

Capstone Project

Brian Deng

1. Objective

To collect local venue data and analysis to recommend new restaurant openings.

2. Business Problem

Let's say I work for a business consulting firm to help clients find the best place to open a new restaurant in a specific city. Our professional services include providing the enterprise with appropriate data on local venues and competitors. These insights will help clients successfully grow their business in their target markets.

All businesses seeking for expansion would be interested in this project as well.

One of the challenges is to collect accurate and detailed local restaurant data in the city. The project will import restaurant and category data, process and analysis to recommend the best expansion locations.

3. Data

Three sets of data are required in the project:

Client Information

Neighborhood Data

Local Venue Data

Client information is detailed business information such as the restaurant name and the type of food they serve. In this project, I assume that there is an East Asian fast-food chain “Asia Express”, which operates more than 100 branches in Canada and is looking for opportunities to expand south to the States.

It chose New York City as their landing point and found our company as their wise location selector. Restaurant category will be one of the most important factors in our analysis and recommendation process.

Neighborhood data is a dataset containing New York City neighborhoods. It was given by the capstone course from Coursera, which lists the names of neighborhoods in New York City and their borough, latitude and longitude coordinates. A specific neighborhood in this dataset will be selected as the ideal new branch location for Asia Express.

Local venue data describes all restaurants in each neighborhood. The dataset was obtained from Foursquare API, a dataset that stores each neighborhood and its category. The recommendation is then implemented by analysing the quantity and quality of competitors

3. Methodology

3.1 Neighborhood Data

Neighborhood data was given by the capstone course. I fetched the file from https://cocl.us/new_york_dataset. I stored the data into a dataset named “neighborhoods”, it lists Borough, Neighborhood, Latitude and Longitude.

```
neighborhoods.head()
```

	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

3.2 Local Venue Data - East Asian Restaurants

Local venue data in this project represents the list of all restaurants in each neighborhood. Since the client offers East Asian food only, we need to specify the category to Japanese Restaurant, Chinese Restaurant and Korean Restaurant.

The data can be obtained from Foursquare API, using neighborhood’s latitude and longitude coordinates. A list of local venue returned and waited to be processed. The project determines the venue’s category and only keep the venue from East Asian restaurant categories.

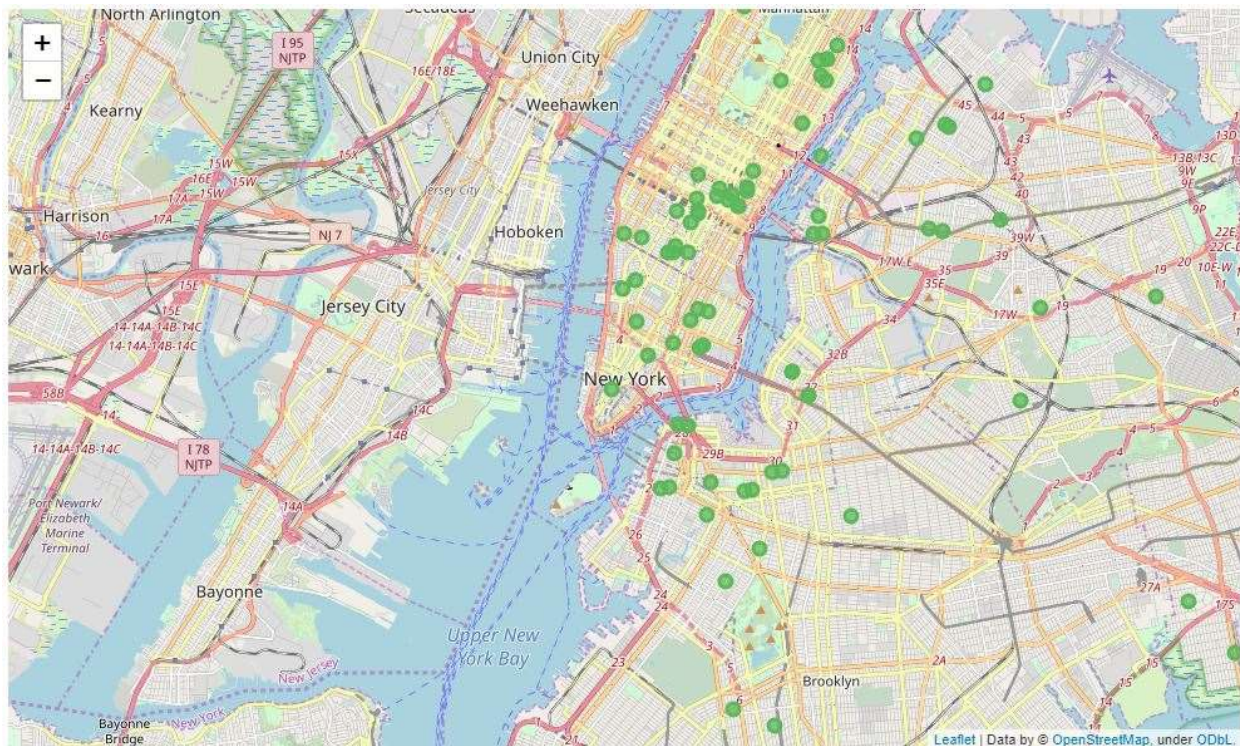
The local venue data is saved into a dataset named “restaurants”, it has serval columns, such as Borough, Neighborhood, Category, Name, Likes, Rating, ID, Latitude and Longitude.

```
restaurants.head()
```

