

Capstone project: Battle of Neighborhoods – Korean restaurants in NYC

This post provides an overview of my final capstone project for the IBM Data Science Professional Certificate.

This post includes the following sections—Introduction, Data, Methodology, Results & Discussion, and Conclusion.

Part 1: Introduction

This Capstone project aims to help entrepreneurs and tourists with their questions regarding the best location to open their Korean restaurants or where to get the best local Korean food.

New York City is one of the world's most ethnically diverse cities and proudly hosting a wide range of international cuisines. Part of the main force driving the regional culinary scene in NYC is the international immigration. In 2018, more than four million immigrants (foreign-born individuals) comprised 23% of the New York population (*American Immigration Council data*). In addition, New York is one of the top three metro areas in the US with the largest number of Korean population. The rise of Korean food in the US comes from increasing East Asian influences on the American diet. In addition, with a large local Korean population and international tourists from Asia, Korean food received a lot of attention and popularity in recent years, which can be seen in social media and fast growing number of Korean restaurants in NYC.

Given that, the purpose of this study is to help entrepreneurs and tourists for their venue/food consideration, For a potential restaurant owner, location is key to business success. For a tourist, time is precious on a trip so restaurant-specific insights from this analysis could help avoid further research.

Part 2: Data

This study requires the following data:

1. New York data that contains all neighborhoods, boroughs, latitude, longitude based on https://cocl.us/new_york_dataset which includes the raw geo data. Coordinates are based on geocoder class of Geopy client
2. Restaurants in each neighborhood of New York: All data related to locations and quality of Korean restaurants come from the FourSquare API using the Request library in Python API

Part 3: Methodology

Overview

New York City data is from https://cocl.us/new_york_dataset and processed into a dataframe. FourSquare API provides all venues data which will be filtered by Korean restaurants. Ratings,

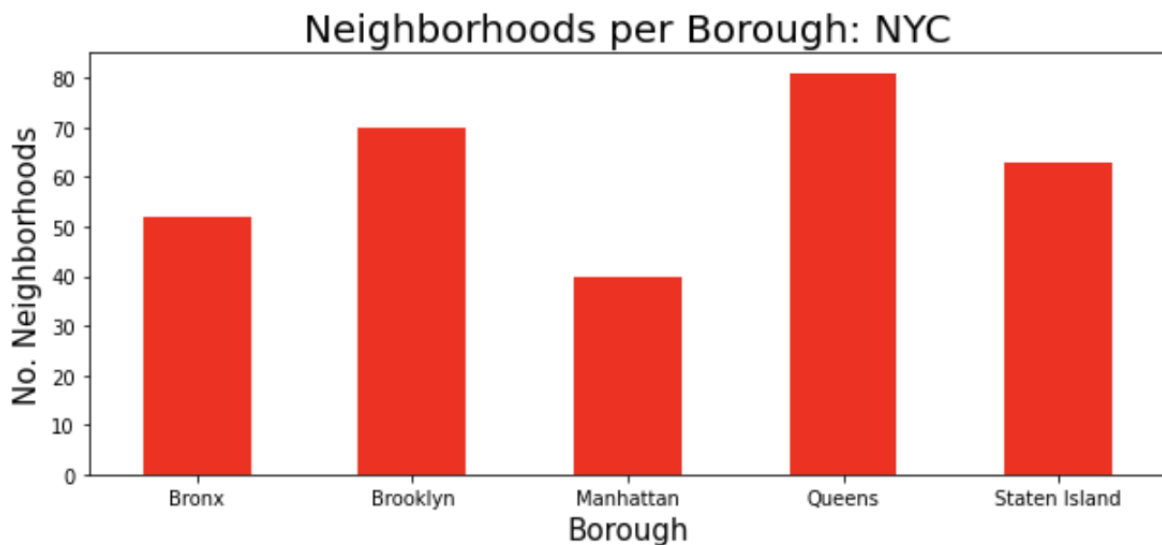
tips, and likes by users will be incorporated into a dataframe. Data will be sorted based on rankings and presented in charts and maps using various libraries.

