IBM Applied Data Science Capstone Project

Title: The Battle of Neighborhoods

Introduction:

New York City is a large and ethnically diverse metropolis. It is the largest city in the United States. The city is made up of 5 boroughs: Manhattan, Brooklyn, Queens, the Bronx and Staten Island, which were grouped together in 1898. The population of New York City in 2020 is approximately 8.5 Million. The New York City has been growing faster than the region over the last decade. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States. New York is the most linguistically diverse city in the world. Around 800 languages are spoken in this city. English remains the most widely spoken language.

NYC has with its diverse culture, offers diverse food items. There are many restaurants in New York City, serving cuisines like Chinese, Indian, and Italian etc.

Problem Statement:

The Background

Client: A leading Italian businessman wants to expand his business. He is keen to identify the prime location spots in New York City to open his restaurant.

The Client is Interested in getting the answers to the below questions:

1. Identifying the neighborhoods and boroughs of New York City with highest and Lowest number of Italian Restaurants.
2. Which neighborhood/borough provides the most ideal location to open a new Italian Restaurant?

Data

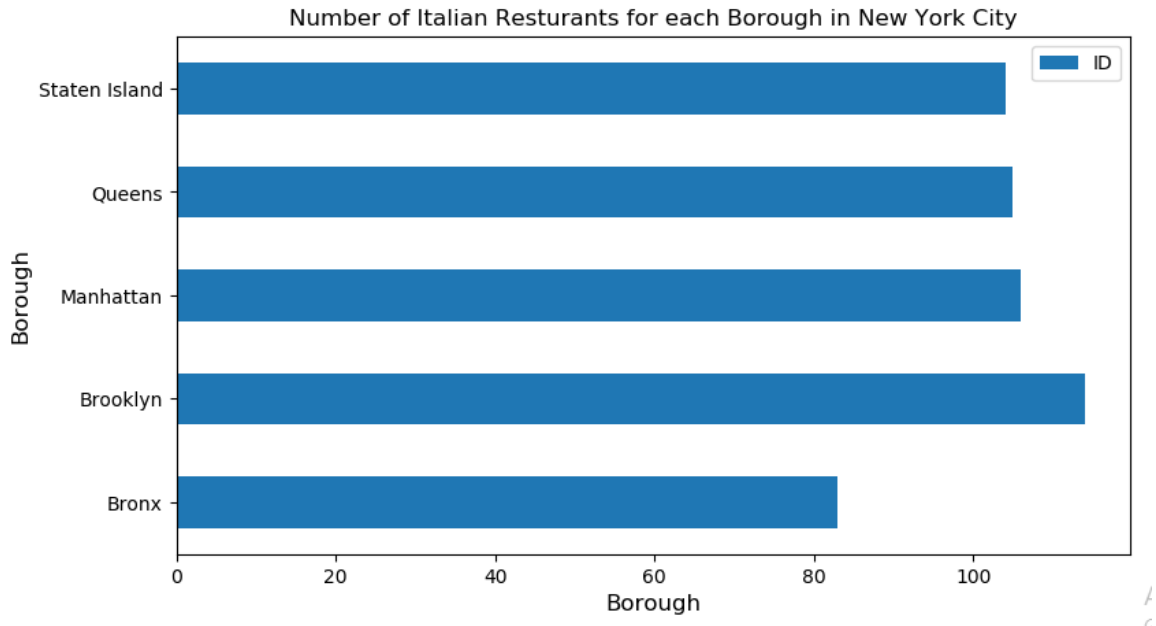
For this project we need the following data to explore the New York City:

1. Dataset containing list of Boroughs and neighborhoods along with their latitude and longitude.
    a. **Data source**: https://cocl.us/new_york_dataset
    Description: The above-mentioned dataset contains the information of boroughs and neighborhoods of NYC. We will use this data set to analyze various neighborhoods of New York City
2. Italian restaurants in each neighborhood of New York City.
    a. **Data source**: Foursquare API
    Description: The Foursqaure API will provide all the venues in each neighborhood. We can filter out the Italian restaurants using the category filter as required.
3. Geo Locations
    a. **Data source:** geopy.geocoders (Python Library)
    Description: This library would enable us to map the latitude and longitude information of each neighborhood within the borough.