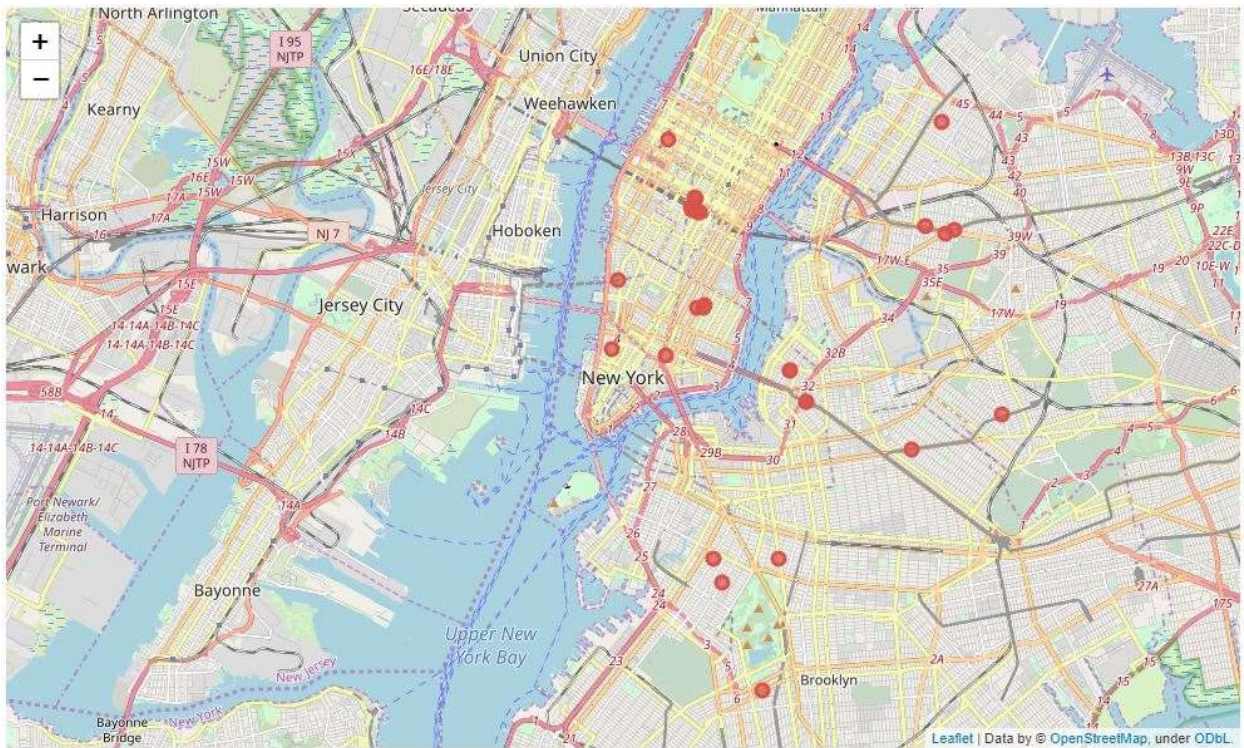
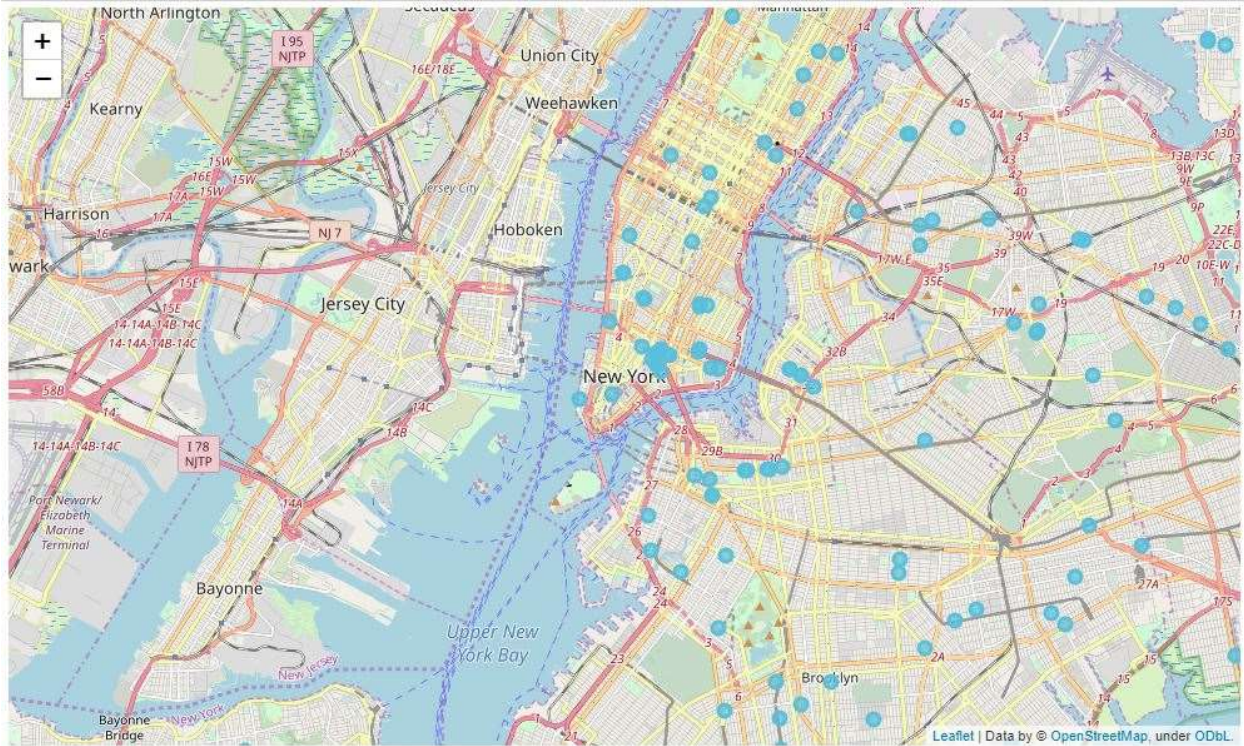
	Borough	Neighborhood	Category	Name	Likes	Rating	ID	Latitude	Longitude
0	Bronx	Co-op City	Chinese Restaurant	Guang Hui Chinese Restaurant	NaN	NaN	4c9d5f2654c8a1cd2e71834b	40.876651	-73.829092
1	Bronx	Eastchester	Chinese Restaurant	Xing Lung Chinese Restaurant	NaN	NaN	4dabc3dc93a04642f09ccabd	40.888785	-73.831226
2	Bronx	Kingsbridge	Chinese Restaurant	Guan Hin Restaurant	NaN	NaN	4e3b11c98877b00cfc27a483	40.883705	-73.897896
3	Bronx	Kingsbridge	Chinese Restaurant	Yeung Hing Chinese Restsurant	NaN	NaN	4da39def540ea1cdfb3b95de	40.885175	-73.899405
4	Bronx	Norwood	Chinese Restaurant	Happy Dragon	NaN	NaN	4ce313fc438b224bbc6c80a3	40.880410	-73.883442