Details:

Imported New York City data from https://cocl.us/new_york_dataset and used pandas to transform the data in the table into a dataframe containing name of the 306 neighborhoods in NYC. The first five rows as below:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

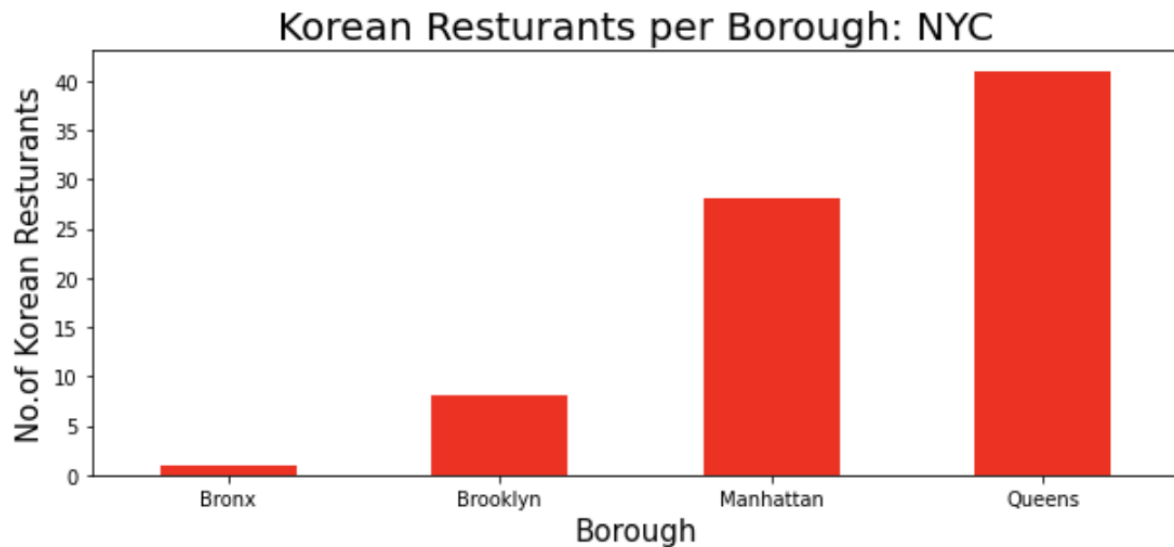
Then, let's group all the neighborhoods by borough and visualize it in a bar chart. Queens has the largest number of neighborhoods in NYC while Manhattan has the least.



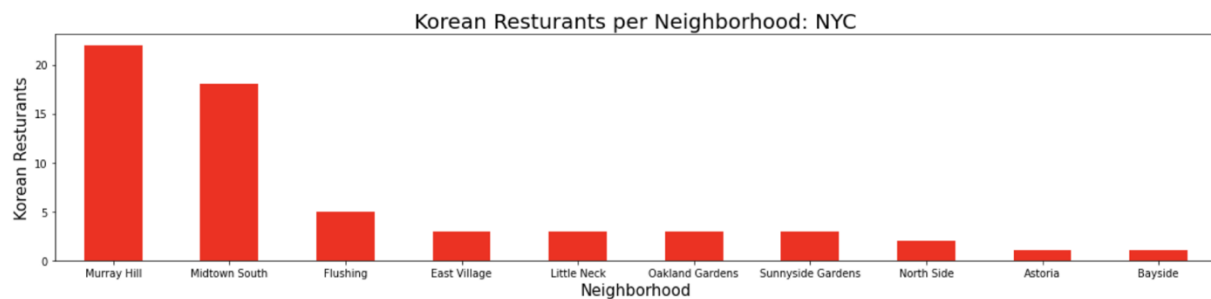
Next, we use FourSquare API to generate a list of Korean restaurants in NYC and the dataframe contains names of the 78 Korean restaurants in NYC. The last five rows as below:

	Borough	Neighborhood	ID	Name
73	Manhattan	Flatiron	591cdaf26f0aa22c76a53ff3	Cote
74	Queens	Sunnyside Gardens	4c0c2765a1b32d7f49279cf0	Sik Gaek
75	Queens	Sunnyside Gardens	57ab54da498e7d61e0c1fe0e	Pelicana Chicken
76	Queens	Sunnyside Gardens	5630104d498e08f2c7217877	Doma
77	Bronx	Kingsbridge Heights	4be747b7bcef2d7fdb1b06e5	KBBQ

Then, let's group all the Korean restaurants by borough and visualize it in a bar chart below. Queens has the largest number of Korean restaurants in NYC while Staten Island has no Korean restaurant based on FourSquare data.



Next, let's zoom in to see how many Korean restaurants the top 10 neighborhoods have. Murray Hill has the highest number of Korean restaurants in NYC and is closely followed by Midtown South. Both neighborhoods are actually located in Manhattan borough and results make sense given both neighborhoods are located in / close to Koreantown in NYC.



