

Coursera - Capstone Project

The Battle of Neighborhoods

Sanjiv Tomar

Introduction/Business Problem

New York is the most populous city in the United States and quite famous for the business and tourism. People do visit NY from different part of the world and enjoy different types of cuisine while their stay in New York. In order to offer different varieties of food, Mr. X is looking for to open a 'Caribbean Restaurant' in New York. There are 5 boroughs in NY, Mr. X is unsure as where to open the restaurant. His requirements for opening the restaurant are:

- As part of the business strategy, the area should be famous for different varieties of food.
- There should be more Caribbean Restaurants in the neighborhood so that Mr. X can give good competition to other restaurants.

The goal of the analytic solution is to explore all 5 boroughs in New York and help Mr. X in finding the most appropriate place to open the 'Caribbean Restaurant' which should meet the above conditions.

Data Section/Requirements

1. **Foursquare API:** Foursquare location data is the key data requirement for solving this problem as this will help in exploring different boroughs in New York and understanding people choices of different varieties of food.
2. **NY Boroughs and Zip Codes:** The Zip codes for all 5 boroughs will be scraped from NYC health website: <https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm>. This will be the base data for exploring each ZIP code of all boroughs.
3. **Geospatial data:** Geographical coordinates (latitude & longitude) to be pulled from publicly shared data available at http://cocl.us/Geospatial_data

Data acquisition and cleaning

- ❑ Borough, Neighborhood and Zip attributes of New York city was scraped from <https://www.health.ny.gov/statistics/cancer/registry/appendix/neighborhoods.htm>
 - Total 178 rows were extracted and 3 columns
 - Massive amount of data cleaning was done such as moving column values from one to another and updating the correct data types etc
- ❑ In order to pull latitude and longitude data, the data was scraped from <https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/table/>
 - The output was merged with 178 rows pulled in first step
- ❑ Venue data such as venue longitude, latitude, category against each borough zip was pulled from Foursquare API