3.3 Map Plot

The folium library can be used to plot venue locations on a map. A New York City map with Japanese restaurants was generated in the project and shown below.

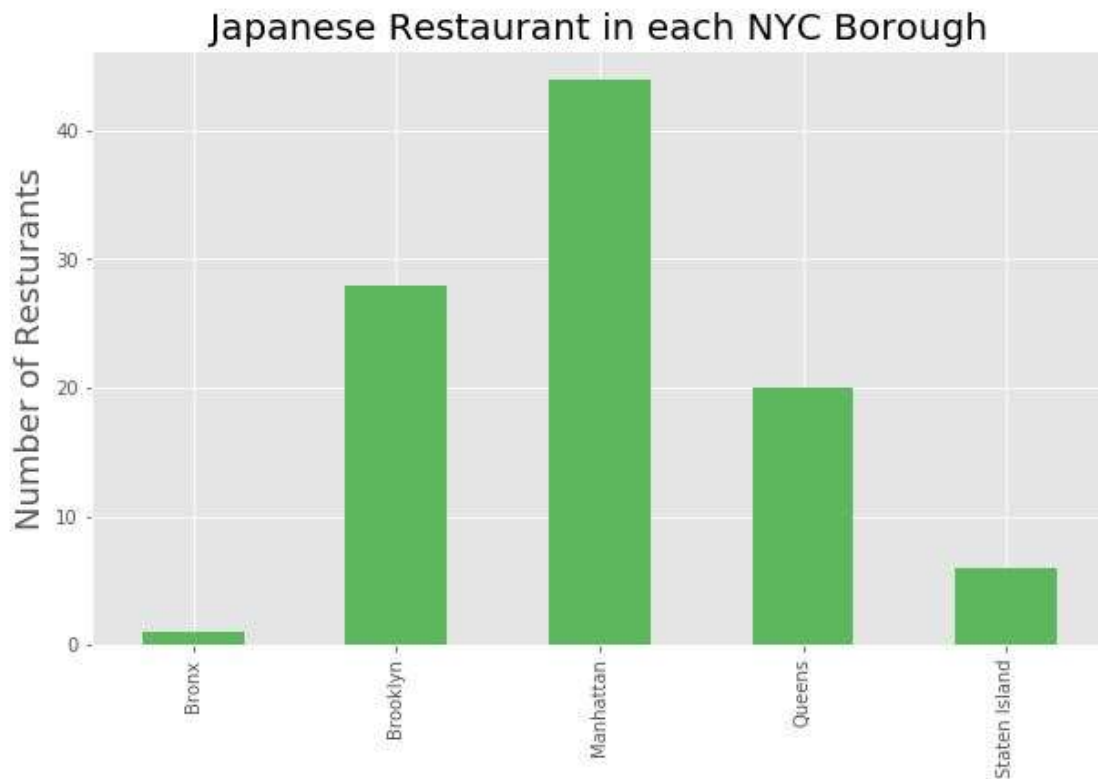


A map with Chinese and Korean restaurants can be generated with the same way, they are showing the distributions of different competitors. This visualization helps the analysis to make the location recommendation. These two maps are shown below.



