Coursera Capstone

IBM Applied Data Science Capstone

Opening a Japanese Restaurant in NYU

Jingwen Pei

Date: 12/05/2020

Introduction

New York City's demographics show a large and ethnically diverse

Metropolis, that is to say, NYC is the largest city in the United States with a relatively

long history of international immigration and in 2018 was home to nearly 8.5 million

people, accounting for over 40% of the population of New York State. In the

meanwhile, a slightly lower percentage of the New York metropolitan area, it is

home to approximately 23.6 million. Over the last decade, the city has been growing

faster than the region. The New York region continues to be by far the leading

metropolitan gateway for legal immigrants admitted into the United States.

This final project explores the best locations for Japanese restaurants throughout

the city of New York. Potentially the owner of the new Japanese restaurant can

have great success and consistent profit. However, as with any business, opening a

new restaurant requires serious considerations and is more complicated than it

seems from the first glance. In particular, the location of the restaurant is one of

the most important factors that will affect whether it will have success or a failure.

So our project will attempt to **answer the questions** from both investor side (i.e.

"Where should the investor open a Japanese Restaurant?") and consumer side (i.e.

"Where should I go If I want great Japanese food?")

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?

Target Audience and some demographic facts

This project is particularly useful to developers and investors looking to open or invest in a Japanese restaurant in 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 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:

1. New York City data containing the neighborhoods and boroughs.

2. Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data.

3. Venue data, particularly data related to restaurants. We are going to use

this data to perform further analysis of the neighborhoods.

Data Source and extracting methods

New York City data containing the neighborhoods and boroughs will be obtained

from the open data source: https://cocl.us/new_york_dataset. After it, we will get

the geographical coordinates of the neighborhoods (latitude and longitude) using

Python Geocoder package. Finally, we will use Foursquare API to get the venue data

for the neighborhoods defined at the previous step. Foursquare has one of the

largest databases of 105+ million places and over 125,000 developers use this

application. Foursquare API provides many categories of the venue data; we are

particularly interested in the restaurant data to solve the business problem defined

above.

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.

Results

The results of our analysis showed below: 1. We see that Queens has the highest
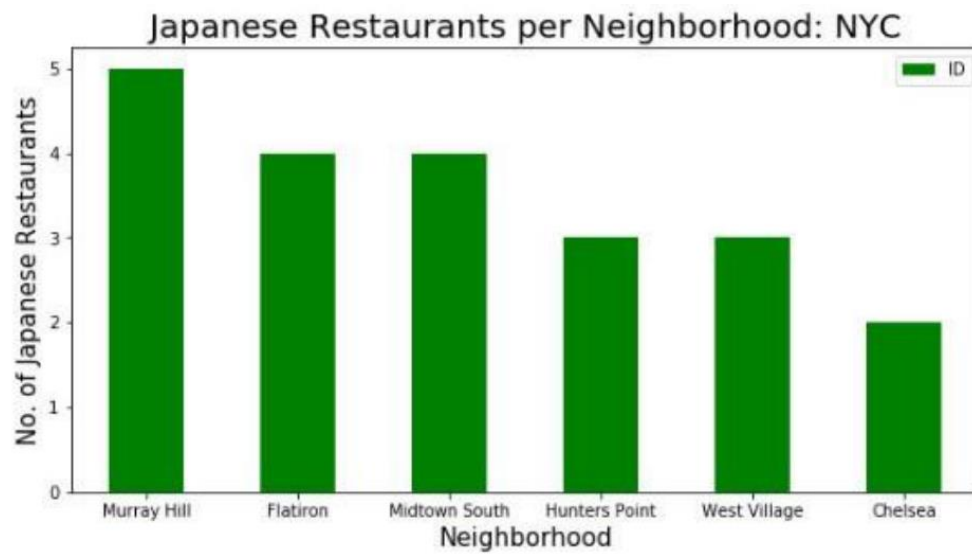
number of Neighborhoods.
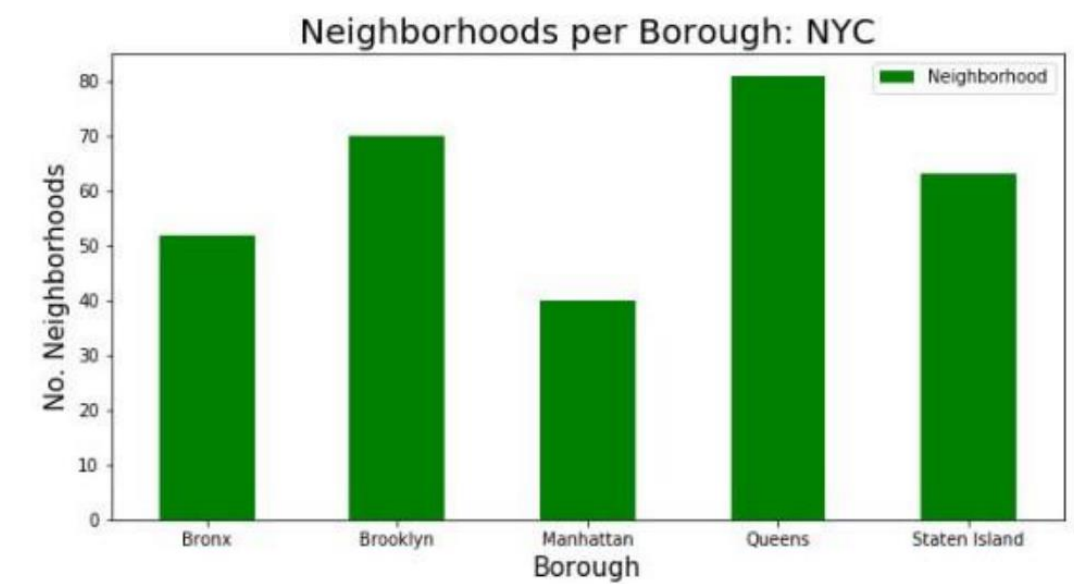
2. Although Manhattan had the least number of neighborhoods, it does have the

