Capstone Project -The Battle of Neighborhoods

Introduction

New York City is the most populated city in the United States with a population of 17,990,455. The city is divided into five boroughs, Manhattan, Brooklyn, the Bronx, Queens, and Staten Island. The city's physical setting is a complex assortment of lands and islands.

Being a multi-cultural city, New York comprises a diverse range of restaurants, each belonging to various categories like French, Indian, Chinese and Italian. New York is the hub of interactions between ethnicities and brings many opportunities for entrepreneurs to start or grow their business. The objective of this project is to use Foursquare location data to determine what might be the suitable neighborhood in New York to open an Italian restaurant.

Data

To complete this project, the following data will be needed:

- List of Boroughs & Neighborhoods in New York City with coordinates, latitude & longitude: https://cocl.us/new_york_dataset
- Italian restaurants in each neighborhood of the New York City: Foursquare API.
- GeoSpace data to get the New York Borough boundaries: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

Methodology

To analyze and find the best neighborhood in New York City in order to start an Italian restaurant, data set from https://cocl.us/new_york_dataset is used to explore various neighborhoods of the New York City. Foursquare API is used to find all venues for each neighborhood, determine ratings, tips and count for each Italian Restaurants. The ranking of neighborhoods is visualized using foliam library. The following questions will be addressed.

1. What is / are the best location(s) for Italian cuisine in New York City?
2. In what Neighborhood and/or borough should I open an Italian restaurant to have the best chance of being successful?

Libraries used

- pandas and numpy for handling data.
- geopy to get co-ordinates of City of New York.
- folium to visualize the results on a map
- request module for using FourSquare API.

```
import pandas as pd
import numpy as np
import requests
#from bs4 import BeautifulSoup
import os
!conda install -c conda-forge folium=0.5.0 --yes
import folium
#!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
%matplotlib inline
import seaborn as sns
```

Functions used

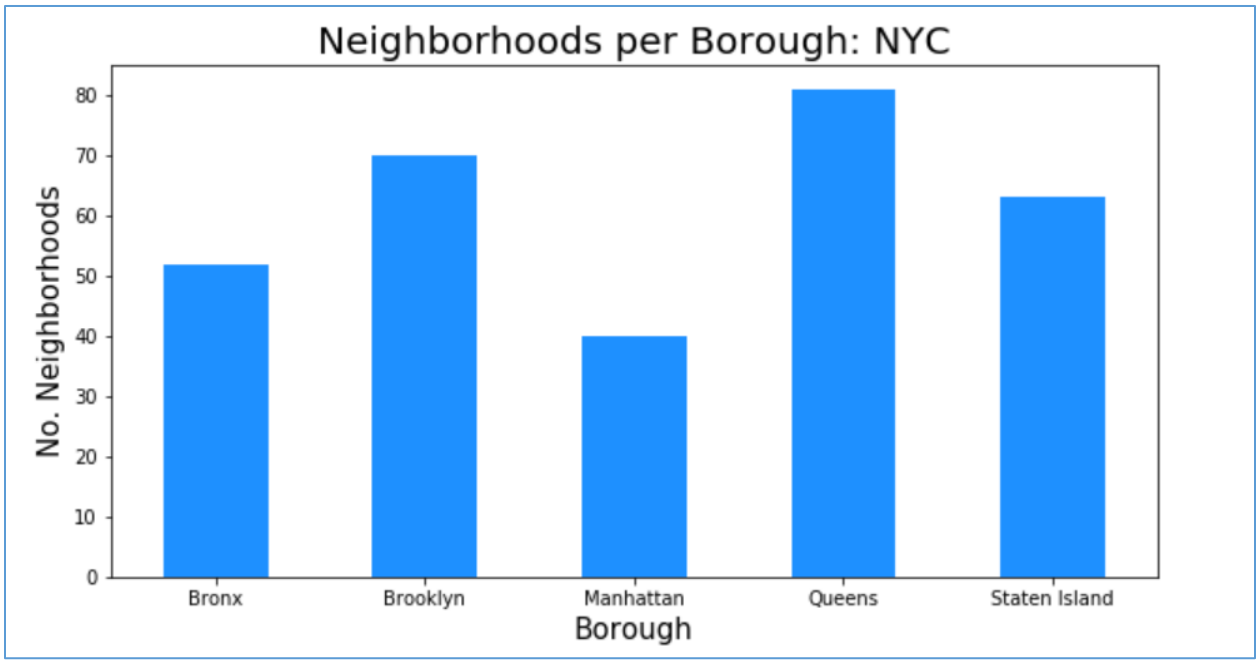
1. Function to get the geocodes, that is, latitude and longitude of a given location using geopy.

```
def geo_location(address):
    # get geo location of address
    geolocator = Nominatim(user_agent="foursquare_agent")
    location = geolocator.geocode(address)
    latitude = location.latitude
    longitude = location.longitude
    return latitude,longitude
```

