COURSERA - IBM Applied Data Science Capstone Project

Battle of the Neighborhoods (Week 5, Part 2) – Final Assignment

Analysis Project: What is the Best Location in New York City to Open a Japanese Restaurant?

by Anna Wong

1. INTRODUCTION

According to the **World Population Review (2020**), **New York City** is the **#1 most populated US city** with **8.6+ million people.**

One of the great metropolitan cities in the world, it is comprised of **5 boroughs**: **Manhattan** (the most well-known and identified with the city), **Brooklyn, Bronx, Queens**, and **Staten Island**.

New York City represents a significant **global cross-road of diverse cultures** and a destination for **unlimited cuisine choices**; there are numerous dining options for even the most discriminating or adventurous foodie.

In recent years, especially in the digital age, many Americans have been **aspiring to eat more healthfully**. In a city such as New York, **Japanese food**, for example, has **wide appeal with umami options**, such *as sushi, sashimi, tonkotsu, edamame, miso, and udon/soba/ramen*.

2. BUSINESS PROBLEM

Opening and **sustaining** a **successful restaurant in New York City** is **extremely competitive**. It requires **great financial investment, favorable customers ratings/reviews,** and **market demand**. **Location** is also an important determining factor in whether such a relentlessly demanding business succeeds or fails.

Japanese food is one of the most popular cuisines and can be found in just about every US city. This analysis, in particular, aims to **determine an ideal location** (i.e. borough and neighborhood), **to launch a Japanese restaurant by a potential business investor (client)**. However, to minimize financial risk, it is essential, with data science, to take a more comprehensive look across all 5 New York City boroughs. The study was conducted with relevant data to provide **an informed recommendation** before the client can make a prudent decision to proceed with his business plans.

To **determine the most advantageous New York City neighborhood to open a Japanese restaurant**, the following questions needed to be asked:

- How many **neighborhoods are** located within each of the **5 boroughs in New York City**?
- Which boroughs and neighborhoods have the **most Japanese restaurants**?
- How many **Japanese restaurants** are located within the most populous borough?
- What is the average number of **"likes", "tips", and "rankings"** of the Japanese restaurants in these areas?

3. DATA SOURCES & DESCRIPTION

To best analyze the problem, it was necessary to consider **data on all 5 New York City boroughs: neighborhoods, latitude/longitude, restaurants, and customer tips/rankings/likes**.

The following data sources were leveraged:

1) https://cocl.us/new_york_dataset [from nyu_2451_34572-geojson. json from https://s3-api.us-geo.objectsstorage.softlayer.net or https://geo.nyu.edu/catalog/nyu_2451_34572]
for New York City data with **boroughs** (Manhattan, Brooklyn, Bronx, Queens, and Staten Island), **neighborhoods, latitudes**, and **longitudes**.
Data was sourced and organized into a **pandas dataframe**.
Latitude and longitude coordinates of neighborhoods were needed to plot the map and get the venue data.
Venue data, as it relates to restaurants, would be used to perform further analysis of the neighborhoods.

2) https://foursquare.com/
for New York City Japanese restaurants – locations and customer rankings/tips/likes (using **FourSquare API** to filter out relevant columns of information).

In sum, the collected data was sorted based on rankings (boroughs/neighborhoods), using dataframes to describe statistics and plotted into visual graphs/maps (Python/Folium libraries to visualize location data). **Informed insights** could then be drawn to make an appropriate recommendation to the business investor seeking to open the restaurant.

4. METHODOLOGY

4.1 Process Overview

• Data was collected from https://cocl.us/new_york_dataset and cleansed for processing into a dataframe.

• FourSquare https://foursquare.com/ was used to locate all venues and then filtered by Japanese restaurants.

"Ratings", "Tips", and "Likes" by customers were counted and added to the dataframe.

To pull in data for the venues, for example, the **Foursquare API** was set up to create a list of places within a **specified radius and limit**. FourSquare supplied the names, locations, and venue type of the surrounding businesses.

• Data was sorted based on rankings.

• Finally, the data was also visually assessed using **graphs and mapping** from various **Python/Folium libraries.**

4.2 Defining the Data Sets

The project/analysis was initiated by downloading the critical dependencies such as **importing and loading** the **Python** and **Folium** libraries. The code is displayed below:

```
import pandas as pd
import numpy as np

pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import requests
from bs4 import BeautifulSoup
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
#!conda install -c conda-forge folium

import os
!pip install folium
import folium # map rendering library

# Matplotlib and associated plotting modules
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
import matplotlib as mp
import re
import csv
%matplotlib inline

print('Libraries imported.')
```

Additionally, the **geopy library** was defined and leveraged to yield the **latitude and longitude** coordinates:

```
def geo_location(address):
    # get geo location of address
    geolocator = Nominatim(user_agent="ny_explorer")
    location = geolocator.geocode(address)
    latitude = location.latitude
    longitude = location.longitude
    return latitude,longitude
```

Moreover, **several data sets** were established as follows:

- "**Get Venue":** The first data set involved **defining a function** with the **FourSquare API** to extract and explore the **venue (restaurant) data**. It contains the venue recommendations for each of the boroughs and neighborhoods to analyze Japanese restaurants. There is a limit of **100 venue recommendations** per neighborhood plus a **radius of 1000 meters** around the geographic coordinates (for a **given latitude and longitude**) for each neighborhood.

```
def get_venues(lat,lng):
    #set variables
    radius=1000
    LIMIT=100
    CLIENT_ID = [redacted] your Foursquare ID
    CLIENT_SECRET = [redacted] your Foursquare Secret
    VERSION = '20200605' # Foursquare API version

    #url to fetch data from foursquare api
    url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

    # get all the data
    results = requests.get(url).json()
    venue_data=results["response"]['groups'][0]['items']
    venue_details=[]
    for row in venue_data:
        try:
            venue_id=row['venue']['id']
            venue_name=row['venue']['name']
            venue_category=row['venue']['categories'][0]['name']
            venue_details.append([venue_id,venue_name,venue_category])
        except KeyError:
            pass

    column_names=['ID','Name','Category']
    df = pd.DataFrame(venue_details,columns=column_names)
    return df
```

- "**Get Venue Details**": A function was defined to yield **venue details** such as **tips, ratings, and likes** for a given **venue ID** for *ranking purposes*.

```
def get_venue_details(venue_id):

    CLIENT_ID = [redacted] # your Foursquare ID
    CLIENT_SECRET = [redacted] # your Foursquare Secret
    VERSION = '20200605' # Foursquare API version

    #url to fetch data from foursquare api
    url = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}'.format(
    venue_id,
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION)
    print (url)

    # get all the data
    results = requests.get(url).json()
    print(results)
    if(results['meta']['code']==200):
        venue_data=results['response']['venue']
        venue_details=[]
        try:
            venue_id=venue_data['id']
            venue_name=venue_data['name']
            venue_likes=venue_data['likes']['count']
            venue_rating=venue_data['rating']
            venue_tips=venue_data['tips']['count']
            venue_details.append([venue_id,venue_name,venue_likes,venue_rating,
            venue_tips])
        except KeyError:
            pass

        column_names=['ID','Name','Likes','Rating','Tips']
        df = pd.DataFrame(venue_details,columns=column_names)
        return df
```

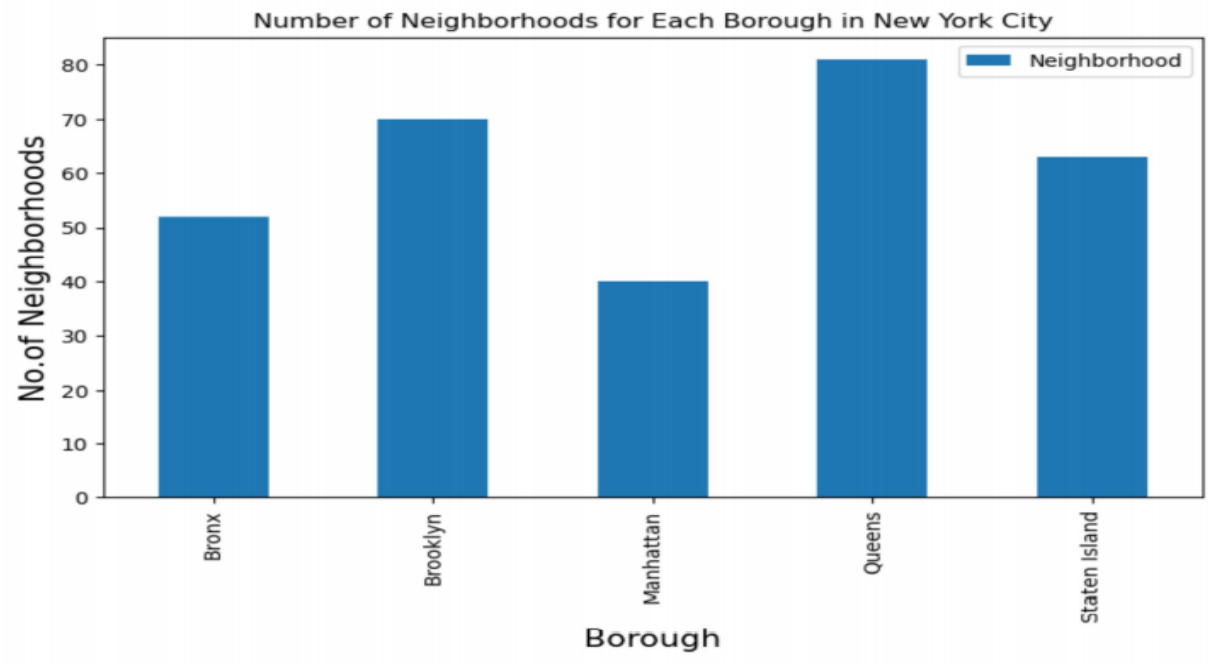
- "**New York City Data**": This data set is defined to yield the **geographical coordinates (latitude and longitude)** of the **5 NYC boroughs** *(Manhattan, Brooklyn, Queens, Bronx, and Staten Island)* and their **corresponding neighborhoods**.

```
def get_new_york_data():
    url='https://cocl.us/new_york_dataset'
    resp=requests.get(url).json()
    # all data is present in features label
    features=resp['features']

    # define the dataframe columns
    column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']
    # instantiate the dataframe
    new_york_data = pd.DataFrame(columns=column_names)

    for data in features:
        borough = data['properties']['borough']
        neighborhood_name = data['properties']['name']

        neighborhood_latlon = data['geometry']['coordinates']
        neighborhood_lat = neighborhood_latlon[1]
        neighborhood_lon = neighborhood_latlon[0]

        new_york_data = new_york_data.append({'Borough': borough,
                                              'Neighborhood': neighborhood_name,
                                              'Latitude': neighborhood_lat,
                                              'Longitude': neighborhood_lon}, ignore_index=True)

    return new_york_data
```

