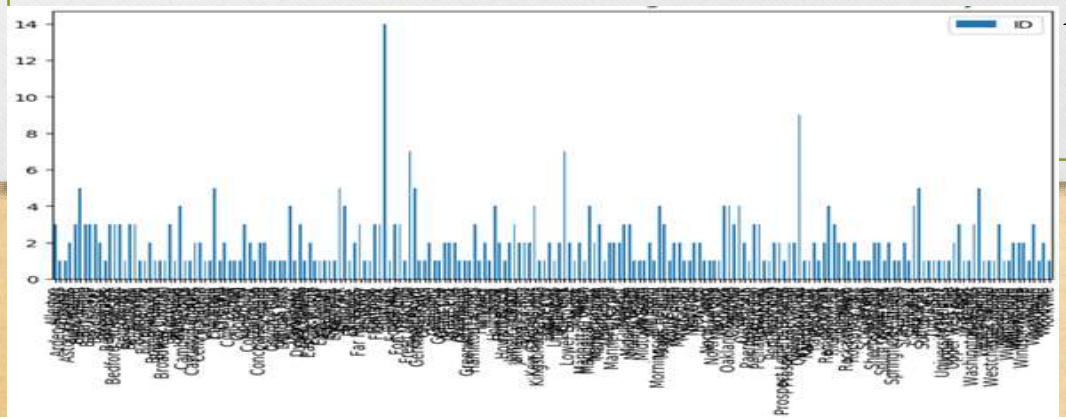
Restaurant Location Analysis

Applied Data Science Capstone

Week 5 Peer-Graded Assignment : Capstone Project - The Battle of Neighborhoods

Kenny Lu



1. Introduction

- **A description of the problem and a discussion of the background.**
- New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States.
- Throughout its history, New York City has been a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighborhoods on the Lower East Side. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world.
- With it's diverse culture , comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese , Indian , French etc.
- So as part of this project , we will find suitable locations to open a Chinese Restaurant in the New York City.

2. Data

- **A description of the data and how it will be used to solve the problem**

a) Population Data

- Data source : https://en.wikipedia.org/wiki/Demographics_of_New_York_City
- Description: We will get population data from this wikipedia page

b) Demographics Data

- Data source : https://worldpopulationreview.com/us-cities/new-york-city-population
- Description: We will get Demographics data from this page to find out where the people who match our market

2. Data (1)

- **A description of the data** **.. continue**

c) DOHMH Farmers Market Data

- Data source : https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2
- Description: We will get the data of Farmers Markets, and we will use this data to figure out where you will be sourcing your food and supplies

d) New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.

- Data source : 'https://cocl.us/new_york_dataset
- Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of New York City.

2. Data (2)

- **A description of the data** **.. continue**

e) Chinese Restaurants in each neighborhood of the New York City.

- Data source : **Foursquare API**
- Description : By using this API we will get all the venues in each neighborhood. We can filter these venues to get only Chinese restaurants.

f) GeoSpace data

- Data source : https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm
- Description : By using this geo space data we will get the New York Borough boundaries that will help us visualize choropleth map.

