Selecting Locations for a Sushi Restaurant in New York City with Foursquare API

Applied Data Science Capstone Coursera Course Week 5 Peer-Graded Project Report

By Brandon Gallemore February, 2021

Table of Contents

1. Introduction

1.1 New York City

New York City is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km^2), New York City is also the most densely populated major city in the United States. It comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that's among the world's major commercial, financial and cultural centers. Its iconic sites include skyscrapers such as the Empire State Building and sprawling Central Park. It is multicultural and provides numerous business opportunities. It is a global hub of business and commerce. The city is a major center for a multitude of business, trade, media, and the arts. This also means that the market is highly competitive. As it is a highly developed city the cost of doing business is also great. Thus, any new business venture or expansion needs to be analyzed carefully. Insights derived from analysis will improve understanding of the business environment and risks which will aid in strategically targeting the market. This will help reduce risk and ensure return on investment will be reasonable.

1.2 Business Problem

New York City is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history. Sushi restaurants are popular in the United States. However, given the risk of pursuing a food industry restaurant care must be taken. Interested parties in this project would be any who would be pursuing opening a Japanese or sushi restaurant in New York City.

1.3 Location Selection

Regions will be selected based on the number of Sushi restaurants in the area, demographics of the region, and average rating of sushi restaurants in the area. The Bronx followed by Manhattan and Queens are the most ethnically diverse of the boroughs. Queens and Brooklyn are the most populous. Given diversity and population density Queens is a candidate.

2. Data

2.1 Data Sources and Extraction Methods

To determine the best location(s) for the prospective sushi bar the resources from the Wikipedia site Demographics of New York City will be used. Web scraping techniques were used to get NYC population density and demographics data from Wikipedia. Preliminary finding indicates that Queens being the second most populous urban area in New York City (NYC), behind Brooklyn; and the most ethnically diverse urban area in NYC with the highest Asian ethnic minority population.

The second resource used will be Foursquare API to obtain venue data for neighbourhoods. The search will focus on Japanese and sushi restaurants to aid in solving the business problem. Also https://cocl.us/new_york_dataset was used. These datasets were used to explore various neighbourhoods and sushi restaurant venues in each neighbourhood.

2.2 Data Required

1. List of neighborhoods and boroughs in New York City
2. Latitude and longitude coordinates of neighborhoods and boroughs
3. Venue data focused on related to sushi restaurants

3. Methodology

3.1 Loading Data

After loading applicable libraries, web scraping techniques were used to get NYC population density and demographics data from the Wikipedia site Demographics of New York City using Python requests. This data includes population, GDP, and land area by boroughs and counties as well as ethnicity and race by jurisdiction.

Next a function was defined to interact with FourSquare API to obtain the top 100 venues within a radius of 1000 metres for a given latitude and longitude (Image 3.1.2). Next functions were developed for ranking venues. Another function was created to obtain borough and neighborhood data with longitudes and latitudes.

```
# Display head
sushi_rest_ny.head()
```

	Borough	Neighborhood	ID	Name
0	Bronx	Pelham Parkway	4c45c922f97fbe9a9e02bd30	Kiraku Hibachi & Sushi
1	Bronx	Country Club	4bbfc5a974a9a5935519cff6	Sushi Q
2	Bronx	Morris Park	4c45c922f97fbe9a9e02bd30	Kiraku Hibachi & Sushi
3	Bronx	Morris Park	4e25c5e0483bbcc48dc8531d	Namū Sushi
4	Bronx	North Riverdale	4baa9c97f964a52088793ae3	Tokyo House

Image 3.1.1 - Dataframe with Sushi Restaurants arranged by borough and neighborhood of New York City and example Python script.

3.2 Exploring Data

Once data was prepared, the population density and demographics dataframe scraped from Wikipedia showed that Brooklyn followed by Queens were the most populous boroughs and Bronx, Manhattan, and Queens were the most ethnically diverse (Image 3.2.1).

```
df = pd.DataFrame(jurisdictions, columns=["jurisdiction","%_white", "%_black_or_african_amercian", "%_Asian", "%_other", "%_mixed_race", "%_hispanic_latino_of_other_race"])
df.head()
```

	jurisdiction	%_white	%_black_or_african_amercian	%_Asian	%_other	%_mixed_race	%_hispanic_latino_of_other_race
0	Queens	44.1	20.0	17.6	12.3	6.1	25.0
1	Manhattan	54.4	17.4	9.4	14.7	4.1	27.2
2	Bronx	29.9	35.6	3.0	25.7	5.8	48.4
3	Staten Island	77.6	9.7	5.7	4.3	2.7	12.1
4	NYC Total	44.7	26.6	9.8	14.0	4.9	27.0

Image 3.2.1 - Demographics of New York City boroughs and preceding python script.