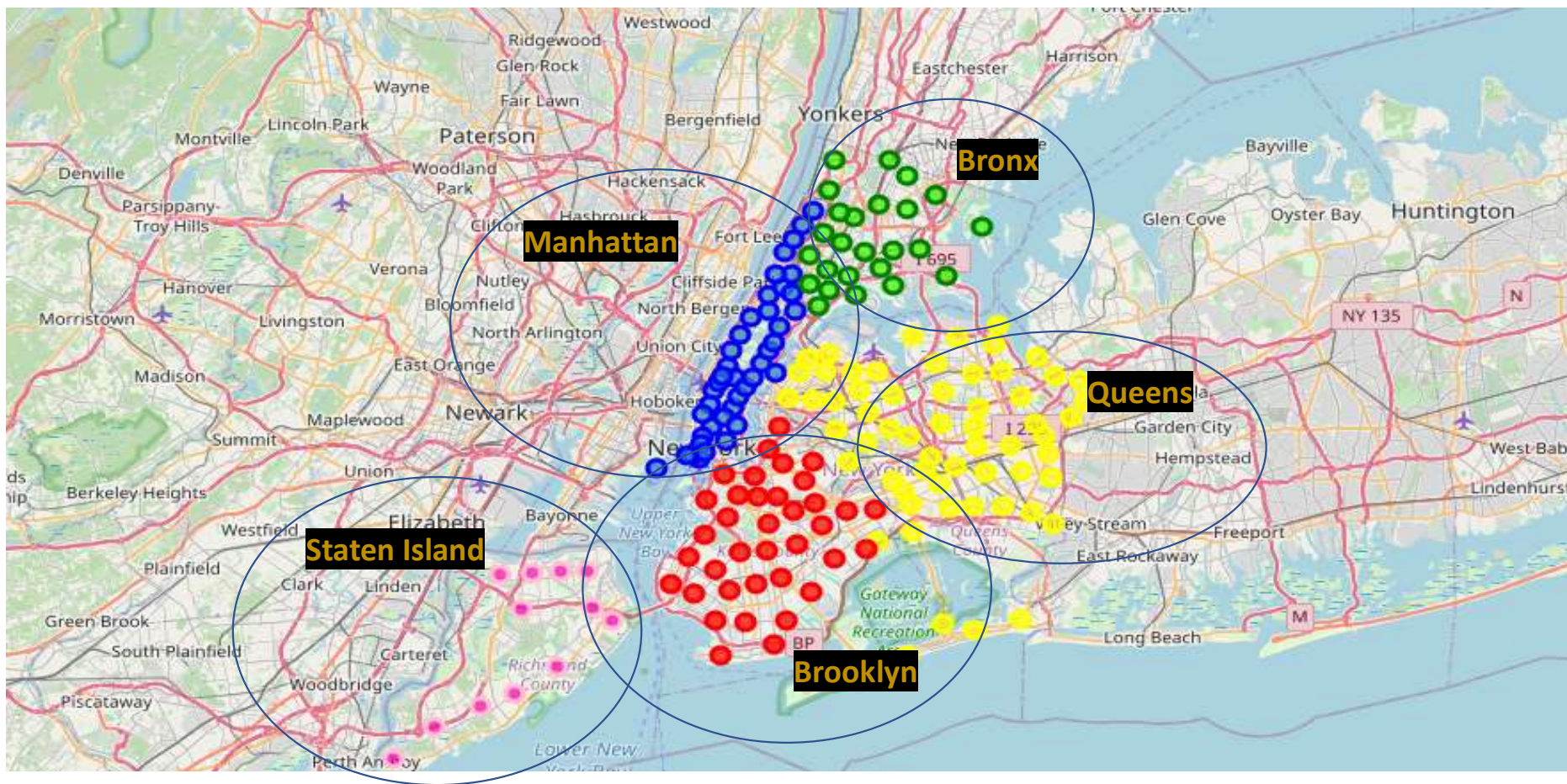
Boroughs Dataframe

	Borough	Neighborhood	Zip	Latitude	Longitude
0	Bronx	Central Bronx	10453	40.853017	-73.91214
1	Bronx	Central Bronx	10457	40.846745	-73.89861
2	Bronx	Central Bronx	10460	40.840950	-73.88036
3	Bronx	Bronx Park and Fordham	10458	40.864166	-73.88881
4	Bronx	Bronx Park and Fordham	10467	40.872265	-73.86937