4.3 Identifying the Number of Neighborhoods per Borough

After running the code for the above data set, it was determined that there was a total of **306 NYC neighborhoods** across the **5 boroughs**.

```
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Neighborhoods for Each Borough in New York City')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Neighborhoods', fontsize=15)
#giving a bar plot
new_york_data.groupby('Borough')['Neighborhood'].count().plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```



Based on the data in the above chart:

- **Queens** is the borough with the **largest number of neighborhoods**, with a **total of 80.**
- **Manhattan**, the most well-known borough and most identified with NYC, has the **least** with a **total of 40**.

It was interesting to see that even the other boroughs of **Brooklyn (2nd most with 70),** Bronx (3rd most), and Staten Island (4th most) still outnumber Manhattan.

4.4 Identifying the Boroughs & Neighborhoods with Japanese Restaurants

A **neighborhood list** was then prepared that contained all NYC restaurants before **clustering only Japanese restaurants:**

```
# prepare neighborhood list that contains Japanese restaurants
column_names=['Borough', 'Neighborhood', 'ID','Name']
japanese_rest_ny=pd.DataFrame(columns=column_names)
count=1
for row in new_york_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    japanese_resturants=venues[venues['Category']=='Japanese Restaurant']
    print('(',count,'/',len(new_york_data),')','Japanese Resturants in '+Neighborhood+', '+Borough+':'+str(len(japanese_resturants)))
    for resturant_detail in japanese_resturants.values.tolist():
        id, name , category=resturant_detail
        japanese_rest_ny = japanese_rest_ny.append({'Borough': Borough,
                                                    'Neighborhood': Neighborhood,
                                                    'ID': id,
                                                    'Name' : name
                                                   }, ignore_index=True)
    count+=1
```

Our analysis further included such information as the **Venue ID (of the restaurant), Name of the Restaurant, Likes, Ratings, and Tips**:

```
japanese_rest_stats_ny.head()
```

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Riverdale	503cfaffe4b066d39de5005a	Aoyu Japanese Restaurant	32	9.0	19
1	Bronx	Riverdale	4b0b311af964a520642e23e3	Palace of Japan	38	8.5	26
2	Bronx	Kingsbridge	503cfaffe4b066d39de5005a	Aoyu Japanese Restaurant	32	9.0	19
3	Bronx	Kingsbridge	4b0b311af964a520642e23e3	Palace of Japan	38	8.5	26
4	Bronx	City Island	4dbdf3d790a02849cbd675be	Ohana Japanese Habachi Seafood & Steakhouse	34	7.7	18

After **further segmentation**, it was discovered that there were **191 Japanese restaurants out of 306 NYC neighborhoods**.

```
japanese_rest_ny.shape
```

```
(191, 4)
```

```
japanese_rest_stats_ny.to_csv('japanese_rest_stats_ny.csv', index=False)
```

```
japanese_rest_stats_ny_csv=pd.read_csv('japanese_rest_stats_ny.csv')
```

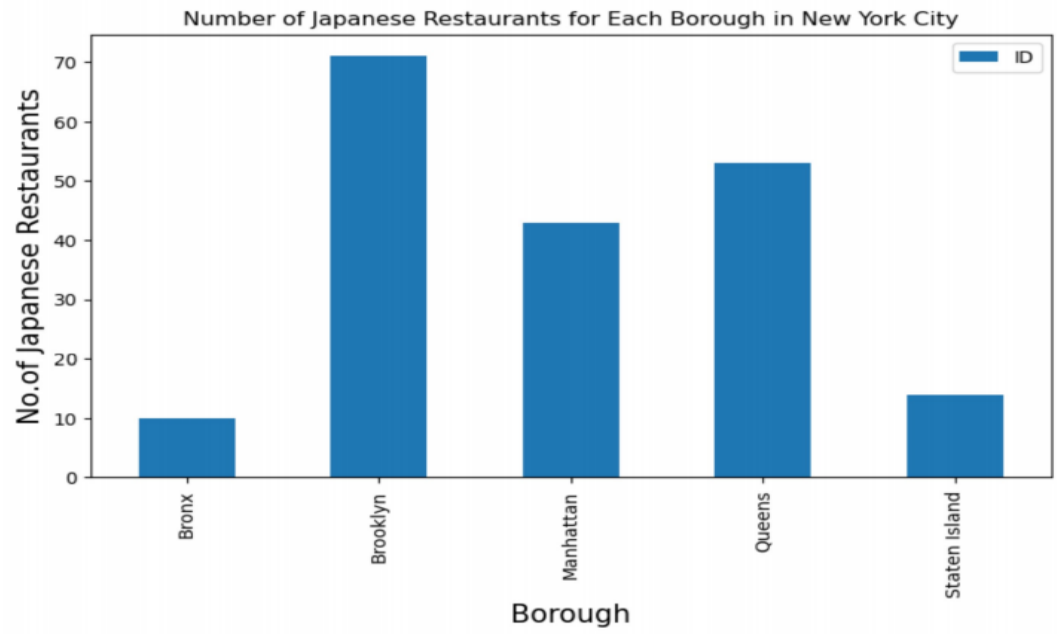
```
japanese_rest_stats_ny_csv.shape
```