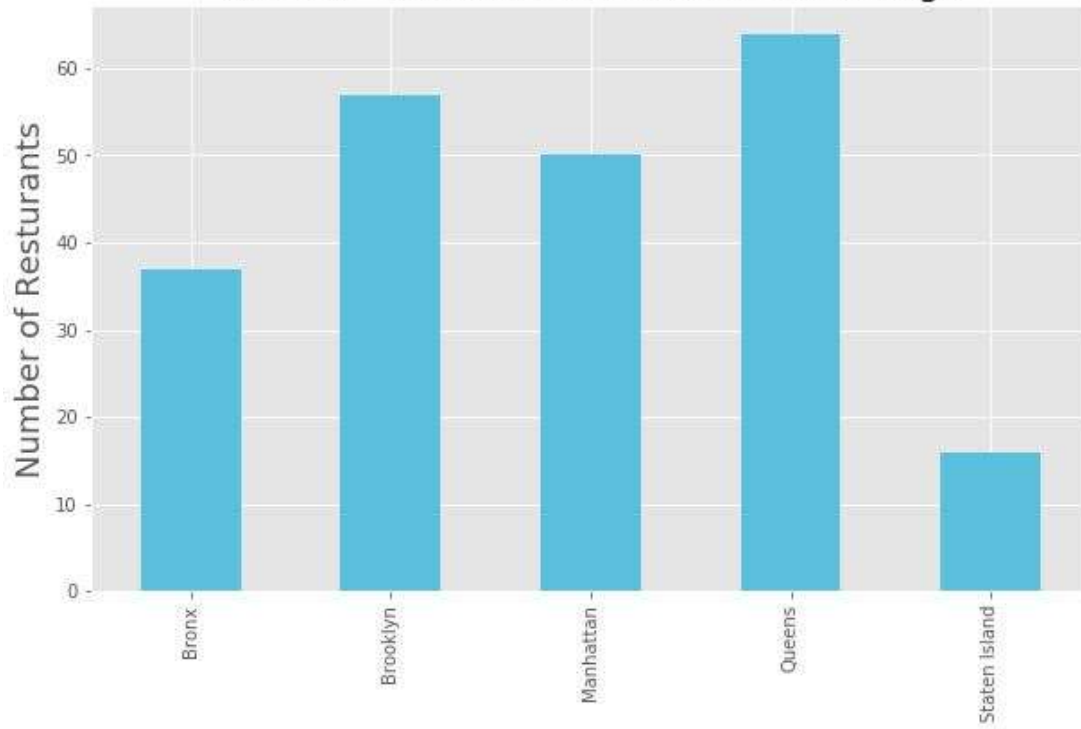
3.4 Bar Plot

With the list of restaurants with different category, different bar plot for East Asia restaurants in each New York City borough can be visualized by bar plot. Japanese restaurants in each NYC borough was generated in the project and shown below.

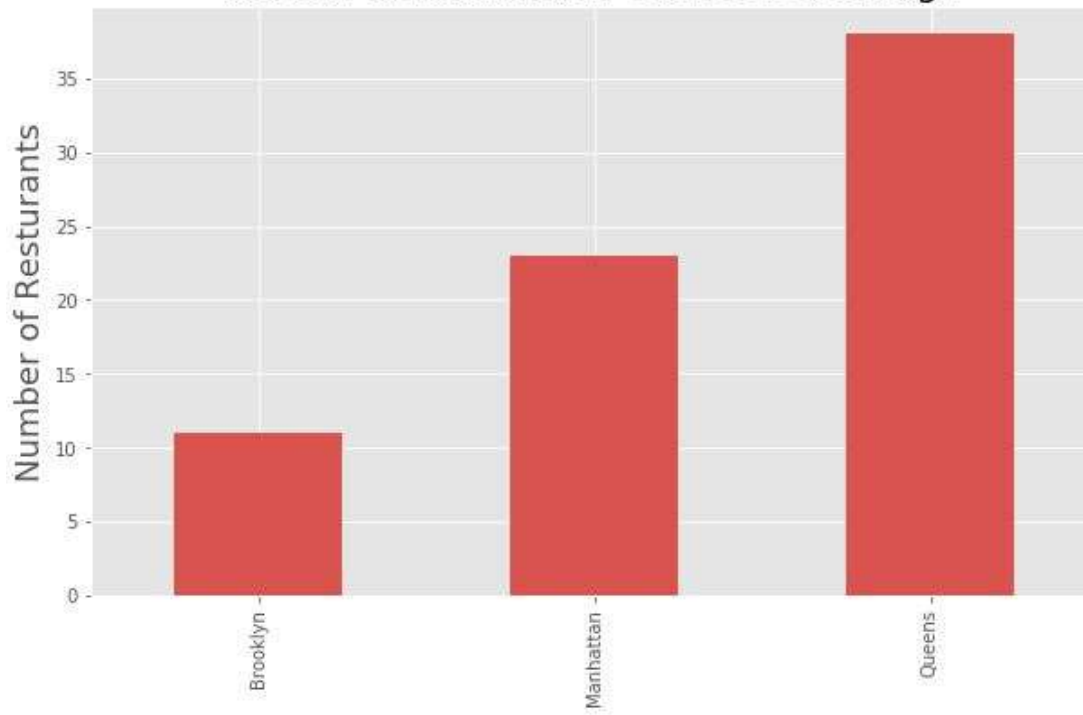


Bar plots with Chinese and Korean restaurants can be generated with the same way, they are showing the distributions of restaurants in each borough. This analysis the competitors in different area and helps the project to make the location recommendation. These two bar plots are shown below.