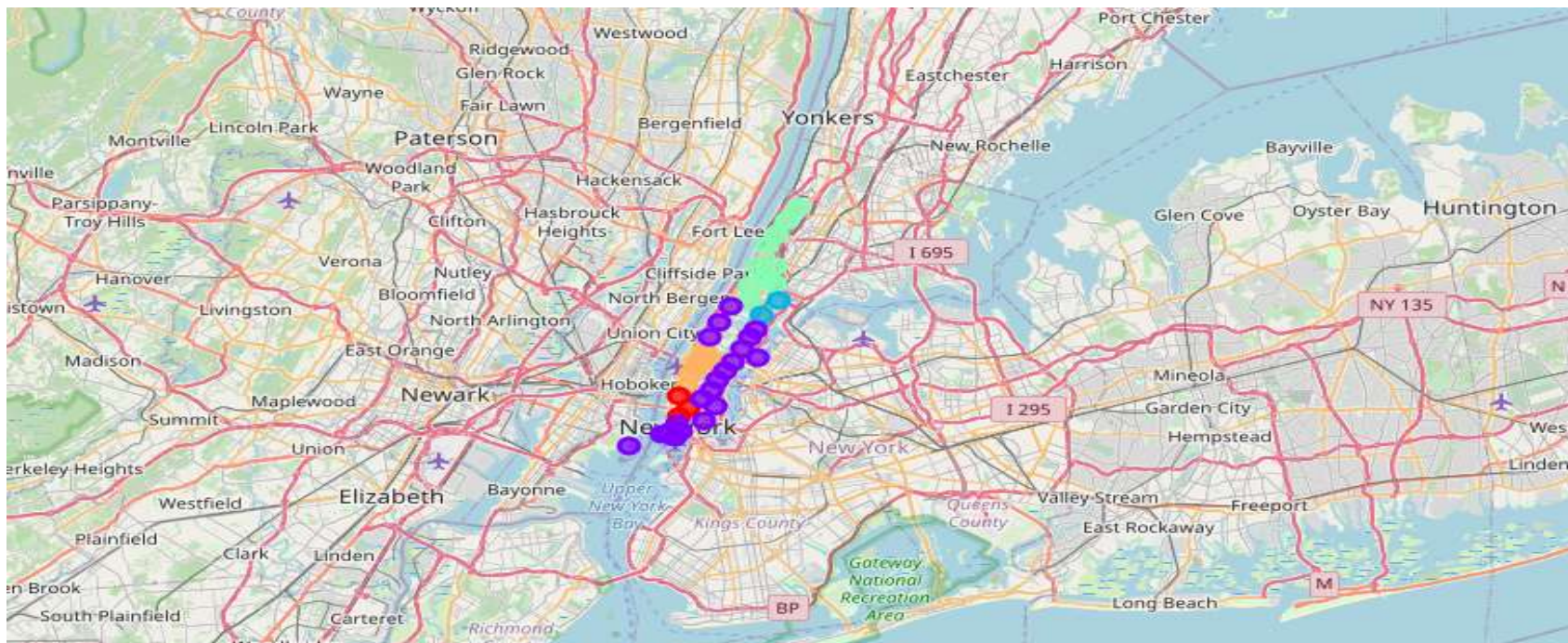
How many neighborhood in each Borough?

```
Queens      61
Manhattan   41
Brooklyn     37
Bronx        25
Staten Island 12
Name: Borough, dtype: int64
```