```
(191, 7)
```

4.4.1 Number of Japanese Restaurants for Each Borough in New York City

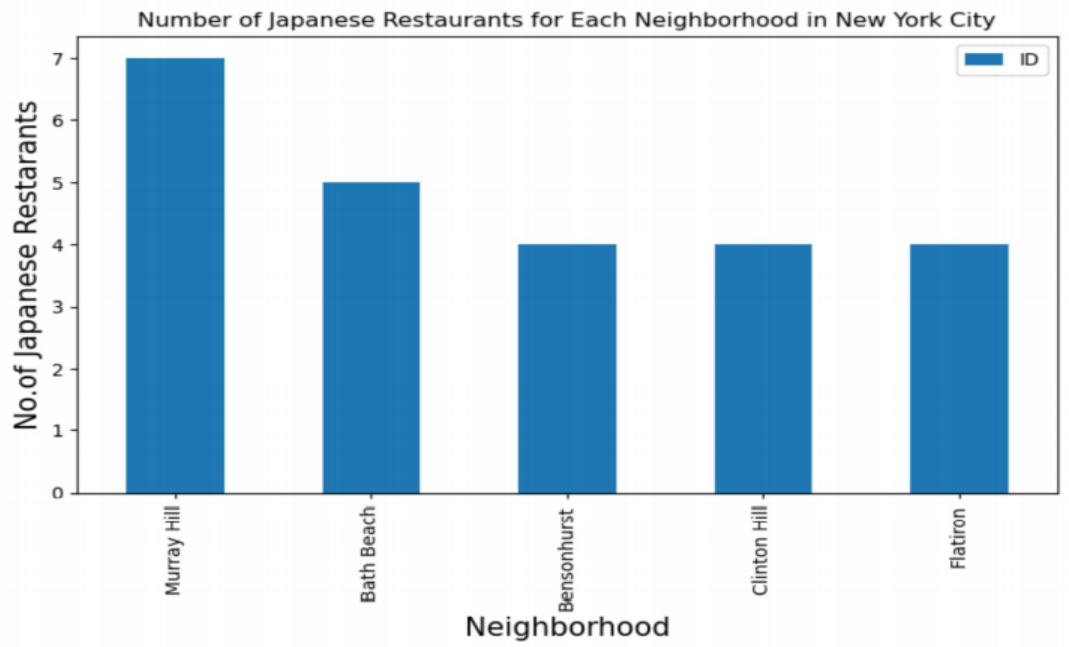
By plotting these results on a graph, one could better visualize the **number of Japanese restaurants** for **each borough in New York City**: As seen in the bar chart below, **Brooklyn** has the **most Japanese restaurants**, *followed by Queens, Manhattan, Staten Island, and the Bronx*.

```
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Japanese Restaurants for Each Borough in New York City')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Japanese Restaurants', fontsize=15)
#giving a bar plot
japanese_rest_ny.groupby('Borough')['ID'].count().plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```



To better understand these findings, one must ***further*** **drill down from the borough level to the <u>neighborhood level</u>. Murray Hill** has the most Japanese restaurants in the neighborhood with 7, followed by **Bath Beach with 5**. **Bensonhurst, Clinton Hill,** and **Flatiron each have 4**.

```
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Japanese Restaurants for Each Neighborhood in New York City')
#On x-axis
plt.xlabel('Neighborhood', fontsize = 15)
#On y-axis
plt.ylabel('No.of Japanese Restarants', fontsize=15)
#giving a bar plot
japanese_rest_ny.groupby('Neighborhood')['ID'].count().nlargest(5).plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```



Upon further analysis as seen below, **Murray Hill** is the name of a neighborhood ("same name") that is found in **<u>2 different boroughs</u>**: **Manhattan** (total of 5) and **Queens** (total of 2).

```
japanese_rest_ny[japanese_rest_ny['Neighborhood']=='Murray Hill']
```

	Borough	Neighborhood	ID	Name
81	Manhattan	Murray Hill	4a99b4f4f964a520f62f20e3	Kajitsu
82	Manhattan	Murray Hill	559cbaa6498eaa4e8d884811	Tempura Matsui
83	Manhattan	Murray Hill	591caee89deb7d0f69be77a2	Omusubi Gonbei
84	Manhattan	Murray Hill	49db8b67f964a520d85e1fe3	Aburiya Kinnosuke
85	Manhattan	Murray Hill	5ad925da2f97ec3e4b17eba2	Nonono
123	Queens	Murray Hill	4bc8eee83740b713fcbe5d65	Northern Sushi
124	Queens	Murray Hill	5cf07afa66f3cd002c8b3ae6	Izakaya Mew

The **other neighborhoods from the bar chart are detailed below. They include 3 neighborhoods in Brooklyn: Bath Beach has (5 restaurants), Bensonhurst, and Clinton Hill (4 restaurants each).** The other **identified neighborhood in Manhattan** is **Flatiron** with **4 Japanese restaurants** as well.

