Coursera Capstone

IBM Applied Data Science Capstone

The Battle of Neighborhoods: Opening a Pizza Place in New York City

By: Mohamad Farid bin Gani

December 2020

Introduction

Pizza can be traced back to Naples, Italy in late 18th century. Although the origin of the word pizza is not clear, the Italians are credited with coining the term pizza. According to Carol Helstoski, pizza was brought to the United States with Italian immigrants in the late nineteenth century and first appeared in areas where Italian immigrants concentrated. Lombardi's, the country's first pizzeria is opened in 1905.

And fast forward, pizza has been one of America's favorite Italian foods since it was introduced to the United States in the early 1900's. Since then, it has become increasingly popular across the country. Every restaurant has its unique way of preparing pizza and each one is always coming out with new and appetizing products.

As for this project, I will explore the best locations for pizza place around the city of New York. **The target audience for this would be entrepreneur or potential investors who are looking to open a pizza place at New York City.** Since pizza is one of the most famous foods in United States, the business owner potentially can have great success and encouraging response from the people. That being said, opening a new restaurant requires serious considerations need to be taken; such as location, target customers, competitors, and others. In general, the location of the premise is one of the crucial factors that will affect whether it will make it or break it. Hence, a good decision making is very important here, since there are a lot of pizza place in New York City.

This project will provide an answer for the business problems that investors need to be considered; Which part of New York should I focus on to open a pizza place? What neighborhood would get a good reception to open a pizza place?

Data

**To solve the problem, we will use data from the following source:**

1. https://cocl.us/new_york_dataset - Data containing the neighborhoods and boroughs of New York City.
2. Python Geocoder - Data of latitude and longitude coordinates of the neighborhoods.
3. Foursquare API - Data containing the venue and location of several restaurants at New York City.

First, we need to get the data of neighborhoods and boroughs of New York City, which will be obtain from this source; https://cocl.us/new_york_dataset. Second, using Python Geocoder, we need to label those neighborhoods and boroughs with geographical coordinates, which is latitude and longitude. Lastly, with databases of more than 100 million places, we will work with Foursquare API to get the data that containing all data related regarding pizza place at New York City. Some of the Python packages that will be used to handle the data includes; Pandas, NumPy, Geopy, Folium, and others.

Methodology

As stated previously, data will be collected from this link, https://cocl.us/new_york_dataset . Then we will process it into a dataframe. Then we will use data from Foursquare to locate all venues in New York City. After then, we will filter the result by looking for venue of pizza place. Also, we will use data of ratings, tips, and likes from the users of Foursquare and make another dataframe. All the data is prepared for analyze purpose. To get a better picture, we will use visual using Python library, namely matplotlib and Folium.

Result

Neighborhoods for Every Borough in NYC
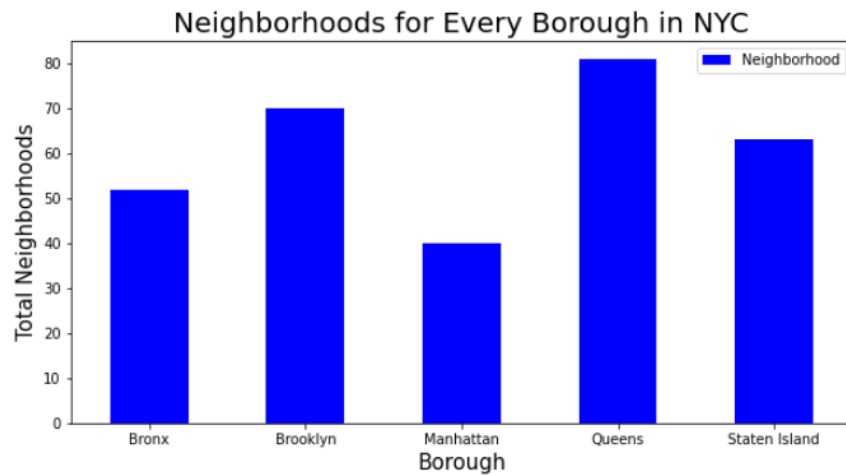
After extract data, we can see that Queens have the most neighborhoods, while Manhattan has the least neighborhood.