Visualizing all 5 boroughs in a single map



Analyzing k-Means clusters of Manhattan

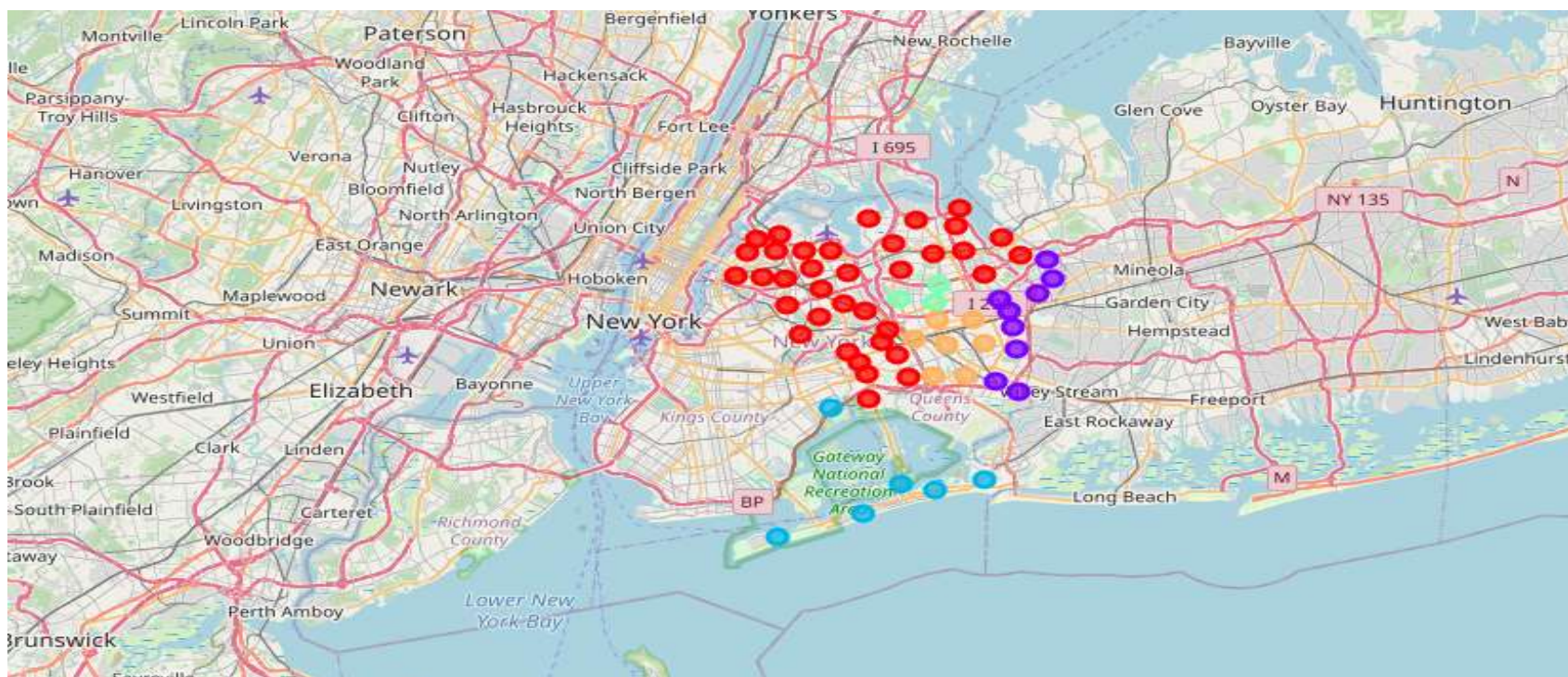


The cluster map of Manhattan looks pretty good. The each colored clusters have been segmented into:

- Red: Italian, American restaurants and coffee shops.
- Purple: Park, gym, coffee shops and wine bars.
- Blue: Mexican restaurants, pharmacy, super markets etc
- Light blue: Chinese restaurants, Coffee Shops, Grocery Stores etc
- Orange: Theater, Coffee Shops, Hotel etc