Using the data from FourSquare, the number of sushi restaurant venues per neighborhood was obtained. Bar Graphs were built using Python scripts to illustrate the number of sushi restaurants by borough (Image 3.2.2)

and then neighborhood (Image 3.2.3) showing over 50 restaurants in Manhattan and Brooklyn with the greatest density in the Sheepshead Bay neighborhood.

```
from matplotlib import pyplot as plt
plt.style.use('ggplot')

plt.figure(figsize=(9,5), dpi = 100)

# title
plt.title('Number of Sushi Restaurants in NYC by Borough')

#On x-axis
plt.xlabel('Borough', fontsize = 15)

#On y-axis
plt.ylabel('No.of Sushi Restaurants', fontsize=15)

#giving a bar plot
sushi_rest_ny.groupby('Borough')['ID'].count().plot(kind='bar')

#legend
plt.legend()

#displays the plot
plt.show()
```

Image 3.2.2 - Sushi restaurants by borough and preceding Python script.

```
from matplotlib import pyplot as plt
plt.style.use('ggplot')

plt.figure(figsize=(9,5), dpi = 100)

# title
plt.title('Number of Sushi Restaurants in NYC by Neighborhood')

#On x-axis
plt.xlabel('Neighborhood', fontsize = 15)

#On y-axis
plt.ylabel('No.of Sushi Restaurants', fontsize=15)

#giving a bar plot
sushi_rest_ny.groupby('Neighborhood')['ID'].count().nlargest(5).plot(kind='bar')

#legend
plt.legend()

#displays the plot
plt.show()
```

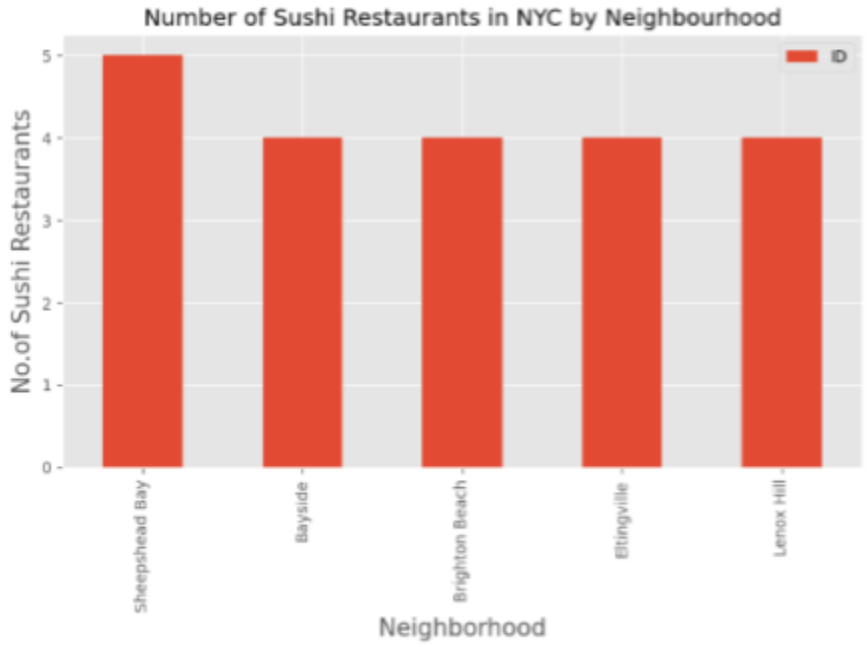

Image 3.2.3 - Sushi restaurants by neighborhood and preceding Python script.