2. Function to get the New York City data such as Boroughs, Neighborhoods along with their latitude and longitude.

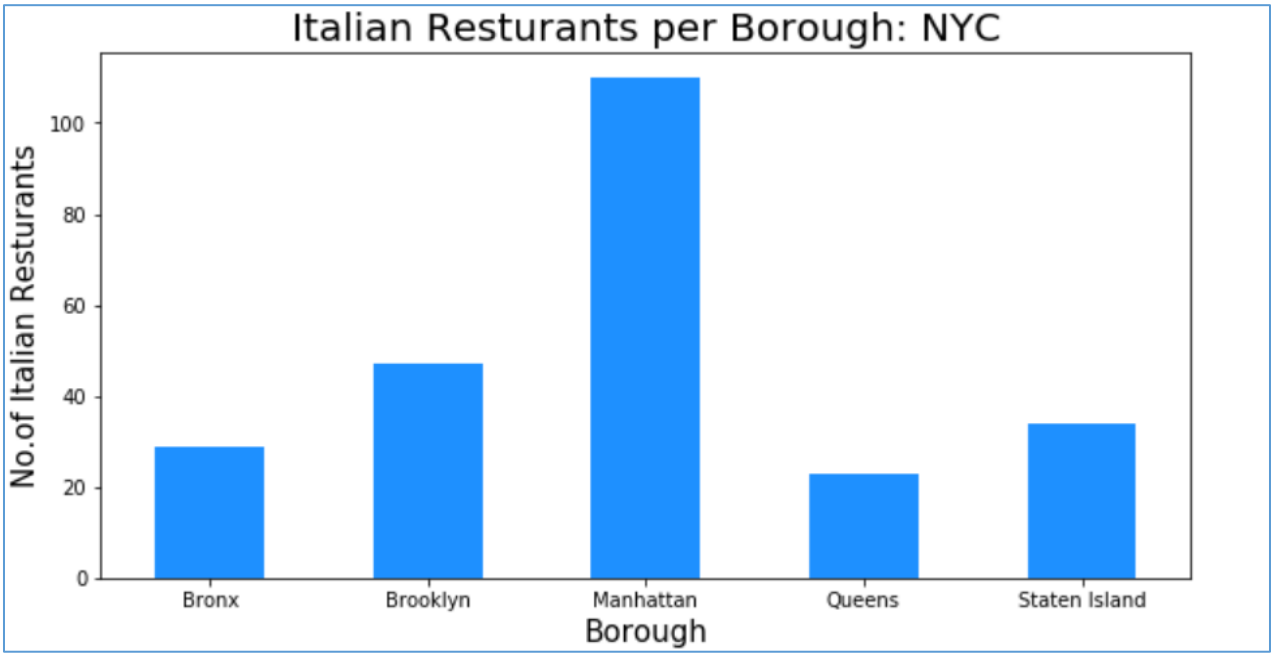
```
def get_new_york_data():
    url='https://cocl.us/new_york_dataset'
    resp=requests.get(url).json()
    # all data is present in features label
    features=resp['features']

    # define the dataframe columns
    column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']
    # instantiate the dataframe
    new_york_data = pd.DataFrame(columns=column_names)

    for data in features:
        borough = data['properties']['borough']
        neighborhood_name = data['properties']['name']

        neighborhood_latlon = data['geometry']['coordinates']
        neighborhood_lat = neighborhood_latlon[1]
        neighborhood_lon = neighborhood_latlon[0]

        new_york_data = new_york_data.append({'Borough': borough,
                                              'Neighborhood': neighborhood_name,
                                              'Latitude': neighborhood_lat,
                                              'Longitude': neighborhood_lon}, ignore_index=True)

    return new_york_data
```

From determining New York City data, the following plot indicates number of neighborhood per borough.



