THE BATTLE OF NEIGHBOURHOOD

SUBMITTED BY: SHWETA SHARMA

INTRODUCTION

New York (NY), is the most densely populated city in the United States. As per the 2018 estimates, the population of 8,398,748 distributed over a land area of about 302.6 square miles. New York City consists of five boroughs; these boroughs are the separate county of the State of New York. The five boroughs are Brooklyn, Queens, Manhattan, Bronx, and Staten Island – were consolidated into a single city in 1898. New York is also known for its diversity in terms of the immigrant population from different countries. There are over 800 languages spoken in New York City. For this reason, New York is often recognized as the world's most linguistically diverse city.

According to the Census Bureau's American Community Survey, 51 percent of the population in New York speaks only English. The remaining 49 percent speak other languages, although there are areas in the outer boroughs in which up to 25% of people speak English as an alternate language, and have limited or no English language fluency. English is least spoken in neighborhoods such as Flushing, Sunset Park, and Corona.

It is also known for its diverse culture amongst those there is one which is called food culture. There are many restaurants in New York City, which belongs to the different cuisines like Chinese, Indian, French, Mexican, Korean, etc.

In the part of this project, I will visualize and analyze all major parts of New York City which have Korean restaurants for recommendation to open Korean restauranttar. Also to get Korean and non Korean customer as this place has advantage of diversity.

Data

The following data :

· New York City data contains the list of Boroughs, Neighborhoods along with their latitude and longitude.

▫ *Data source: https://cocl.us/new_york_dataset*

▫*Description: This dataset has all required information. I will use this dataset to explore various neighborhoods of New York City.*

· Korean restaurants in each neighborhood of New York City.

▫ *Data source: Foursquare API*

▫*Description: with the help of API, I will acquire the information of all the venues in each neighborhood. After this step, I will extract information about only Korean restaurants.*

· Geo Space data

▫ *Data source: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm*

▫*Description: The geospace data will help to get information about the New York Borough boundaries to visualize choropleth map.*

Methodology

Analysis (Tools)

To Analysis the data I will use some built-in and non built- in fuctions:

· *pyforest to import all data science libraries for handling data.*

· *Request module to use Foursquare API.*

· *Geopy to get co-ordinates of the City of New York.*

Approach to find the Result:

· Collect the new York city data from **https://cocl.us/new_york_dataset**
· Using Foursquare API to find all venues for each neighborhood.
· Extract information about all venues that are Korean Restaurants.
· Find out Tips, rating, and like count for each Korean Restaurants with the help of Foursquare API.
· Use rating for each restaurant to sort that data.
· Visualize the Ranking of neighborhoods by using folium library(python)

Questions that can be asked with this datasets

· *First I go with, which areas have potential Korean Restaurant Market?*

· *Then list out all those areas, which are lack of Korean Restaurant?*

· *What is the best location in New York City for Korean Cuisine?*

· *Which is the best place if I preferred Korean Cuisine?*

In this project, I use single line code PYFOREST to import all libraries of data science. Then I go for To get the geocodes with define function i.e latitude and longitude of a given location using geopy. After this step, I have use the define a function to interact with Four Square API and get top 20 venues within a radius of 1000 Meters for a given latitude and longitude with return function to get venue id , venue name and category.After getting the details venue id etc..next In this step with the help of define function I extract the information about venue details like rating ,likes,names, tip counts for a given venue id. which will be used for ranking. Next i take a

help of define a function to get the New York City data such as Boroughs, Neighborhoods along with their latitude and longitude. With request library I got the New York dataset which show its has 306 different neighbourhood in the NEW YORK CITY.

It shows that there is 306 different Neighborhoods in New York

	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

Number of Neighborhood for each Borough in New York City

The above figure showed that the how many neighbourhood in each borough such Bronx,Brooklyn etc..has in the New York city. This data shows that QUEENS has the highest Neighbourhood.
After this I have prepare the neighborhood list that contains Korean Restaurant.