Next lists of restaurants were prepared to explore venues by ‘Ratings’ and ‘Likes’. A dataframe and .csv file was created with venue, location, and ratings data and graphs were built displaying average ratings by borough and data indicated that the highest ratings for sushi restaurants were in Brooklyn followed by the Bronx with high ranking venues dispersed across neighborhoods. Finally, Folium was used to create a map of Brooklyn with 9 venues with a rating ≥8.5.

```
# Display head
sushi_rest_ny.head()
```

	Borough	Neighborhood	ID	Name
0	Bronx	Pelham Parkway	4c45c922f97fbe9a9e02bd30	Kiraku Hibachi & Sushi
1	Bronx	Country Club	4bbfc5a974a9a5935519cff6	Sushi Q
2	Bronx	Morris Park	4c45c922f97fbe9a9e02bd30	Kiraku Hibachi & Sushi
3	Bronx	Morris Park	4e25c5e0483bbcc48dc8531d	Namū Sushi
4	Bronx	North Riverdale	4baa9c97f964a52088793ae3	Tokyo House

Image 3.2.4 - Python script to display head of .csv file.

```
ny_neighborhood_stats
```

	Neighborhood	Average Rating
11	Boerum Hill	8.6
15	Brooklyn Heights	8.6
26	Downtown	8.6
76	North Side	8.9
79	Park Slope	8.6
81	Pelham Parkway	8.0
84	Prospect Heights	8.9
85	Prospect Lefferts Gardens	8.5
95	South Side	8.9
109	Williamsburg	8.9

Image 3.2.4 - Average rating by neighborhood and preceding Python script.

```
# Display 'Average Rating' in descending order with merged 'Latitude' and 'Longitude'
ny_neighborhood_stats.sort_values(['Average Rating'],ascending=False).head(10)
```

	Borough	Neighborhood	Latitude	Longitude	Average Rating
3	Brooklyn	North Side	40.714823	-73.958809	8.9
6	Brooklyn	Prospect Heights	40.676822	-73.964859	8.9
8	Brooklyn	South Side	40.710861	-73.958001	8.9
9	Brooklyn	Williamsburg	40.707144	-73.958115	8.9
0	Brooklyn	Boerum Hill	40.685683	-73.983748	8.6
1	Brooklyn	Brooklyn Heights	40.695864	-73.993782	8.6
2	Brooklyn	Downtown	40.690844	-73.983463	8.6
4	Brooklyn	Park Slope	40.672321	-73.977050	8.6
7	Brooklyn	Prospect Lefferts Gardens	40.658420	-73.954899	8.5
5	Bronx	Pelham Parkway	40.857413	-73.854756	8.0

Image 3.2.5 - Average rating by neighborhood with coordinates and preceding Python script.

