The Battle of Neighborhoods

1.Introduction and Business Problem

The city of New York is the most populous city in the United States. New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. Over the last decade the city has been growing faster and is by far the leading metropolitan gateway for legal immigrants admitted into the United States. With its diverse culture, comes diverse food items. The City of New York is famous for its excellent cuisine. Its food culture includes an array of international cuisines like Chinese, Indian, French etc. influenced by the city's immigrant history.

In this project, we will list and visualize all the neighborhoods of New York City that has great Italian restaurants.

Target Audience

This would interest anyone who wants to find great Italian restaurants in New York city.

2.Data

For this project we need the following data:

1.New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude

- Data source: https://cocl.us/new_york_dataset
- Description: This data set contains the required information which we will use to explore various neighborhoods of New York city.

2. Italian restaurants in each neighborhood of New York city

- Data source: Fousquare API
- Description: By using this API we will get all the venues in each neighborhood. We can filter these venues to get only Italian restaurants.

3.GeoSpace data

- Data source : https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm
- Description: By using this geo space data we will get the New york Borough boundaries that will help us visualize choropleth map.

3.Methodology

- Collect the New York city data from https://cocl.us/new_york_dataset
- Using Foursquare API we will find all venues for each neighborhood.
- Filter out all venues that are Italian Restaurants.
- Find rating, tips and likes for each Italian restaurant using Foursquare API.
- Using rating for each restaurant, we will sort that data.
- Visualize the Ranking of neighborhoods using folium library(python)

Questions that can be asked using the above-mentioned datasets

1. What is the best location in New York City for Italian cuisine?
2. Which areas have potential Italian Restaurant market?
3. Which areas lack Italian restaurants?
4. Which is the best place to stay if you prefer Italian cuisine?

We will import the required libraries for python.

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

```
In [1]: from bs4 import BeautifulSoup
        import requests
        import pandas as pd
        import numpy as np
        import os
        from sklearn.cluster import KMeans
        !conda install -c conda-forge folium=0.5.0 --yes
        import folium # plotting library
        from geopy.geocoders import Nominatim
        import matplotlib.cm as cm
        import matplotlib.colors as colors
        import matplotlib.pyplot as plt

        Solving environment: done

        ## Package Plan ##
```

Next, we use the Foursquare API.

This API has a database of more than 105 million places. This project will use Four-square API as its prime data gathering source. This API provides the ability to perform location search, location sharing and details about a business.

Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 700.

We define a function to interact with Foursquare API and get top 100 venues within a radius of 700 meters for a given latitude and longitude. Below function will return us the venue id, venue name and category.

In [208]:
```
def get_venues(lat,lng):

    #set variables
    radius=1000
    LIMIT=100
    CLIENT_ID = os.environ['CLIENT_ID'] # your Foursquare ID
    CLIENT_SECRET = os.environ['CLIENT_SECRET'] # your Foursquare Secret
    VERSION = '20180605' # Foursquare API version

    #url to fetch data from foursquare api
    url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

    # get all the data
    results = requests.get(url).json()
    venue_data=results["response"]['groups'][0]['items']
    venue_details=[]
    for row in venue_data:
        try:
            venue_id=row['venue']['id']
            venue_name=row['venue']['name']
            venue_category=row['venue']['categories'][0]['name']
            venue_details.append([venue_id,venue_name,venue_category])
        except KeyError:
            pass

    column_names=['ID','Name','Category']
    df = pd.DataFrame(venue_details,columns=column_names)
    return df
```

Now define a function to get the venue details like the ratings, tips for a given venue id. This will be used for ranking.

In [209]:
```
def get_venue_details(venue_id):

    CLIENT_ID = os.environ['CLIENT_ID'] # your Foursquare ID
    CLIENT_SECRET = os.environ['CLIENT_SECRET'] # your Foursquare Secret
    VERSION = '20180605' # Foursquare API version

    #url to fetch data from foursquare api
    url = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}'.format(
            venue_id,
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION)

    # get all the data
    results = requests.get(url).json()
    venue_data=results['response']['venue']
    venue_details=[]
    try:
        venue_id=venue_data['id']
        venue_name=venue_data['name']
        venue_likes=venue_data['likes']['count']
        venue_rating=venue_data['rating']
        venue_tips=venue_data['tips']['count']
        venue_details.append([venue_id,venue_name,venue_likes,venue_rating,venue_tips])
    except KeyError:
        pass

    column_names=['ID','Name','Likes','Rating','Tips']
    df = pd.DataFrame(venue_details,columns=column_names)
    return df
```

Next, we define another function to get the New York city data such as Boroughs, Neighborhoods along with their latitude and longitude.