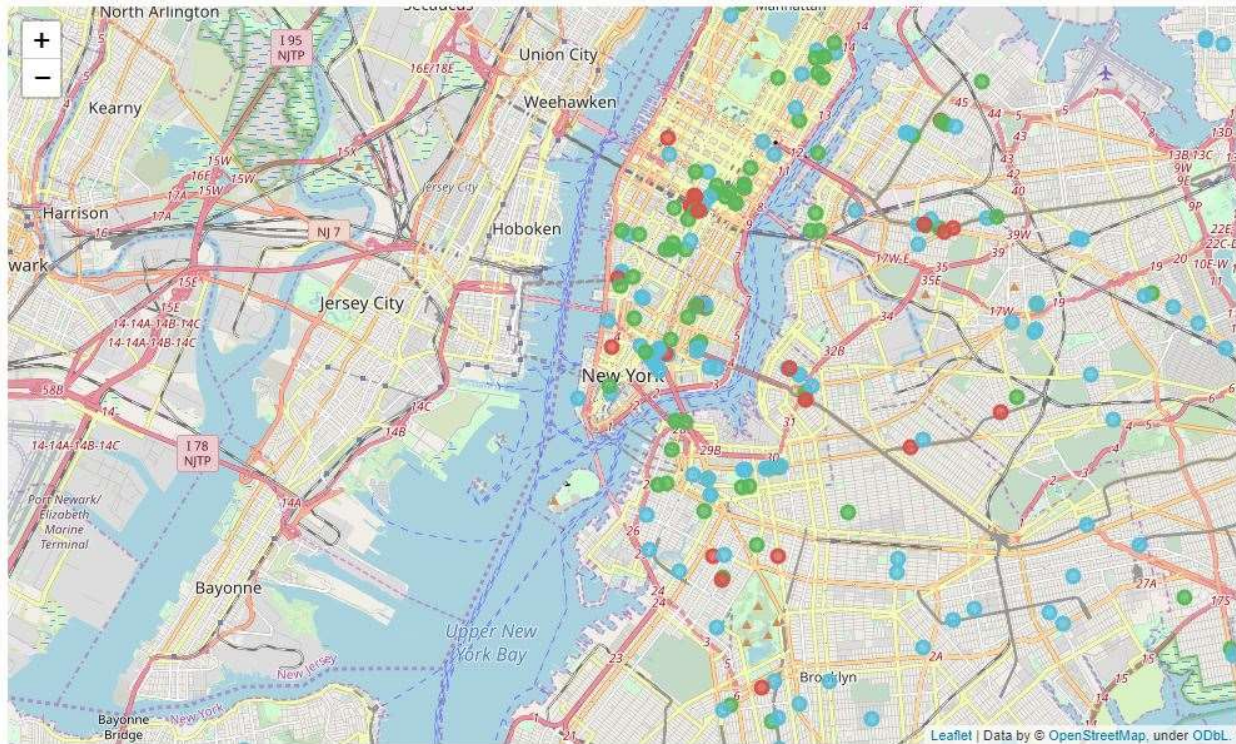
Chinese Restaurant in each NYC Borough



Korean Restaurant in each NYC Borough

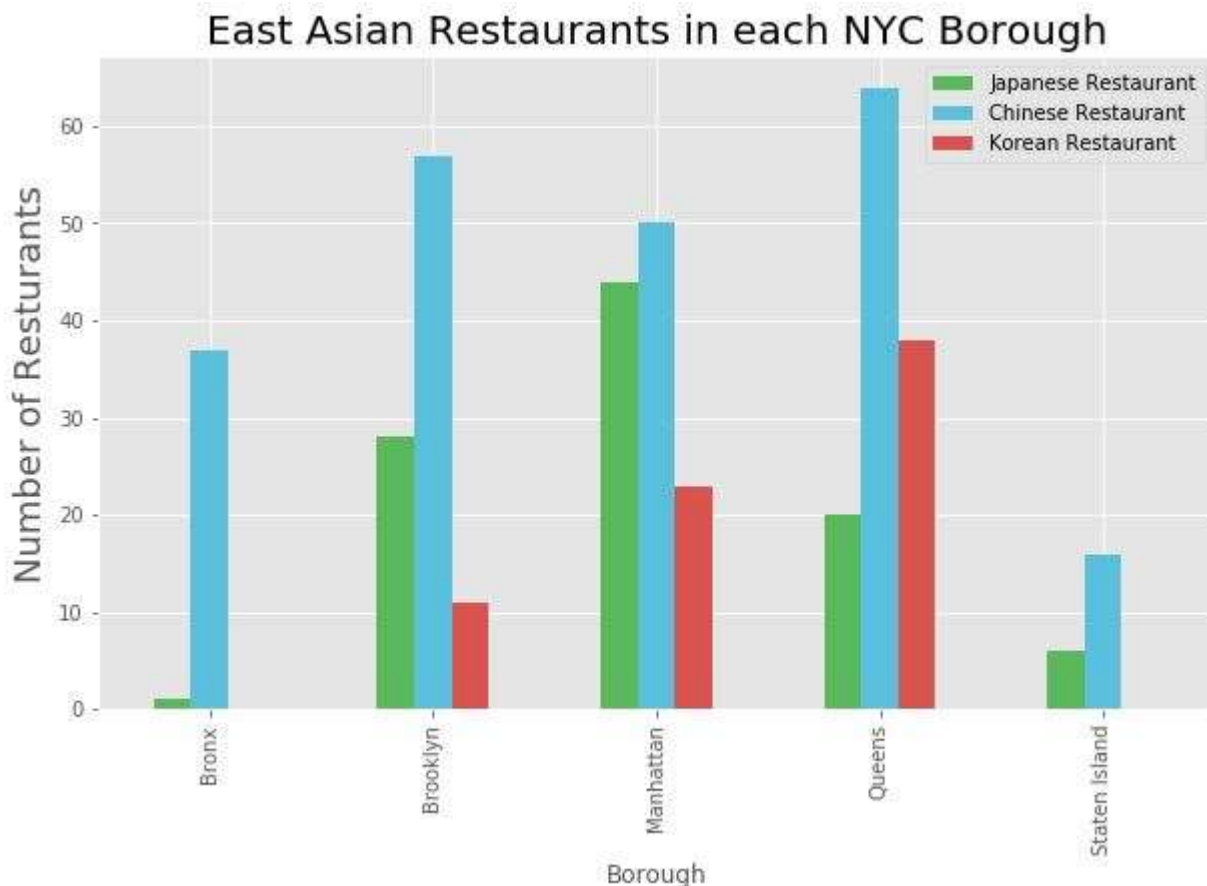


3.5 Aggregated Map and Bar Plot



Based on information from different restaurants data at hand, we can generate the map of New York City with all East Asia restaurants using the same method. The restaurants are superimposed in different colors at the top of the map to indicate different restaurants categories. The aggregated map is shown above.

Based on information from different restaurants data listed in the neighborhood dataset, we can generate the bar plot with all East Asia restaurants using the same method. The restaurants are visualized in different colors shown in the legend to indicate different restaurants categories. The aggregated bar plot is shown below.