Next step is to get likes, ratings, and tips data for all Korean restaurants in NYC using FourSquare API and data will be cleaned and processed as a new dataframe. The last five rows of data as below:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
73	Manhattan	Flatiron	591cdaf26f0aa22c76a53ff3	Cote	251	8.3	58
74	Queens	Sunnyside Gardens	4c0c2765a1b32d7f49279cf0	Sik Gaek	79	8.1	42
75	Queens	Sunnyside Gardens	57ab54da498e7d61e0c1fe0e	Pelicana Chicken	24	7.2	7
76	Queens	Sunnyside Gardens	5630104d498e08f2c7217877	Doma	16	7.0	6
77	Bronx	Kingsbridge Heights	0	0	0	0.0	0

We then calculated count, mean, standard deviation, min, 25%-75% quartiles, and max of the following user data: Likes, Ratings, and Tips and summary table below:

	Likes	Rating	Tips
count	78.000000	78.000000	78.000000
mean	157.038462	6.976923	41.961538
std	260.096508	2.655302	61.311307
min	0.000000	0.000000	0.000000
25%	10.250000	6.925000	3.000000
50%	26.000000	7.900000	8.500000
75%	177.000000	8.400000	57.500000
max	1228.000000	9.200000	283.000000

Using the latest dataframe, we then find out BCD Tofu House has the most likes, Jongro BBQ has the highest rating, and Woorijip has the most tips.

```

Borough           Manhattan
Neighborhood      Midtown South
ID                517563f9498e0a4859d08fe4
Name              BCD Tofu House
Likes            1228
Rating           8.9
Tips             205
Name: 58, dtype: object

```

```

Borough           Manhattan
Neighborhood      Midtown South
ID                540f86da498e020149fa7676
Name              Jongro BBQ
Likes             848
Rating            9.2
Tips             158
Name: 54, dtype: object

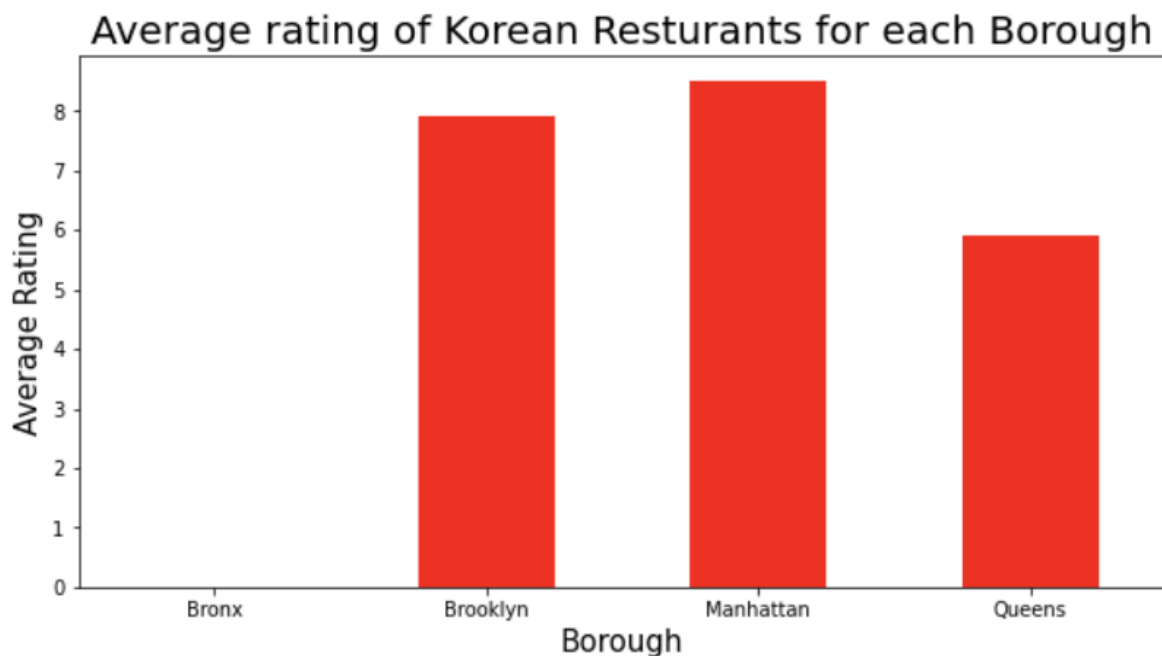
```

```

Borough           Manhattan
Neighborhood      Midtown South
ID                49becfd3f964a520e0541fe3
Name              Woorijip
Likes             929
Rating            8.4
Tips             283
Name: 56, dtype: object

