CAPSTONE PROJECT: BATTLE OF NEIGHBORHOODS

SOHAN KUNDU

MAY 25, 2020

INTRODUCTION

New York City (**NYC**), often called **The City** or simply **New York** (**NY**), 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), New York is also the most densely populated major city in the United States.[6] Located at the southern tip of the U.S. state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass.[10] With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. Home to the headquarters of the United Nations,[12] New York is an important center for international diplomacy.[13][14]

Situated on one of the world's largest natural harbors, New York City is composed of five boroughs, each of which is a county of the State of New York.[15] The five boroughs—Brooklyn, Queens, Manhattan, the Bronx, and Staten Island—were consolidated into a single city in 1898.[16] The city and its metropolitan area constitute the premier gateway for legal immigration to the United States. As many as 800 languages are spoken in New York,[17] making it the most linguistically diverse city in the world. New York is home to more than 3.2 million residents born outside the United States,[18] the largest foreign-born population of any city in the world as of 2016.[19][20] As of 2019, the New York metropolitan area is estimated to produce a gross metropolitan product (GMP) of $2.0 trillion. If the New York metropolitan area were a sovereign state, it would have the 8th largest economy in the world. New York is home to the highest number of billionaires of any city in the world.

People from all over the globe tend to migrate to the New York city to enjoy the metropolis benefits of the city. Among all those migrated, are sportspersons, athletes, gym rats and keep-fit-fanatics. Most of them while migrating seem to be confused about which part of the city they should settle. Their major preference is well accessible gyms or fitness centres. This project will explore the city of New York and try to solve this problem.

PROBLEM

This project will explore the data and find solutions as follows:

- Visualize the city of New York on a map
- Explore and visualize the boroughs and neighborhoods in the city
- Explore Gyms and Fitness Centers in each borough
- Explore Gyms and Fitness Centers in each neighborhood

- Analyze the borough consisting the largest number of Gyms and Fitness Centers
- Analyze the neighborhood consisting the largest number of Gyms and Fitness Centers
- Select the preferred borough
- Visualize the borough and its neighborhood on maps
- Find possible solutions of settling in the best possible neighborhood

DATA SECTION

For this project we require the following data:

❖ **New York City data**

- Data Source: https://cocl.us/new_york_dataset.
- Data Description: This JSON dataset contains data of all the boroughs and neighborhoods of New City and also their latitude and longitude.

❖ **Gyms and Fitness Centers in each neighborhood data**

- Data Source: Foursquare API
- Data Description: By using the Foursquare API, we can get the details of all Gyms and Fitness Centers within a radius of a given latitude and longitude.

❖ **Geospatial data**

- Data Source: Geocoders
- Data Description: By using Nominatim from geopy.geocoders, we can get the latitude and longitude of a given address.

METHODOLOGY

1. We begin my importing all the libraries like Numpy, Pandas, JSON, Matplotlib, Geocoders, Folium, etc. Once the libraries are successfully imported, we proceed to the next steps.

2. **Data Preprocessing**

   We download the JSON file containing the New York City data in our system from https://cocl.us/new_york_dataset. We then open the JSON file from the Jupyter Notebook and load the data. We get the data in the following format:

   ```
   Out[3]: {'type': 'FeatureCollection',
            'totalFeatures': 306,
            'features': [{'type': 'Feature',
             'id': 'nyu_2451_34572.1',
             'geometry': {'type': 'Point',
             'coordinates': [-73.84720052054902, 40.89470517661]},
             'geometry_name': 'geom',
             'properties': {'name': 'Wakefield',
             'stacked': 1,
             'annoline1': 'Wakefield',
             'annoline2': None,
             'annoline3': None,
             'annoangle': 0.0,
             'borough': 'Bronx',
             'bbox': [-73.84720052054902,
             40.89470517661,
             -73.84720052054902,
             40.89470517661]}},
            {'type': 'Feature',
             'id': 'nyu_2451_34572.2'
   ```