In [210]:
```
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
```

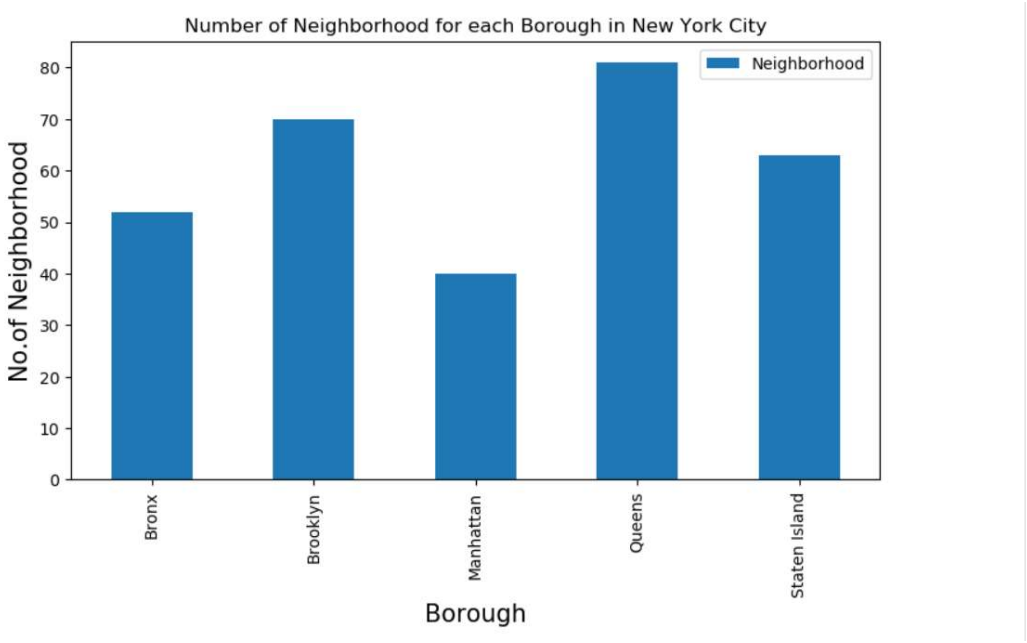
In [212]:
```
new_york_data.head()
```

Out[212]:

	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

We find that there are total of 306 different Neighborhoods in New York.

We can visualize the number of neighborhoods in each Borough of New York city.



We see that Queens has highest number of neighborhoods.

Next, we will find Italian restaurants for each Neighborhood

In [11]:
```
# prepare neighborhood list that contains Italian restaurants
column_names=['Borough', 'Neighborhood', 'ID','Name']
italian_rest_ny=pd.DataFrame(columns=column_names)
count=1
for row in new_york_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    italian_restaurants=venues[venues['Category']=='Italian Restaurant']
    print('(',count,'/',len(new_york_data),')','Italian Restaurants in '+Neighborhood+', '+Borough+':'+str(len(italian_restaurants)))
    for restaurant_detail in italian_restaurants.values.tolist():
        id, name , category=restaurant_detail
        italian_rest_ny = italian_rest_ny.append({'Borough': Borough,
                                                  'Neighborhood': Neighborhood,
                                                  'ID': id,
                                                  'Name' : name
                                                 }, ignore_index=True)
    count+=1
```

```
( 288 / 306 ) Italian Restaurants in Egbertville, Staten Island:1
( 289 / 306 ) Italian Restaurants in Roxbury, Queens:0
( 290 / 306 ) Italian Restaurants in Homecrest, Brooklyn:0
( 291 / 306 ) Italian Restaurants in Middle Village, Queens:1
( 292 / 306 ) Italian Restaurants in Prince's Bay, Staten Island:1
( 293 / 306 ) Italian Restaurants in Lighthouse Hill, Staten Island:1
```

Now that we have got all the Italian restaurants in New York city, we can analyze it.