3. Methodology

Approach

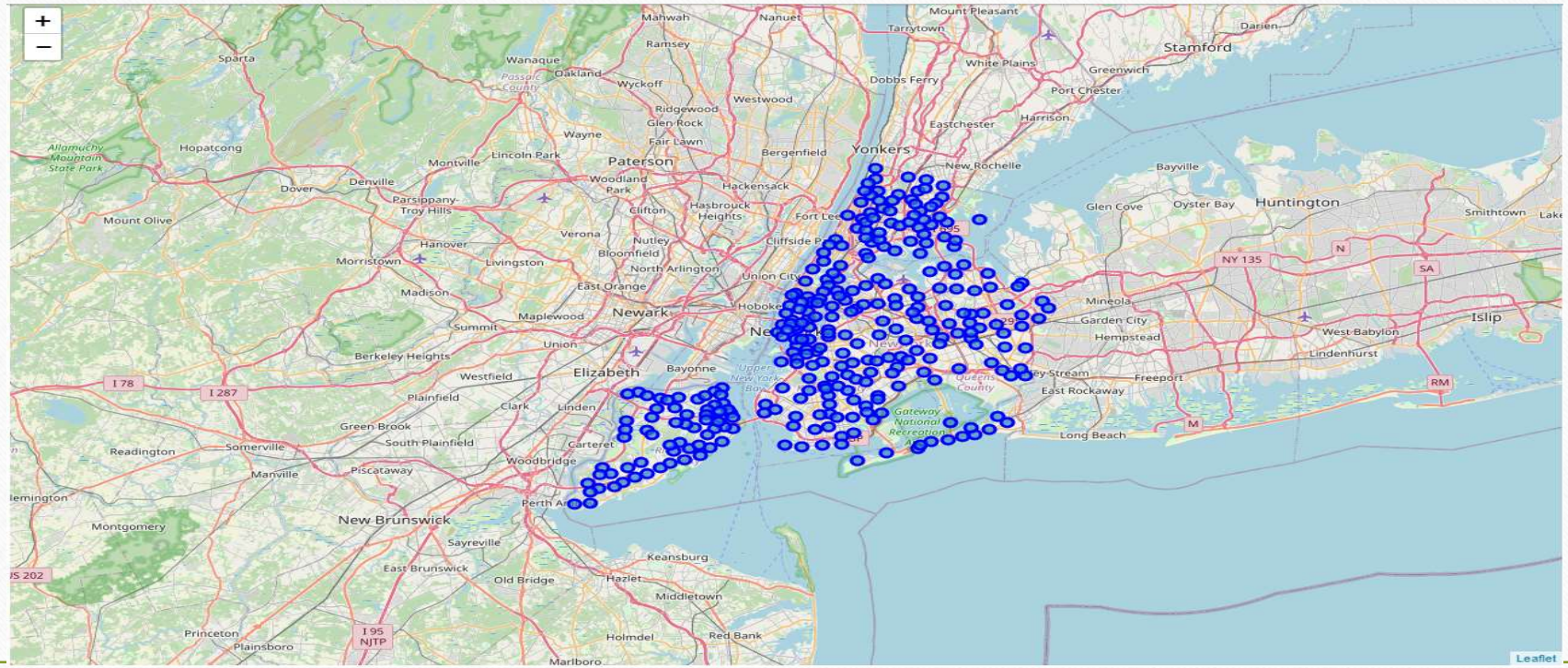
The workflow of the project starts with the web scraping and data wrangling. Using the Beautiful Soup library, the postal code and the neighborhood is processed to derive the latitude and longitude of the New York City neighborhood, latest population data and demographics data of New York City were also being scrapped.

These data had been stored into Pandas dataframe, with necessary data cleaning and wrangling, all needed data were transform into preferable format.

With libraries: Seaborn, Matplotlib and Folium Map, the processed data were used to do the data visualization.

3. Methodology (1)

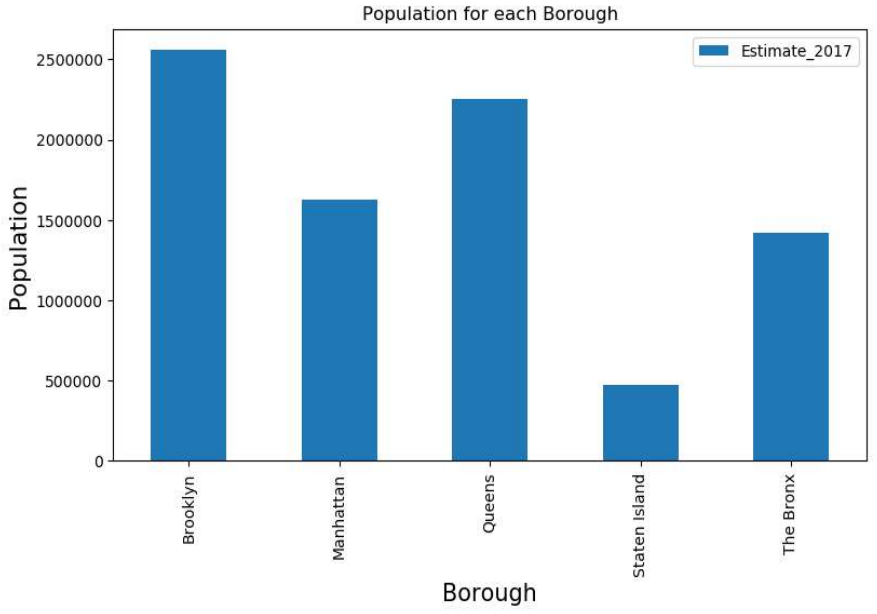
- **Map of New York with neighborhoods superimposed on top**