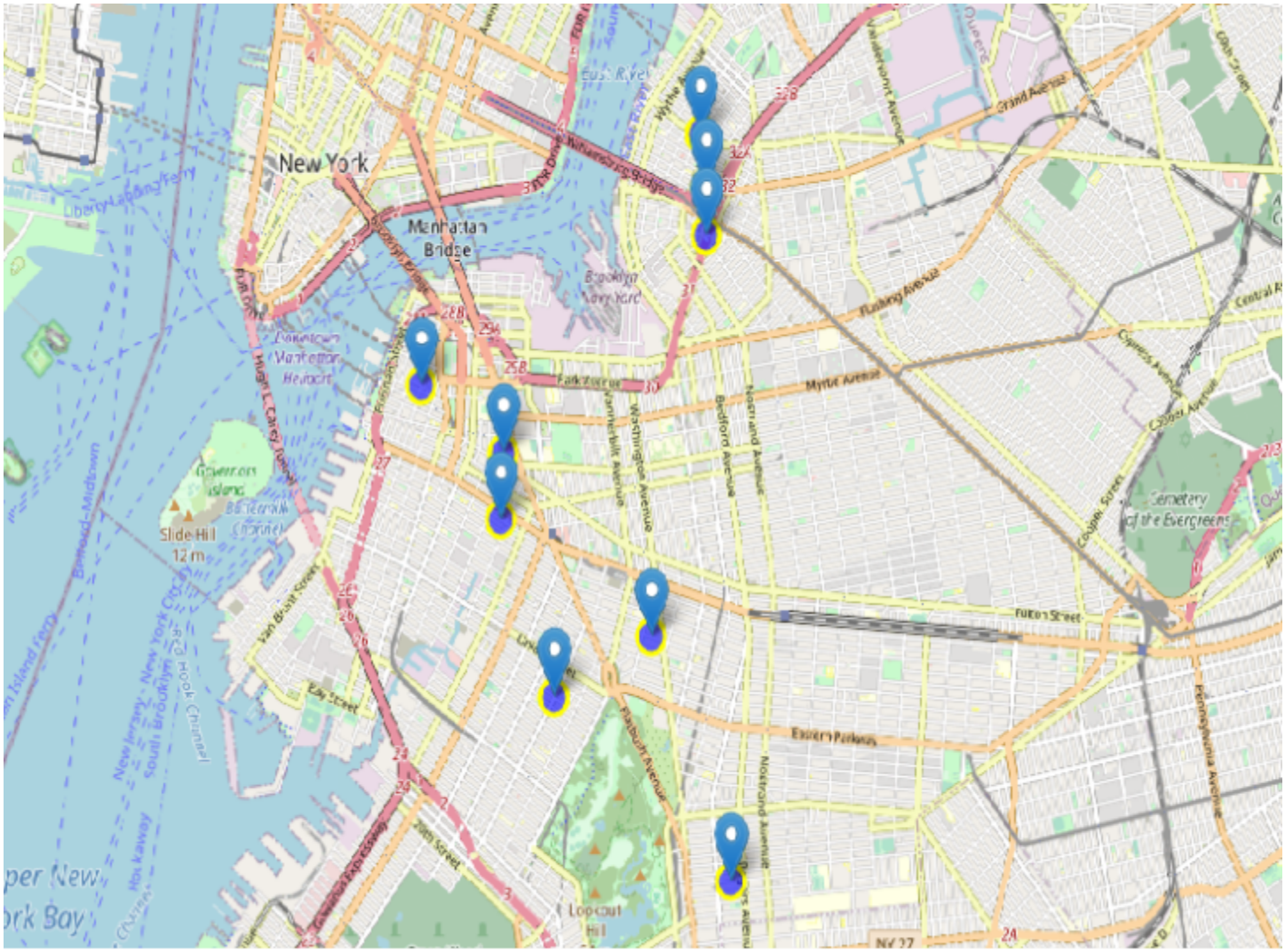

Image 3.2.6 - Map of Brooklyn with markers of 9 highest rated sushi restaurants.

4. Results

The analysis enabled the discovery and visual description of the following:

1. Neighborhoods and boroughs of New York City
2. Demographics and population density of New York City boroughs
3. Number and geographic coordinates of sushi restaurant venues in New York city by neighborhoods and boroughs
4. ‘Ratings’ and ‘Likes’ of sushi restaurant venues
5. Locations of highest rated sushi restaurant venues

9 of the top 10 sushi restaurants by ‘Rating’ were located in the Brooklyn borough which has the highest population density. However, it was shown that ethnic diversity was greater in other boroughs.

	jurisdiction	%_white	%_black_or_african_amercian	%_Asian	%_other	%_mixed_race	%_hispanic_latino_of_other_race
0	Queens	44.1	20.0	17.6	12.3	6.1	25.0
1	Manhattan	54.4	17.4	9.4	14.7	4.1	27.2
2	Bronx	29.9	35.6	3.0	25.7	5.8	48.4
3	Staten Island	77.6	9.7	5.7	4.3	2.7	12.1
4	NYC Total	44.7	26.6	9.8	14.0	4.9	27.0

Image 4.1.1 - Demographics of New York City boroughs.

