

**The Battle of Neighborhoods: Coursera Capstone Project**  
**Opening a new Japanese Restaurant in Manhattan and Queens,**  
**New York City.**

**Tongxi Zhao**  
**April 30<sup>th</sup>, 2021**

# 1. Introduction

New York city is the most populous city in the United States, the core of the New York metropolitan area, and one of the largest cities in the world. It is an international metropolis with great influence on the global economy, commerce, finance, media, politics, education, and entertainment. New York City consists of five boroughs: Manhattan, Queens, Brooklyn, Bronx, and Staten Island, each of which is a county of the state of New York.

New York is also the most densely populated major city in the United States. In 2019, there were 8.419 million people living in the New York city according to United States Census Bureau, Eurostat<sup>1</sup>.

New York city is the headquarters of the United Nations, so it is also considered the center of world diplomacy. New York city is also known as the “cultural capital of the United States”. People describe the New York City as being diversity, including diverse people coming from the everywhere across the world, diverse languages, diverse food culture etc. According to 2010 Census<sup>2</sup>, the New York City has become the home to more than one million Asian Americans, it contains the highest total Asian population.

In addition, according to the 2000 census, over half of the 37,279 people of Japanese ancestry in the U.S. state of New York lived in New York City. As of 2012, the New York City metropolitan area was home to the largest Japanese community on the East Coast of the United States<sup>3</sup>. After Italian, Chinese and Mexican, Japanese food is probably the most popular ethnic cuisine in the United States. Based on these real facts, to open a new Japanese restaurant is becoming the planned business plan in New York City, and this final project explores the possible neighborhoods to start new Japanese restaurants throughout the Queens and Manhattan, because these two boroughs are having taste food from around the globe, people are more likely to explore delicacy here.

## 1.1 Business Problem

I will be leveraging the Foursquare location data to find numbers of popular or people favored Japanese restaurants for each neighborhood in Queens and Manhattan, I will also look at the Japanese restaurants’ frequency in venue category for each

---

<sup>1</sup> [https://datacommons.org/place/geoId/3651000?utm\\_medium=explore&mprop=count&popt=Person&hl=en](https://datacommons.org/place/geoId/3651000?utm_medium=explore&mprop=count&popt=Person&hl=en)

<sup>2</sup> [http://www.ameredia.com/resources/demographics/asian\\_american.html](http://www.ameredia.com/resources/demographics/asian_american.html)

<sup>3</sup> [https://books.google.com/books?id=O\\_9B1qXjohsC&pg=PA105#v=onepage&q&f=false](https://books.google.com/books?id=O_9B1qXjohsC&pg=PA105#v=onepage&q&f=false)

neighborhood in Queens and Manhattan. The project assumes that the neighborhood with larger number of favored Japanese restaurants are not optimal choice for opening a new Japanese restaurant because the competition in these neighborhoods could be fierce. The assumption is very naïve without considering other dimensions, such as demographics, geographic preferences, traffic routes, and so on due to the limitation of data availability and time. To improve the results, I will be implementing K-means clusters model to have a cluster on types of venue category with string 'Restaurant', the cluster results will indicate the similar neighborhoods and ranking of types of food in these neighborhoods. By combining the numbers of favored Japanese Restaurants and clustered restaurants, I will suggest the possible neighborhoods to open a new Japanese restaurant.

## 2. Data:

### 2.1 Data Downloaded from:

[https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork\\_data.json](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json)

It is a New York city data with city features, the features in *json* file as Python dictionary key provide a list of neighborhoods. To extract other desired dictionary keys, I can create a *Pandas Dataframe* with columns including boroughs in New York Cities, neighborhoods in each boroughs and geographic coordinates for each neighborhood.

### 2.2 Japanese restaurants in Queens and Manhattan's Neighborhoods of New York City

Source: Foursquare API

By using this API, I will get all the venues in the Queens and Manhattan's neighborhood. I can filter these venues to get Japanese restaurants.