Pizza Place for Every Borough in New York City
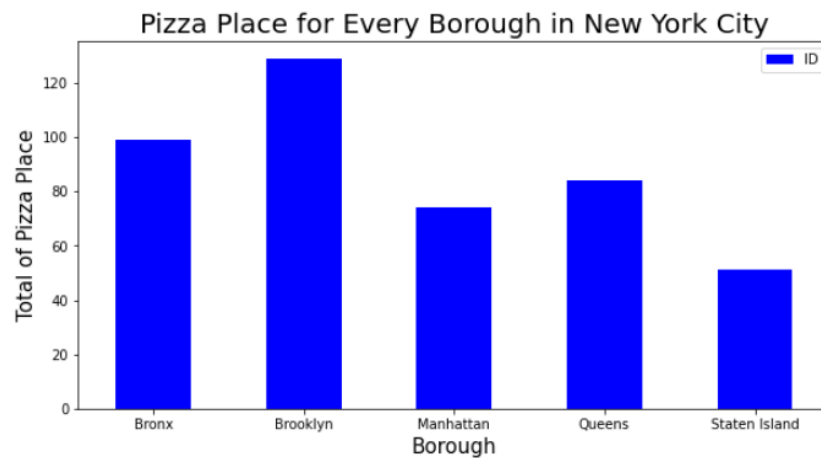
While Queens has the most neighborhood, it seems like there are not much pizza place here, compared to Brooklyn and Bronx. Here shows that Brooklyn has the most pizza place here.

Among neighborhood, we will calculate their average likes, rating, and tips based on number of pizza place at each neighborhood.
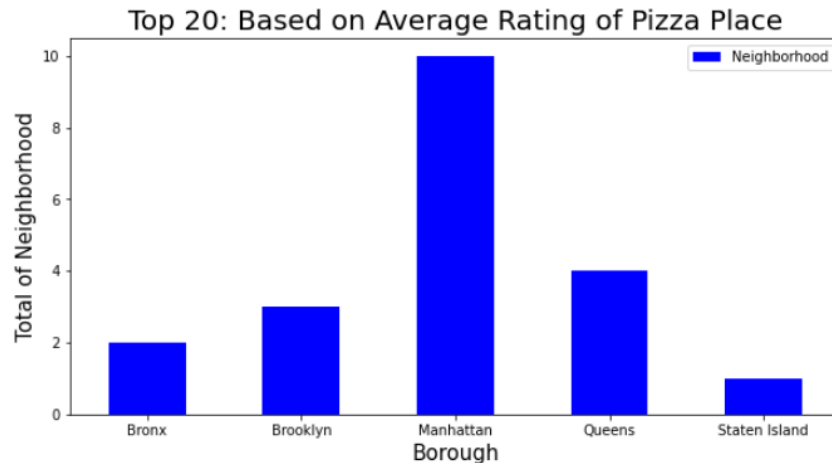
And from 20 neighborhood that have highest value of average rating, 10 of them is from Manhattan. Which is suprising given that Manhattan is the 2nd lowest number of pizza place among borough.
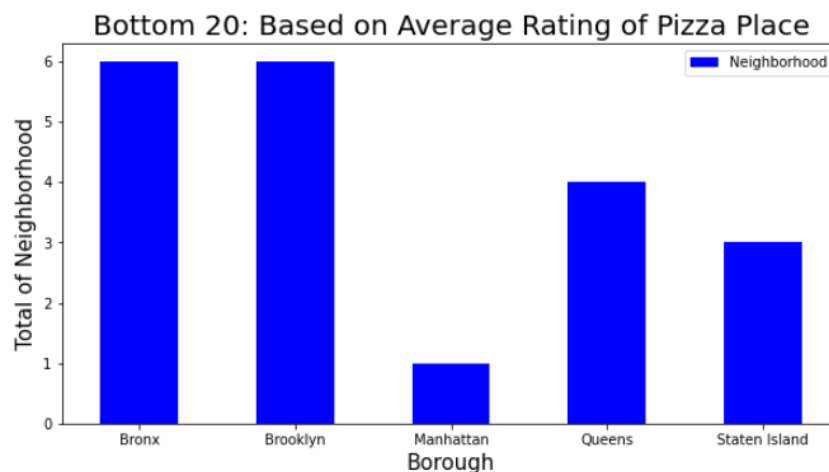
Quality over quantity. While Brooklyn and Bronx have the most pizza place, their neighborhood have some of the lowest average rating given by user of Foursquare. An observation; in terms of rating, Manhattan have the best rated pizza place around their neighborhood.