However, the map doesn't show any restarants chain for Carribean foods

Analyzing k-Means clusters of Queens

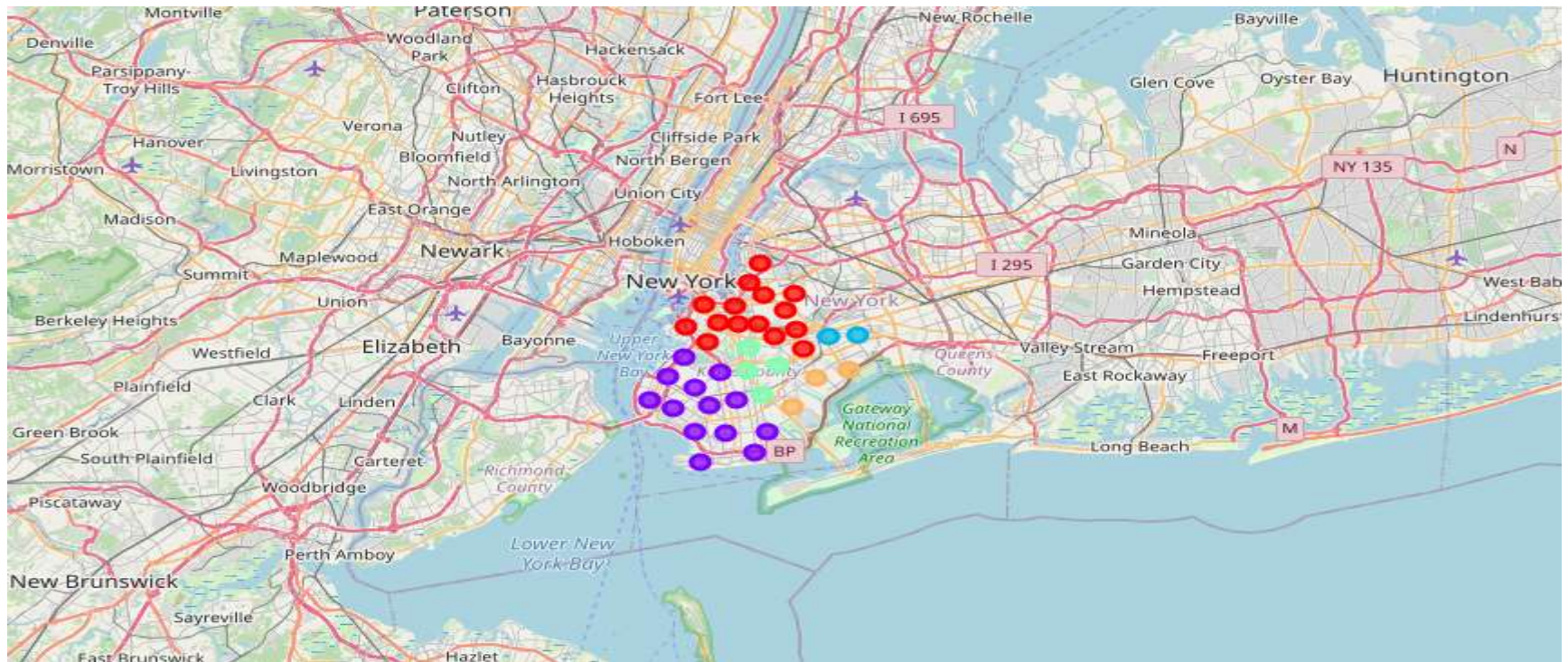


In Queens map, the red marker cluster has come out very dense and strong one. Let's explore all clusters in Queens map:

- Red: Chinese & Korean Restaurants, Pizza/Bakery Shops, Pharmacy, Supermarket etc
- Purple: Caribbean Restaurant, Fast Food Restaurant, Sandwich/Pizza Place Pharmacy etc
- Blue: Beach, Deli/Bodega, Department/Discount Stores etc
- Light blue: Sandwich/Pizza place, Ice Cream Shops, Bank etc
- Orange: Chinese Restaurants, Donut Shops, Bus Stations etc

Interestingly second cluster shows Caribbean restaurants, we need to further explore this later