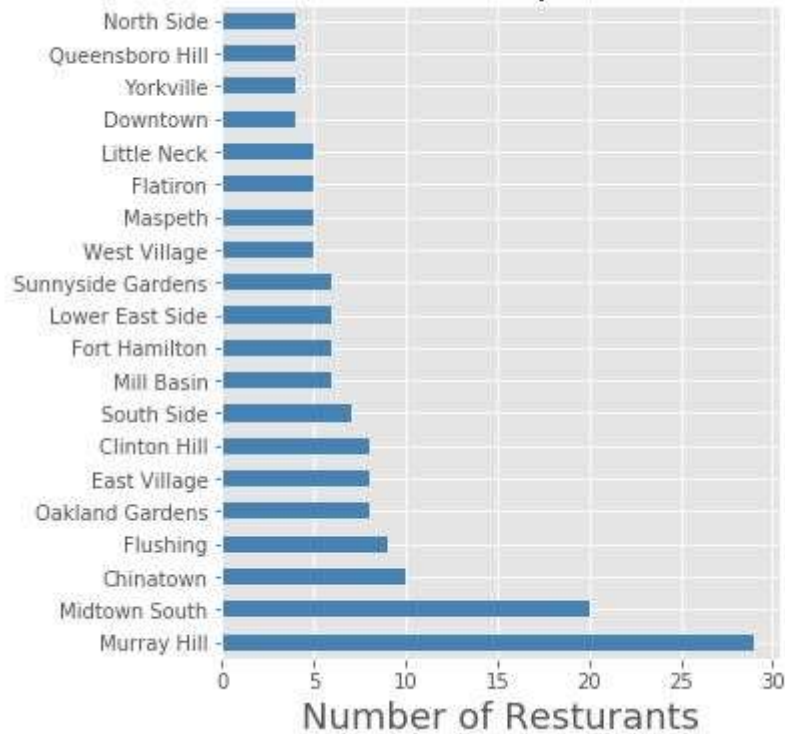


3.6 Rank neighborhood by restaurants quantity

A horizontal bar plot is shown below, indicates the top 20 New York City neighborhood containing a greater number of East Asian restaurants.

Assume that our algorithm is to find the neighborhood with the greatest quantity of East Asia restaurants, then the recommendation turns to be Murray Hill, because there are more than 25 restaurants offering similar food type in the neighborhood. The common type of restaurants will attract the most population who prefer to dine out.

East Asian Resturants in Top 20 NYC Neighborhood



3.7 Get restaurants rating

Using the Foursquare API again to get the details of each restaurant including the number of Likes and Rating. The corresponding columns were created already, the Null values will be changed to exact numbers from the result of Foursquare request.

The dataset “restaurants” is filled out completely, some of the contents are shown below. All the restaurants with no rating are dropped, for numeric analysis propose.

```
restaurants.head()
```

	Borough	Neighborhood	Category	Name	Likes	Rating	ID	Latitude	Longitude
5	Bronx	Pelham Parkway	Chinese Restaurant	Mr. Q's Chinese Restaurant	9.0	7.4	4b9d6b45f964a52078ab36e3	40.855790	-73.855455
6	Bronx	Bedford Park	Chinese Restaurant	Choi Yuan - Chinese Restaurant	6.0	6.9	4e2e08021838f1c552b6b8eb	40.873078	-73.889086
7	Bronx	Bedford Park	Chinese Restaurant	Hung Hing Chinese Restaurant	5.0	5.9	4e4d0387bd413c4cc66dfd72	40.871181	-73.886759
11	Bronx	Fordham	Chinese Restaurant	Panda Express	12.0	7.6	566f33e7498e44c2501bda81	40.863001	-73.900894
20	Bronx	Parkchester	Chinese Restaurant	Sabor Latino Seafood Restaurant	6.0	7.1	4b89b62df964a520ff4c32e3	40.836366	-73.853034

3.8 Rank neighborhood by average rating

If our strategy is changed to find the neighborhood with the greatest average ratings of East Asia restaurants, then the recommendation turns to be Windsor Terrace, because the average user rating is more than 8 in this neighborhood. The better service location of restaurants will attract the most population who prefer to dine out.

East Asian Restaurants Rating in Top 20 NYC Neighborhood



A horizontal bar plot is shown above, indicates the top 20 New York City neighborhood containing a greater average rating of East Asian restaurants.

4. Results

The result recommendation is Murray Hill if we assume our strategy is to find the neighborhood with the most similar competitors. There are more than 25 restaurants serving East Asian food in this neighborhood.

The result recommendation is Windsor Terrace if our algorithm is changed to find neighborhood with the greatest average ratings of East Asian restaurants. The neighborhood has an average user rating of more than 8.

5. Discussion

New York City has 306 different neighborhoods and about 395 different restaurants serving East Asia cuisine. Considering the details of each restaurant, the data is huge and difficult to analysis. We fetch data from different data source and apply different algorithms. The data is easily stored in the Python data structure and has greatly assisted our analysis.

The observation I noticed during the project was that different business logic or conditions/algorithms can make the result completely different.

6. Conclusion

We can use Python to import data from different data source and then process the data in preparation for analysis. Different conditions can then be applied, and the result determined using different algorithms.

The Foursquare API can be easily used to get important local venue information to help us with our analysis. It is difficult to obtain detailed venue data in real life. This concludes the report.