Capstone Project - The Battle of Neighborhoods

Introduction/Business Problem

Comparing the neighborhoods of the two cities i.e. New York City & Toronto and determining whether they are similar or dissimilar?

The intent behind the problem statement:

- To know both the cities better?
- Comparing whether they are similar or diverse?
- Concluding the best one for a person to roam around and to explore neighborhoods?

Data Collection & Processing:

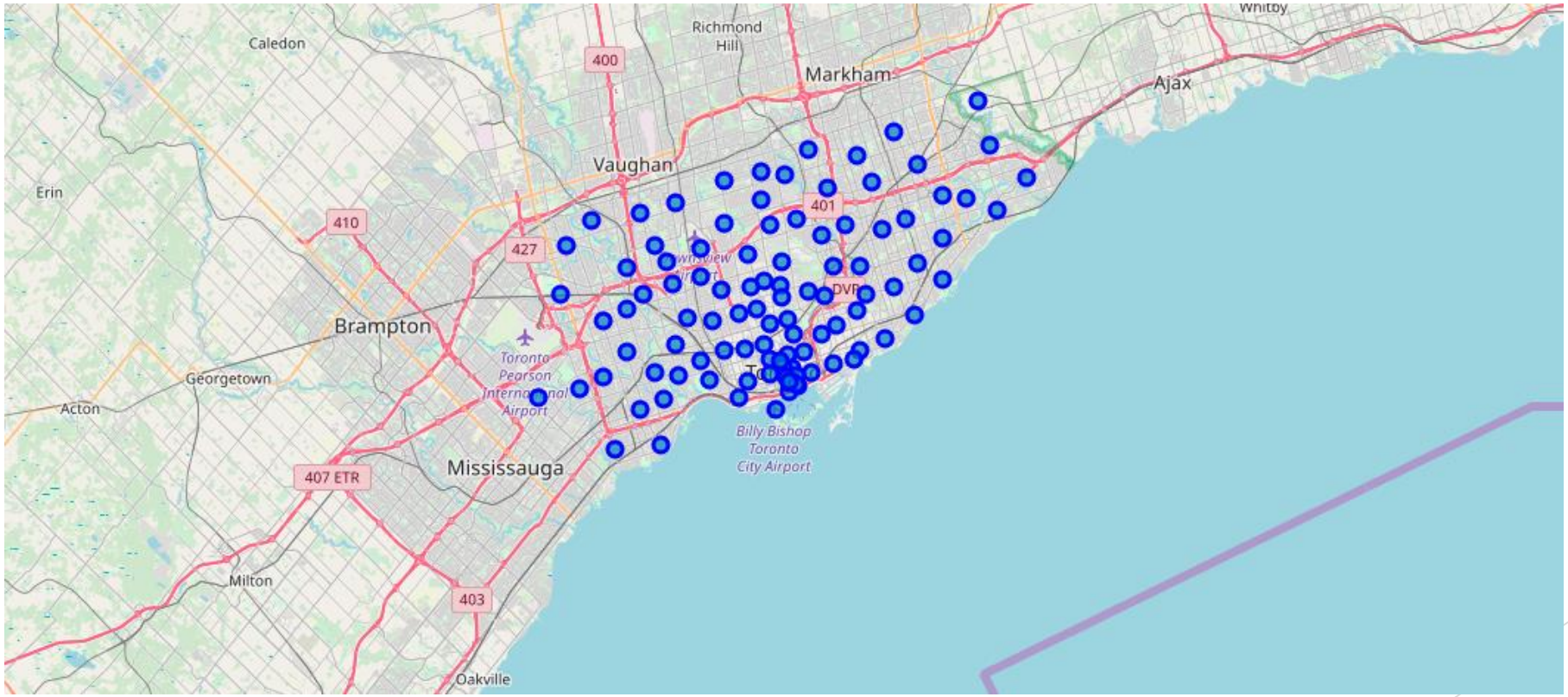
Toronto

- Scrapping the data from the web url.
- Converting the data into pandas dataframe by standardizing the data.
- Appending the Geospatial coordinates in resultant dataframe.
- Mapping the neighborhoods of Toronto on the map of Toronto.

