The Battle of Neighborhoods

Capstone Project

1. Introduction

1.1 Background

New York is a major central city for diversity since many people from different cultural atmospheres have brought their families and dreams to New York City. The city has consistently seen people from around the world move to the city and call it home. It has been a center for trade and economic growth. It is the most populous city in the United States with an estimated 2019 population of 8,336,817. New York is known world wide as a cultural melting pot. While other states have had immigration surges, none have compared to the diversity and sheer number of immigrants that have made their way to the City. So these number of varies cultures combined to create a great diversity for itself. Since People from all over the world tend to come up here, we can see some many of their cultural aspects Transport, Food, Clothing, and so on...

1.2 Problem

Since there are so many restaurants or cuisines in NewYork city, it is quite difficult to set up a restaurant and be successful. Several factors are to be considered before opening a restaurant in a particular area. Some of them are:

- Market Places
- Competitors in particular location
- Aiding places that make people come to restaurants like Gym, Entertaining Public places
- Population
- Demographics

And so on...

The objective of this project is to find the best neighborhood to open a Italian restaurant with low risk and high success rate and to explore competitors in that Neighborhood by applying exploratory data analysis.

1.3 Target Audience

Target Audience are the ones who want to open an Italian Cuisine restaurant and want to know about the many criteria that leads to come to a data driven conclusion for selecting the best neighborhood for opening one. But some of the criteria we are gonna be focusing on in our project is as follows,

- Best Neighborhood for opening Restaurant (Italian Cuisine) in New York.
- Area which lack Italian Restaurants.
- Neighborhoods with best Italian Cuisine.
- Boroughs with Potential Italian Restaurants.

This report can also help tourists and the people who recently shifted to Newyork who want to know which place is good for Italian restaurants to try in Newyork

2. Data

Since we will be focusing on New York for opening our restaurant, we will be gathering the data for the same and the three main data we are gonna use are the New York Neighborhood data, data about the cuisines in New York and the Geo spacial data of boundaries of the Neighborhoods in New York.

2.1 Data 1

New York Neighborhood Data which will be used to get to know about the various neighborhoods which are gonna take into consideration for our Project. This data is in json format and it is converted to dataframe using pandas library and performed analysis on the data using basic techniques. From the data, we found that there are 5 boroughs and 306 neighborhoods in NewYork

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

2.2 Data 2

Geo-spacial data of the New York to get a better understanding of the neighborhoods in it and their corresponding locations in the Folium map would make certain things clear for the Project. This will be achieved using the acquired data and visualize the same using Choropleth maps.

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

From the the link provided above, geojson file of NewYork boundaries is downloaded and used for visualizing boroughs with average rating through CHOROPLETH maps

2.3 Data 3

Data from the Foursquare API is acquired to know about the various restaurants in each neighborhood of New York city especially Italian cuisine, which will be used for acquiring the information regarding same for all the venues of each neighborhood. In this data we consider only Italian restaurants by filtering venue categories. Using the foursquare API we can also get the specific venue details like ratings,tips and likes which are used for finding best restaurants in the neighborhood

Data source : https://foursquare.com/developers/apps

3. Methodology

3.1 Exploratory Data Analysis

1. Download and explore newYork city dataset

In order to segment the neighborhoods of New York City, a dataset is required that contains the 5 boroughs and the neighborhoods, that exist in each borough, with respective latitude and longitude coordinates. This dataset is downloaded using the mentioned URL.

Once the .json file is downloaded, it is analyzed to understand the structure of the file. A python dictionary is returned by the URL and all the relevant data is found to be in the features key, which is basically a list of the neighborhoods. The dictionary is transformed, into a pandas dataframe, by looping through the data and filling the dataframe rows one at a time.

As a result, a dataframe is created with Borough, Neighborhood, Latitude and Longitude details of the New York City's neighborhood.

Upon analysis, it is found that the dataframe consists of 5 boroughs and 306 neighborhoods.