   We notice that there are many unnecessary data and the dataset is not properly formatted. Thus, we extract only the necessary features from the JSON data and discard all other features. We then normalize the JSON data using the inbuilt Pandas function and convert it to a Pandas dataframe. The dataset contains data of the New York City like Borough, Neighborhood, Latitudes and Longitudes. So, we give similar column names to the converted Pandas dataframe. The first five rows of our dataframe looks like:

 | | 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 |

3. **Exploratory Data Analysis**

   Now that we have successfully cleaned our data and extracted the required features, we procees to the following steps. We check the shape of our dataframe and find out that New York City has 5 boroughs and 306 neighborhoods. Using Geocode we can find the Latitude and Longitude of New York City by passing the address. Once we have found the Latitude and Longitude of New York City, we can plot a map of the city using Folium with markers having labels of Neighborhoods and Boroughs. The plotted map looks like the following image:
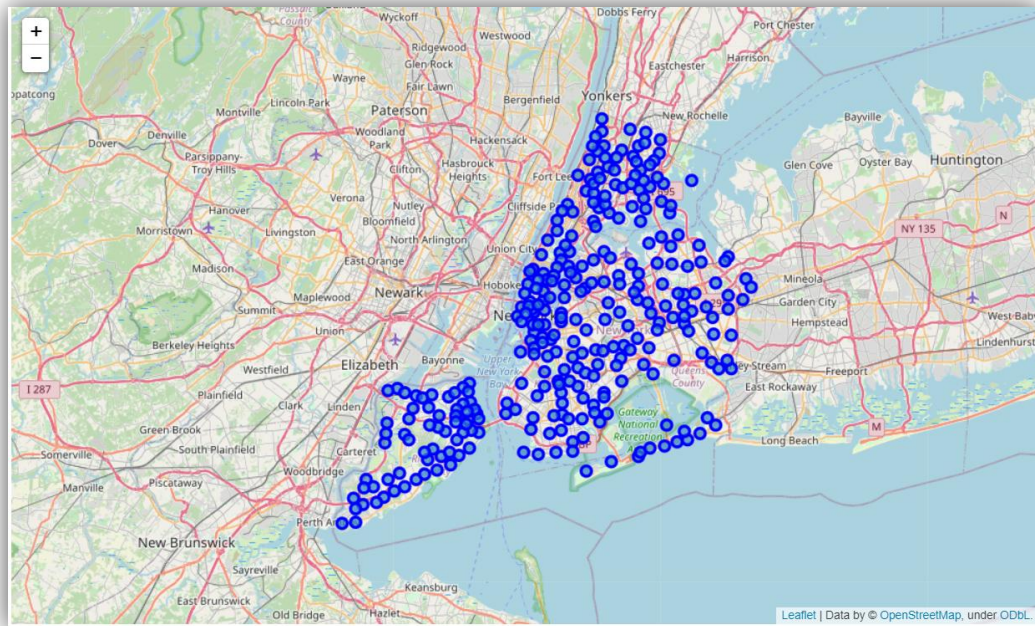
Next, we define a function which takes in latitudes and longitudes of a place and return the names and IDs of all gyms and fitness centers within a specified radius of the place. We do so by using the Foursquare API and passing our credentials to the URL. With the help of the aforesaid function, we can find the names and IDs of all gyms and fitness centers within each neighborhood. The names and IDs are stored in a dataframe along with the Borough and Neighborhood. We drop the unnecessary columns. The resultant dataframe is saved locally as a CSV file for further use.

