IBM DATA SCIENCE

CAPSTONE PROJECT

BEST LOCATION TO OPEN A JAPANESE RESTAURANT IN NEW YORK

by Abay Mirmanov

Introduction

New York is the most populous city in the US. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2), New York City is also the most densely populated major city in the United States.Located at the southern tip of the U.S. state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities.

Business Problem

The objective of this Capstone project is to analyze and select the best locations in the city of New York to open a new Japanese restaurant. Using Data Science methodology and instruments such as Data Analysis and Visualization, this project aims to provide solutions to answer the business question: Where in the city of New York, should the investor open a Japanese Restaurant?

This project is particularly useful to developers and investors looking to open or invest in a Japanese restaurantin the city of New York. Overall, New York is a great place to open a restaurant with an ethnical cuisine. As New York is the most diverse city in the world (800 languages are spoken in New York). With its diverse culture, comes diversity in the food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Indian, French, etc. Why did we decide to focus on Japanese cuisine in our project? Now when the idea of a healthy lifestyle conquered the minds of people all over the country, Japanese 3 restaurants became extremely popular, as they offer a healthy alternative to regular American eating habits.

Data

To solve the problem, we will need the following data:

- New York City data containing the neighborhoods and boroughs.
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data.
- Venue data, particularly data related to restaurants.

We are going to use this data to perform further analysis of the neighborhoods. This project will require using of many data science skills, from web scrapping (open source dataset), working with API (Foursquare), data cleaning, data wrangling, to map visualization (Folium). In the next Methodology section, we will discuss and describe any exploratory data analysis that we did, any inferential statistical testing that we performed, and what machine learning techniques were used.

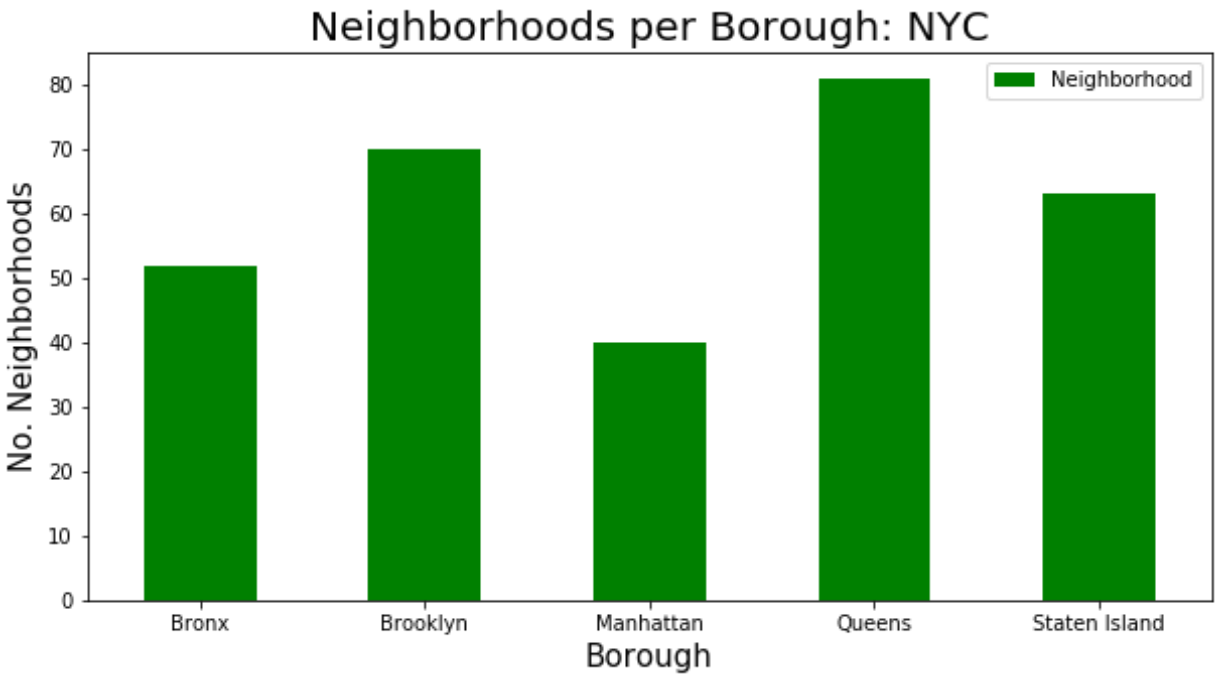
Methodology

- Data will be collected from https://cocl.us/new_york_dataset and cleaned and processed into a dataframe.
- FourSquare be used to locate all venues and then filtered by Japanese restaurants. Ratings, tips, and likes by users will be counted and added to the dataframe.
- Data will be sorted based on rankings.
- Finally, the data be will be visually assessed using graphing from Python libraries.