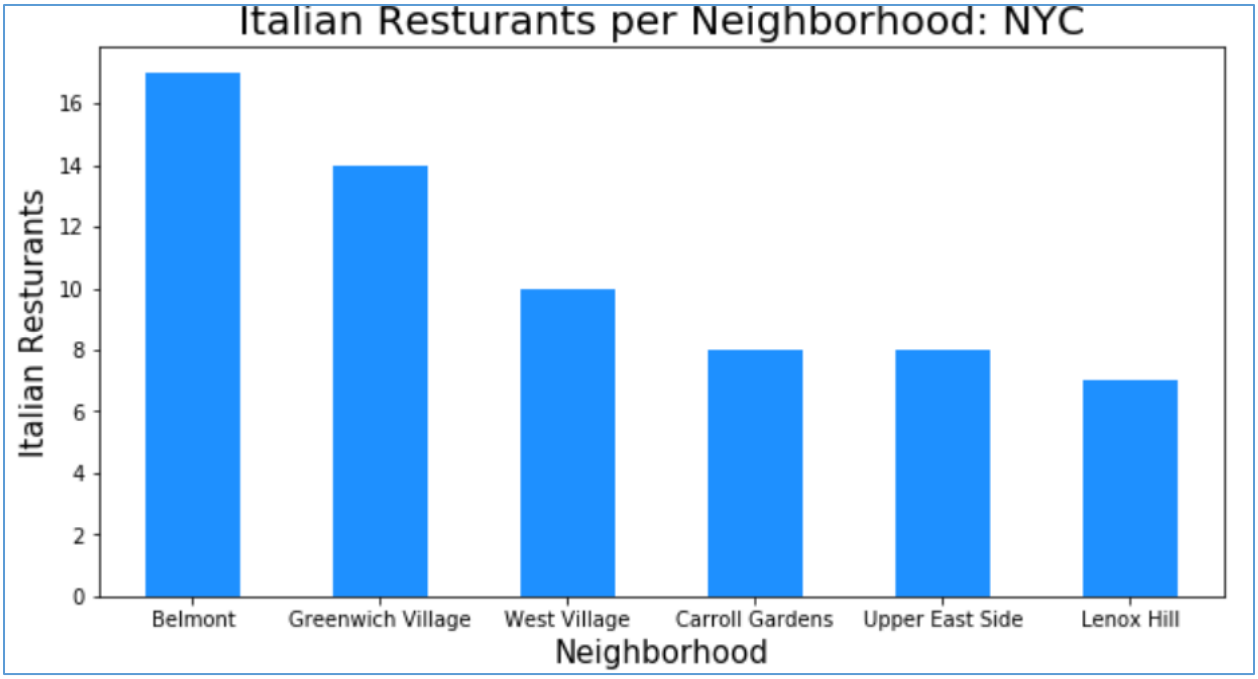
3. Code to collect number of Italian restaurants for each Borough is given below and it's observed that Manhattan has the most Italian restaurants.

```
# queens has most neighborhoods
# prepare neighborhood list that contains italian resturants
column_names=['Borough', 'Neighborhood', 'ID','Name']
italian_rest_ny=pd.DataFrame(columns=column_names)
count=1
for row in ny_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    italian_resturants=venues[venues['Category']=='Italian Restaurant']
    print('(',count,'/',len(ny_data),')','Italian Resturants in '+Neighborhood+', '+Borough+':'+str(len(italian_resturants)))
    print(row)
    for resturant_detail in italian_resturants.values.tolist():
        id, name , category=resturant_detail
        italian_rest_ny = italian_rest_ny.append({'Borough': Borough,
                                                  'Neighborhood': Neighborhood,
                                                  'ID': id,
                                                  'Name' : name
                                                 }, ignore_index=True)
    count+=1
```



4. Code to determine the number of Italian restaurants per neighborhood. Its observed that Belmont has the highest number.

```
NOofNeigh = 6 # top number for graphing all the same past 6
italian_rest_ny.groupby('Neighborhood')['ID'].count().nlargest(NOofNeigh).plot.bar(figsize=(10,5), color=clr)
plt.title('Italian Resturants per Neighborhood: NYC', fontsize = 20)
plt.xlabel('Neighborhood', fontsize = 15)
plt.ylabel('Italian Resturants', fontsize=15)
plt.xticks(rotation = 'horizontal')
plt.show()
```