This data have got all the Korean restaurants in New York city for analyzation.

```
     Borough      Neighborhood                        ID  \
0    Brooklyn     Brighton Beach   4c9d5c0303133704a96f5ed5

1    Brooklyn     Prospect Heights  4fa162e0e4b0badc81404a51

2    Brooklyn     Williamsburg     52bf3053498e754b09a440b5

  3 Manhattan     East Village     578bec6c498e3150fc369f3b

  4 Manhattan     Manhattan Valley  56a14149498e9983c0199038
                                Name
1   Cafe At Your Mother-in-Law
2              Kimchi Grill
3                    Dotory
4          Thursday Kitchen
5                     Mokja
```


Number of Korean Resturants for each Borough in New York City

The above figure shows that among each borough of New York City the QUEENS has the largest number of Korean restaurants

Number of Korean Resturants for each Neighborhood in New York City
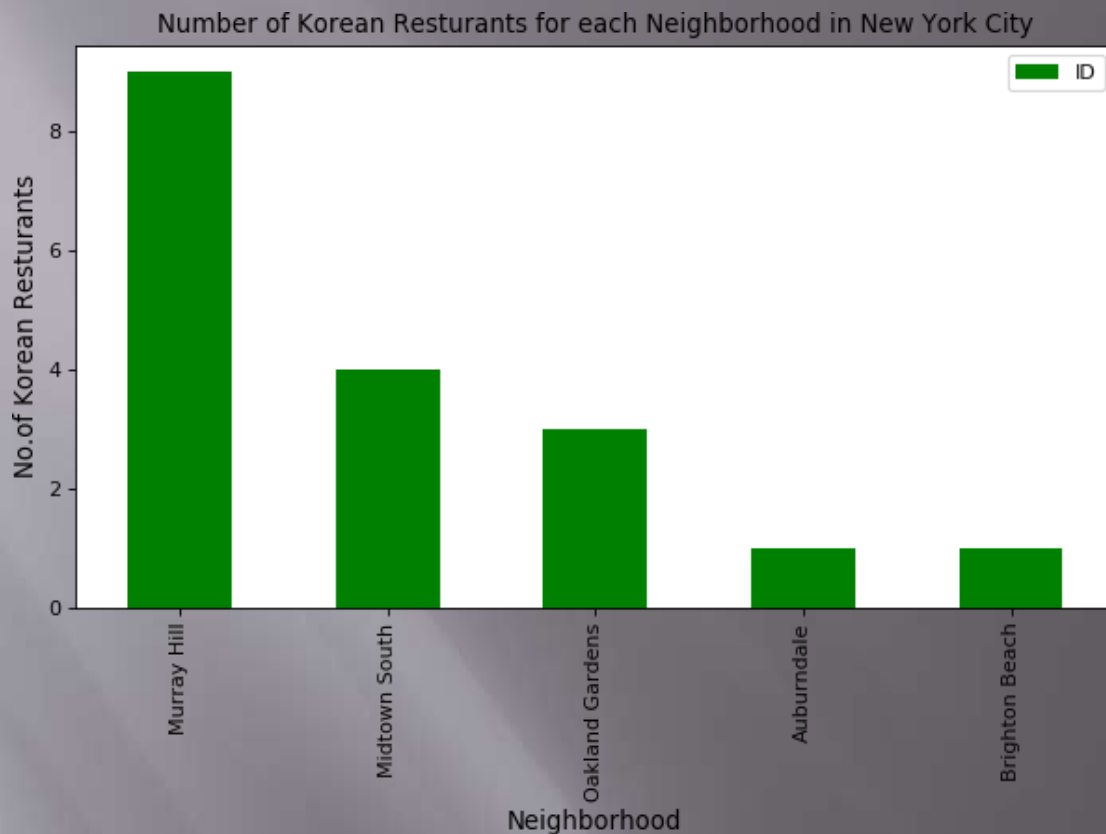
So Murray Hills in Queens has the largest number of Korean Restaurants with a total count of 9.