```
japanese_rest_ny[japanese_rest_ny['Neighborhood']=='Bath Beach']
```

	Borough	Neighborhood	ID	Name
42	Brooklyn	Bath Beach	51f42bb3498e2681264e00a1	Kasumi Sushi & BBQ Restaurant
43	Brooklyn	Bath Beach	4c326b9fa0ced13ac126156e	Ichi Sushi
44	Brooklyn	Bath Beach	4b787863f964a52080d02ee3	Shiki Japanese Hibachi
45	Brooklyn	Bath Beach	513e7ec8e4b07f4fa1b53b3d	Fuki Sushi Japanese Restaurant
46	Brooklyn	Bath Beach	4bad45c9f964a52089403be3	Chikurin

```
japanese_rest_ny[japanese_rest_ny['Neighborhood']=='Bensonhurst']
```

	Borough	Neighborhood	ID	Name
13	Brooklyn	Bensonhurst	51f42bb3498e2681264e00a1	Kasumi Sushi & BBQ Restaurant
14	Brooklyn	Bensonhurst	4b787863f964a52080d02ee3	Shiki Japanese Hibachi
15	Brooklyn	Bensonhurst	513e7ec8e4b07f4fa1b53b3d	Fuki Sushi Japanese Restaurant
16	Brooklyn	Bensonhurst	4c326b9fa0ced13ac126156e	Ichi Sushi

```
japanese_rest_ny[japanese_rest_ny['Neighborhood']=='Clinton Hill']
```

	Borough	Neighborhood	ID	Name
50	Brooklyn	Clinton Hill	591f7c5f35f9836382455eca	U-Gu
51	Brooklyn	Clinton Hill	5739c66c498e7ef6085cec4f	Karasu
52	Brooklyn	Clinton Hill	5aa09a586eda026624b25ecb	66S Fusion
53	Brooklyn	Clinton Hill	4e3cadcb1f6e844231dbe017	Waza

```
japanese_rest_ny[japanese_rest_ny['Neighborhood']=='Flatiron']
```

	Borough	Neighborhood	ID	Name
178	Manhattan	Flatiron	5bd7a5561c675b0039be9acc	HALL
179	Manhattan	Flatiron	4f7b61f6e4b07aaa5d524842	OOTOYA 大戸屋 (OOTOYA)
180	Manhattan	Flatiron	581a10901df6b32e66ec3a07	Sugarfish
181	Manhattan	Flatiron	55805bb5498e8356d675aa22	Yakiniku Futago

4.5 Identifying Japanese Restaurants by Venue ID, Restaurant Name, Likes, Ratings, and Tips Among Neighborhoods/Boroughs

The analysis further included the **Venue ID** (of the restaurant), **Name of the Restaurant**, **Likes**, **Ratings**, and **Tips.** For example:

```
japanese_rest_stats_ny.head()
```

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Riverdale	503cfaffe4b066d39de5005a	Aoyu Japanese Restaurant	32	9.0	19
1	Bronx	Riverdale	4b0b311af964a520642e23e3	Palace of Japan	38	8.5	26
2	Bronx	Kingsbridge	503cfaffe4b066d39de5005a	Aoyu Japanese Restaurant	32	9.0	19
3	Bronx	Kingsbridge	4b0b311af964a520642e23e3	Palace of Japan	38	8.5	26
4	Bronx	City Island	4dbdf3d790a02849cbd675be	Ohana Japanese Habachi Seafood & Steakhouse	34	7.7	18

Select columns have **object dtypes** rather than **float dtypes**; hence, the **pandas.Series.astype** was used to convert the "**Likes**", "**Tips**", and "**Ratings**" **object dtypes** for **consistency** in the data analysis.

```
japanese_rest_stats_ny['Likes']=japanese_rest_stats_ny['Likes'].astype('float64')
```

```
japanese_rest_stats_ny['Tips']=japanese_rest_stats_ny['Tips'].astype('float64')
```

```
japanese_rest_stats_ny['Rating']=japanese_rest_stats_ny['Rating'].astype('float64')
```

```
japanese_rest_stats_ny.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 191 entries, 0 to 190
Data columns (total 7 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Borough       191 non-null    object 
 1   Neighborhood  191 non-null    object 
 2   ID            191 non-null    object 
 3   Name          191 non-null    object 
 4   Likes         191 non-null    float64
 5   Rating        191 non-null    float64
 6   Tips          191 non-null    float64
dtypes: float64(3), object(4)
memory usage: 10.6+ KB
```

Next, the identified Japanese restaurants in NYC were analyzed for their **maximum Likes, Ratings, and Tips**. The analysis, for example, revealed that in the borough of **Brooklyn**:

1) The neighborhood of **Williamsburg** has the same restaurant with the maximum number of **"Likes" and "Tips".**

2) The neighborhood of **Greenpoint** has a restaurant with the maximum **"Ratings".**

```
# Resturant with maximum Tips
japanese_rest_stats_ny.iloc[japanese_rest_stats_ny['Tips'].idxmax()]
```

```
Borough                          Brooklyn
Neighborhood                 Williamsburg
ID              4c7f0887fb74236a7727f9b9
Name                         Samurai Mama
Likes                                 898
Rating                                  9
Tips                                  247
Name: 30, dtype: object
```

```
# Resturant with maximum Likes
japanese_rest_stats_ny.iloc[japanese_rest_stats_ny['Likes'].idxmax()]
```

```
Borough                          Brooklyn
Neighborhood                 Williamsburg
ID              4c7f0887fb74236a7727f9b9
Name                         Samurai Mama
Likes                                 898
Rating                                  9
Tips                                  247
Name: 30, dtype: object
```

```
# Resturant with maximum Rating
japanese_rest_stats_ny.iloc[japanese_rest_stats_ny['Rating'].idxmax()]
```

```
Borough                          Brooklyn
Neighborhood                    Greenpoint
ID              5e4c841b1485b40007d77e6e
Name                      Rule Of Thirds
Likes                                 46
Rating                               9.1
Tips                                   6
Name: 18, dtype: object
```

The Japanese restaurants in neighborhoods with the **highest average ratings** are primarily located in **Brooklyn** and the **Bronx**.