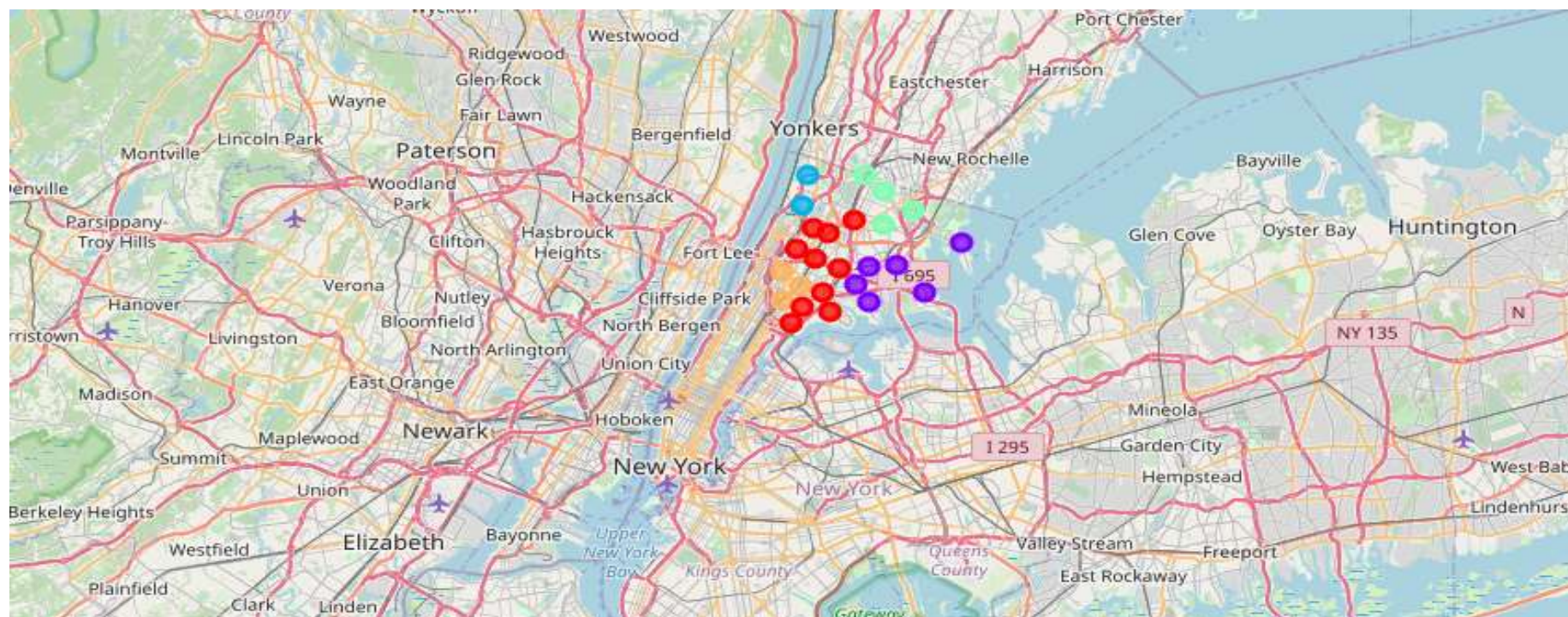
Analyzing k-Means clusters of Brooklyn



Now let's look at the clusters Brooklyn:

- Red: Bar, Caribbean Restaurant, Coffee Shop, Chicken Point, Mexican Restaurant etc
- Purple: Pizza Place, Italian/Chinese Restaurant, Pharmacy, Bakery, Mobile Shops etc
- Blue: Pizza Place, Supermarket, Discount Store, Playground etc
- Light blue: Caribbean Restaurant, Chinese Restaurant, Bank, Mobile Shops etc
- Orange: Chinese Restaurant, Pizza Place, Italian/American Restaurant, Super Market, Pharmacy etc