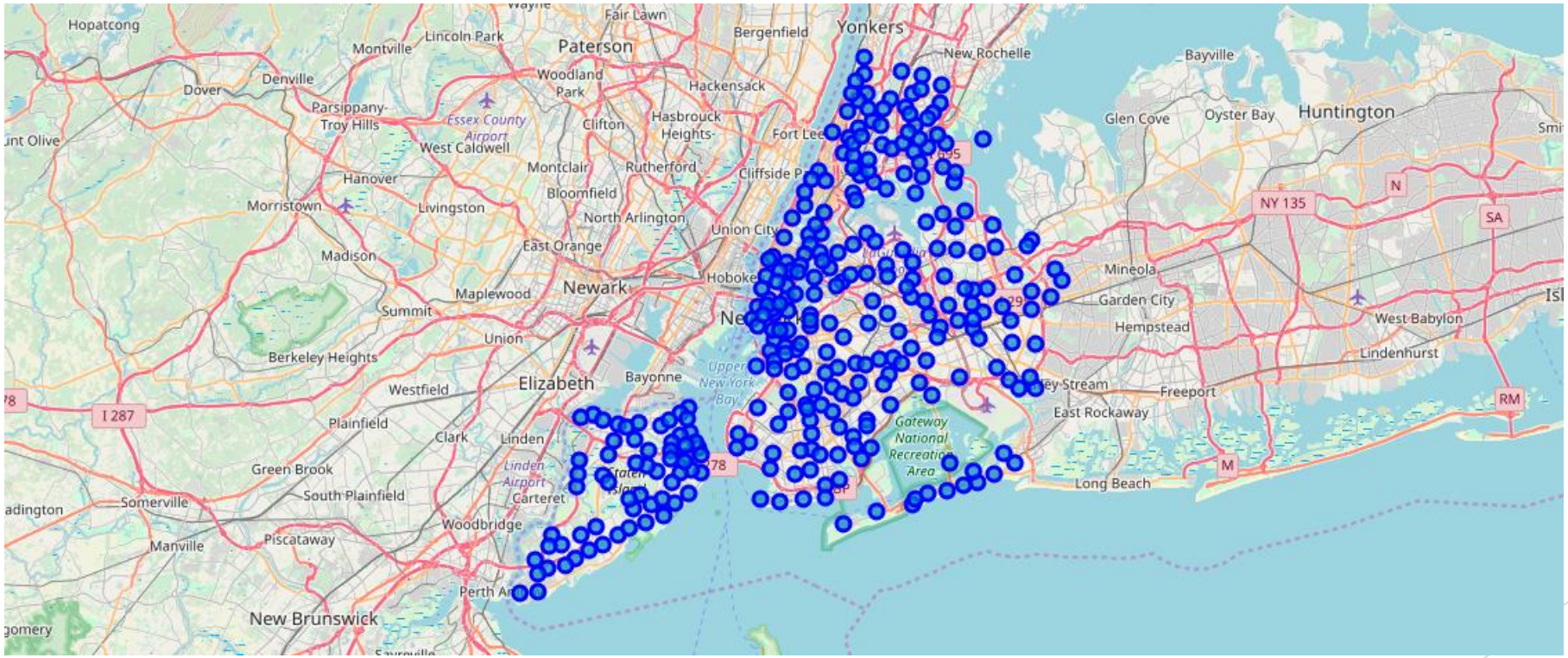
New York City

- Downloading the data from dataset.
- Extracting the features set i.e. Borough, Neighborhood, latitiude & Longitude.
- Standardizing the pandas dataframe.
- Using geopy library mapping the neighborhoods of New York City on the map of New York.

Exploring Neighborhoods: Toronto



Exploring Neighborhoods: New York City



Methodology:

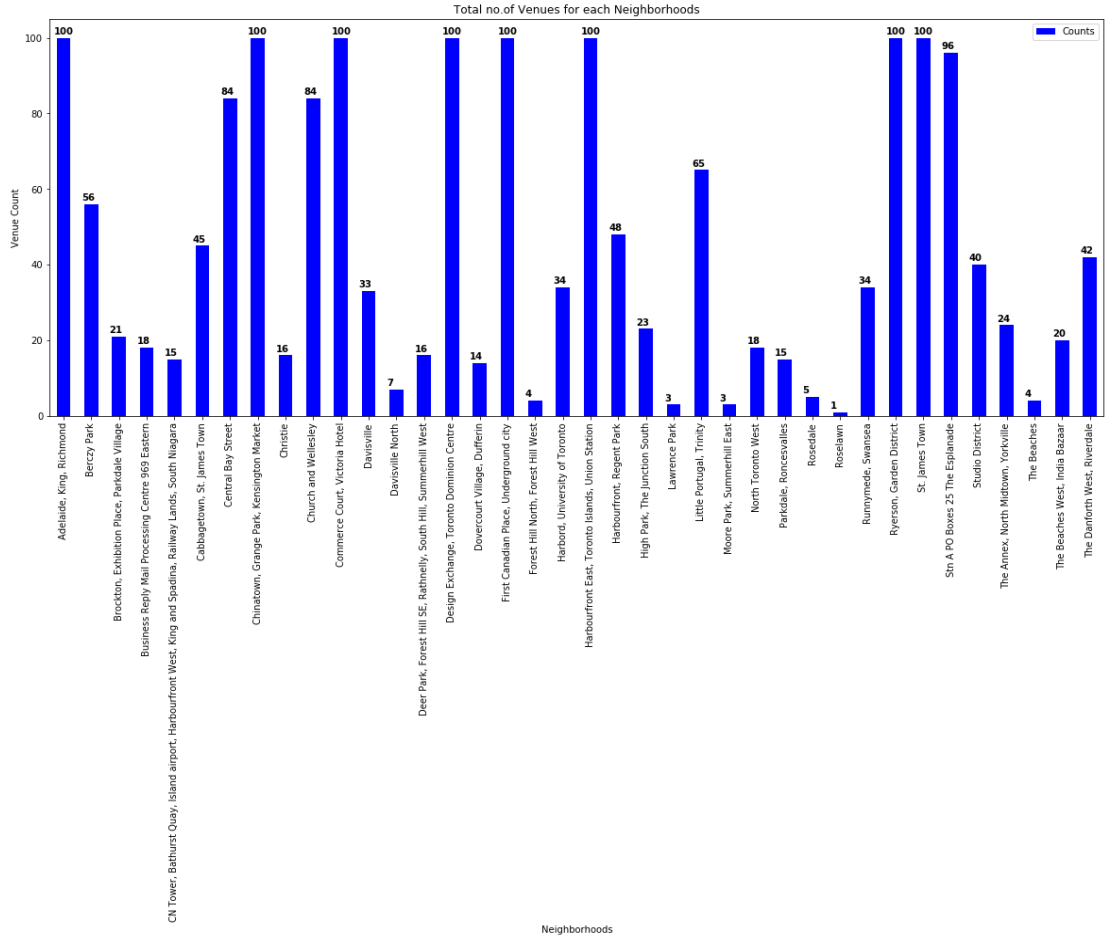
Toronto

- To concise down the result I have selected boroughs of Old Toronto only.
- With the help of forsquare API I have limit the count till 100 most common venues in the vicinity of 500 meters for each neighborhoods.
- Visualize the count of venues for each neighborhoods.

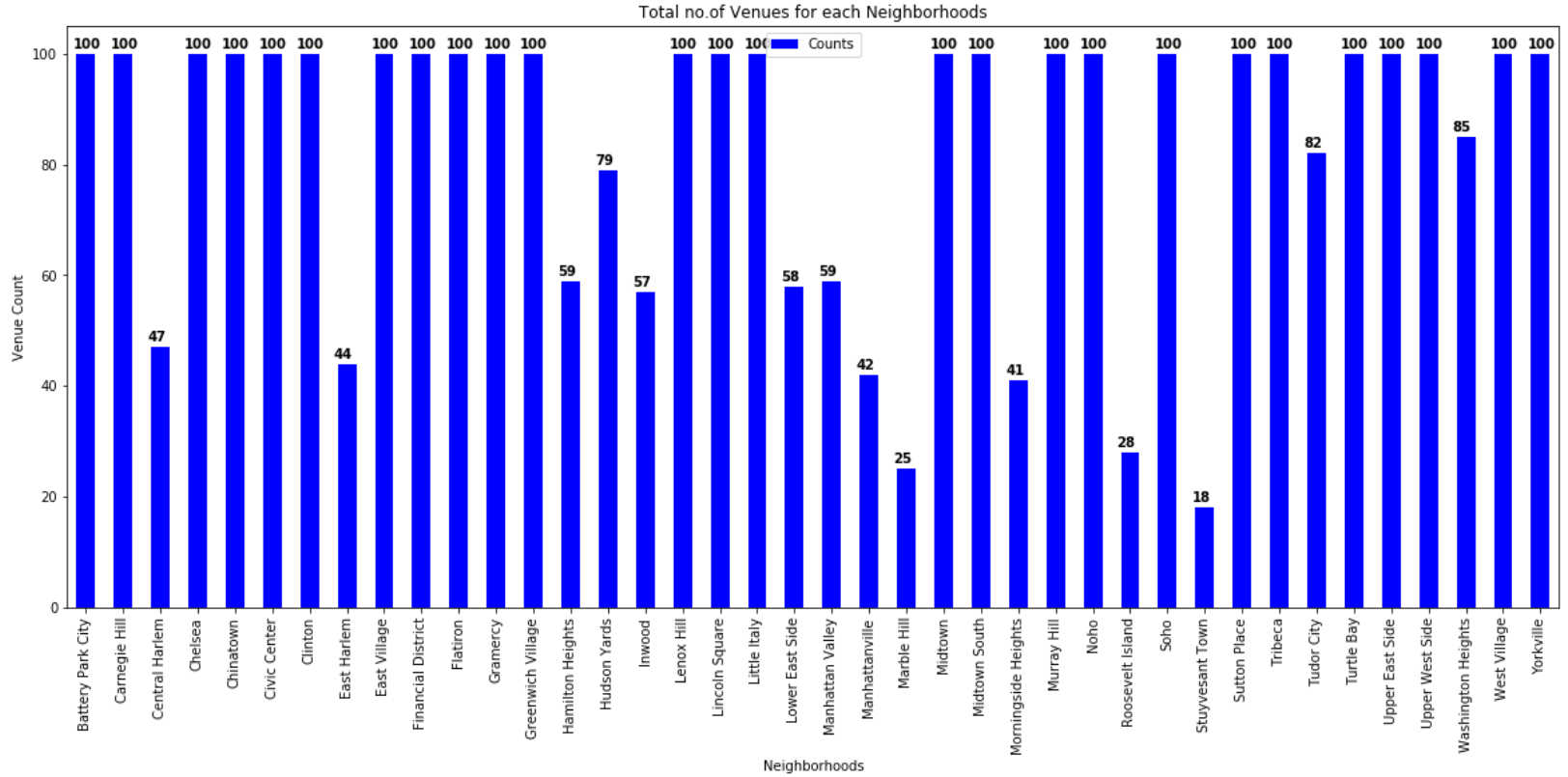
New York City

- To concise down the result I have selected Manhattan as the only borough for New York City.
- With the help of forsquare API I have limit the count till 100 most common venues in the vicinity of 500 meters for each neighborhoods.
- Visualize the count of venues for each neighborhoods.