	Borough	Neighborhood	ID	Name
11	Queens	Murray Hill	4b830e44f964a520ebf4 30e3	Hahm Ji Bach –
12	Queens	Murray Hill	4b8f0931f964a520d045 33e3	Mapo BBQ
13	Queens	Murray Hill	49e10e0bf964a5208961 1fe3	Mad For Chicken
14	Queens	Murray Hill	4baeabe6f964a52037cd 3be3	Kum Sung Chik Naengmyun
15	Queens	Murray Hill	4c7ad3dda86837048bee 144d	Geo Si Gi Restaurant
16	Queens	Murray Hill	5518a1a8498e4c7fe626 8b88	Jongro BBQ
17	Queens	Murray Hill	5590598a498eb1c3f975 b6a6	Mr. Tofu
18	Queens	Murray Hill	56db53cc498eed9b6862 1019	SGD TofuHouse
19	Queens	Murray Hill	4ad3bfabf964a52016e6 20e3	Han Joo

After that I have prepared list with borough, neighbor, Id, name, likes, ratings tips.
Also assigned the zero value to those restaurants which doesn't exist in Four-Square API to extract the accurate results.

	Borough	Neighborhood	ID \
0	Brooklyn	Brighton Beach	4c9d5c0303133704a96f5ed5
1	Brooklyn	Prospect Heights	4fa162e0e4b0badc81404a51
2	Brooklyn	Williamsburg	52bf3053498e754b09a440b5
3	Manhattan	East Village	578bec6c498e3150fc369f3b
4	Manhattan	Manhattan Valley	56a14149498e9983c0199038

	Name	Likes	Rating	Tips
0	Cafe At Your Mother–in–Law	29	7.5	28
1	Kimchi Grill	289	8.4	108
2	Dotory	138	8.5	44
3	Thursday Kitchen	274	8.9	72
4	Mokja	23	7.9	6

So we got data for all restaurants now i will save this data to a CSV sheet. In case we by mistake modify it. It better to refer to saved data sheet CSV if required
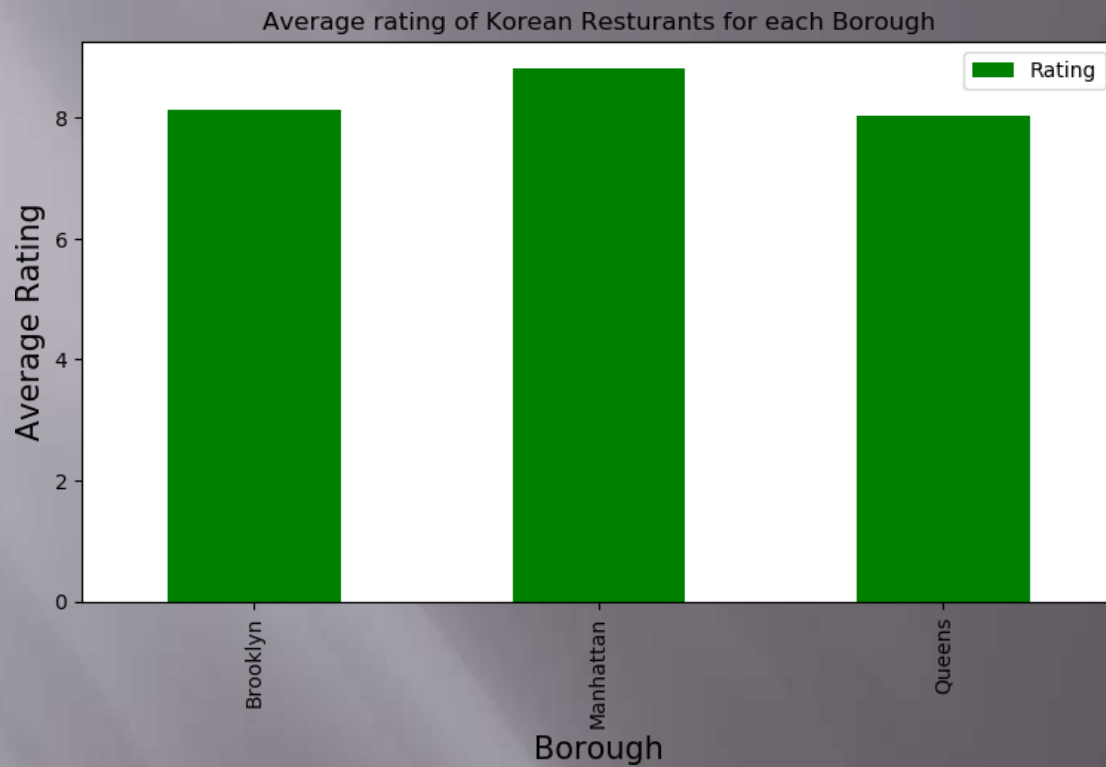
	Borough	Neighborhood	ID \
1	Brooklyn	Brighton Beach	4c9d5c0303133704a96f5ed5
2	Brooklyn	Prospect Heights	4fa162e0e4b0badc81404a51
3	Brooklyn	Williamsburg East	52bf3053498e754b09a440b5
4	Manhattan	Village	578bec6c498e3150fc369f3b
5	Manhattan	Manhattan Valley	56a14149498e9983c0199038