Further, 'geopy' library is used to get the latitude and longitude values of New York City, which was returned to be Latitude: 40.71, Longitude: -74.01

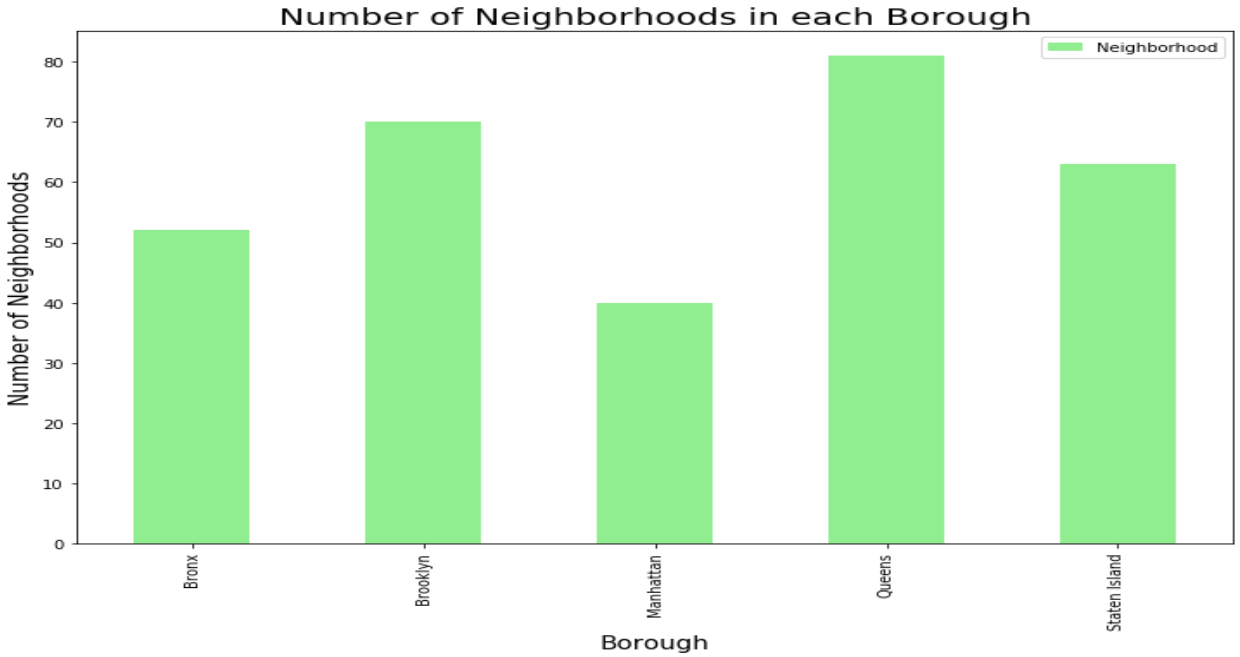
2. Two functions are created to return dataframe of venue id, name and category and the other dataframe consists of venue ratings,likes and tips in addition to venue id and name

3. Through looping filtered the dataframe which consists of only the restaurants of Italian cuisine

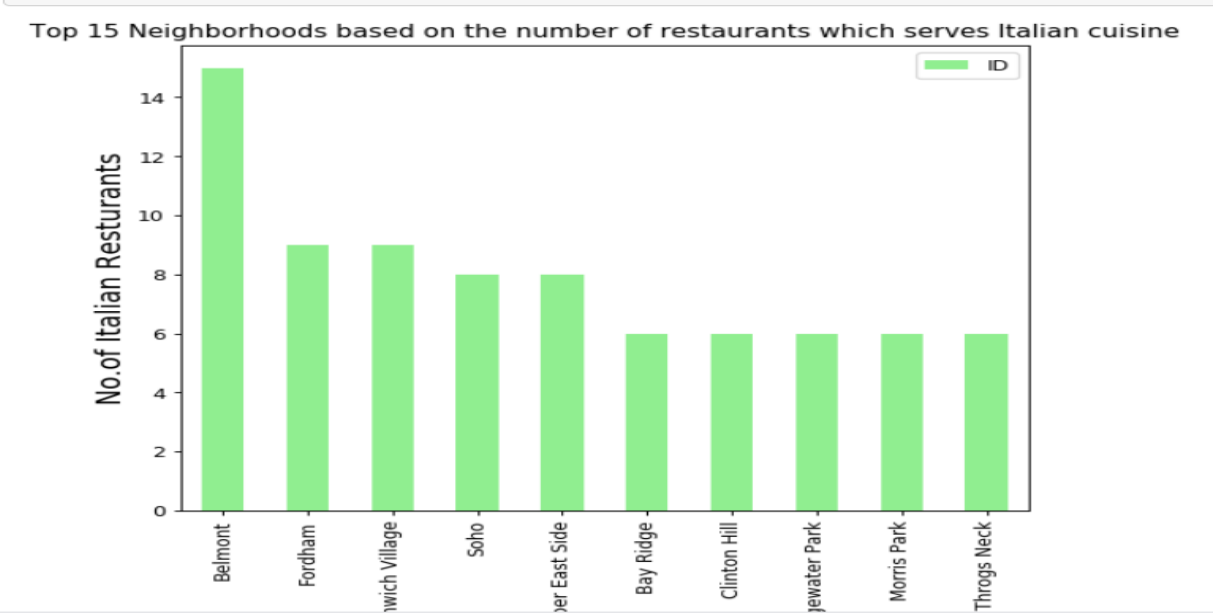
4. Dataframe is grouped by borough and and rating is averaged and sorted values by 'rating ' and the same is done for Neighborhood too