Top 20: Based on Average Likes of Pizza Place
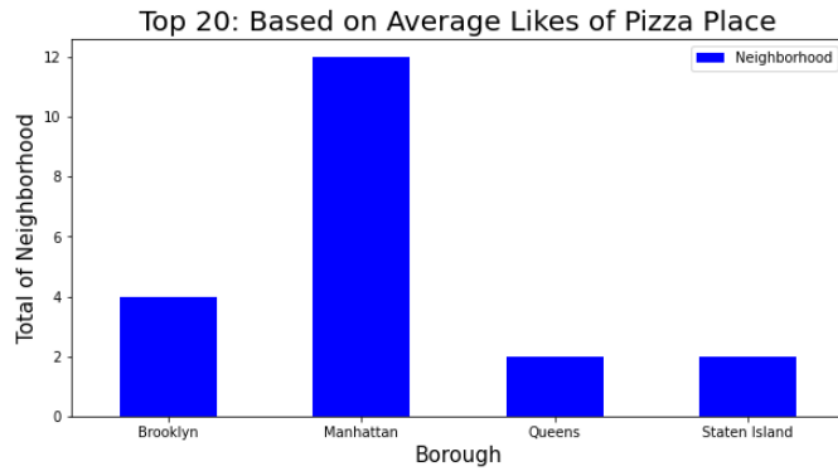
For highest like, no Bronx here, although they have more pizza place than Manhattan, Queens, and Staten Island. From top 20 of neighborhood, 12 of them is from Manhattan.

Bottom 20: Based on Average Likes of Pizza Place
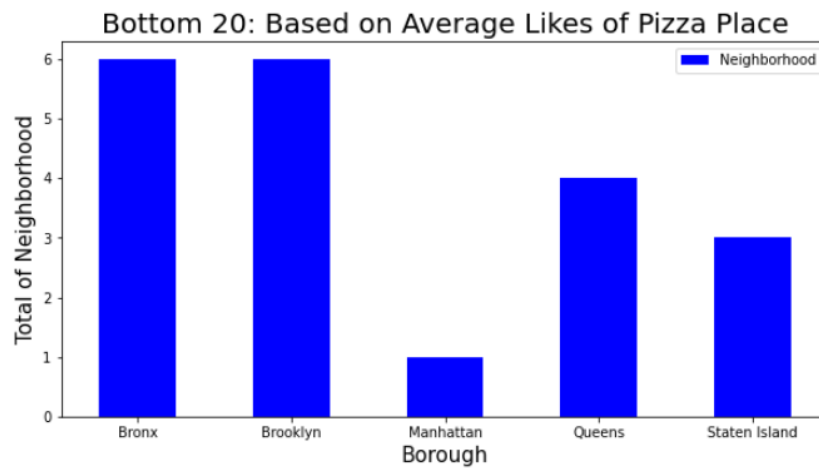
A similar pattern of bar chart compared to previous chart of average rating of bottom 20. Bronx and Brooklyn have 6 neighborhood that have least average like of pizza place. New business of pizza place should take advantage of these neighborhood.

