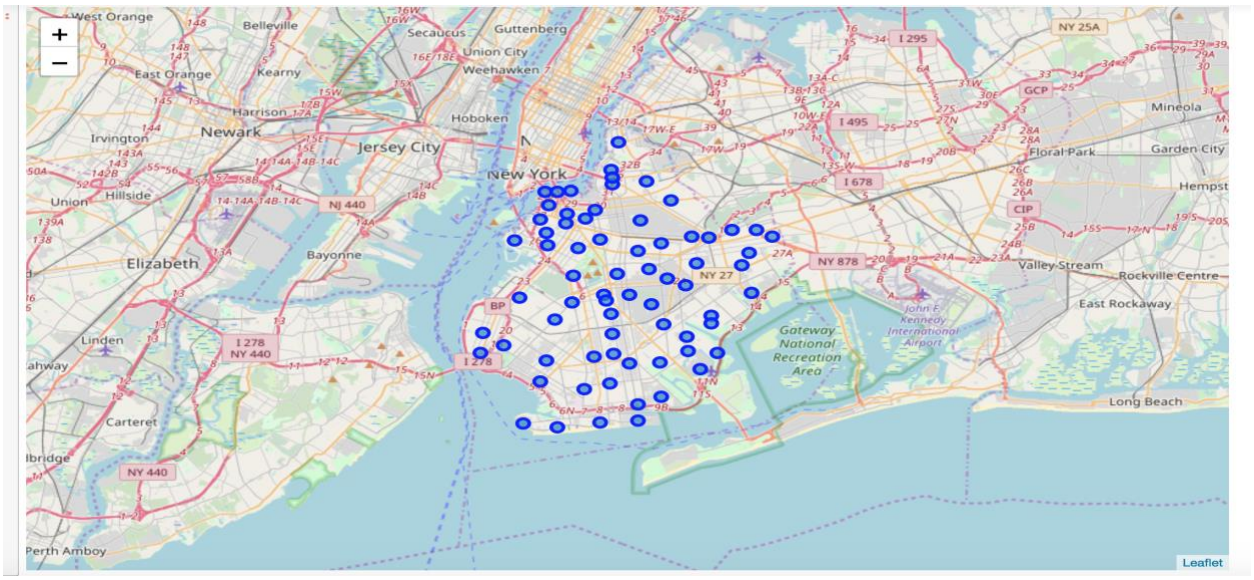
- New York City and its neighborhood Map using the Foursquare location data about venues:



- New York City Borough and Neighborhoods Bar chart

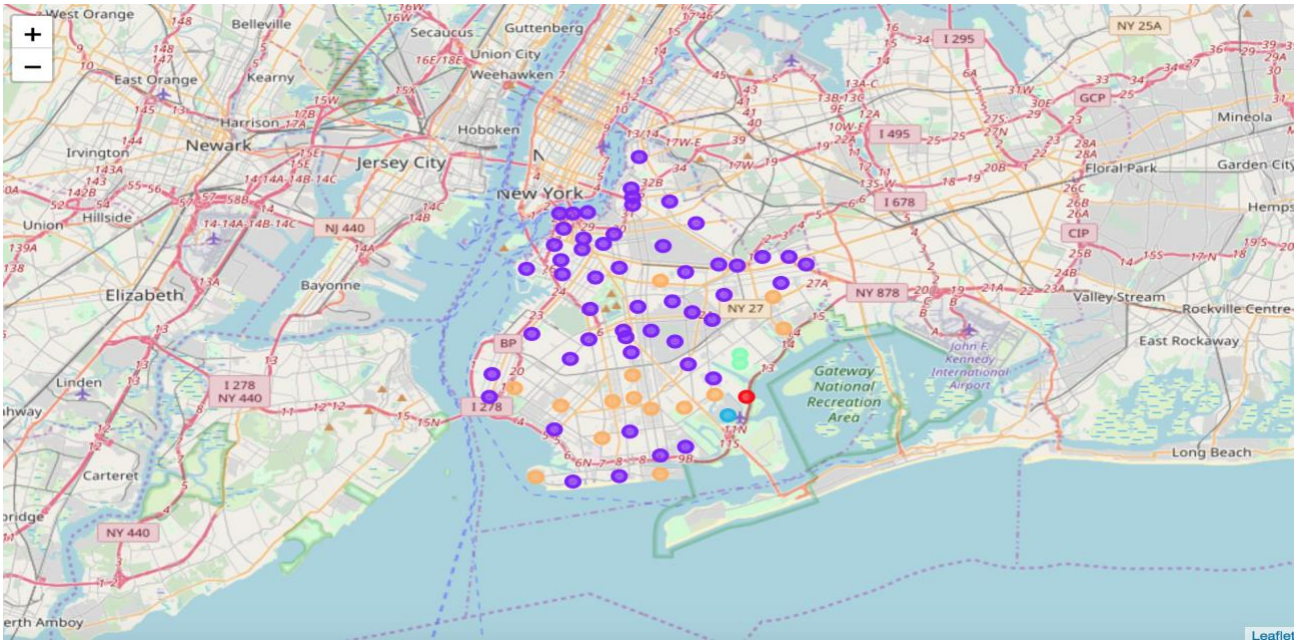


- Visualization of Brooklyn Neighborhood



4. Results

Map showing the cluster of neighborhoods in Brooklyn.



Cluster Number	Cluster Description
Cluster 1	Harbor, farm, and sports fields (Red circle)
Cluster 2	Small shops and Asian restaurants (Purple circles)
Cluster 3	Pool, lake, and farm (Light green circle)
Cluster 4	Asian restaurant (Blue circle)
Cluster 5	General shopping store, restaurants, supermarket, mall (Orange circles)

5. Discussion

It's interesting to find that the cluster 1 (Bergen Beach) and cluster 3 (Mill Island) have only one neighborhood. In general, relatively, different than Manhattan which makes the clustering result very different though in the same city.

6. Conclusion

The clustering results are a good illustration to get a general idea or make a good judgment analysis of the Brooklyn borough in terms of different venues. K-means clustering, location data, and mapping visualization: these all technologies made this project fun to work on!