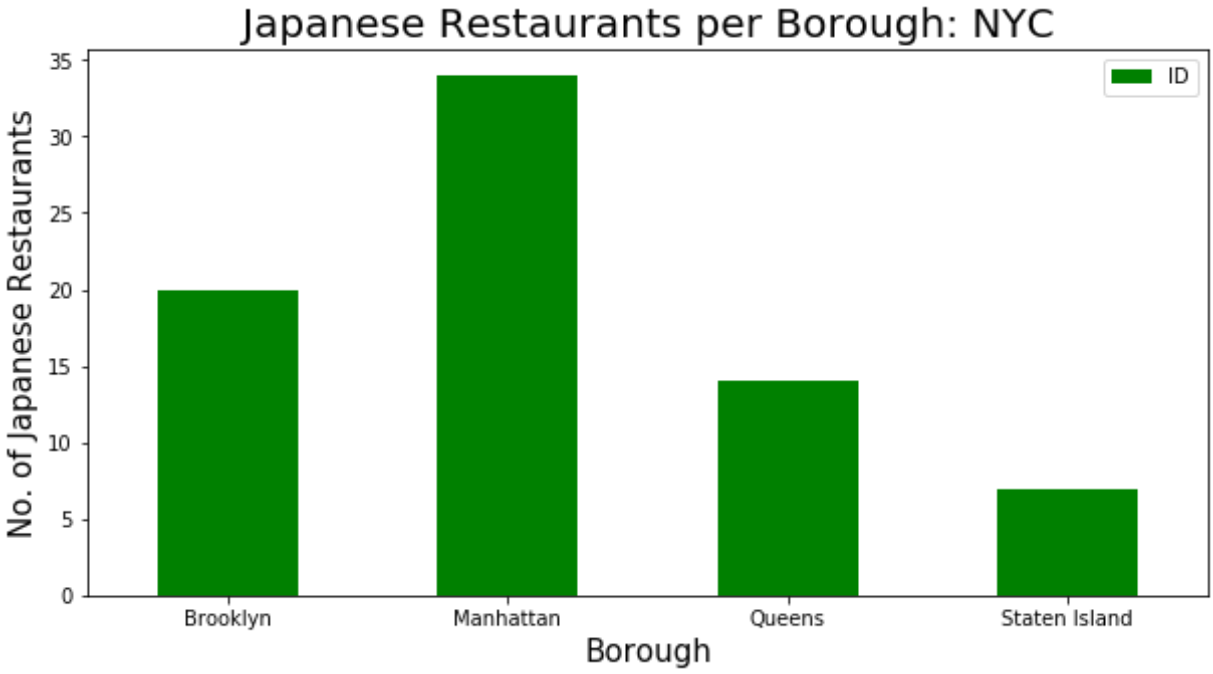
Results

The results of our analysis showed below:

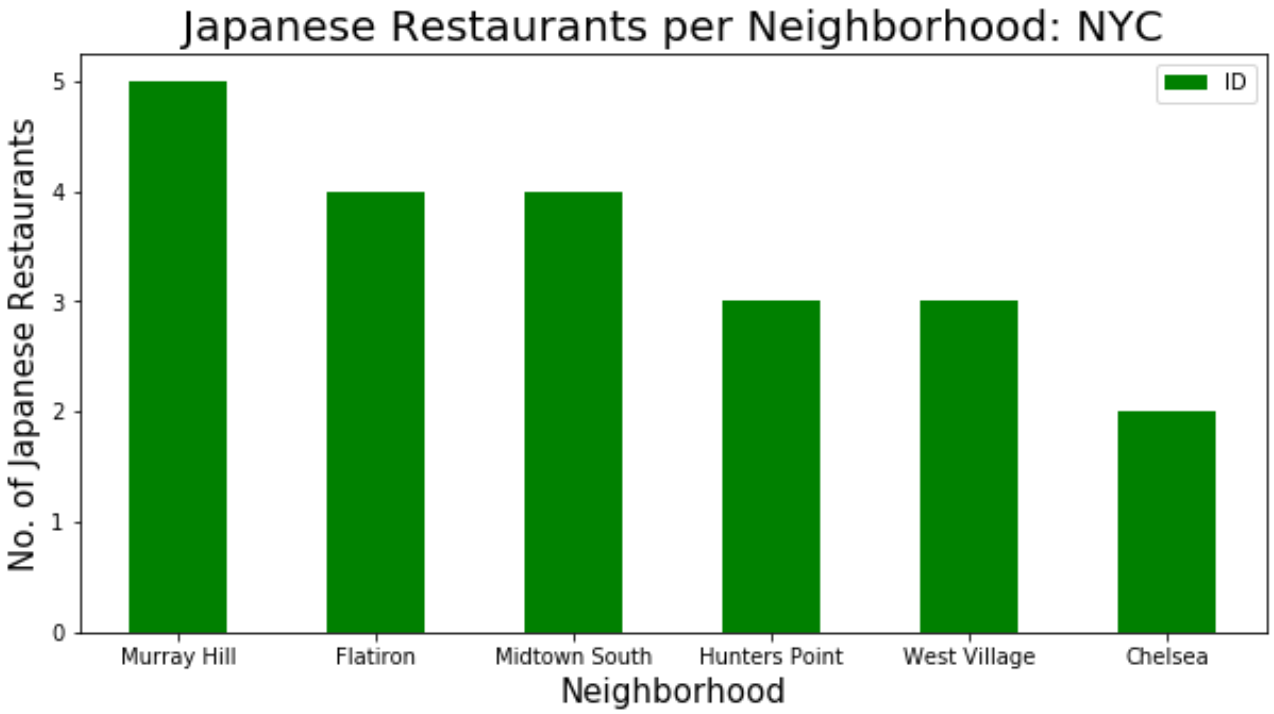
1. We see that Queens has the highest number of Neighborhoods.



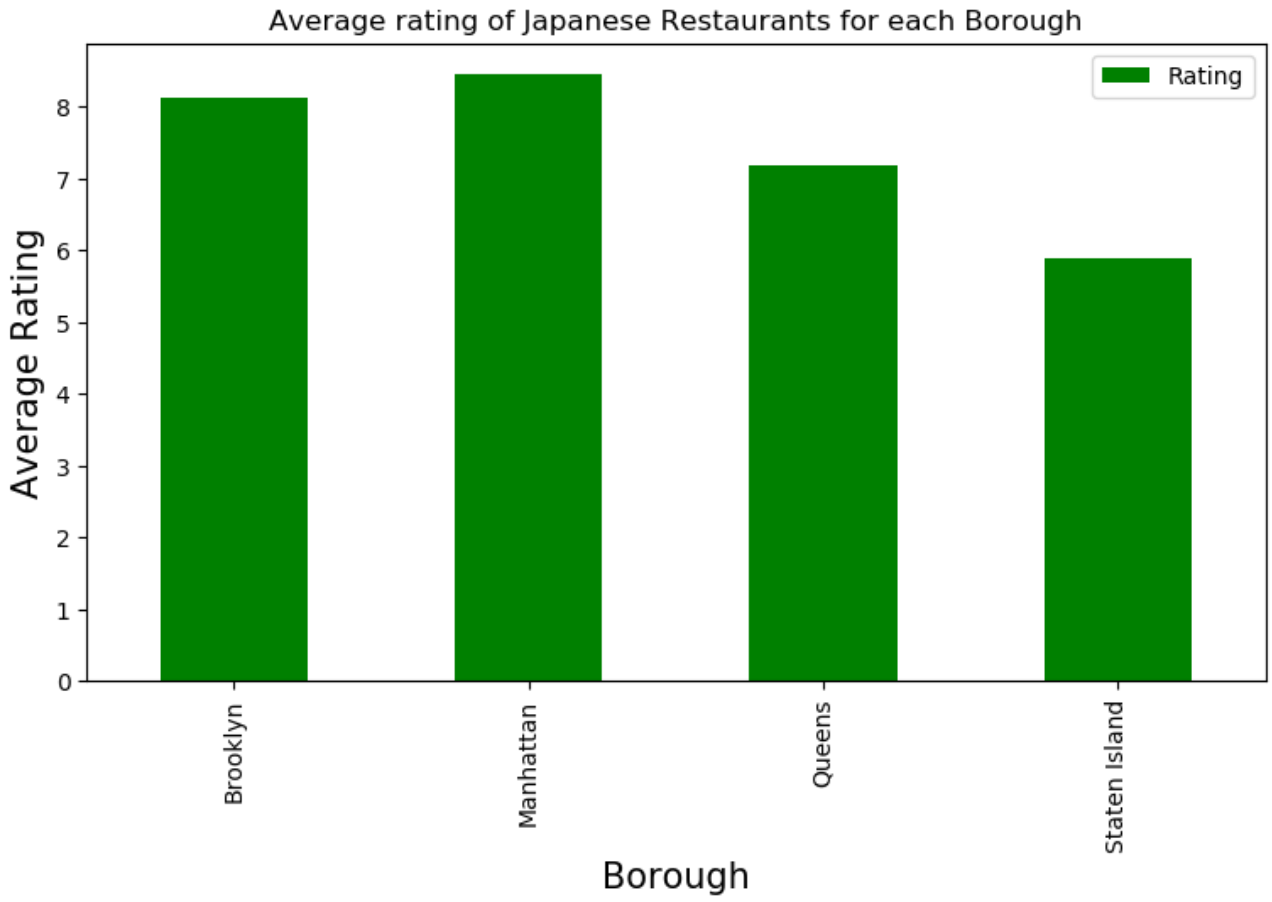
2. Although Manhattan had the least number of neighborhoods, it does have the highest number of Japanese restaurants.