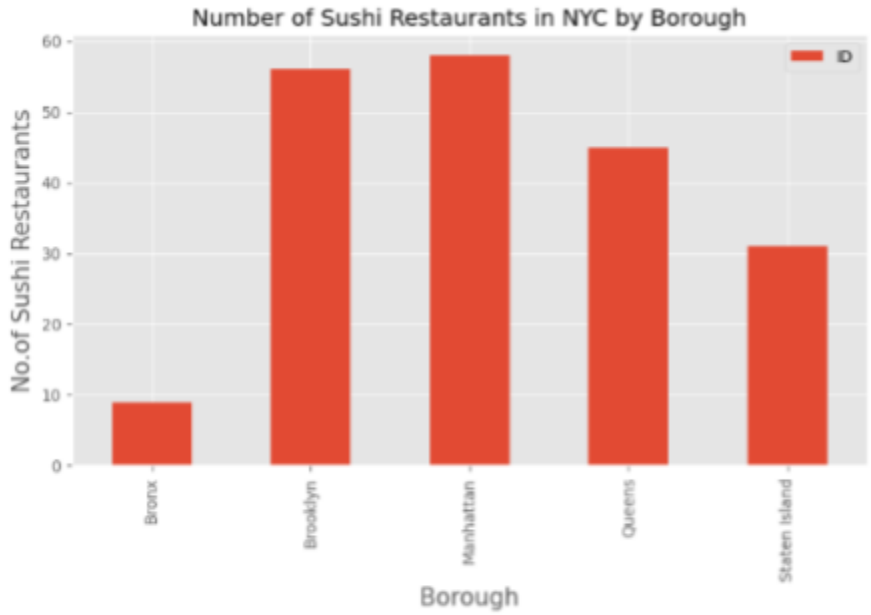

Image 4.1.2 - Sushi restaurants by borough.

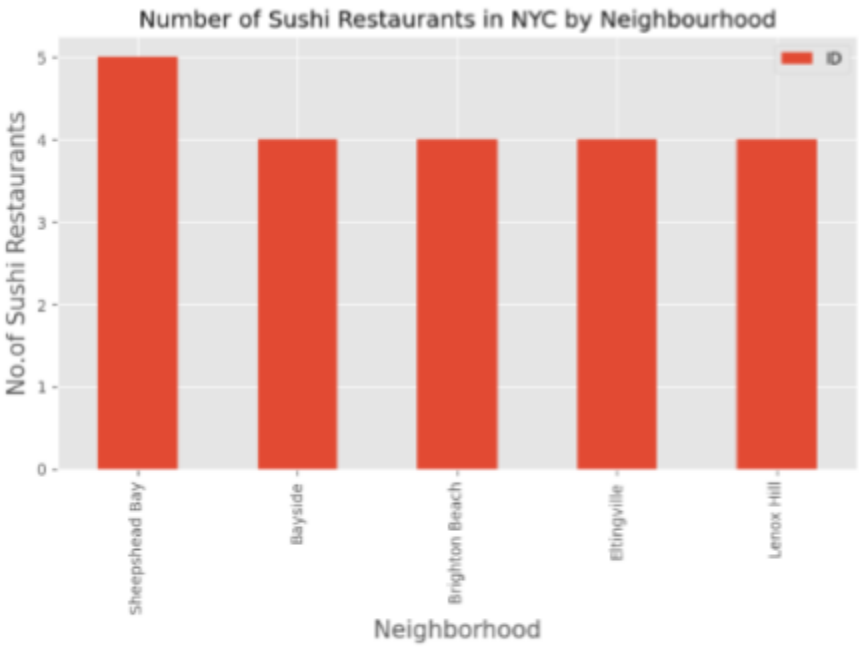

Image 4.1.3 - Sushi restaurants by neighborhood.

	Borough	Neighborhood	Latitude	Longitude	Average Rating
3	Brooklyn	North Side	40.714823	-73.958809	8.9
6	Brooklyn	Prospect Heights	40.676822	-73.964859	8.9
8	Brooklyn	South Side	40.710861	-73.958001	8.9
9	Brooklyn	Williamsburg	40.707144	-73.958115	8.9
0	Brooklyn	Boerum Hill	40.685683	-73.983748	8.6
1	Brooklyn	Brooklyn Heights	40.695864	-73.993782	8.6
2	Brooklyn	Downtown	40.690844	-73.983463	8.6
4	Brooklyn	Park Slope	40.672321	-73.977050	8.6
7	Brooklyn	Prospect Lefferts Gardens	40.658420	-73.954899	8.5
5	Bronx	Pelham Parkway	40.857413	-73.854756	8.0

Image 4.1.4 - Average rating by neighborhood with coordinates.