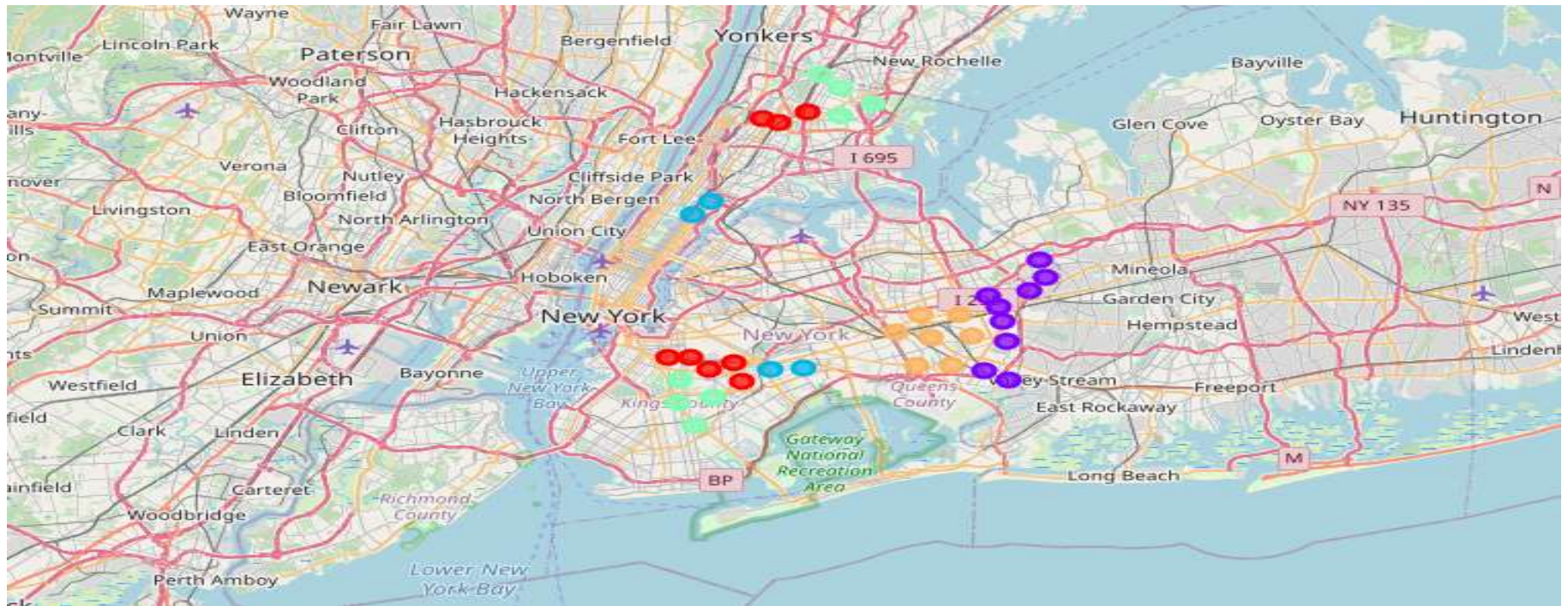
Analyzing k-Means clusters of Bronx



Now let's look at the clusters of Bronx:

- Red: Pizza Place, Donut/Coffee Shop, Pharmacy, Latin/Chinese Restaurant, Bank, Caribbean/Chinese Restaurant etc
- Purple: Latin American Restaurant, Bar, Pharmacy, Pizza/Donut Shop, Park etc
- Blue: Pizza Place, Gym, Mexican/Spanish Restaurant etc
- Light blue: Pizza Place, Caribbean Restaurant, Pharmacy, Chinese Restaurant, Gas Station etc
- Orange: Fast Food Restaurant, Ice Cream Shop, Pizza Shop, Bar etc

Visualizing the map only for Caribbean Restaurants



```
df_carb.shape
```

 $(36, 16)$

Location of Caribbean Restaurants basis on Rank



- Icon with Green color are most popular ones where Caribbean Restaurant is located
- Pink is the Second Rank and Orange is Third Rank
- The red icon indicates fourth rank which should be ruled out basis on less popularity

Rank 1 venues for suggesting the location of Caribbean Restaurant.

Borough	Neighborhood	Cluster Labels	Latitude	Longitude	Zip
Queens	Southeast Queens	1	40.742944	-73.70956	11004
Queens	Southeast Queens	1	40.756983	-73.71480	11005
Queens	Southeast Queens	1	40.693538	-73.73574	11411
Queens	Southeast Queens	1	40.670138	-73.75141	11413
Queens	Southeast Queens	1	40.662538	-73.73514	11422
Queens	Southeast Queens	1	40.732239	-73.72108	11426
Queens	Southeast Queens	1	40.728235	-73.74782	11427
Queens	Southeast Queens	1	40.719981	-73.74127	11428
Queens	Southeast Queens	1	40.708833	-73.73903	11429
Brooklyn	Flatbush	3	40.649059	-73.93304	11203
Brooklyn	Flatbush	3	40.627946	-73.94552	11210
Brooklyn	Flatbush	3	40.662892	-73.95509	11225
Brooklyn	Flatbush	3	40.645256	-73.95553	11226

Our Recommendation

Conclusion

Purpose of this project was to identify the best space for opening the Caribbean Restaurant amongst 5 boroughs i.e. Manhattan, Brooklyn, Queens, Bronx and Staten Island. When we plotted the neighborhoods of all 5 boroughs in a single map, we found that Staten Island has relatively low population and less density so that is not the optimal choice. So we ruled out Staten Island from our analysis and focused on remaining 4 boroughs. We identified the top 10 venues of each borough and segmented them using K-Means clustering. The clustering output helped us to segment the boroughs by common venue categories. We identified the venues where Caribbean Restaurants located and analyzed them using Folium map. We finally identified Southeast Queens and Flatbush in Brooklyn where Caribbean Restaurants are popular however Queens has relatively large number of restaurants which may give stiff competition so Flatbush-Brooklyn is the optimal choice.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc