

Applied Data Science Capstone

Capstone Project - The Battle of Neighborhoods - Report

1. Introduction

New York is a city located in Bronx County New York. With a 2020 population of 8,323,338, it is the largest city in New York and the largest city in the United States. New York is currently declining at a rate of -0.45% annually but its population has increased by 1.81% since the most recent census, which recorded a population of 8,175,133 in 2010. New York reached it's highest population of 8,475,976 in 2016. Spanning over 468 miles, New York has a population density of 27,709 people per square mile.

The city features five separate boroughs: Staten Island, The Bronx, Brooklyn, Queens, and Manhattan. As many as 800 languages are spoken throughout New York City, making it the most diverse city in the world when it comes to linguistic multiplicity.

According to the most recent ACS, the racial composition of New York City was:

- White: 42.67%
- Black or African American: 24.27%
- Other race: 15.12%
- **Asian: 13.95%**
- Two or more races: 3.51%
- Native American: 0.43%
- Native Hawaiian or Pacific Islander: 0.05%

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, and French etc.

Our client had set their target on the **Asian (13.95%)** of New York City population, and wish to find suitable locations to open a **Chinese Restaurant** in the New York City.

2. Data

For this project we need the following data:

- **Population Data**
 - Data source : https://en.wikipedia.org/wiki/Demographics_of_New_York_City
 - Description: We will get population data from this wikipedia page
- **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
- **DOHMH Farmers Market**
 - 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
- **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.
 - An extract of the json is as follows:

```
{'type': 'Feature',  
'id': 'nyu_2451_34572.306',  
'geometry': {'type': 'Point',  
'coordinates': [-74.08173992211962, 40.61731079252983]},  
'geometry_name': 'geom',  
'properties': {'name': 'Fox Hills',  
'stacked': 2,  
'annoline1': 'Fox',  
'annoline2': 'Hills',  
'annoline3': None,  
'annoangle': 0.0,  
'borough': 'Staten Island',  
'bbox': [-74.08173992211962,  
40.61731079252983,  
-74.08173992211962,  
40.61731079252983]}}
```
- **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.
 - An extract of an API call is as follows:

```
'categories': [{ 'id': '4bf58dd8d48988d110941735',  
'name': 'Chinese Restaurant',
```

```
'pluralName': 'Chinese Restaurants',  
'shortName': 'Chinese',  
'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/Chinese_',  
'suffix': '.png'},  
'primary': True}],  
'verified': False,  
'stats': {'tipCount': 17},  
'url': 'http://eccorestaurantny.com',  
'price': {'tier': 4, 'message': 'Very Expensive', 'currency'}
```

- **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.

Questions that can be asked using the above mentioned datasets

- Why we choose New York City?
- Where is the proximity to Supplier?
- Where is best location in New York City for Chinese Cuisine?
- Which areas have potential Chinese Restaurant Market?
- Which all areas lack Chinese Restaurants?
- Which is the best place to stay if I prefer Chinese Cuisine?

So as part of this project, we will analyze and visualize all major neighborhood of New York City that has great Chinese Restaurants.

3. Methodology

Business Understanding:

Our project goal is to do the Restaurant Location Analysis, so that identify the optimum location for the new restaurant business in New York City.

Analytic 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.

We further to use Foursquare API to explorer neighborhoods in Brooklyn & Manhattan, we use the explore function to get the most common venue categories in each neighborhood, and then use the features to group the neighborhoods into clusters. We use the k-means clustering algorithm to complete this.

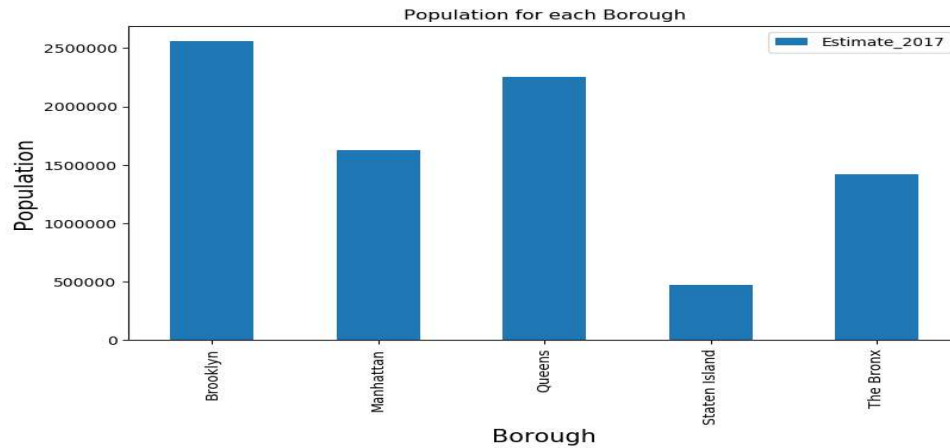
With libraries: Seaborn, Matplotlib and Folium Map, the processed data were used to do the data visualization.

Exploratory Data Analysis:

Population Data of New York City

New York City neighborhood has a total of 5 boroughs and 306 neighborhoods.

	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

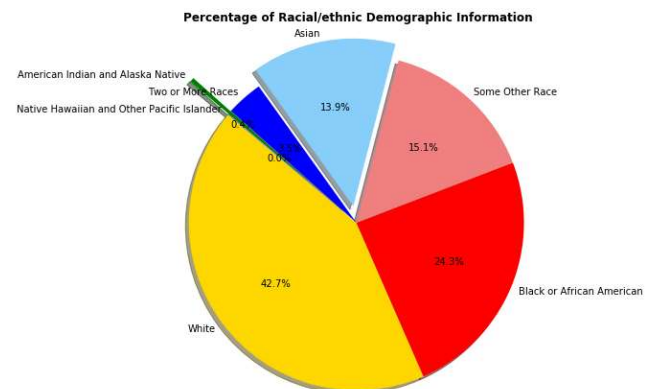
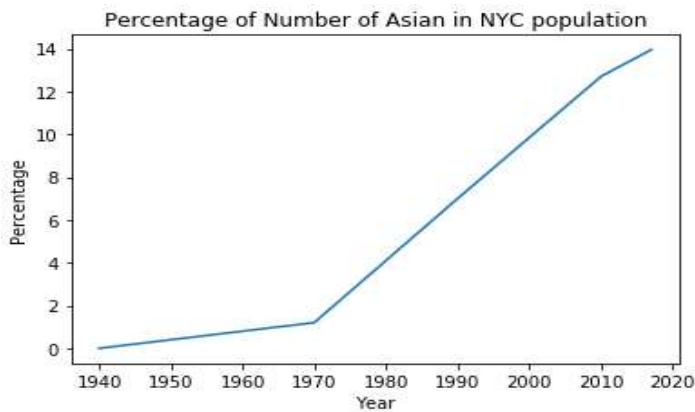


From above figures, we will have the insight for the data:

- Brooklyn is the city's most populous borough
- Queens is geographically the largest borough and second populous borough

Demographics Data of New York City

Between 1900 and 2010, New York City's total Asian population increased by over one-hundred-and-fifty times over the same period. With the latest figures, total Asian population in New York City had been increased to 13.95% of New York City total population.



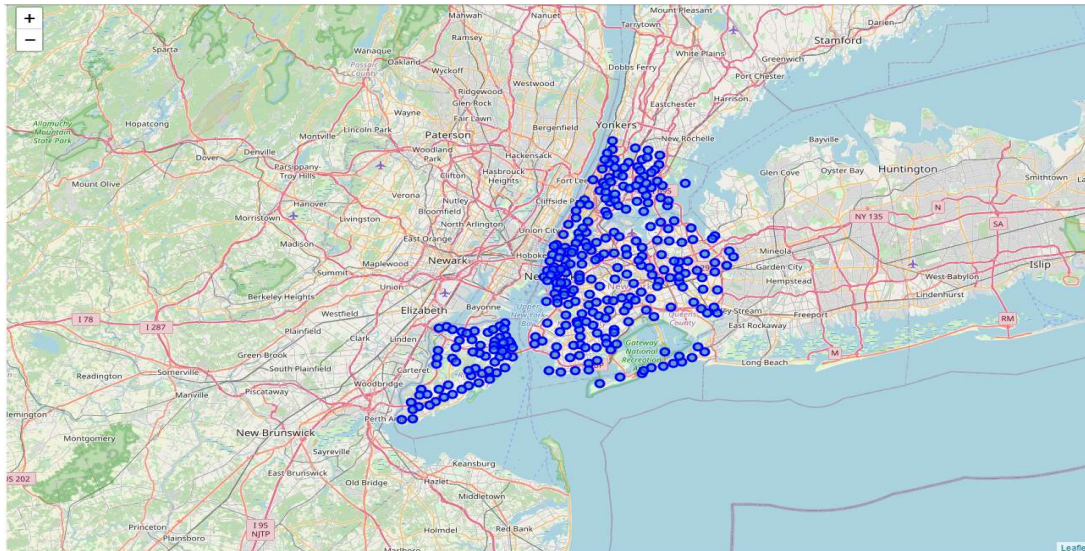
Map of New York with neighborhoods superimposed on top

New York City has all 5 boroughs and 306 neighborhoods. In this project, we firstly to collect the New York City data from https://cocl.us/new_york_dataset, transform the data into a pandas dataframe and then use geopy to get New York City Geographical Coordinates data and finally use Folium libraries to create a map of New York City with neighborhoods superimposed on the top.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Let's make sure that the dataset has all 5 boroughs and 306 neighborhoods.

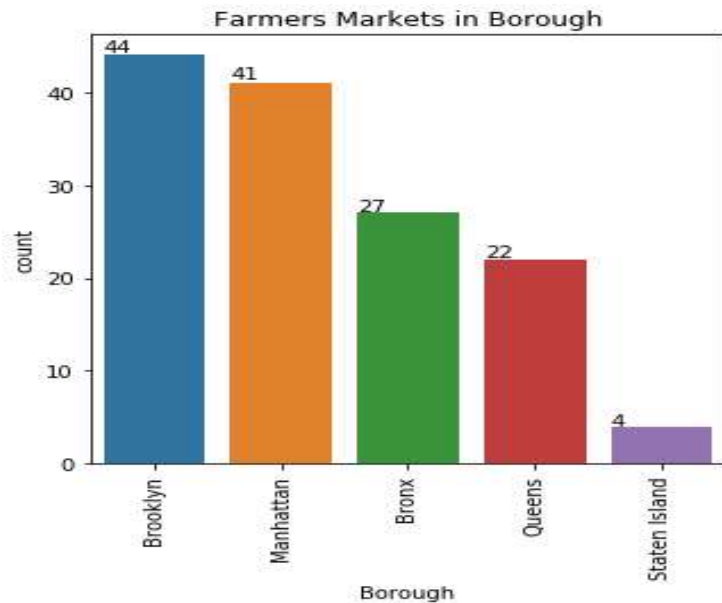
Map of New York with neighborhoods superimposed on top



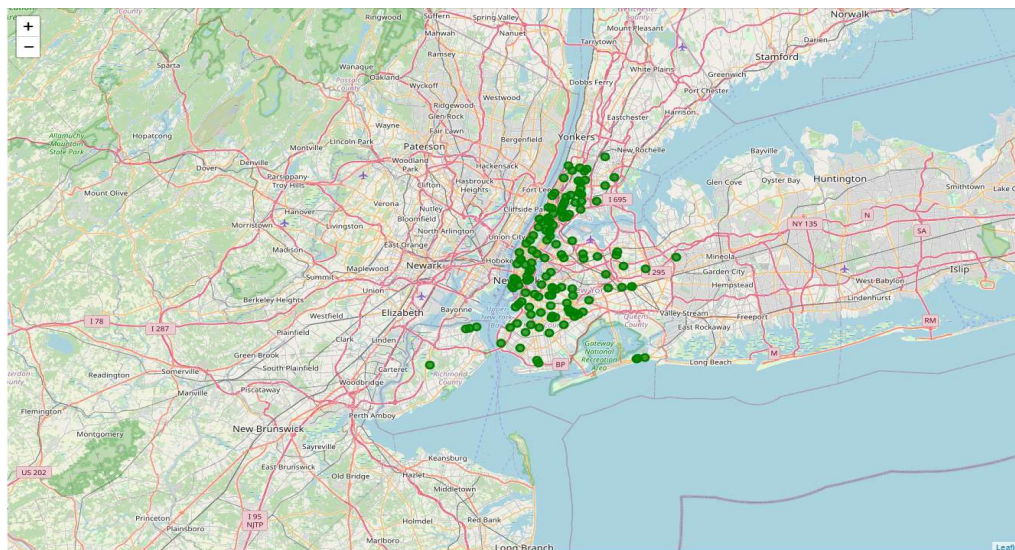
Farmers Markets in Borough

There are totally 138 farmers in New York City.

Manhattan and Brooklyn has highest numbers of Farmers Markets. There are 44 Farmers Markets in Borough Brooklyn while 41 in Manhattan

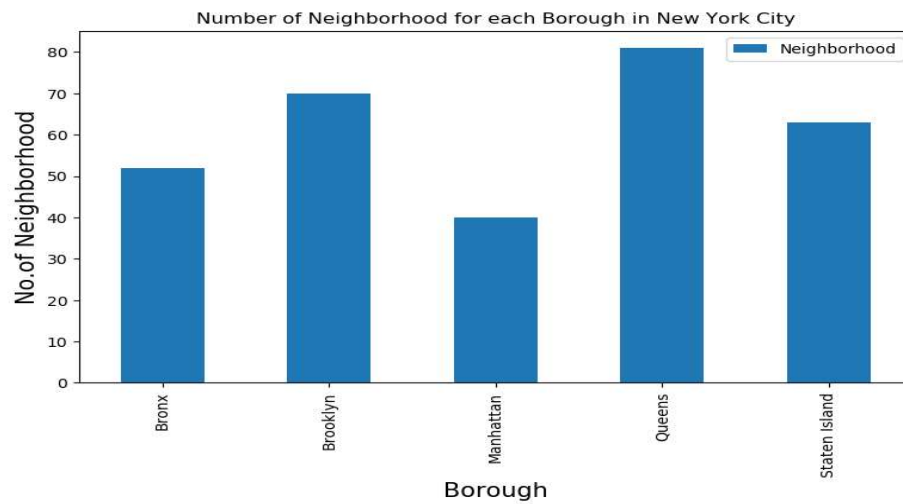


Map of New York with Markets Name & Borough of Farmers Market superimposed on top



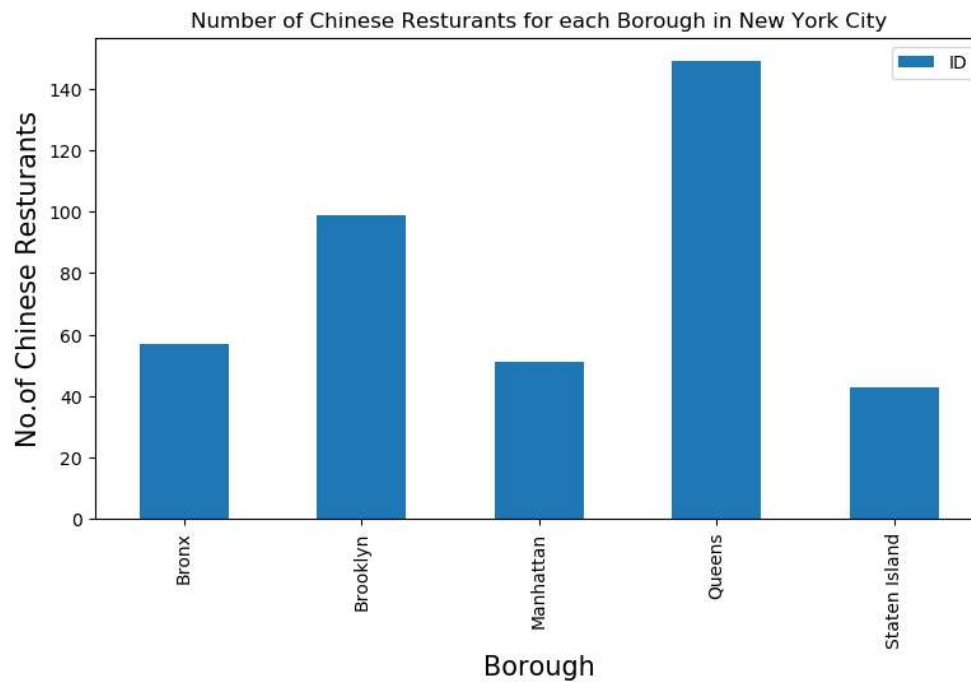
Number of Neighborhood for each Borough in New York City

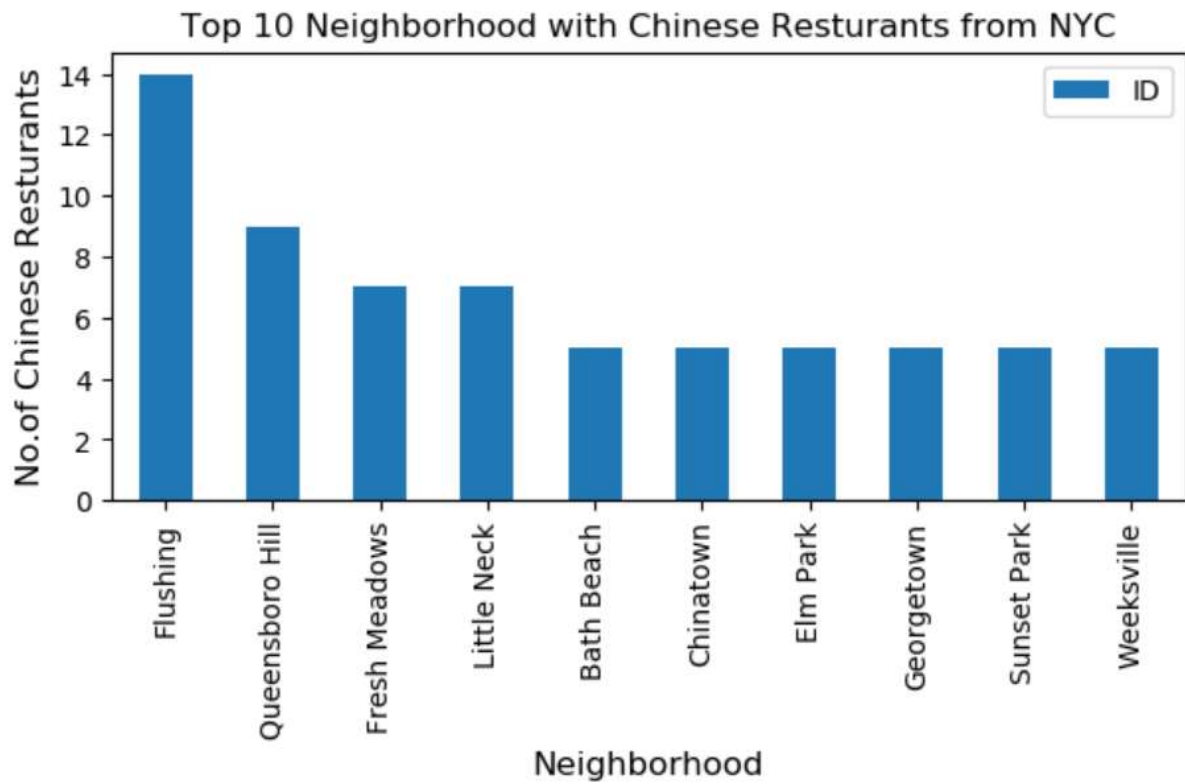
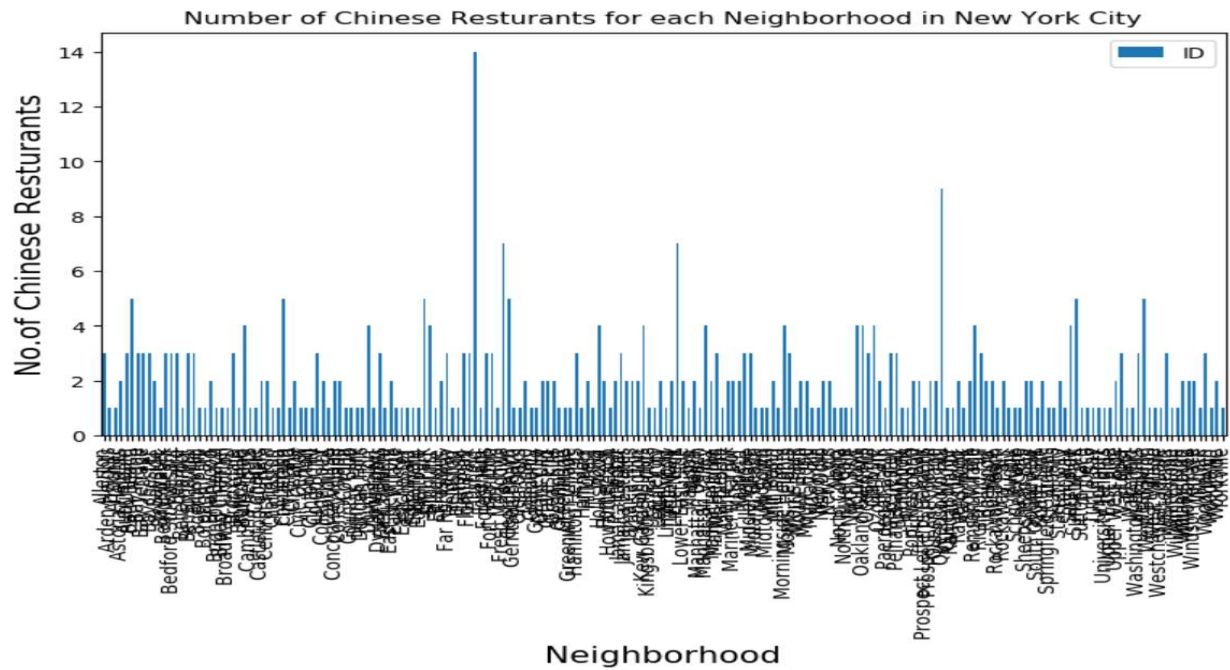
Queens has highest number of neighborhoods. Brooklyn and Manhattan has the second and third one.



Chinese restaurants for each Neighborhood

We got 410 Chinese Restaurants across New York City. Queens, Manhattan and Brooklyn has the most numbers of Chinese Restaurants.





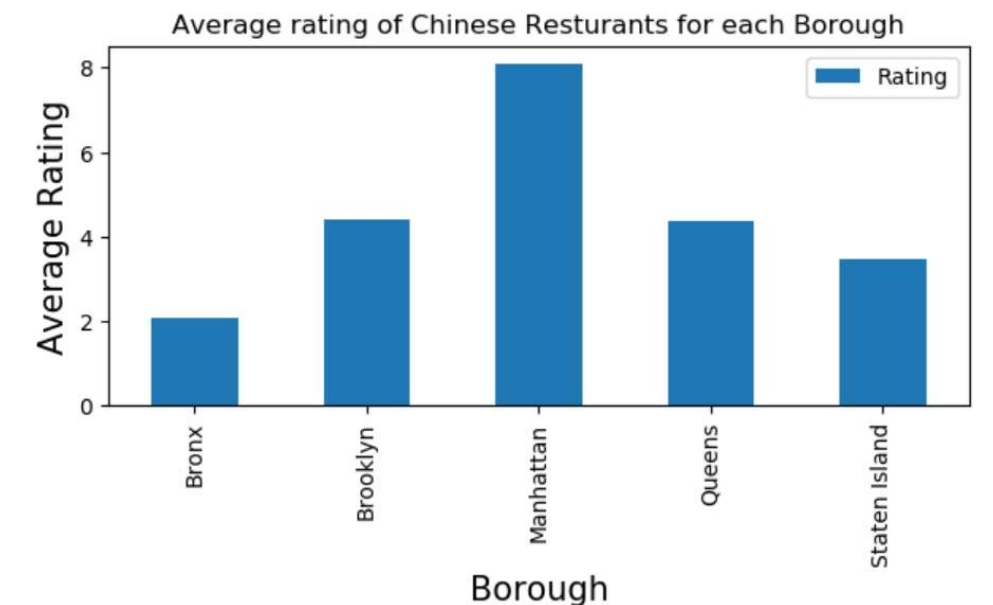
Flushing in Queens has the highest number of **Chinese Restaurants** with a total count of 14.

Borough **Queens** and **Brooklyn** has the highest number of Chinese Restaurants.

Borough	Neighborhood	
Queens	Flushing	14
	Queensboro Hill	9
	Fresh Meadows	7
	Little Neck	7
Brooklyn	Bath Beach	5
	Georgetown	5
	Sunset Park	5
	Weeksville	5
	Chinatown	5
Manhattan	Chinatown	5
Staten Island	Elm Park	5
Name: ID, dtype: int64		

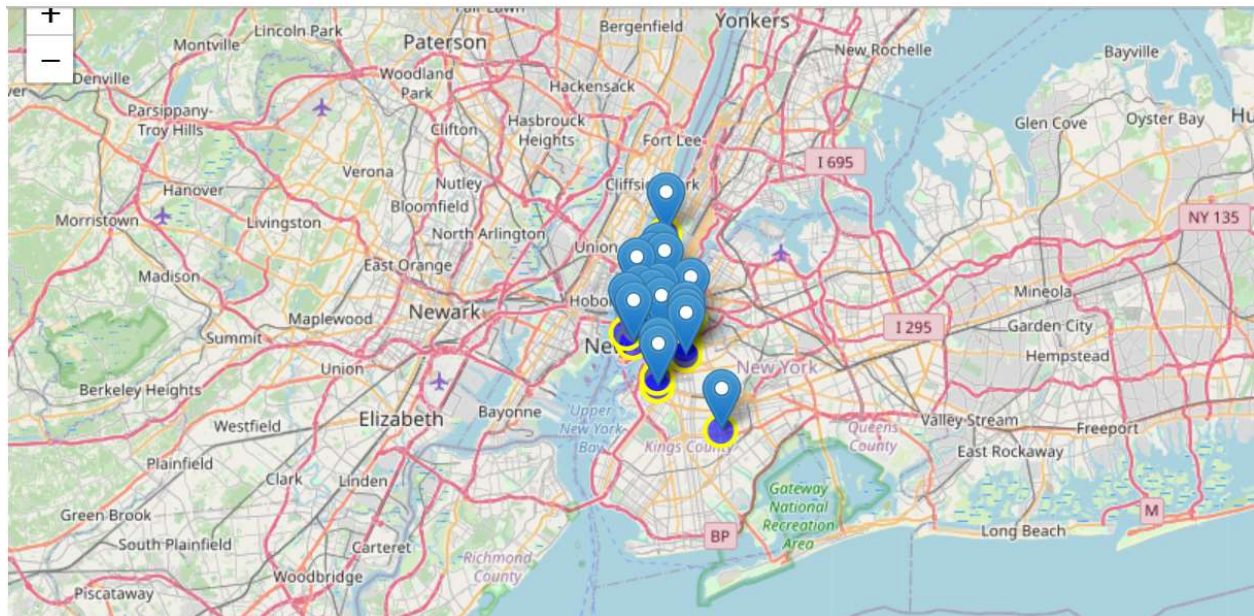
These are the **average rating** of Chinese Restaurants for each neighborhood in each **Borough**. The highest three are in Borough Brooklyn and Manhattan.

	Borough	Neighborhood	Average Rating
58	Brooklyn	Greenpoint	9.00
67	Brooklyn	North Side	8.90
95	Manhattan	Lower East Side	8.85
100	Manhattan	Midtown South	8.80
91	Manhattan	Greenwich Village	8.80
99	Manhattan	Midtown	8.80
85	Manhattan	Chelsea	8.80
102	Manhattan	Murray Hill	8.80
103	Manhattan	Noho	8.70
74	Brooklyn	Rugby	8.70
89	Manhattan	East Village	8.70
108	Manhattan	Upper West Side	8.60
38	Brooklyn	Boerum Hill	8.60
48	Brooklyn	Downtown	8.60
81	Brooklyn	Williamsburg	8.50
105	Manhattan	Tribeca	8.50
87	Manhattan	Civic Center	8.50
77	Brooklyn	South Side	8.50
116	Queens	Bayswater	8.40
110	Manhattan	West Village	8.40



Map of New York - Top 18 Rate Chinese Restaurants superimposed on top

There are 18 restaurants with rate over than 8.5.



Map of Brooklyn/Manhattan with neighborhoods superimposed on top

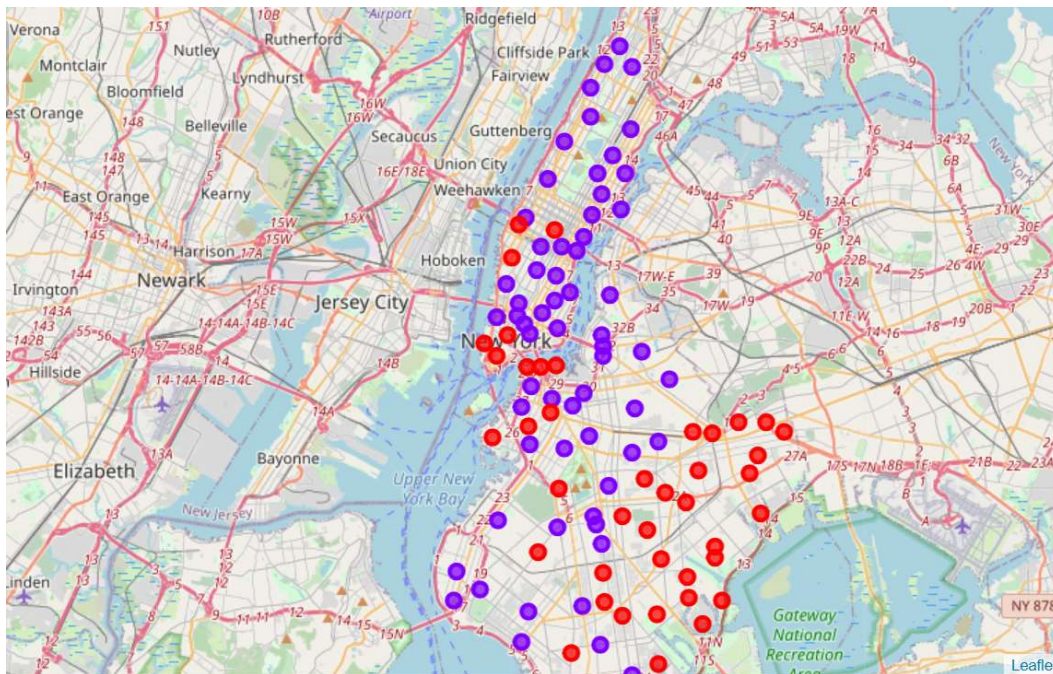


	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

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.



- 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

1. The Farmers Markets in Bronx, Queens and Staten Island may need to increase
2. In Manhattan and Brooklyn the rating of Chinese Restaurants is high in general
3. In Queens, the rating of Chinese Restaurants is lower than Manhattan and Brooklyn
4. 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.