3. Methodology (2)

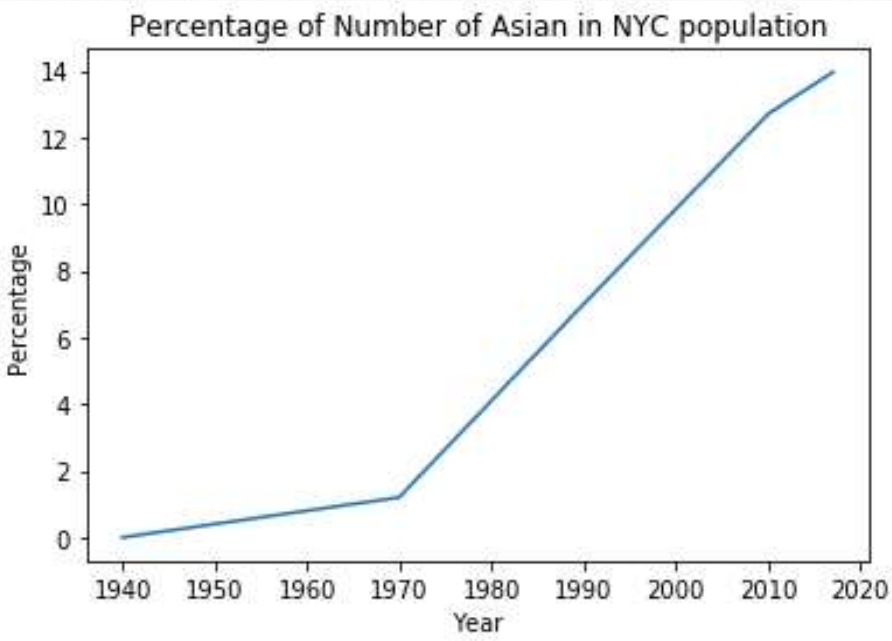
- **Population Data of New York City**

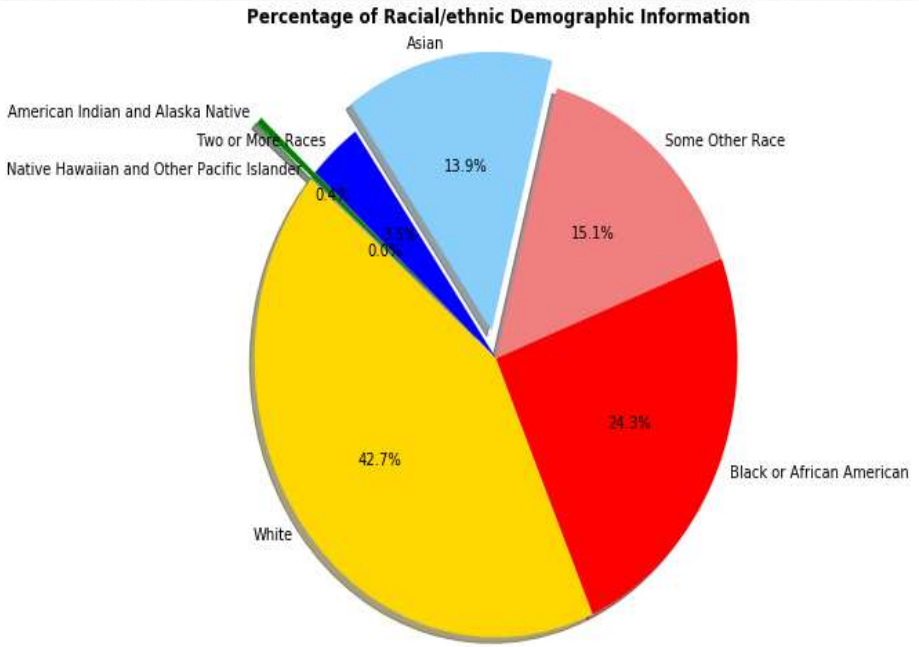
	Borough	County	Estimate_2017	GDP	square_miles	square_km	persons_sq_mi
0	The Bronx	Bronx	1,418,207	42.695	30,100	42.10	109.04