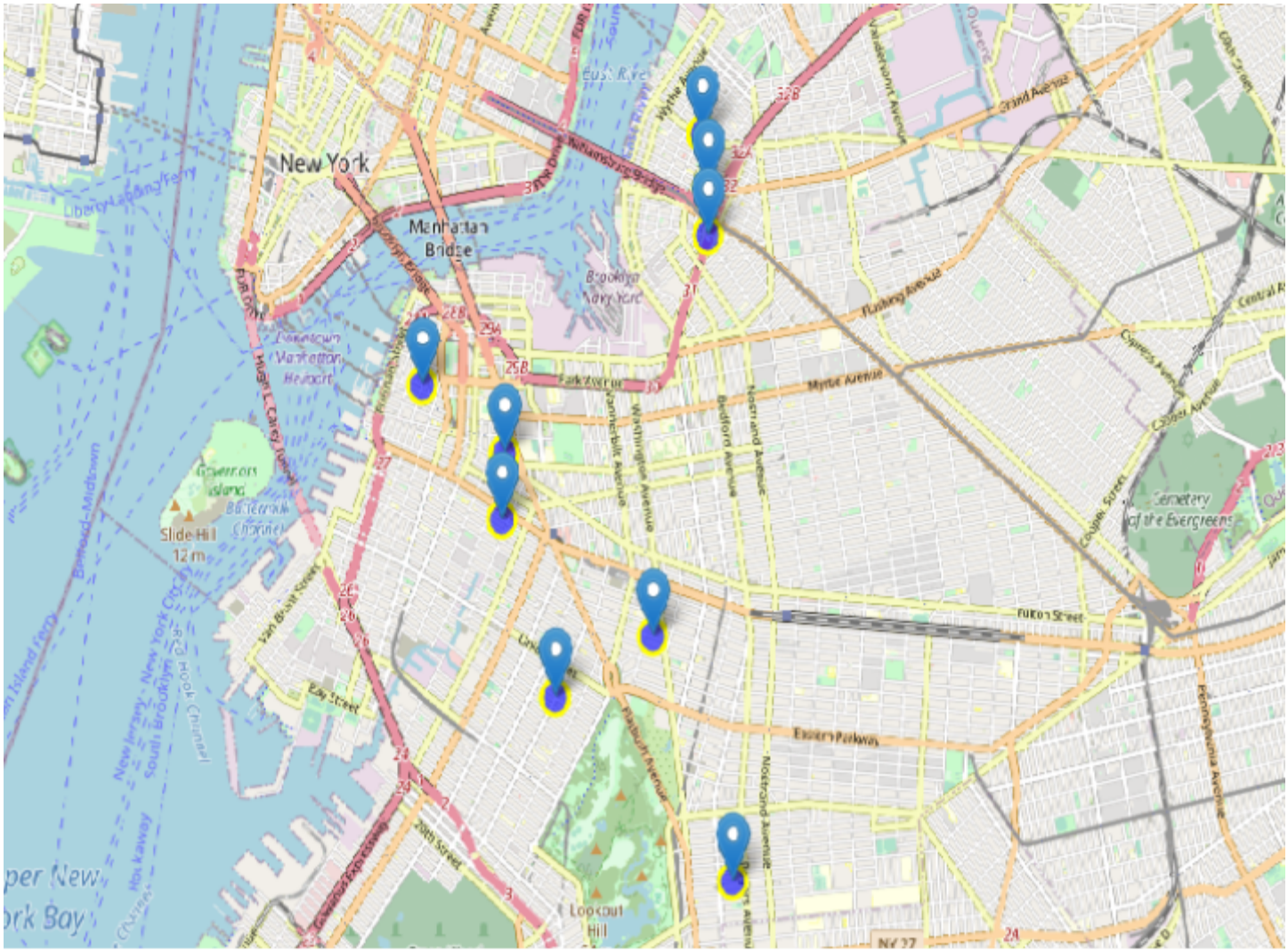

Image 4.1.5 - Map of Brooklyn with markers of 9 highest rated sushi restaurants.

5. Conclusion, Discussion, and Recommendations

Using a combination of datasets from Foursquare, data scraped from Wikipedia, and the https://cocl.us/new_york_dataset dataset, analysis and description of sushi restaurants in by neighborhood and borough in New York City was able to be performed. This revealed The highest rated sushi restaurants to be located in Brooklyn which is also the most population dense borough. There were two clusters of 3 venues in the North side/South Side/Williamsburg and Downtown/Brooklyn Heights/Boerum Hill neighborhoods of Brooklyn.

This was an introductory look into opening a sushi restaurant in New York City and does not account for many factors. Improving this could include Foursquare venue data for ‘Japanese Restaurant’ venue data. This analysis could be expanded to include other data including property crime at restaurants, other restaurant venues nearby, rental prices, location of food suppliers, and other nearby highly rated venues. The demographics are not specific to the country which is a limitation of data scraping from the wikipedia page.

With these limitations, the current data would suggest Brooklyn as a possible source for further analysis given the current density of highly rated sushi venues suggesting an existing consumer base as well as population density.