3. The Murray Hill in Manhattan has the highest number of Japanese Restaurants with a total count of 5.



4. Manhattan has the highest average rating for Japanese Restaurants. Brooklyn goes right behind the Manhattan with the average rating of its Japanese restaurants only a little lower than Manhattan. Two neighborhoods with the highest average rating for Japanese Restaurants are Cobble Hill and North Side (both located in Brooklyn).



Top 10 neighborhoods with Japanese Restaurants with the best average rating are:

1. Cobbie Hill
2. North Side
3. Flatiron
4. East Village
5. Hunter Point
6. Fort Greene
7. Boerum Hill
8. Downtown
9. Midtown South
10. Chelsea

	Neighborhood	Average Rating
9	Cobble Hill	9.100000
36	North Side	9.100000
15	Flatiron	8.975000
12	East Village	8.950000
25	Hunters Point	8.933333
16	Fort Greene	8.900000
1	Boerum Hill	8.800000
10	Downtown	8.800000
33	Midtown South	8.800000
6	Chelsea	8.750000

Discussion Section

Based on the results of our analysis, I would state that Manhattan and Brooklyn are the best locations for Japanese cuisine in NYC. To have the best shot of success, I would open a Japanese restaurant in Brooklyn. Brooklyn has multiple neighborhoods with average ratings exceeding 8.0 on a scale of 1.0 to 10.0 and has less amount of Japanese restaurants than Manhattan, making competition easier. In addition, we should keep in mind, that real estate prices in Brooklyn are much cheaper than in Manhattan. Particularly, I would recommend considering opening a Japanese Restaurant either in Cobble Hill or in North Side, because both of these neighborhoods have the highest rating for Japanese restaurants

Conclusions

In the project we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing data analysis, and lastly providing recommendations to the investors/developers. During the project, we applied different data science methods and instruments to get the answer to our main question: “Where in the City of New York, should the investor open a Japanese Restaurant?” The findings of this project will help the relevant investor better understand the advantages and disadvantages of different New York neighborhoods/boroughs in terms of opening a Japanese restaurant.

In a fast moving world, there are many real life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to cluster neighborhoods in New York based on the most common venues in those neighborhoods. Similarly, data can also be used to solve other problems, which most people face in metropolitan cities.