3.2 Visualization

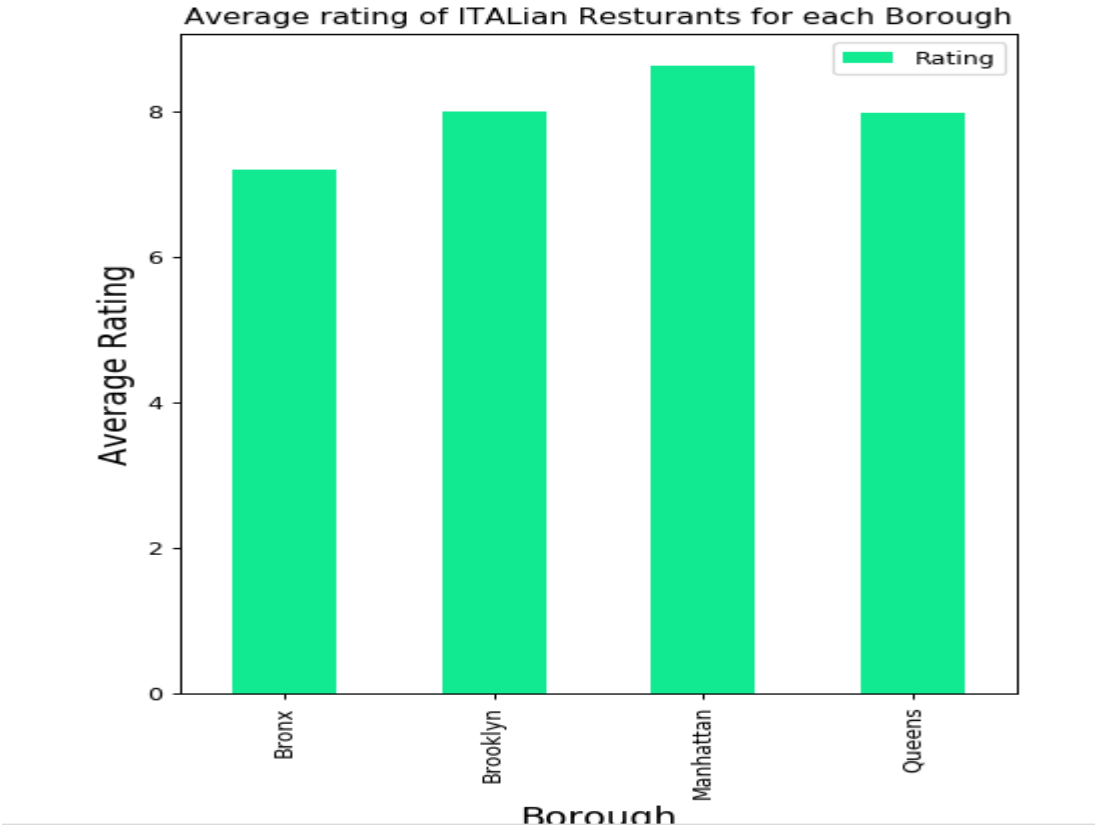
1. Boroughs in New York charted based on the number of Neighborhoods



