Capstone Project - The Battle of Neighborhoods

Table of contents

Business Problem

The aim of this project is to find a safe and secure location for opening of Italian restaurants in the city of New York. New York is the most populous city in the United States with an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km2) which makes it the most densely populated major city in the United States(Wikipedia). New York City has the largest population of Italian Americans in the United States of America as well as North America. Over 1.3 million Italians and Italian-Americans live in the greater New York City metro area, with about 800,000 living within one of the five New York City boroughs. For this big population the need to find and enjoy Italian cuisine is on the rise. We will make use of our data science tools to analyze data and focus on the safest borough and explore its neighborhoods to find the best Italian restaurants in the city of New York. Also we are going to cluster neighborhoods in order to recommend the best locations of opening Italian restaurant and the best places to enjoy Italian cuisine.

Data Selection & Source

Based on definition of our problem, factors that will influence our decision are:

Finding boroughs to include boundaries, latitude, longitude, restaurants, and restaurant ratings
Finding the most common venues for easy access
Selecting the right neighborhood within the borough

Data source: https://cocl.us/new_york_dataset & https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

To explore and target recommended locations across different neighborhood boundaries we will access data through Foursquare API interface and arrange them as a data frame for visualization.

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which we can send requests. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood Names, Latitude, Longitude
2. Venue Name, ID, Ratings, Tips, Likes

Methodology

The Methodology section will describe the main components of our analysis and predication system. The Methodology section comprises six stages:

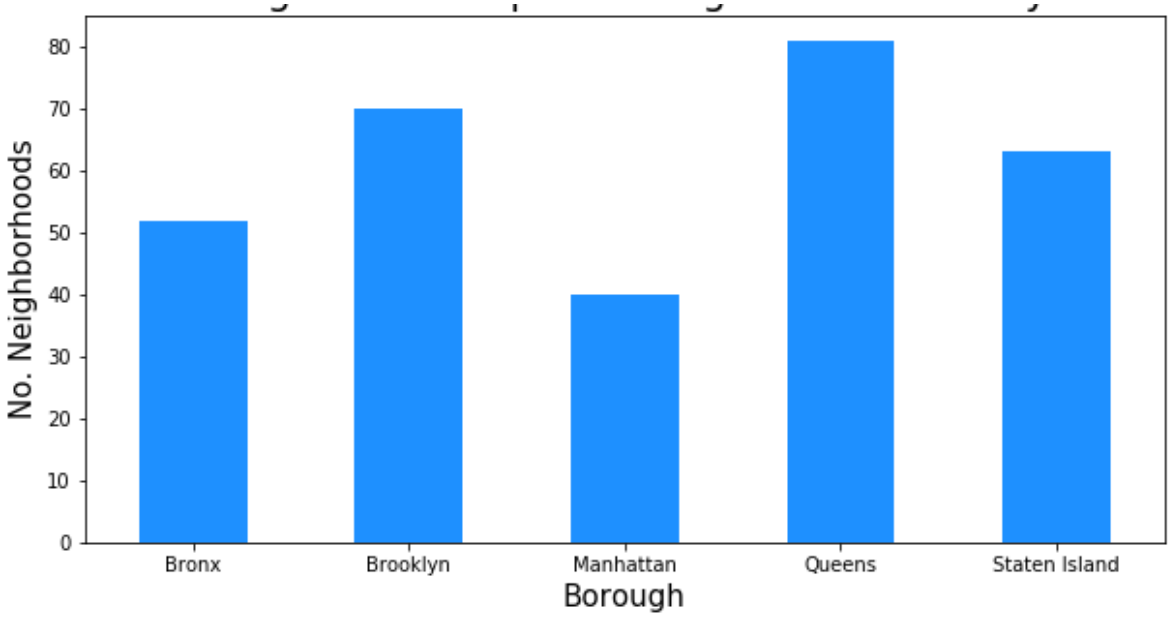
1. Collect Inspection Data
2. Using Foursquare locating all venues filtered by Italian restaurants. Ratings, tips, and likes by users
3. Explore and Understand Data
4. Data preparation and preprocessing
5. Modeling
6. Data Visualization using various Python libraries.