```
ny_neighborhood_stats.sort_values(['Average Rating'],ascending=False).head(10)
```

	Neighborhood	Average Rating	
22	Cobble Hill	9.10	Brooklyn
83	Park Slope	8.85	
39	Fort Greene	8.80	
61	Kingsbridge	8.75	Bronx
98	Spuyten Duyvil	8.75	
92	Riverdale	8.75	
11	Brooklyn Heights	8.65	Brooklyn
49	Greenpoint	8.65	
110	Williamsburg	8.50	
85	Prospect Heights	8.40	

The **average ratings for Japanese restaurants in the NYC boroughs** were explored as well:

```
ny_borough_stats= japanese_rest_stats_ny.groupby('Borough',as_index=False).mean()[['B
orough','Rating']]
ny_borough_stats.columns=['Borough','Average Rating']
```

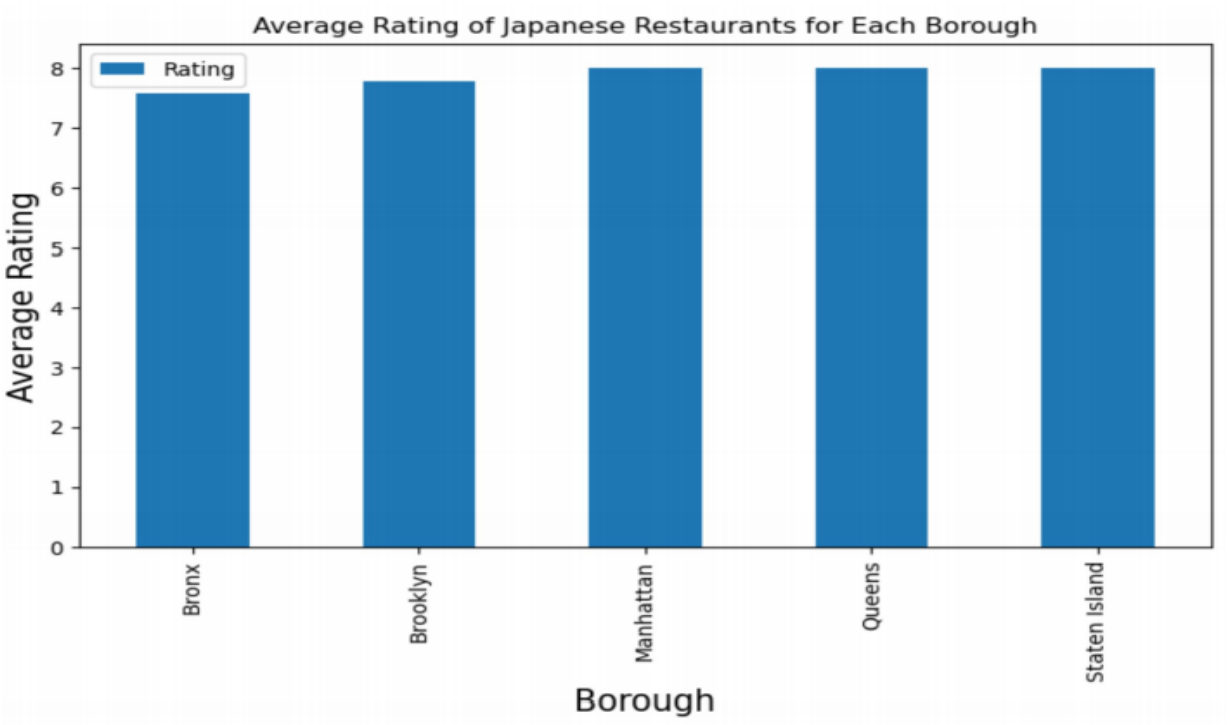
There was a **3-borough, average rating tie** of 8.0 among **Manhattan, Queens,** and **Staten Island**. This was followed by **7.78** in Brooklyn and **7.57** in the Bronx.

```
ny_borough_stats.sort_values(['Average Rating'],ascending=False).head()
```

	Borough	Average Rating
2	Manhattan	8.000000
3	Queens	8.000000
4	Staten Island	8.000000
1	Brooklyn	7.780282
0	Bronx	7.570000

The **average rating of Japanese restaurants** in **each NYC borough** was visualized as follows:

```
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Average Rating of Japanese Restaurants for Each Borough')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('Average Rating', fontsize=15)
#giving a bar plot
japanese_rest_stats_ny.groupby('Borough').mean()['Rating'].plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```



From the above bar chart, the restaurants in the boroughs of **Manhattan, Queens,** and **Staten Island** are **tied** for the highest average rating if 8.0 for Japanese restaurants in the NYC boroughs. **Brooklyn's restaurants** averaged at 7.78 and the **Bronx restaurants** averaged at 7.57.