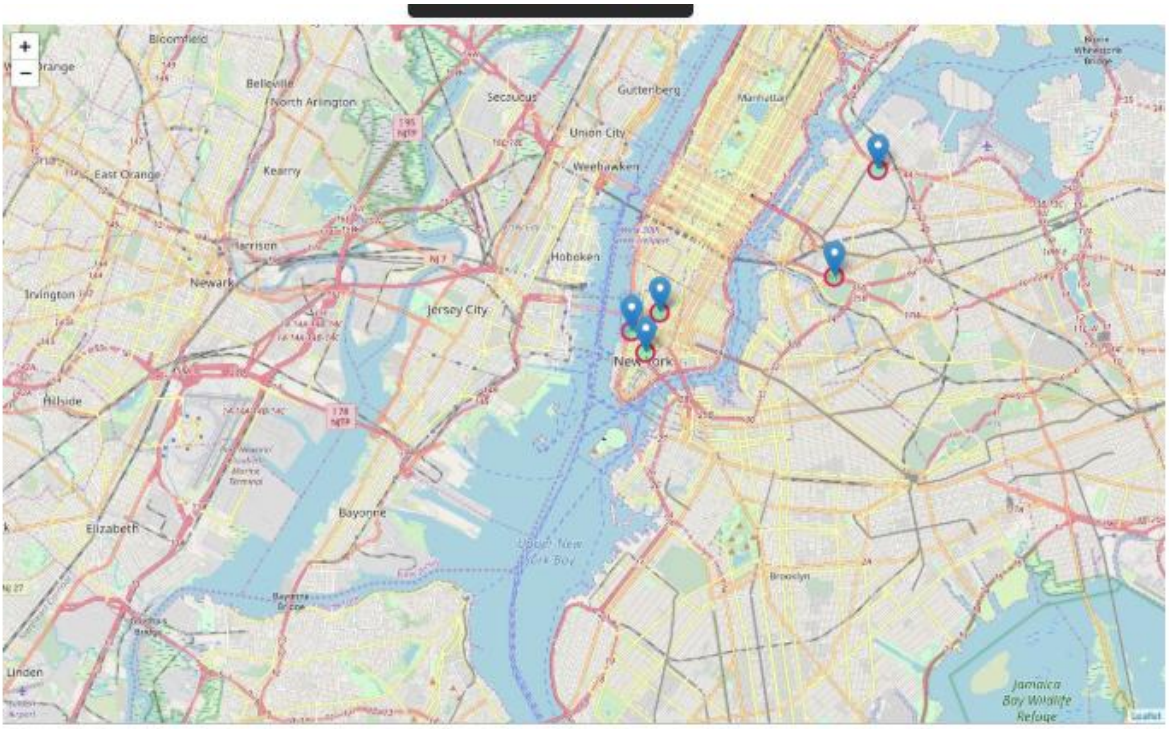
2. Neighborhoods in New York and the corresponding number of restaurants with Italian cuisines



3. Boroughs vs average rating for the Italian cuisine restaurants in the same in New York



4. Neighborhoods which has top rated Italian restaurants in New York city.



4. Results

1. Out of 5 boroughs in NewYork, Queens and Brooklyn boroughs has highest number of neighborhoods and Manhattan has the least
2. Out of 295 Italian restaurants in neighborhoods of NewYork, Manhattan has highest number of restaurants despite having the minimum neighborhoods compared to other boroughs
3. Among neighborhoods, Belmont(borough-Bronx) has more number of restaurants of Italian cuisine
4. "Rubirosa Ristorante" which is in Manhattan is more liked and has more number of tips
5. Top 5 neighborhoods with restaurants of average rating more than 9 are Downtown,Boerum Hill,Central Harlem and Greenpoint
6. Manhattan has the highest average rating of 8.6 among boroughs
7. Area with less Italian restaurants is Queens and has average rating around 8
8. In Top 5 neighborhoods of average rating >9, 4 neighborhoods belong to borough "Brooklyn" which means brooklyn has highly rated restaurants of Italian cuisine in NewYork

5. Discussion

Since the problem of this project is to find best location to open Italian restaurant, some of the suggestions can be made to the target audience which are as follows:

- Menu and tip can be taken from top rated restaurants before opening a restaurant
- Manhattan and Brooklyn has highly rated restaurants which means more number of customers are in that area who likes Italian cuisine. So the restaurant can be opened here but the investor has competitor risk. This can be reduced by doing competitor analysis which means by findiing out the factors that the competitors draw customers. This can be an extension to this project
- If the investor want less competitor risk and high successful rate, Queens is the best area to open Italian restaurant as there are few restaurants when compared to other boroughs and has average rating of 8.

6. Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing exploratory data analysis and providing recommendation to the target audience. The results of this project can be improved and made more inquisitive by using a current New York City's dataset along with API platforms which are more interested in Food Venues (like Yelp, etc.). The scope of this project can be expanded to understand the competitors(competitor analysis) by focusing on the features which help them to draw customers