	Name	Likes	Rating	Tips
1	Cafe At Your Mother–in–Law	29	7.5	28
2	Kimchi Grill	289	8.4	108
3	Dotory	138	8.5	44
4	Thursday Kitchen	274	8.9	72
5	Mokja	23	7.9	6

We see that values like Tips and likes are string values. We need to convert them into float for further analysis.Also extract the data of restaurant with the highest rating, likes and tips to visualize neighborhood and Borough with maximum average rating of restaurants. In the next step we need to visualize neighborhood and each Borough with maximum average rating of restaurants.

	Neighborhood	Average Rating
7	Midtown South	9.025000
3	East Village	8.900000
0	Auburndale	8.600000
12	Williamsburg	8.500000
4	Flushing	8.400000
10	Prospect Heights	8.400000
11	Sunnyside Gardens	8.300000
8	Murray Hill	8.111111
6	Manhattan Valley	7.900000
2	College Point	7.800000

	Borough	Average Rating
1	Manhattan	8.816667
0	Brooklyn	8.133333
2	Queens	8.023529

Average rating of Korean Resturants for each Borough

Next I consider all the neighborhoods with average rating greater or equal 8.8 to visualize on map. I also added in next step the longitude and latitude with this dataset to show data on map.

	Borough	Neighborhood	Latitude	Longitude	Average Rating
1	Manhattan	East Village	40.727847	–73.982226	8.900
2	Manhattan	Midtown South	40.748510	–73.988713	9.025