1	Brooklyn	Kings	2,559,903	91.559	35,800	70.82	183.42
2	Manhattan	New York	1,628,706	600.244	368,500	22.83	59.13
3	Queens	Queens	2,253,858	93.310	41,400	108.53	281.09
4	Staten Island	Richmond	476,143	14.514	30,500	58.37	151.18



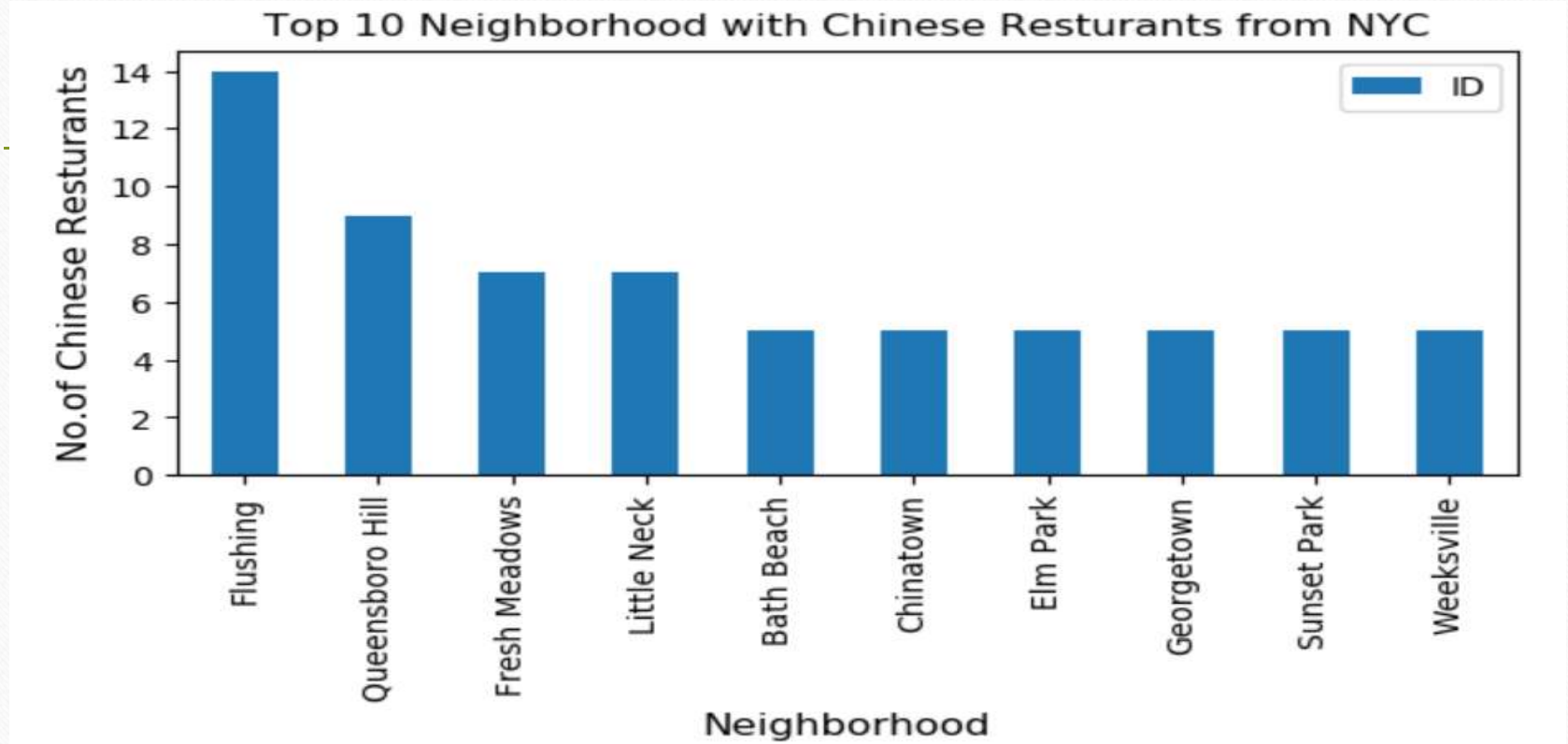
3. Methodology (3)

- **Demographics Data of New York City**



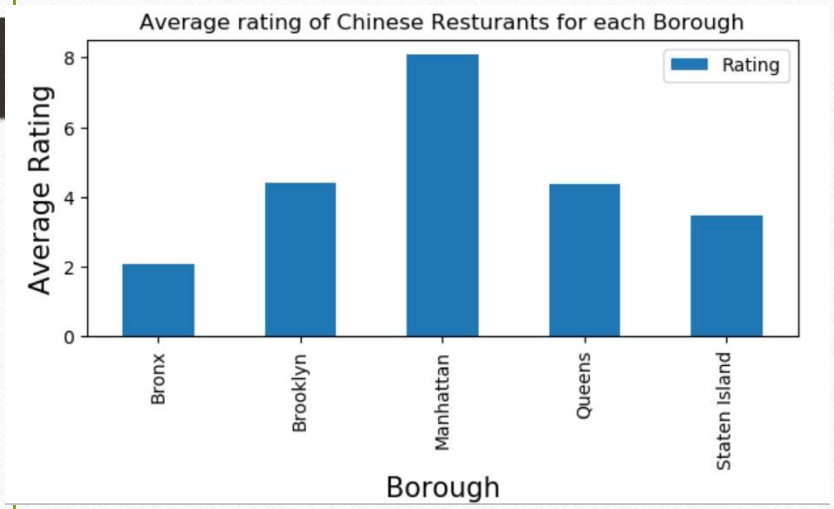


3. Methodology (4)

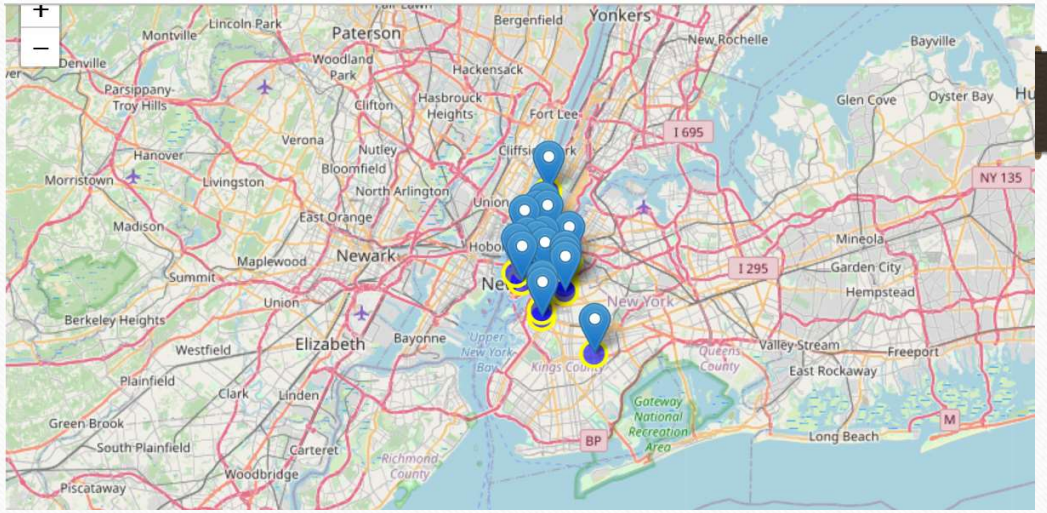


3. Methodology (5)

These are the **average rating** of Chinese Restaurants for each neighborhood in each **Borough.** The highest three are in Borough Brooklyn and Manhattan.