	Neighborhood	Average Likes	Average Rating	Average Tips	Borough	Latitude	Longitude
58	Elmhurst	62.000000	9.200000	25.00	Queens	40.744049	-73.881656
32	Chinatown	239.000000	9.100000	61.00	Manhattan	40.715618	-73.994279
67	Forest Hills Gardens	194.000000	9.100000	95.00	Queens	40.714611	-73.841022
129	North Side	284.800000	9.020000	58.20	Brooklyn	40.714823	-73.958809
24	Carnegie Hill	176.000000	9.000000	61.50	Manhattan	40.782683	-73.953256
155	Soho	794.500000	8.950000	127.50	Manhattan	40.722184	-74.000657
46	Cypress Hills	21.000000	8.900000	13.00	Brooklyn	40.682391	-73.876616
57	Elm Park	225.000000	8.900000	110.00	Staten Island	40.630147	-74.141817
70	Fresh Meadows	14.000000	8.900000	9.00	Queens	40.734394	-73.782713
81	Greenwich Village	321.000000	8.900000	111.00	Manhattan	40.726933	-73.999914
75	Gowanus	108.000000	8.800000	24.50	Brooklyn	40.673931	-73.994441
111	Midtown	188.000000	8.800000	44.50	Manhattan	40.754691	-73.981669
169	Tudor City	143.000000	8.800000	51.00	Manhattan	40.746917	-73.971219
105	Manhattan Valley	111.500000	8.700000	52.00	Manhattan	40.797307	-73.964286
131	Oakland Gardens	27.000000	8.700000	10.00	Queens	40.745619	-73.754950
53	East Village	249.750000	8.675000	92.75	Manhattan	40.727847	-73.982226
100	Little Italy	533.666667	8.666667	85.00	Manhattan	40.719324	-73.997305
126	Noho	88.500000	8.625000	21.50	Manhattan	40.723259	-73.988434
19	Bronxdale	47.000000	8.600000	27.00	Bronx	40.852723	-73.861726
180	Woodlawn	21.000000	8.600000	8.00	Bronx	40.898273	-73.867315

Here is the top 20 neighborhood with highest average rating of pizza place given by user on Foursquare. Seems like most of it located at Manhattan.

	Borough	Average Likes	Average Rating	Average Tips
0	Bronx	13.658228	7.162025	7.645570
1	Brooklyn	112.765766	7.612613	42.648649
2	Manhattan	172.958333	8.193056	52.694444
3	Queens	27.111111	7.519444	13.333333
4	Staten Island	26.500000	7.450000	13.968750

Here is the figure for average likes, rating, and tips for 5 borough in New York City.

Discussion

On average, Manhattan has the best rated pizza place. The fact that they have less number of pizza place, shows the quality over quantity. In terms of quality, Manhattan and Brooklyn have better pizza place compared to other borough.

Assuming the number of likes from Foursquare is from visitors, people love to go to pizza place at Manhattan and Brooklyn. Judging from the number of likes, there are fewer visitors at Bronx, Queens, and Staten Island compared to the other 2. Especially for Bronx, since they have the 2nd highest of pizza place among 5 boroughs. Apparently, Bronx has the least likes for pizza place. In terms of visitors, again, Manhattan and Brooklyn is the place to go.

Conclusion

As a result, I would suggest to open the pizza place at Manhattan and Brooklyn. Seems like this borough get encouraging responses from the visitors. However, as both boroughs have the highest average rating, one need to wonder if our products and services can compete with another pizza place that already operating here. A safety measure, if the businessman do not want to compete with those establish pizza place, one can start at neighborhood that have a lower average rating; which are Lower East Side (Manhattan) and Madison, Rugby, Gravesend, Gerritsen Beach, City Line, Homecrest (Brooklyn).

Worth to mention that all of the above analysis is depending on the accuracy of data on Foursquare. A more comprehensive analysis and future work would need to incorporate data from other external databases.