### 2.3 All type of restaurants in in Queens and Manhattan's neighborhoods of New York City.

Source: Foursquare API

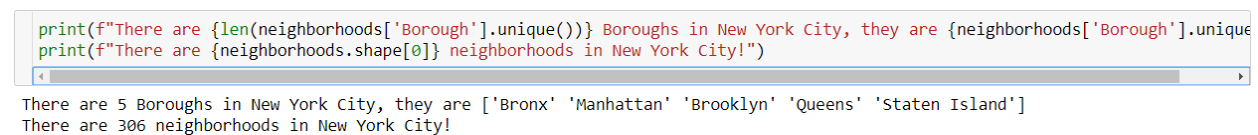
By using this API, I will get all the venues in the Queens and Manhattan's neighborhood. I can filter these venues to get all type of restaurants.

Given the geographic coordinates of each neighborhood, the project explores the popular venues in each neighborhood using Foursquares API. The venues information dataset will be leveraged to analyze the business problem.

### 3. Methodology

#### 3.1 Group by neighborhood and check the mean of the frequency of occurrence of Japanese restaurant.

The first methodology to suggest a location to open a Japanese restaurant is a basic statistical group method which is obtaining the mean value of the occurrence of Japanese restaurant in each neighborhood. The assumption is that a neighborhood with many Japanese restaurants is much more competitive, it will be more business pressure if a new restaurant open in the neighborhood.



```
print(f"There are {len(neighborhoods['Borough'].unique())} Boroughs in New York City, they are {neighborhoods['Borough'].unique()}")
print(f"There are {neighborhoods.shape[0]} neighborhoods in New York City!")
```

There are 5 Boroughs in New York City, they are ['Bronx' 'Manhattan' 'Brooklyn' 'Queens' 'Staten Island']  
There are 306 neighborhoods in New York City!

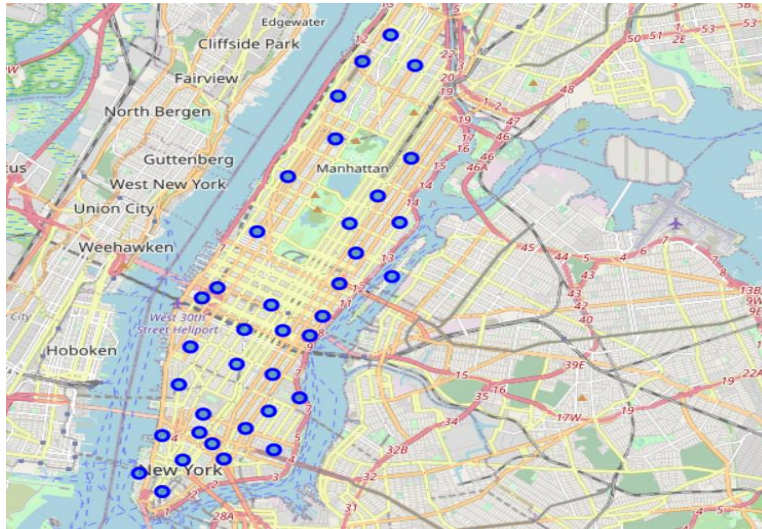
There are 5 boroughs in New York City, they are Bronx, Manhattan, Brooklyn, Queens, Staten Island. And there are 306 neighborhoods in New York City. As mentioned in the introduction, I only look at the locations in Queens and Manhattan, because these two boroughs are having taste food from around the globe, people are more likely to explore delicacy here.

##### 3.1.2 Start with Manhattan.

I start with the neighborhoods in Manhattan, there are 40 neighborhoods in Manhattan Boroughs. The neighborhoods are distributed as the below map.

```
print(f"There are {len(manhattan_data['Neighborhood'].unique())} Neighborhoods in Manhattan Boroughs")
```

There are 40 Neighborhoods in Manhattan Boroughs



I rely on the API to call the top 100 venues with a radius 500 meters for each neighborhood, the information obtained from API call are reliable since the venues are rated or commented by local people or residents or tourists.

```
print(f"There are {manhattan_venues.shape[0]} top venues within neighborhoods in Manhattan")
manhattan_venues.head()
```

There are 3225 top venues within neighborhoods in Manhattan

There are 3,225 top venues data points available within a radius 500 meters for each neighborhood in Manhattan, it means that for some neighborhoods, the API does not collect enough numbers of top 100 venues. Then I check how many venues were returned for each neighborhood and find out how many unique categories can be curated from all the returned venues. There is total 329 unique venue category in Manhattan.

```
print('There are {} uniques categories.'.format(len(manhattan_venues['Venue Category'].unique())))
```

There are 329 uniques categories.

I analyze each neighborhood and group rows by neighborhood and take the mean of the frequency of occurrence of each venue category, the new data frame is obviously having 40 rows for different neighborhood and 329 columns for unique venue categories.

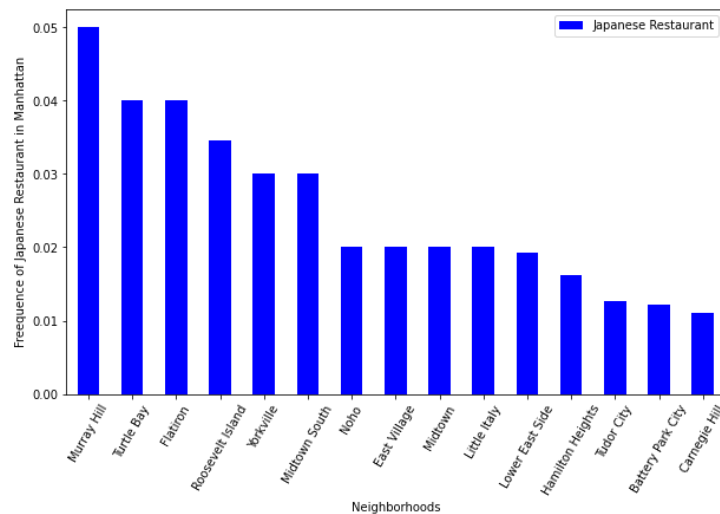
I filter the Japanese restaurant as venue category for each neighborhood and display the frequency of occurrence of Japanese restaurant as venue category in each

neighborhood, and sort it by descending order. The table below on the left shows the top 15 neighborhoods that having Japanese restaurants that people are more likely to visit.

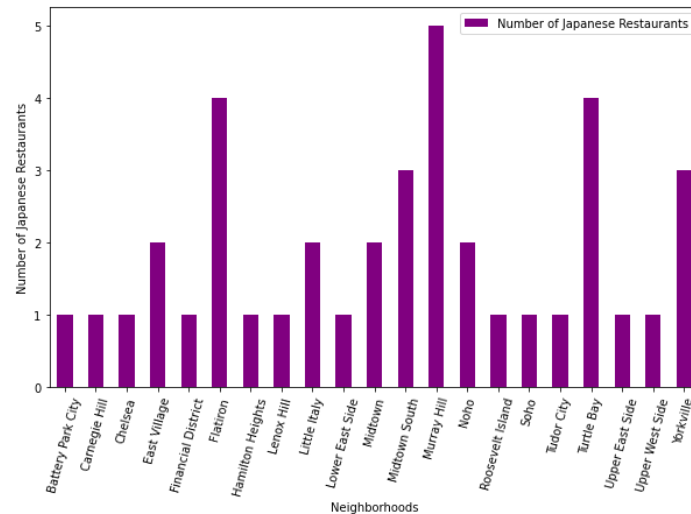
Neighborhood	Japanese Restaurant
Murray Hill	0.050000
Turtle Bay	0.040000
Flatiron	0.040000
Roosevelt Island	0.034483
Yorkville	0.030000
Midtown South	0.030000
Noho	0.020000
East Village	0.020000
Midtown	0.020000
Little Italy	0.020000
Lower East Side	0.019231
Hamilton Heights	0.016129
Tudor City	0.012658
Battery Park City	0.012195
Carnegie Hill	0.011111

Neighborhood	Japanese Restaurant
Tribeca	0.0
Sutton Place	0.0
Stuyvesant Town	0.0
Civic Center	0.0
Chinatown	0.0
Inwood	0.0
Clinton	0.0
Morningside Heights	0.0
East Harlem	0.0
Marble Hill	0.0
Manhattanville	0.0
Gramercy	0.0
Greenwich Village	0.0
Lincoln Square	0.0
Manhattan Valley	0.0

Japanese restaurants within the neighborhoods including Murray Hill, Turtle Bay, Flatiron, Roosevelt Island, Yorkville, Midtown South are popular, thus the new Japanese restaurant should not be opened within these neighborhoods to get rid of fierce competition. The above table on the right shows the neighborhoods having very low frequency of Japanese restaurants visits, or possibly there are no Japanese restaurants in these neighborhoods. If combining the Manhattan map, I suggest to open restaurants in neighborhoods **Lincoln Square** and **Manhattan Valley**, both neighborhoods are on the northwestward of the central park and close to the central park.



The above bar chart shows the neighborhoods with some popular Japanese restaurants, the new openings should be away from these neighborhoods or be away from the adjacent neighborhoods.



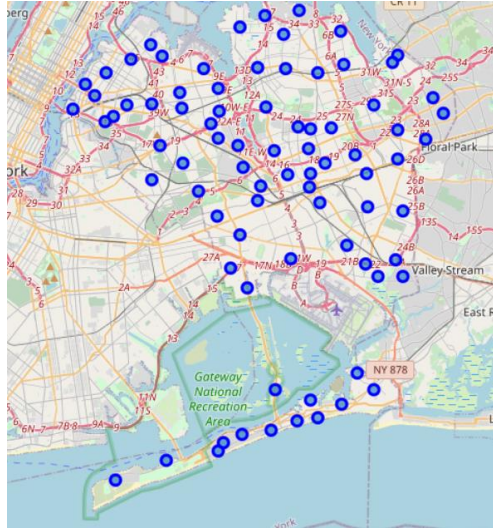
The purple above bar figure shows the number of Japanese restaurants in all neighborhoods, 21 out of 40 neighborhoods in Manhattan has Japanese restaurants, Murray Hill has 5, Turtle Bay and Flatiron have 4.

### 3.1.3 Queens Borough

After analyzing the possible location for Manhattan, I will be looking at Queens. Queens has 81 neighborhoods, it's a relatively large number of neighborhoods.

```
print(f"There are {len(Queens_data['Neighborhood'].unique())} Neighborhoods in Queens Borough")
```

There are 81 Neighborhoods in Queens Borough



The above map shows the distribution of neighborhoods in Queens. The distribution is relatively sparse. Queens has 2,103 top venues within each neighborhood, on average, each neighborhood has about 25 popular venues to be recorded in Foursquares API.

```
print(f"There are {queens_venues.shape[0]} top venues within neighborhoods in Queens.")
queens_venues.head()
```

There are 2103 top venues within neighborhoods in Queens.

Then I check how many venues were returned for each neighborhood and find out how many unique categories can be curated from all the returned venues. There is total 275 unique venue category in Queens.

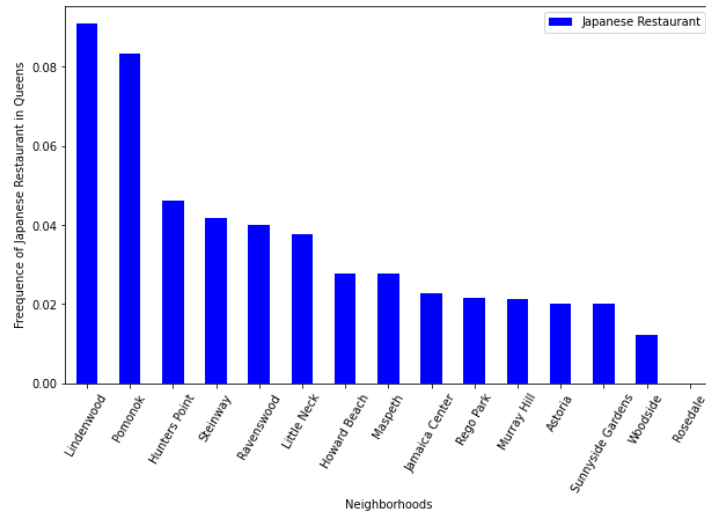
I analyze each neighborhood and group rows by neighborhood and take the mean of the frequency of occurrence of each venue category, the new data frame is obviously having 81 rows for different neighborhoods and 275 columns for unique venue category.



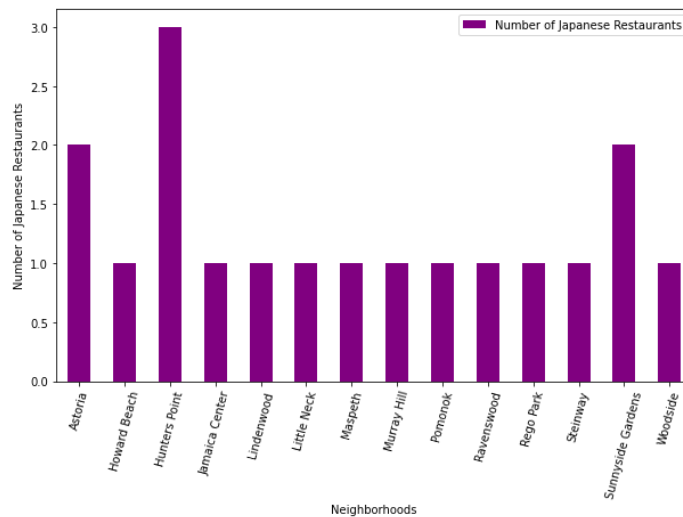
Japanese Restaurant		Neighborhood Japanese Restaurant	
Neighborhood		Neighborhood	Japanese Restaurant
Pomonok	0.100000	Jackson Heights	0.0
Lindenwood	0.090909	Holliswood	0.0
Hunters Point	0.045455	Hollis	0.0
Steinway	0.043478	Hillcrest	0.0
Ravenswood	0.040000	Hammels	0.0
Maspeth	0.029412	Glendale	0.0
Howard Beach	0.028571	Glen Oaks	0.0
Jamaica Center	0.022222	Fresh Meadows	0.0
Rego Park	0.021739	Forest Hills Gardens	0.0
Murray Hill	0.021277	Forest Hills	0.0
Astoria	0.020000	Flushing	0.0
Sunnyside Gardens	0.020000	Floral Park	0.0
Little Neck	0.018519	Far Rockaway	0.0
Woodside	0.012821	Elmhurst	0.0
Rosedale	0.000000	Jamaica Hills	0.0

I filter the Japanese restaurant as venue category for each neighborhood and display the frequency of occurrence of Japanese restaurant as venue category in each neighborhood, and sort it by descending order. The above table on the left shows the top 15 neighborhoods that having Japanese restaurants that people are more likely to visit.

Japanese restaurants within the neighborhoods including Pomonok, Lindenwood, Hunters Point, Steinway, Ravenswood are popular, thus the new Japanese restaurant should not be opened within these neighborhoods to get rid of fierce competition. The above table on the right shows the neighborhoods having very low frequency of Japanese restaurants visits, or possibly there are no Japanese restaurants in these neighborhoods. If combining the Queens map, I suggest to open restaurants in neighborhoods **Jackson Height, Forest Hills, Flushing and Elmhurst**, these neighborhoods are close to the center of Queens and easy to visit after existing the freeways.



The above bar chart shows the neighborhoods with some popular Japanese restaurants, the new opening should be away from these neighborhoods or be away from the adjacent neighborhoods.



The purple above bar figure shows the number of Japanese restaurants in all neighborhoods, 14 out of 81 neighborhoods in Queens has Japanese restaurants, Hunting Points has 3, Astoria and Sunnyside Gardens have 2.

To wrap up, based on the analysis, the **Lincoln Square** and **Manhattan Valley** are possible neighborhoods to open new Japanese restaurant if considering the competition, location, and surrounding point of interests in Manhattan. **Jackson Height, Forest Hills, Flushing, and Elmhurst** are possible best neighborhoods in Queens to open new Japanese restaurants, since they are in the center of Queens, and the location is conveniently accessible.

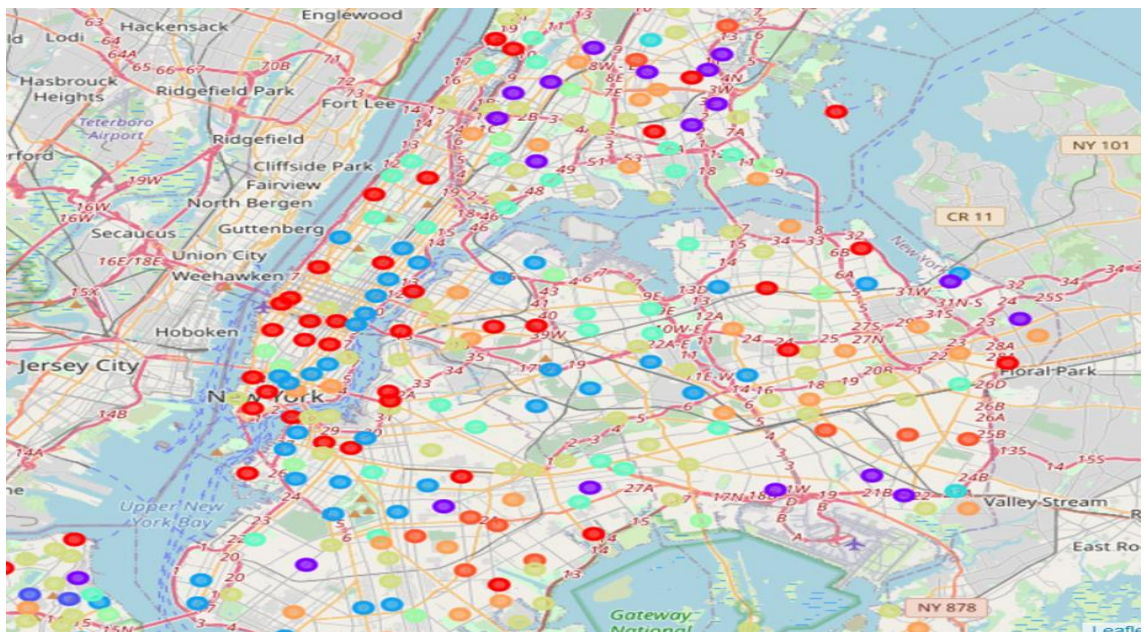
### 3.2 K-means Cluster Analysis

The last methodology I implement is to see all the restaurants in the New York City, and cluster them into 5 groups, to check their locations and rankings of favorable restaurants in each neighborhood.

There are 10,106 popular venues in the New York City, and 437 unique venue categories. Then I find out the columns' names having string "restaurant" to filter out all type of restaurants, the total types of restaurants in the New York City are 93, for examples, Afghan Restaurants, African Restaurant, American Restaurants, Asian Restaurant and so on.