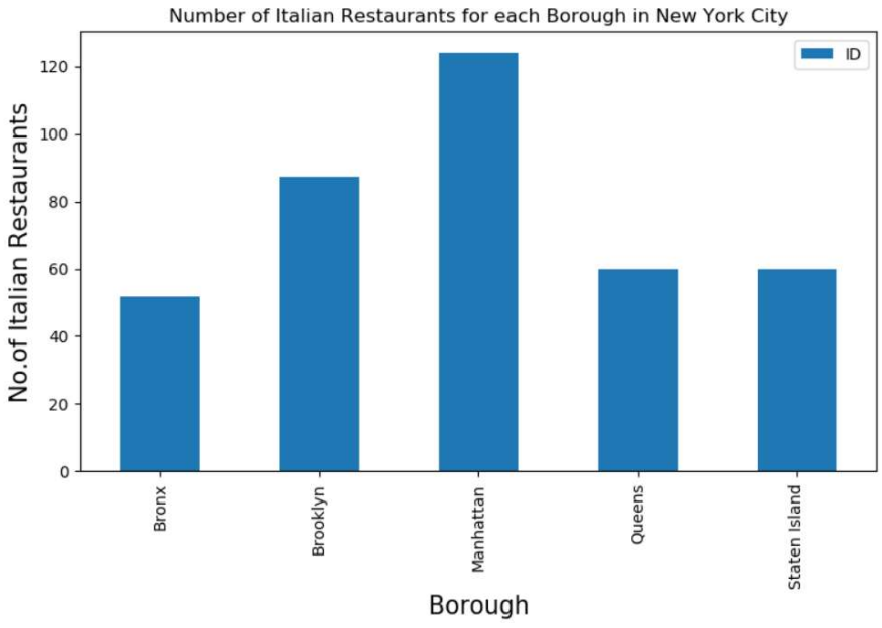
Out[12]:

	Borough	Neighborhood	ID	Name
0	Bronx	Riverdale	55aaee4d498e3cbb70e625d6	Bella Notte Pizzeria
1	Bronx	Woodlawn	511edb6de4b0d58346fd272d	Patrizia's Of Woodlawn
2	Bronx	Woodlawn	4d3cb3026b3d236a066a6364	Rivers Edge
3	Bronx	Pelham Parkway	4bf96ae65317a593d837017f	Enzo's
4	Bronx	Pelham Parkway	4b47f069f964a5208c4426e3	Pasta Pasta

```
n [14]: italian_rest_ny.shape
```

Out[14]: (383, 4)

4.Results



We see that Manhattan has the largest number of Italian restaurants.

We get the ranking of each restaurant for further analysis.

```
In [19]: italian_rest_stats_ny.head()
```

Out[19]:

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Riverdale	55aaee4d498e3cbb70e625d6	Bella Notte Pizzeria	9	7.3	4
1	Bronx	Woodlawn	511edb6de4b0d58346fd272d	Patrizia's Of Woodlawn	18	8.5	14
2	Bronx	Woodlawn	4d3cb3026b3d236a066a6364	Rivers Edge	10	6.8	8
3	Bronx	Pelham Parkway	4bf96ae65317a593d837017f	Enzo's	26	8.7	11
4	Bronx	Pelham Parkway	4b47f069f964a5208c4426e3	Pasta Pasta	9	6.2	8

The top neighborhoods with top average rating of Italian restaurants.

```
In [36]: ny_neighborhood_stats.sort_values(['Average Rating'],ascending=False).head(10)
```

Out[36]:

	Neighborhood	Average Rating
19	Bushwick	9.500000
38	Downtown	9.200000
15	Boerum Hill	9.200000
119	Soho	9.025000
94	Noho	9.000000
61	Hamilton Heights	9.000000
60	Greenwich Village	8.977778
74	Little Italy	8.950000
96	North Side	8.900000
139	West Village	8.875000

The average rating of Italian Restaurants for each Borough.

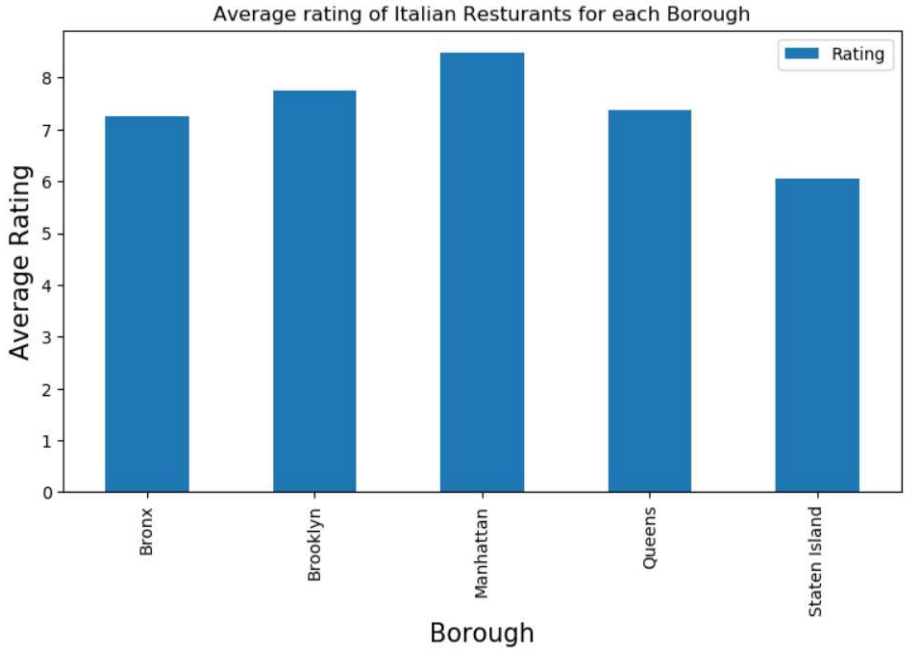
```
In [37]: ny_borough_stats=italian_rest_stats_ny.groupby('Borough',as_index=False).mean()[['Borough','Rating']]
         ny_borough_stats.columns=['Borough','Average Rating']
```

```
In [40]: ny_borough_stats.sort_values(['Average Rating'],ascending=False).head()
```