BeautifulSoup is used to scrape boroughs from Wikipedia and organize a table containing Neighborhoods information of New York City and **Geopy** is to get the geological location of each community board.

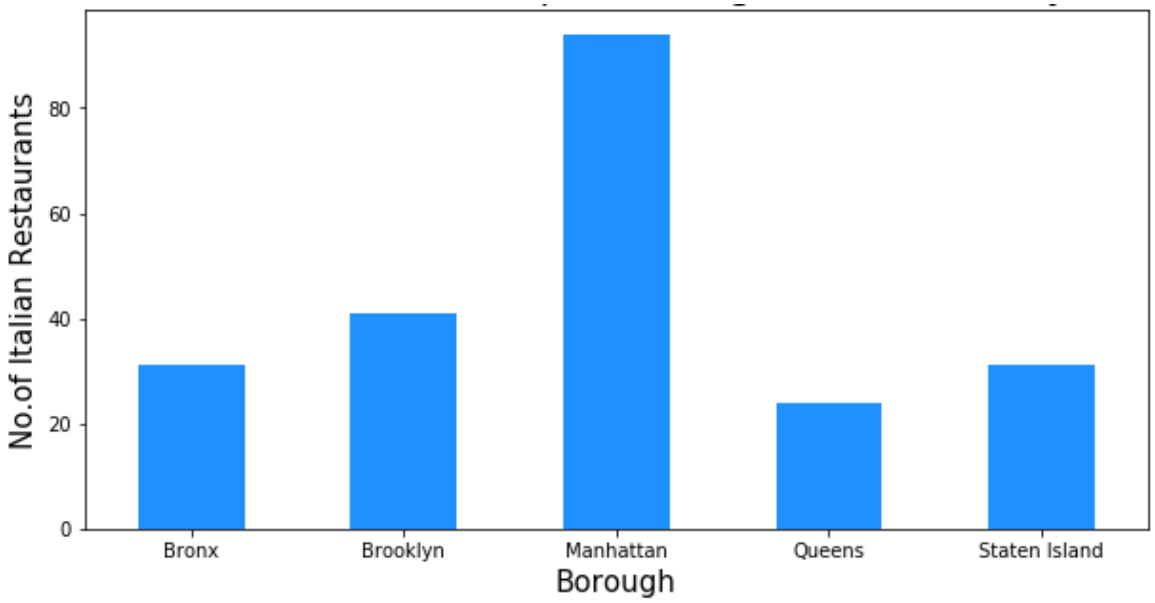
Foursquare API is used to explore the boroughs and segment them. I designed the limit as 100 venues and the radius 400 meters for each borough from their given latitude and longitude information. Here is the header of the result, adding venue id, venue name, category, latitude, and longitude information from Foursquare API.