We can plot the number of Gyms and Fitness Centres for each Borough in New York City on a Bar Graph with the help of matplotlib. The plot looks like the following image:
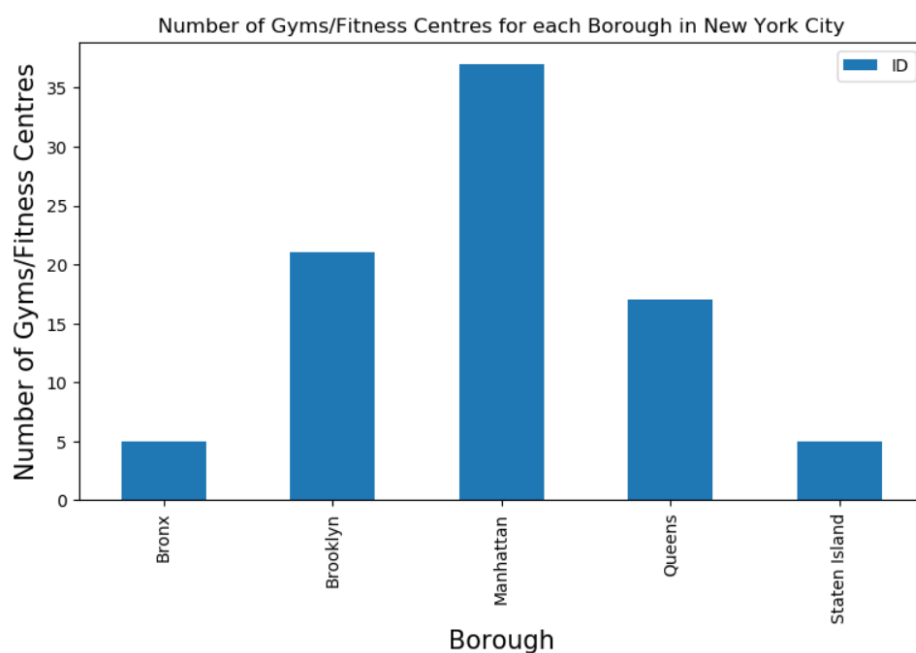
On visualizing the above Bar Graph, we can clearly say that Manhattan has the highest number of Gyms and Fitness Centers.

We can also plot the number of Gyms/Fitness Centres for each Neighborhood in New York City on a Bar Graph with the help of matplotlib. The plot looks like the following image:
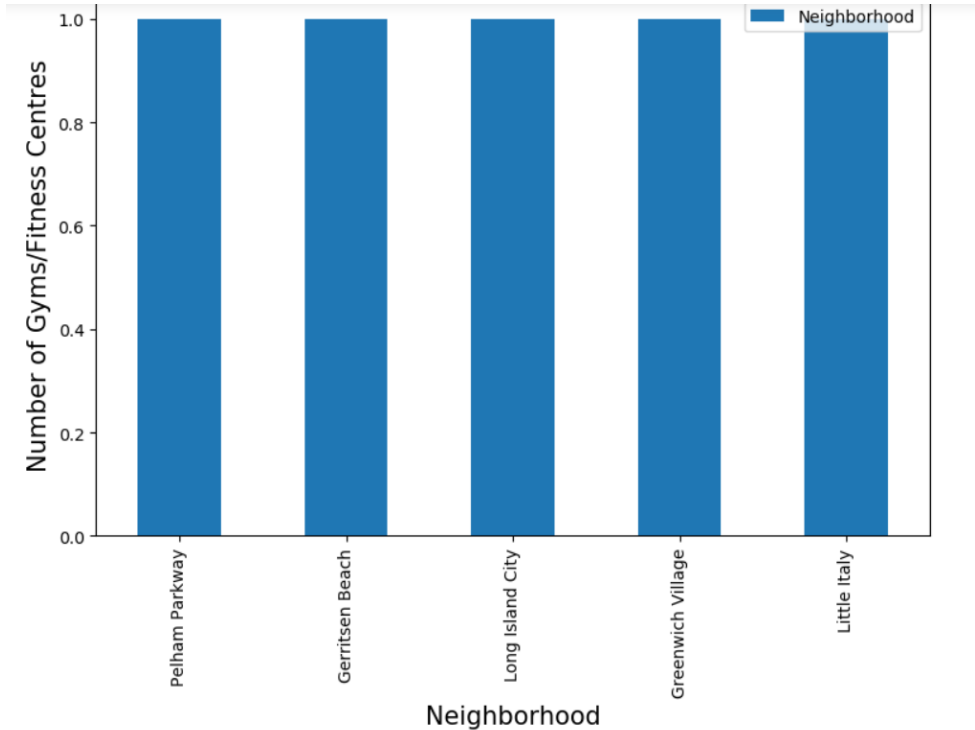
From the above graph, we find that the top 5 neighborhoods have similar frequency and it is difficult for us to conclude which has the highest number. Thus, we analyze the neighborhoods of Manhattan and find out that all the neighborhoods have 1 gym or fitness center each (as shown in the image below).