	Neighborhood	1st Most Common Type Restaurant	2nd Most Common Type Restaurant	3rd Most Common Type Restaurant	4th Most Common Type Restaurant	5th Most Common Type Restaurant	6th Most Common Type Restaurant	7th Most Common Type Restaurant	8th Most Common Type Restaurant	9th Most Common Type Restaurant	10th Most Common Type Restaurant
0	Allerton	Chinese Restaurant	Fast Food Restaurant	Spanish Restaurant	Scandinavian Restaurant	Russian Restaurant	Romanian Restaurant	Restaurant	Ramen Restaurant	Puerto Rican Restaurant	Polish Restaurant
1	Annadale	Restaurant	Sushi Restaurant	Afghan Restaurant	Paella Restaurant	Romanian Restaurant	Ramen Restaurant	Puerto Rican Restaurant	Polish Restaurant	Peruvian Restaurant	Persian Restaurant
2	Arden Heights	Afghan Restaurant	Paella Restaurant	Russian Restaurant	Romanian Restaurant	Restaurant	Ramen Restaurant	Puerto Rican Restaurant	Polish Restaurant	Peruvian Restaurant	Persian Restaurant
3	Arlington	American Restaurant	Afghan Restaurant	Paella Restaurant	Russian Restaurant	Romanian Restaurant	Restaurant	Ramen Restaurant	Puerto Rican Restaurant	Polish Restaurant	Peruvian Restaurant
4	Arrochar	Italian Restaurant	Mediterranean Restaurant	Middle Eastern Restaurant	Polish Restaurant	Afghan Restaurant	Pakistani Restaurant	Russian Restaurant	Romanian Restaurant	Restaurant	Ramen Restaurant

Then I have a data frame showing top 10 most common type restaurants in each neighborhood. The Japanese restaurant is not very popular in each neighborhood compared with other types.



I implement a K-means cluster model to segmenting the neighborhoods having similar favored type restaurants, the K is set up to 10. Then I plot the clustered results on the map. The red dot represents cluster 0, neighborhoods in cluster 0's first most common type of restaurant is American food. The purple dot represent cluster 1, neighborhoods in cluster 1's first most common type of restaurant is fast food. Only 4 neighborhoods in cluster 2, it is ambiguous to say what type of foods they prefer.

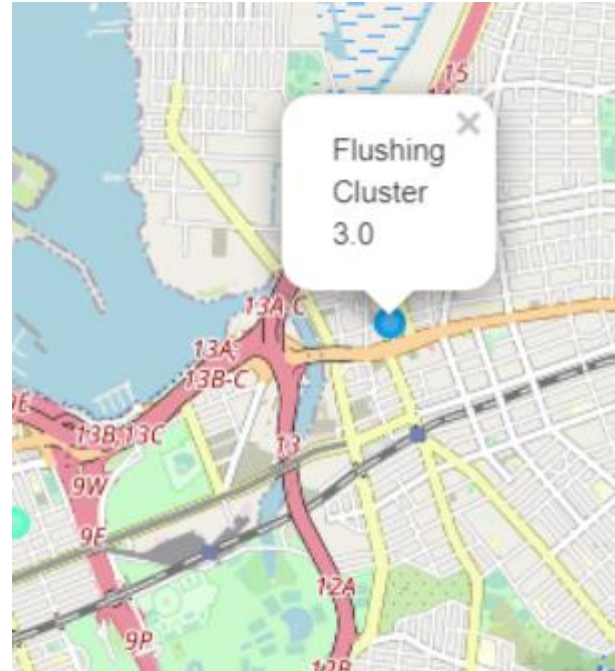
The below table shows the partial results of cluster 3, in the map, the blue, based on the common sense, cluster 3 is much closer to Japanese style food, because Chinese, Asian, Indian, Thai, Sushi, Ramen restaurants are frequently showed up in the table. Opening a Japanese restaurants in these neighborhoods may attract more potential customers, because their taste style tends to be more Japanese or Asian.