5. Code to determine the rankings of restaurants.

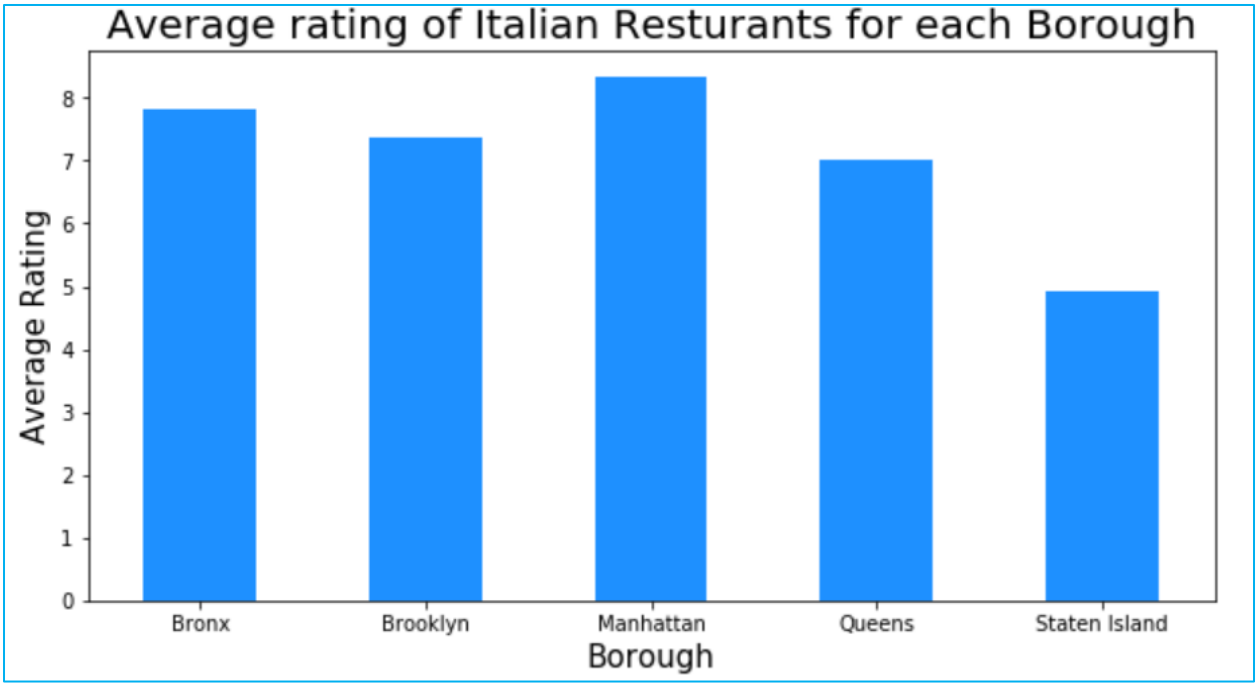
```
column_names=['Borough', 'Neighborhood', 'ID','Name','Likes','Rating','Tips']
italian_rest_stats_ny=pd.DataFrame(columns=column_names)
count=1
for row in italian_rest_ny.values.tolist():
    Borough,Neighborhood,ID,Name=row
    try:
        venue_details=get_venue_details(ID)
        print(venue_details)
        id,name,likes,rating,tips=venue_details.values.tolist()[0]
    except IndexError:
        print('No data available for id=',ID)
        # we will assign 0 value for these resturants as they may have been
        #recently opened or details does not exist in FourSquare Database
        id,name,likes,rating,tips=[0]*5
    print('(',count,'/',len(italian_rest_ny),')','processed')
    italian_rest_stats_ny = italian_rest_stats_ny.append({'Borough': Borough,
                                                'Neighborhood': Neighborhood,
                                                'ID': id,
                                                'Name' : name,
                                                'Likes' : likes,
                                                'Rating' : rating,
                                                'Tips' : tips
                                               }, ignore_index=True)
    count+=1
italian_rest_stats_ny.tail()
```

	Likes	Rating	Tips
count	243.000000	243.000000	243.000000
mean	130.226337	7.482305	53.049383
std	233.410791	2.120301	107.324860
min	0.000000	0.000000	0.000000
25%	15.000000	7.500000	7.000000
50%	39.000000	8.000000	18.000000
75%	116.500000	8.600000	51.500000
max	1431.000000	9.400000	1136.000000

6. Below are the top neighborhoods with top average rating of Italian restaurants.

	Neighborhood	Average Rating
86	Tribeca	9.200000
11	Boerum Hill	9.200000
12	Bushwick	9.200000
44	Hamilton Heights	9.000000
37	Fulton Ferry	8.900000
54	Lower East Side	8.900000
28	Dumbo	8.900000
64	Noho	8.766667
43	Greenwich Village	8.707143
80	Soho	8.700000

7. Average rating of Italian restaurants for each borough.



8. We visualized the map of neighborhoods with an average rating greater than or equal to 8.

Results and Conclusion

Queens and Manhattan comprise the top rated Italian restaurants on average. Bronx and Queens have the lowest number of Italian restaurants. Nevertheless, the neighborhood of Belmont in Bronx has the most Italian Restaurants across all New York City. Even though Manhattan comprises the lowest amount of neighborhoods in all five boroughs, it has the most Italian restaurants. From these observations, I would say that Queens and Manhattan are the best boroughs for Italian cuisine in the New York City. The final location

I would choose to open an Italian restaurant is Queens because even though it has fewest Italian restaurants, it has the good overall rating and this would also be good to avoid competition.