Because the differences among all 5 boroughs were not statistically significant, it was necessary to additionally determine which borough's neighborhoods had **average ratings >8.5** (ex. setting a higher criterion to further separate out higher ranked restaurants among the boroughs).

```
ny_neighborhood_stats=ny_neighborhood_stats[ny_neighborhood_stats['Average Rating']>=
8.5]
```

```
ny_neighborhood_stats=pd.merge(ny_neighborhood_stats,new_york_data, on='Neighborhood
')
```

```
ny_neighborhood_stats=ny_neighborhood_stats[['Borough','Neighborhood','Latitude','Longitude','Average Rating']]
```

```
ny_neighborhood_stats
```

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Brooklyn	Brooklyn Heights	40.695864	-73.993782	8.65
1	Brooklyn	Cobble Hill	40.687920	-73.998561	9.10
2	Brooklyn	Fort Greene	40.688527	-73.972906	8.80
3	Brooklyn	Greenpoint	40.730201	-73.954241	8.65
4	Bronx	Kingsbridge	40.881687	-73.902818	8.75
5	Brooklyn	Park Slope	40.672321	-73.977050	8.85
6	Bronx	Riverdale	40.890834	-73.912585	8.75
7	Bronx	Spuyten Duyvil	40.881395	-73.917190	8.75
8	Brooklyn	Williamsburg	40.707144	-73.958115	8.50

As seen above, these were the **top 9 neighborhoods by borough** with the highest average rating of **8.5 or higher for Japanese restaurants**. The analysis indicated that **Brooklyn dominated** this category with **6 out of the 9 neighborhoods (with the highest rating of 9.10, while other restaurants ranged from 8.50 to 8.85).** The **Bronx** had the remaining 3 on the list with the same average rating of **8.75**.

5. MAPPING OF TOP BOROUGH & CORRESPONDING NEIGHBORHOODS FOR JAPANESE RESTAURANTS

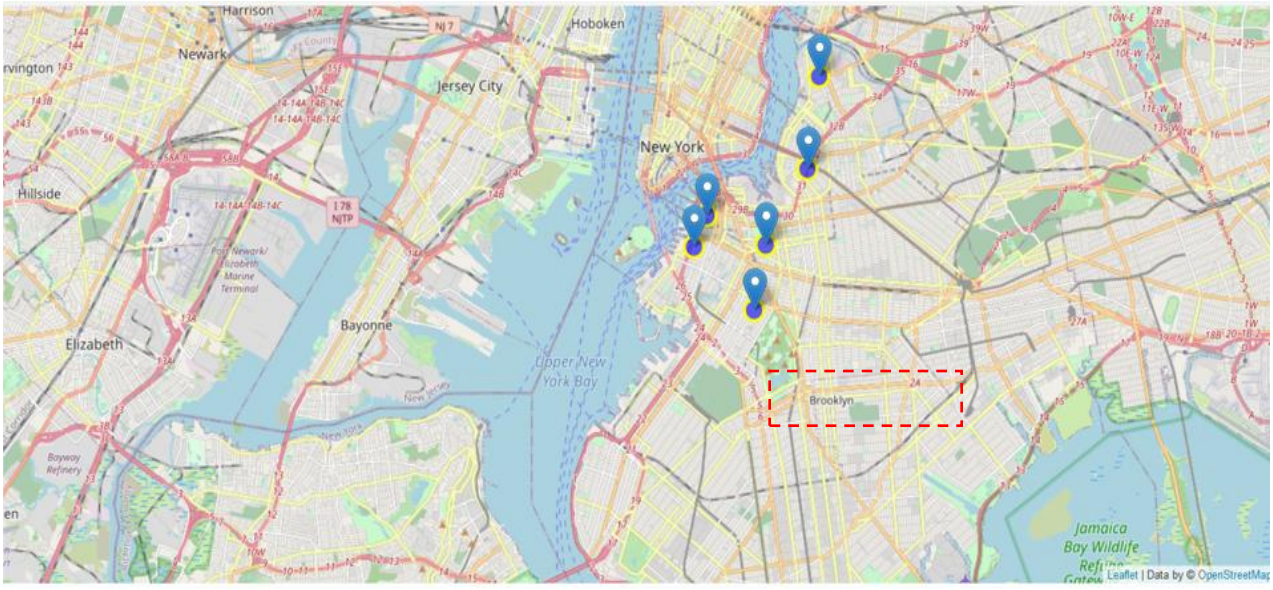
The map below had been plotted to reflect the **top-rated neighborhoods for Japanese restaurants** as dominated by the **borough of Brooklyn**:

```
ny_map = folium.Map(location=geo_location('New York'), zoom_start=12)
```

```
# instantiate a feature group for the incidents in the dataframe
incidents = folium.map.FeatureGroup()

# loop through the neighborhood and add each to the feature group
for lat, lng, in ny_neighborhood_stats[['Latitude','Longitude']].values:
    incidents.add_child(
        folium.CircleMarker(
            [lat, lng],
            radius=10, # define how big you want the circle markers to be
            color='yellow',
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        )
    )
```

```
ny_neighborhood_stats['Label']=ny_neighborhood_stats['Neighborhood']+', '+ny_neighborhood_stats['Borough']+'('+ny_neighborhood_stats['Average Rating'].map(str)+')'
```

```
# add pop-up text to each marker on the map
for lat, lng, label in ny_neighborhood_stats[['Latitude','Longitude','Label']].values:
    folium.Marker([lat, lng], popup=label).add_to(ny_map)
# add incidents to map
ny_map.add_child(incidents)
```