Methodology

1. Collected the New York city location data from "https://cocl.us/new_york_dataset".
2. Used Python's geopy.geocoders library to get the longitude and latitude coordinates for the locations
3. Map the latitude and logitude data to each Neighbourhood
4. Using FourSquare API, extracted the Venue information of Italian Restaurants in each Neighbourhood
5. Determined the number of Italian restaurants in each neighbourhood
6. Based on the number of restaurants per neighbourhood, created bins for the level of competition the client woud face when opening a new restaurant in the neighbourhood
    - **High Competition Neighborhood (denoted by 3)**: No. of Italian restaurant > 9
    - **Average Competition Neighborhood (denoted by 2):** No. of Italian restaurant >5 and <= 9
    - **Low Competition Neighborhood (denoted by 1):** No. of Italian restaurant <= 5
7. Finally, the classification is visualized on New York City's map using Python's Folium library

Horizontal Bar Chart Depicting the Number of Italian Restaurants for each Borough



Classificaion of Neighborhoods based on the Level of Competition (High – 3, Medium – 2, Low – 1)

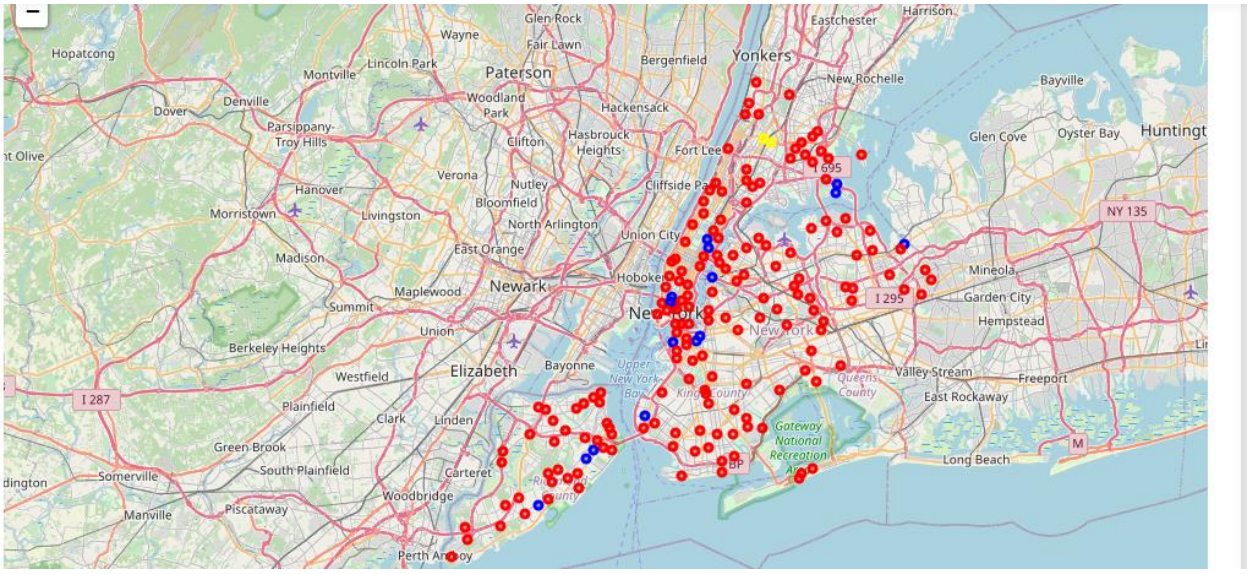
(Important Assumption: Only the Neighborhoods with atleast one Restaurant is considered in classification)

```
In [163]: grouped_data_final.head()
```

Out[163]:

	Borough	Neighborhood	Latitude	Longitude	# of Italian Restaurant	Competition_Level
0	Bronx	Riverdale	40.890834	-73.912585	1	1
1	Bronx	Kingsbridge	40.881687	-73.902818	1	1
2	Bronx	Woodlawn	40.898273	-73.867315	2	1
3	Bronx	Baychester	40.866858	-73.835798	2	1
4	Bronx	Pelham Parkway	40.857413	-73.854756	3	1

Results



Yellow Markers - High Competition Neighborhood (Italian Restaurants >9)
Blue Markers - Medium Competition Neighborhood (Italian Restaurants <=9 and >5)
Red Markers - Low Competition Neighborhood (Italian Restaurants <=5)
Unmarked Neighborhoods – Zero Competition locations

Conclusion

Answering the above Questions

- **Identifying the neighborhoods and boroughs of New York City with highest and Lowest number of Italian Restaurants.**

- ***Highest****: Bronx, Belmont: (14 Italian Restaurants) and Fordham, Belmont: (10 Italian Restaurants)*
- ***Lowest****: All the Neighborhoods unmarked on the New York City's map have no Italian restaurants and the points marked in red have <=5 Italian Restaurants*

- **Which neighborhood/borough provides the most ideal location to open a new Italian Restaurant?**

- Most Ideal Location to open a new Italian Restaurant can be answered with the level of competition the client is looking to accept:
- ***For High and Medium acceptance level of competition:*** *The Neighborhoods marked in yellow and Blue would be ideal.*
- ***For Low and No acceptance level of competition:*** *The Neighborhoods marked in red and unmarked neighborhoods would be ideal.*