highest number of Japanese restaurants.
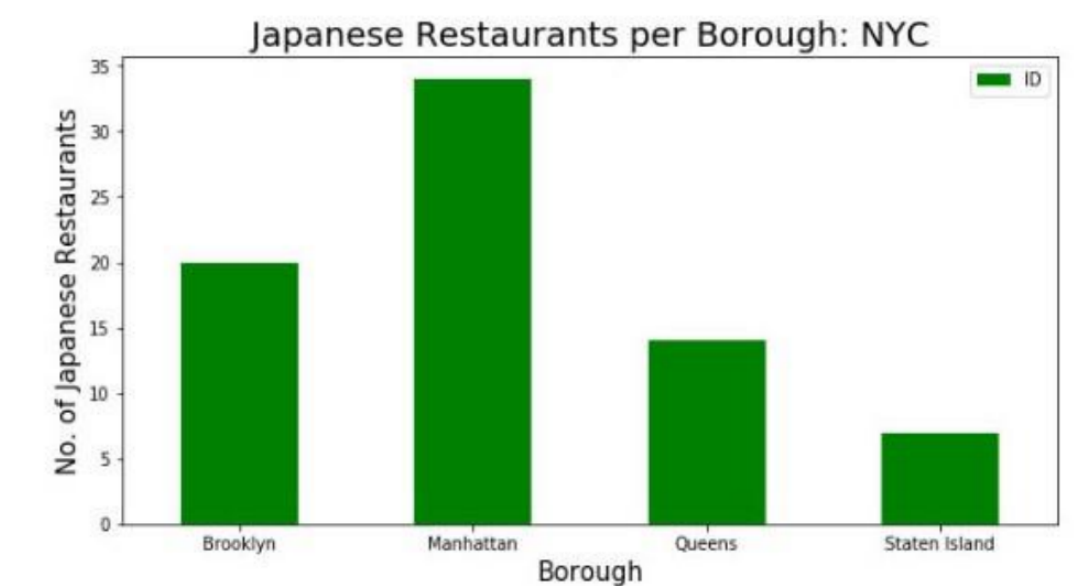
3. The Murray Hill in Manhattan has the highest number of Japanese Restaurants

with a total count of 5.

Japanese Restaurants per Neighborhood: NYC

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).

Average rating of Japanese Restaurants for each Borough

	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

Top Ten neighborhoods with Japanese Restaurants with the best average rating

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.

Limitations and suggestions

For future researches All of the above analysis is depended on the accuracy of Four-Square data. Besides, during this project, we used a free Sandbox Tier Account of Foursquare API that goes with limitations as to the number of API calls and results returned. To get better results, future research work and more comprehensive analysis could consider using a paid account to bypass these limitations as well as incorporating data from other external databases.

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.