	Neighborhood	1st Most Common Type Restaurant	2nd Most Common Type Restaurant	3rd Most Common Type Restaurant	4th Most Common Type Restaurant	5th Most Common Type Restaurant	6th Most Common Type Restaurant	7th Most Common Type Restaurant	8th Most Common Type Restaurant	9th Most Common Type Restaurant	10th Most Common Type Restaurant
46	Bay Ridge	Italian Restaurant	American Restaurant	Greek Restaurant	Chinese Restaurant	Sushi Restaurant	Thai Restaurant	Dim Sum Restaurant	Indian Restaurant	Caucasian Restaurant	Mexican Restaurant
51	Brighton Beach	Russian Restaurant	Eastern European Restaurant	Restaurant	Sushi Restaurant	Seafood Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Persian Restaurant	Romanian Restaurant	Ramen Restaurant
53	Manhattan Terrace	Eastern European Restaurant	Japanese Restaurant	Sushi Restaurant	Afghan Restaurant	Pakistani Restaurant	Romanian Restaurant	Restaurant	Ramen Restaurant	Puerto Rican Restaurant	Polish Restaurant
55	Crown Heights	Sushi Restaurant	Afghan Restaurant	Paella Restaurant	Romanian Restaurant	Restaurant	Ramen Restaurant	Puerto Rican Restaurant	Polish Restaurant	Peruvian Restaurant	Persian Restaurant
58	Windsor Terrace	Chinese Restaurant	American Restaurant	Middle Eastern Restaurant	Italian Restaurant	Sushi Restaurant	Afghan Restaurant	Russian Restaurant	Romanian Restaurant	Restaurant	Ramen Restaurant

Combining with the previous numbers of restaurants and frequency of occurrence analysis, I conclude that best location for a Japanese restaurant in Manhattan is the **neighborhoods between Lincoln Square and Manhattan Valley**, because upper west neighborhood is included in the cluster 3 and is in the middle of Lincoln Square and Manhattan Valley.

In Queens, the best neighborhoods may be Flushing, in cluster 3 and not too many Japanese restaurant there.





## 4. Results

To wrap up, based on my analysis, if we only look at the numbers of restaurants or the frequency of occurrence of venue category, the Lincoln Square and Manhattan Valley are possible neighborhoods to open new Japanese restaurant if considering the competition, location, and surrounding point of interests in Manhattan. Jackson Height, Forest Hills, Flushing, and Elmhurst are possible best neighborhoods in Queens to open new Japanese restaurants, since they are in the center of queens, and the location is conveniently accessible.

In addition, if we refer to the K-means cluster results, the best place to open Japanese restaurant in Manhattan is still the neighborhoods between Lincoln Square and Manhattan Valley. The best neighborhood in Queens is flushing. The cluster analysis provides some further analysis on how much people like a certain type of food in a given neighborhood.

## 5. Discussion

This analysis is subject to a number of shortcomings, in future I suggest adding some other social economic data, demographic data in each neighborhood, social

economic data like income, education may affect people's choice on food types. The population in each neighborhood may affect the business profits. Other factor, such as, whether neighborhood is a business-based region or not, traffic conditions, public transportation may affect the results.

## **6. Conclusion**

The project explores the possible neighborhoods to start new Japanese restaurants throughout the Queens and Manhattan, because these two boroughs are having taste food from around the globe, people are more likely to explore delicacy here. The project uses the Foursquare location data to find numbers of popular or people favored Japanese restaurants for each neighborhood in Queens and Manhattan, looks at the Japanese restaurants' frequency in venue category for each neighborhood in Queens and Manhattan. The project assumes that the neighborhood with larger number of favored Japanese restaurants are not optimal choice for opening a new Japanese restaurant because the competition in these neighborhoods could be fierce. The project implements K-means clusters model to have a cluster analysis on types of venue category with string 'Restaurant', the cluster results will indicate the similar neighborhoods and ranking of types of food in these neighborhoods. By combining the numbers of favored Japanese Restaurants and clustered restaurants, I will suggest the possible neighborhoods to open a new Japanese restaurant are the neighborhoods between Lincoln Square and Manhattan Valley in Manhattan, Flushing in Queens.