	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

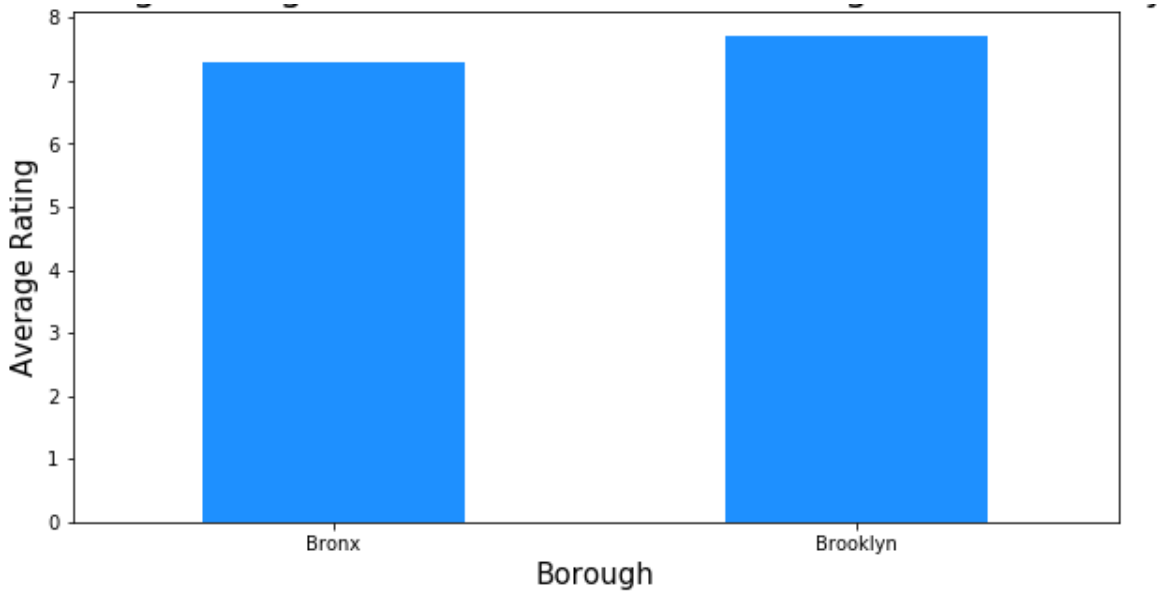
306 rows of data has been returned with Borough, Neighborhood, Latitude and Longitude. Now that we have our initial data, let's start analyzing it. Then we found Queens has most Neighborhoods. We can see below that Manhattan has the least number of neighborhoods in all five boroughs.



After that we prepared neighborhood list that contains Italian restaurants in Boroughs of New York city and saved the information so far to a .csv file due to limited calls on **Foursquare.** As we continue our analysis, we see below that although Manhattan had the least number of neighborhoods, it does have the highest number if Italian restaurants. Additionally, we see how many restaurants the top 6 neighborhoods have. The neighborhood of Belmont has the highest number of Italian restaurants in all of NYC and is actually located in the borough of Bronx and Manhattan.



Then we calculated average rating of Neighborhoods per Borough and got the result that Neighborhoods of Brooklyn has average better rating than Bronx although the Belmont, Neighborhood of Bronx has the most number of Italian restaurants.



	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Brooklyn	Bushwick	40.698116	-73.925258	9.4
1	Bronx	City Island	40.847247	-73.786488	8.2
2	Brooklyn	Cobble Hill	40.687920	-73.998561	8.5
3	Bronx	Morris Park	40.847549	-73.850402	8.5
4	Bronx	Pelham Gardens	40.862966	-73.841612	8.7
5	Bronx	Woodlawn	40.898273	-73.867315	8.6

Results& Decision

Here we can conclude that borough wise Manhattan has most Italian restaurant. On the other hand Bronx has as twice as many in half the number of restaurants in Manhattan. After analyzing the data we found that Bronx's Belmont neighborhood has more restaurants than Manhattan neighborhood Lenox Hill. Considering Likes and Tips, Belmont is much ahead out of other neighborhoods. Most importantly, Bushwick neighborhood has highest average rating than other neighborhood in Brooklyn. In Bronx, Woodlawn and Pelham Garden's restaurant have highest average ratings than other neighborhoods.

Finally, we drew the conclusion that Bronx is the best option to open an Italian restaurants. In Bronx, most restaurants in neighborhood have on average 8.5 rating out of 10. Finally, I would go to Belmont in Bronx for the best Italian food based on 145 likes.

Conclusion

To sum up, finding best places to open Italian restaurant throughout the city in New York we included boundaries, latitude, longitude, restaurants, and restaurant ratings and sorted the most common venues for easy access and finally selected the right neighborhood within the borough .Although in this report, it demonstrates the relations between location and ratings, but ratings might not reflect the operation status of the restaurant. A restaurant with a high rating could still be unprofitable, which is unsuccessful from a business perspective As a final note, all of the above analysis is depended on the adequacy and accuracy of Four Square data. To suggest more practical and profitable ideas, the relationship between customer reactions and financial performance should be evaluated.