```

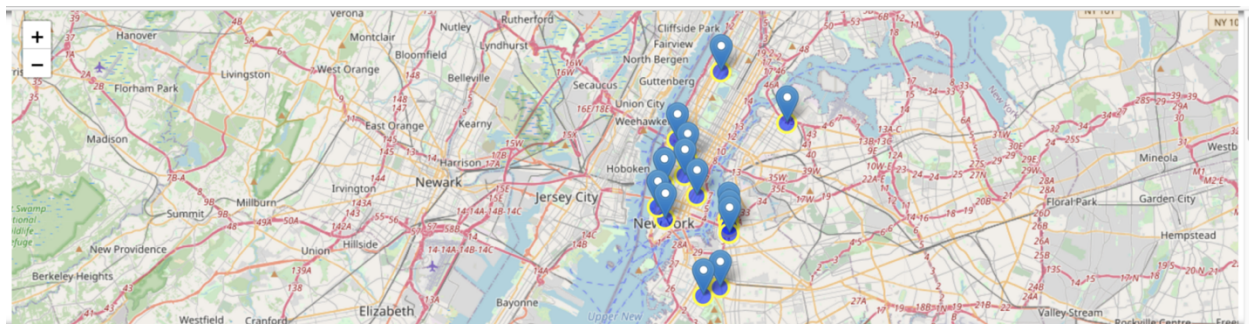
Then, we calculate the average rating of Korean restaurants by Borough and visualize it in a bar chart below. Restaurants in Manhattan have the highest average rating among other boroughs.



To zoom in further, let's map the neighborhoods with an average rating *greater than or equal to 8* after merge dataframes to include latitudes and longitudes. Manhattan again compared to other boroughs has the largest number of neighborhoods with above 8.0 average rating.

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Queens	Astoria	40.768509	-73.915654	8.400000
1	Manhattan	Civic Center	40.715229	-74.005415	8.900000
2	Manhattan	Clinton	40.759101	-73.996119	8.300000
3	Manhattan	East Village	40.727847	-73.982226	8.900000
4	Manhattan	Flatiron	40.739673	-73.990947	8.300000
5	Manhattan	Manhattan Valley	40.797307	-73.964286	8.400000
6	Manhattan	Midtown South	40.748510	-73.988713	8.483333
7	Brooklyn	North Side	40.714823	-73.958809	8.050000
8	Brooklyn	Park Slope	40.672321	-73.977050	8.900000
9	Brooklyn	Prospect Heights	40.676822	-73.964859	8.400000
10	Brooklyn	South Side	40.710861	-73.958001	8.300000
11	Manhattan	Tribeca	40.721522	-74.010683	8.900000
12	Manhattan	West Village	40.734434	-74.006180	8.300000
13	Brooklyn	Williamsburg	40.707144	-73.958115	8.300000

Results can also be clearly seen in the below map generated by Folium.



Part 4: Results & Discussion

This analysis provides an overview of the Korean restaurants in New York City and there are some interesting insights that could help potential restaurant owners and first-time tourists in the city. Let's summarize our findings based on the two kinds of target audience:

- **For people searching for Korean food, Manhattan is the best location in NYC:**
 - Manhattan has the second largest number of Korean restaurants in all of NYC even though it has the least number of neighborhoods.

- Two neighborhoods in Manhattan, Murray Hill and Midtown South, have the highest number of Korean restaurants in NYC and this suggests competition / more options for customers to choose from.
- The Korean restaurants in Manhattan have the highest average rating among other boroughs.
- Manhattan has the largest number of neighborhoods with Korean restaurants above 8.0 average rating in all NYC.
- All three restaurants with the highest rating/likes/tips in all NYC are located in Manhattan.
- **However, for entrepreneurs who want to open a Korean restaurant in NYC, Brooklyn is the best location:**
 - Brooklyn has the second largest number of neighborhoods across all boroughs in NYC, providing a potential large customer base.
 - Brooklyn has approximately a third number of Korean restaurants compared to Manhattan, proving much less local competition.
 - Among three boroughs that have Korean restaurants over 8.0 average rating on FourSquare, Brooklyn ranked #2 with 5 restaurants vs. 8 in Manhattan.
- In terms of limitation/future research opportunity, we could look into additional factors that may affect venue / food choice decisions, such as price range of the restaurant, distance between restaurant and local transportation, eating habits in nearby neighborhoods, etc. With additional tools other than FourSquare API, we could gather more information and conduct a more comprehensive analysis.

Part 5: Conclusion

Data science has many real-life implications where we can use data to conduct analysis and find insights, including this analysis. The results of the study might help an entrepreneur to decide on a potential restaurant location or a first-time visitor to NYC who is interested in trying Korean food.

Throughout this analysis, python libraries are used to clean and process data, FourSquare API is used to provide venue-related data, and folium is used to specifically create maps to show restaurant distribution. With certain limitations of this study discussed above, this analysis aims to show an *overview* of the Korean cuisine scene in NYC and hope the key insights are useful for our target audience.

Find the code in [Github](#).