Out[40]:

	Borough	Average Rating
2	Manhattan	8.481452
1	Brooklyn	7.758621
3	Queens	7.368333
0	Bronx	7.246154
4	Staten Island	6.048333

We can visualize it.

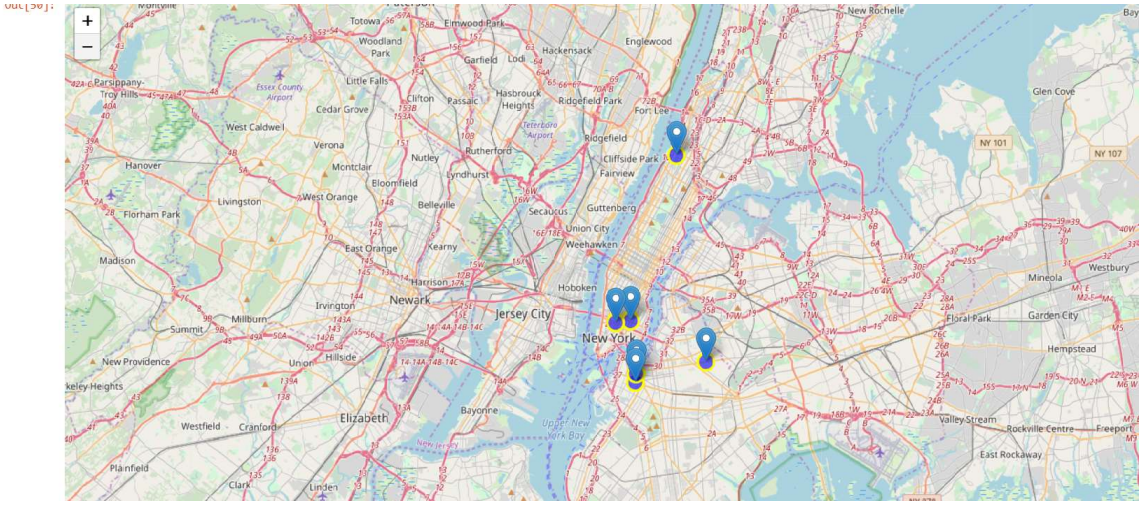


All the neighborhoods in New York city where the average rating of Italian restaurants is greater or equal to 9.0

Out[46]:

	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Brooklyn	Boerum Hill	40.685683	-73.983748	9.200
1	Brooklyn	Bushwick	40.698116	-73.925258	9.500
2	Brooklyn	Downtown	40.690844	-73.983463	9.200
3	Manhattan	Hamilton Heights	40.823604	-73.949688	9.000
4	Manhattan	Noho	40.723259	-73.988434	9.000
5	Manhattan	Soho	40.722184	-74.000657	9.025

We show this data on a map to visualize.



5.Discussion

The aim of this project is to help people find good Italian restaurants in the city of New York. Some of the neighborhoods in the Borough of Brooklyn have high rated Italian restaurants. We find that the Borough of Manhattan has the largest number of Italian restaurants and also tops the average rating of Italian restaurants amongst all the Boroughs.

6.Conclusion

- Best neighborhoods for Italian restaurants are.

  Boerum Hill (Brooklyn), Bushwick (Brooklyn), Downtown (Brooklyn),

  Hamilton Heights (Manhattan), Noho (Manhattan)

- Manhattan has potential Italian Restaurant markets
- Staten Island ranks last in average rating of Italian restaurants.
- Manhattan is the best place to stay if you prefer Italian cuisine.

Limitations

- The ranking is purely on basis of rating of restaurants.
- The accuracy of the data depends on the data provided by Foursquare.