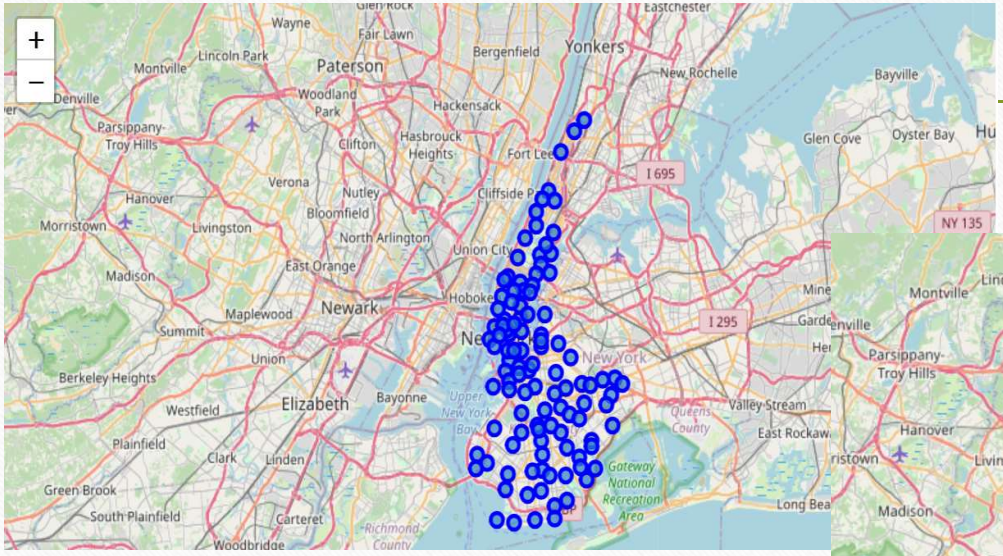


Top 18 Rate Chinese Restaurants superimposed on top

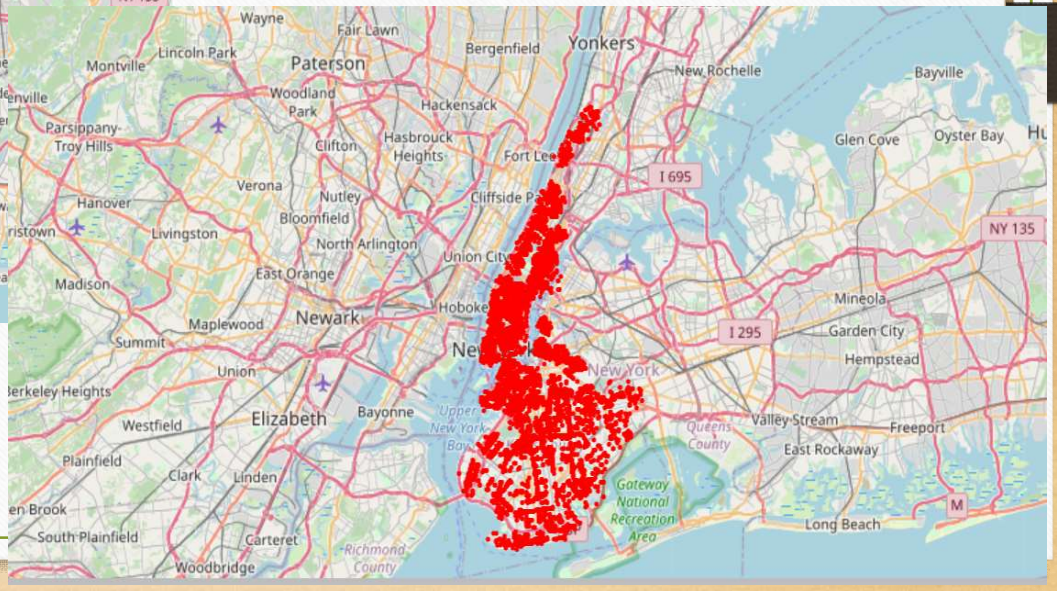


3. Methodology (6)

Brooklyn and Manhattan Visualization



Brooklyn and Manhattan Venue Visualization



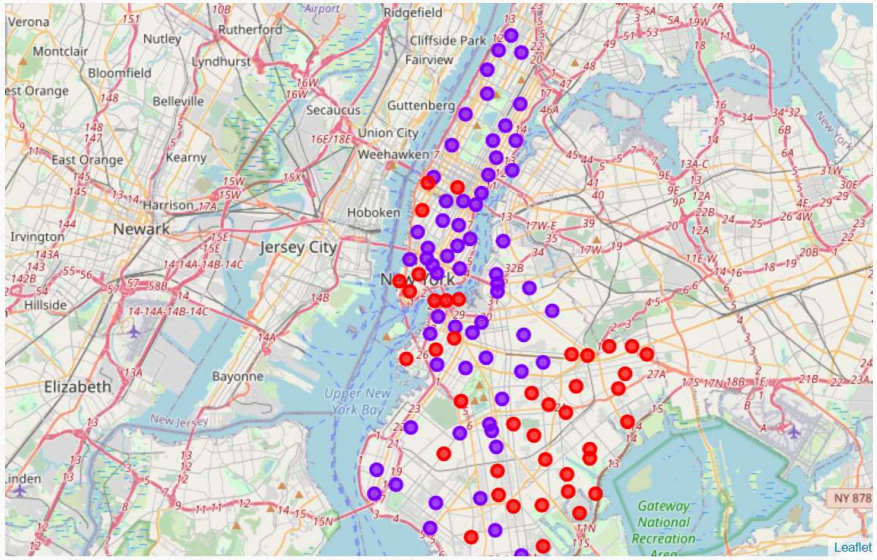
4. Results

- From the analysis of Population data of New York City, we have find that Brooklyn, Manhattan and Queens has the most Borough. Further to analyses the Farmers Markets data, we have that Brooklyn and Manhattan has the most number of Farmers Markets. Also from Foursquare data, Queens, Manhattan and Brooklyn has the most numbers of Chinese Restaurants.
- By analyses the venues details data, we cluster the neighborhoods into two clusters and use K-Means clustering algorithm to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.

4. Results (1)

K-Means Clustering Algorithm

- Cluster0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.
- Cluster1: The Total and Total Sum of cluster1 has highest values. It shows that the markets are saturated. Number of restaurants are very high.
- There are no untapped neighborhoods in Brooklyn and Manhattan.



5. Discussion

- The Farmers Markets in Bronx, Queens and Staten Island may need to increase
- In Manhattan and Brooklyn the rating of Chinese Restaurants is high in general
- In Queens, the rating of Chinese Restaurants is lower than Manhattan and Brooklyn
- In Manhattan and Brooklyn restaurants of many countries are available. The competition is keen, not just only compete with Chinese restaurants.

6. Conclusion

- In this project, the analysis is performed with limited data. Base on the analysis, although Brooklyn and Manhattan has high concentration of restaurant business, and with the most high rating Chinese Restaurants there, but it still is our recommendation location. Bronx, Queens and Staten Island also has good number of restaurants, but is still can be explored.

THANK YOU