6. DISCUSSION: FINDINGS & RESULTS

- **Based on the data set, there are a total of 306 neighborhoods across the 5 boroughs in New York City for consideration.**
- **Queens** is the borough with the **greatest number of neighborhoods with 80**. This is followed by *Brooklyn, Staten Island, Bronx, and Manhattan*.
- In NYC, there are **191 Japanese restaurants**. The borough of **Brooklyn** has the **highest number of Japanese restaurants** with **70**. This is followed by *Queens, Manhattan, Staten Island, and Bronx*.
- The **neighborhoods with the most Japanese restaurants** are noted below:
  - **Murray Hill**: Interestingly enough, this **neighborhood name** can be found in ***2 different boroughs: Manhattan and Queens.*** **Manhattan's Murray Hill** has **5 Japanese restaurants** and **Queens' Murray Hill** has **2 Japanese restaurants**.
  - In **Brooklyn**, the Japanese restaurants are located in the neighborhoods of **Bath Beach (5), Bensonhurst (4), Clinton Hill (4).**
  - The other **Manhattan** neighborhood of **Flatiron** has 4 such restaurants.

- Overall, there is a tie among Manhattan, Queens, and Staten Island with a borough **average rating of 8.0**, followed by **Brooklyn (7.78**), and the **Bronx (7.57).** Statistically and holistically, the differences are quite negligible among all 5 boroughs.

  Hence, one then needs to look at more than just the average ratings at the **borough level**. The key differentiator requires analysis at the **next level of detail** - the **specific neighborhoods (neighborhood level)** within each borough for their average ratings.

- **Brooklyn** has 3 restaurants dominating the categories of **Maximum Likes, Ratings, and Tips**:
  - **Maximum Likes & Tips**: Williamsburg
  - **Maximum Ratings**: Greenpoint

By delving deeper at customer feedback in the form of **"Likes", "Tips", and "Ratings"** by restaurant in each of the identified borough's top neighborhoods, it has been determined that **Brooklyn dominates across these 3 categories**. In fact, it has the **most neighborhoods with average ratings of 8.5 or higher (ex. focused on high quality restaurants)**. The neighborhood of **Cobble Hill** garnered the highest average rating of **9.10**. The other 5 Brooklyn neighborhoods averaged between **8.50** and **8.80**.

The borough of the **Bronx** came in second with 2 neighborhoods **Spuyten Duyvil** and **Riverdale** with an average rating of **8.75** but Brooklyn is more highly recommended.

More importantly, there were **no neighborhoods** from the other boroughs of Manhattan, Bronx, or Queens with results that ranked over **8.5** or more.

7. CONCLUSIONS & RECOMMENDATIONS

Queens is the **borough with the most neighborhoods.** However, **Brooklyn,** with the **2nd greatest number of neighborhoods**, has the **most Japanese restaurants (marketability and demand)** and **most neighborhoods** with the **highest average ratings of 8.5 or more (favorability).**

Based on the data analysis, **Brooklyn** would therefore be the **recommended borough** with a **choice of 6 key neighborhoods** for the client to open a Japanese restaurant in New York City. (See map on page 12 above).

Brooklyn neighborhoods with **highest average ratings** include:

1. ***Cobble Hill: 9.10***
2. ***Park Slope: 8.85***
3. ***Fort Greene: 8.80***
4. ***Greenpoint: 8.65***
5. ***Brooklyn Heights: 8.65***
6. ***Williamsburg: 8.50***

The Bronx would be a **secondary choice for a borough to open a restaurant in the neighborhoods of Kingsbridge, Riverdale, or Spuyten Duyvill**; they each had an average rating of **8.75.**

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Brooklyn	Brooklyn Heights	40.695864	-73.993782	8.65
1	Brooklyn	Cobble Hill	40.687920	-73.998561	9.10
2	Brooklyn	Fort Greene	40.688527	-73.972906	8.80
3	Brooklyn	Greenpoint	40.730201	-73.954241	8.65
4	Bronx	Kingsbridge	40.881687	-73.902818	8.75
5	Brooklyn	Park Slope	40.672321	-73.977050	8.85
6	Bronx	Riverdale	40.890834	-73.912585	8.75
7	Bronx	Spuyten Duyvil	40.881395	-73.917190	8.75
8	Brooklyn	Williamsburg	40.707144	-73.958115	8.50

8. ADDITIONAL CONSIDERATIONS & OBSERVATIONS

The client can choose to compete in the top-rated Brooklyn neighborhood of **Cobble Hill** or face slightly less competition in the other 5 Brooklyn neighborhoods (***Park Slope, Fort Greene, Greenpoint, Brooklyn Heights, and Williamsburg***).

In sum, there is **great demand and a market for Japanese restaurants** in **Brooklyn;** there are **several key neighborhoods** to choose from to favorably establish a new, Japanese restaurant with quality food and service.

As a final note, all of the above analysis is dependent on the **adequacy and accuracy** of **FourSquare data**. A more comprehensive analysis and future work would need to incorporate data from **other external databases to ensure broader statistical data sources.**

Moreover, besides location, the client must also take into account other factors such as *real estate costs and availability, neighborhood safety (crime rate), office vs. residential area proximity, building size (square footage), transportation accessibility, etc.*) (Note: This deeper, detailed analysis with multiple other considerations is out-of-scope for this specific assignment.)

9. APPENDIX

- https://worldpopulationreview.com/us-cities
- https://en.wikipedia.org/wiki/Boroughs_of_New_York_City
- https://cocl.us/new_york_dataset [from nyu_2451_34572-geojson. json from https://s3-api.us-geo.objectsstorage.softlayer.net or https://geo.nyu.edu/catalog/nyu_2451_34572]
- https://foursquare.com/