Visualizing Count of Venues: Toronto



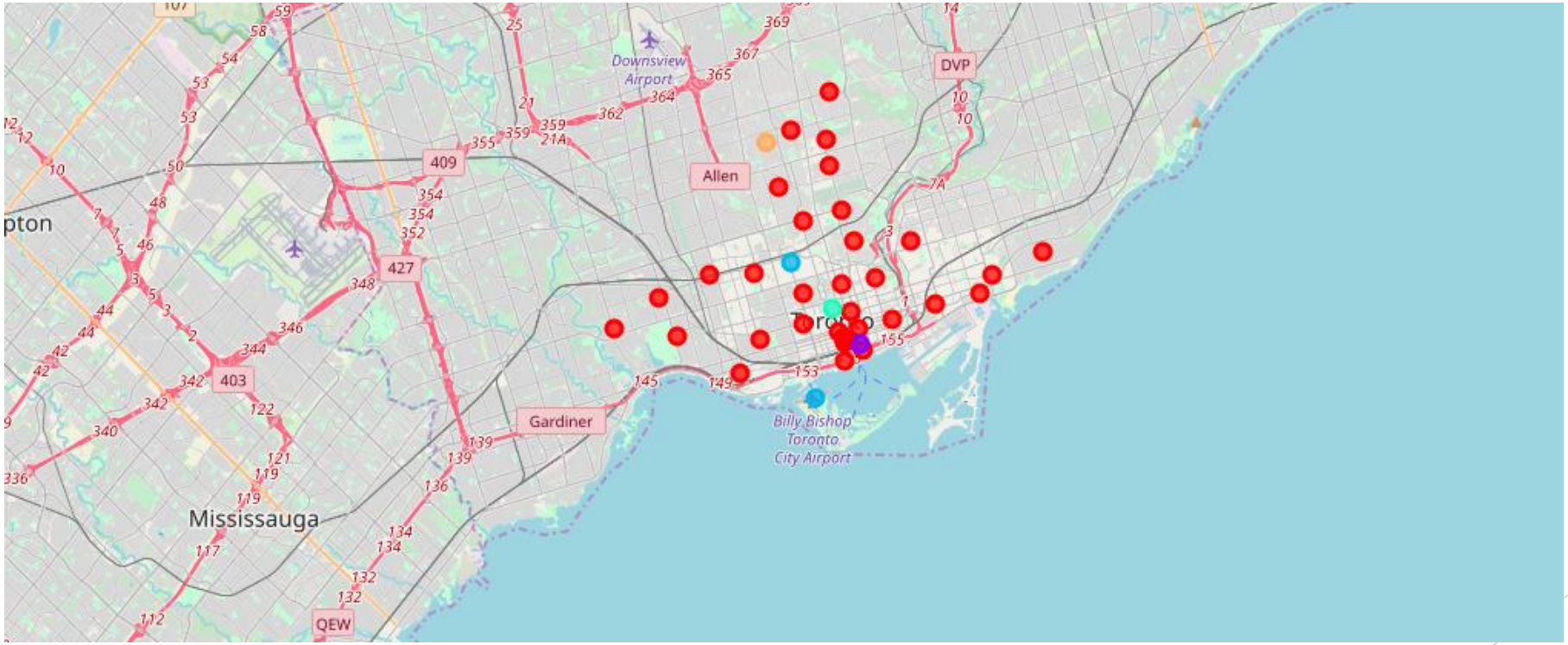
Visualizing Count of Venues: New York City