There are two restaurant in Neighborhood based on average rating greater than more Equal to 8.8
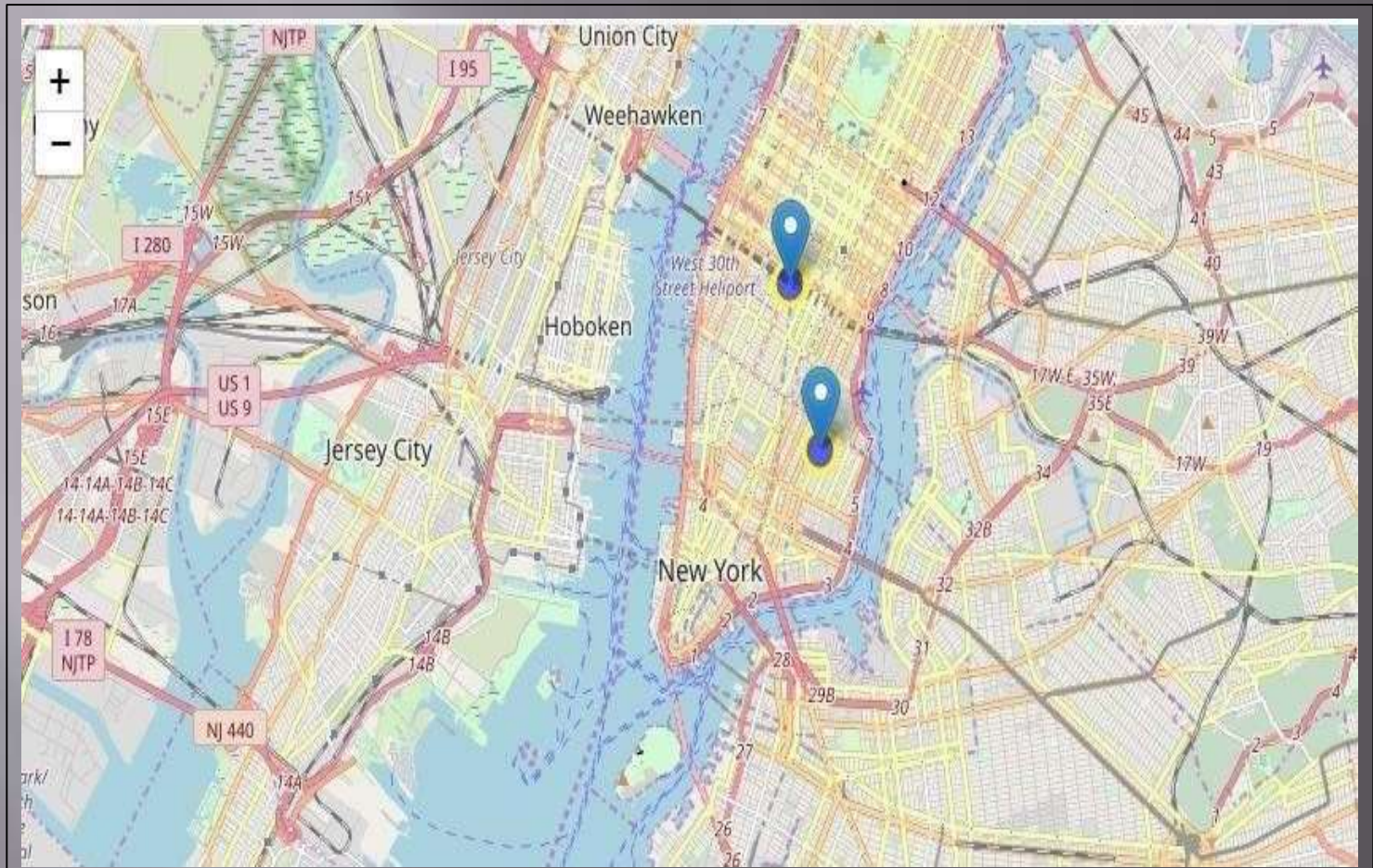
After visualized the Neighborhoods. Next ,visualize boroughs based on average ratings

DISSCUTION:

- **I feel this capstone project has provide the opportunity to understand and apply these data science tools and algorithms in more appropriate manner.**
- **In this project I found New York is cultural diverse city which has many Different cuisine restaurant, which mean this area has more opportunity For Korean restaurant to open.**

CONCULSION (RECOMMENDATION):

East village (Manhattan), Midtown South (Manhattan) are some of the best neighbor- hoods for Korean Cuisine.

Manhattan has potential Korean Restaurant Market.

Brooklynand Queens Ranks stand same inaverage rating standard of Korean Restaurants.

Manhattan is the best place to stay if you prefer Korean Cuisine.