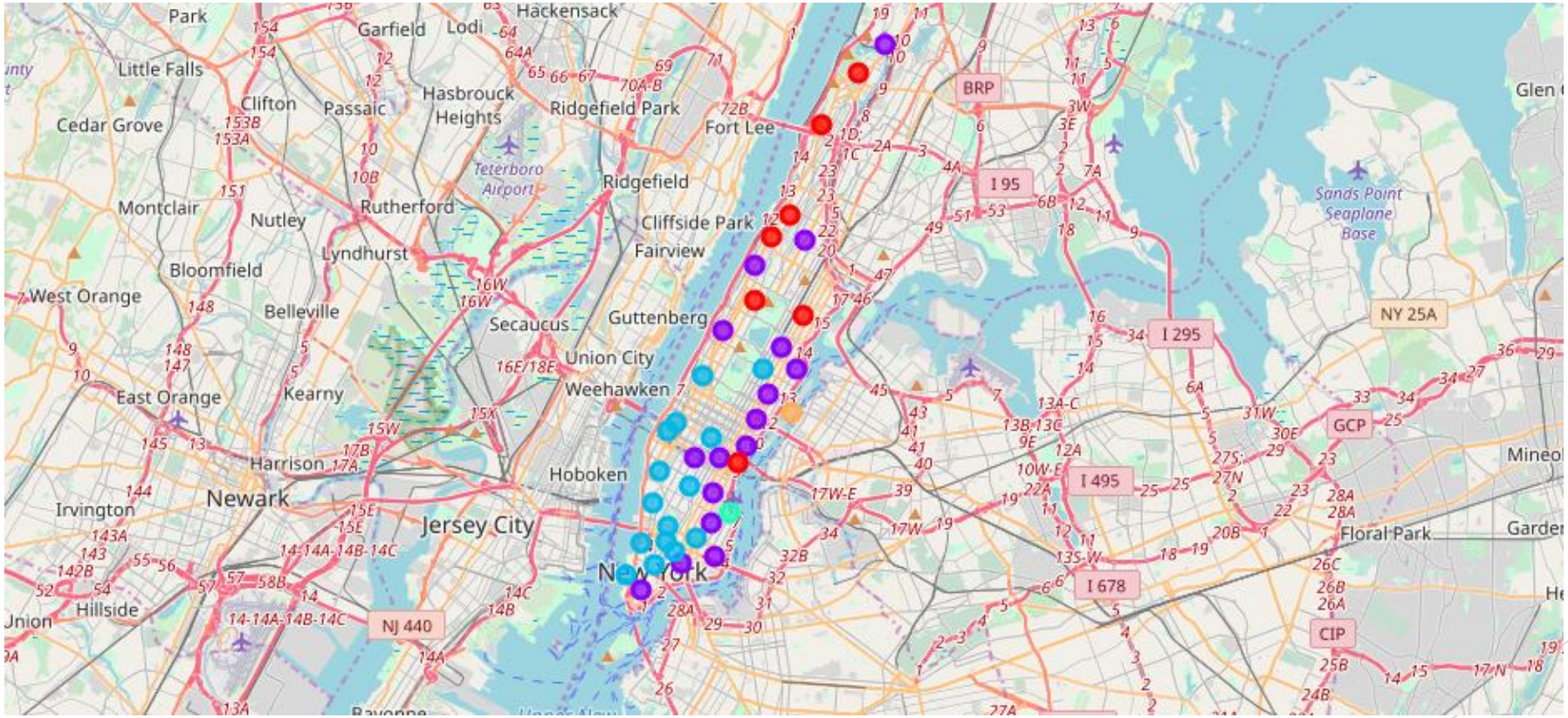
Results:

- Now to narrow down the results, I find out the unique venues from each category.
- Determined the frequency distribution for each venues in their neighborhoods.
- Clustered the Neighborhoods together with the help of k-Means Algorithm based the similarity of the frequency distribution of venues amongst the neighborhoods.
- Visualize the resultant on the map with the help of folium library and interpreted the result.

Clustering the Neighborhoods: Toronto



Clustering the Neighborhoods: New York City



Conclusions:

- Now since I have compared both the cities, it can be concluded that from above scenario that both the cities are quite similar.
- The only difference that I could identify was the no of options and venues available were quite more then those available in the Toronto.
- So if a person needs to roam around or a tourist needs to expore either of one city then the New York is quite a better option as compared to the Toronto city.