Hudson Yards	1
West Village	1
East Village	1
Little Italy	1
Gramercy	1
Flatiron	1
Noho	1
Civic Center	1
Roosevelt Island	1
Midtown	1
Murray Hill	1
Tudor City	1
Lincoln Square	1
Manhattan Valley	1
Chinatown	1
Turtle Bay	1
Yorkville	1
Morningside Heights	1
Greenwich Village	1
Central Harlem	1
Sutton Place	1
Lower East Side	1
East Harlem	1
Clinton	1
Marble Hill	1
Midtown South	1
Chelsea	1
Inwood	1
Lenox Hill	1
Battery Park City	1
Financial District	1
Hamilton Heights	1
Soho	1
Tribeca	1
Upper West Side	1
Upper East Side	1
Carnegie Hill	1

We now plot all the neighborhoods of the borough of Manhattan which has one Gym or Fitness Center on a map of Ney York using Folium. The image of the map is shown below:
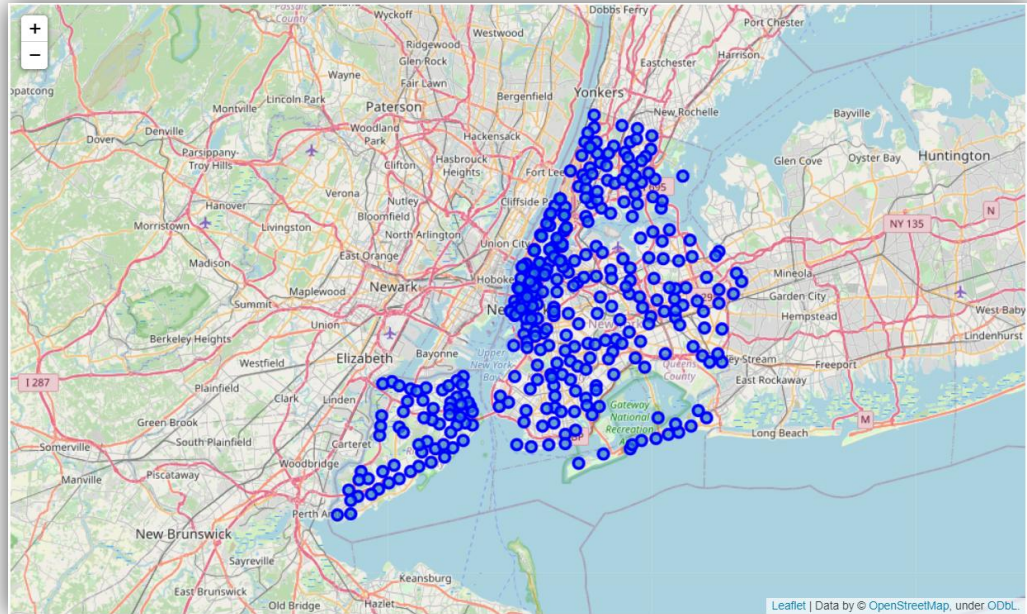
We can further use Foursquare API to find the rating of the all the Gyms and Fitness Centers in Manhattan. We can form a dataframe and analyze it. After visualization, we can select the few top rated Gyms and Fitness Centers. Then, we can suggest settlements in the nearby neighborhoods of the top rated Gyms and Fitness Centers.

CONCLUSION

I have successfully analyzed the data of various Gyms and Fitness Centers all across New York throughout this project based on the Boroughs and Neighborhoods they are located in. After analyzing and visualizing the data, I have come up with the conclusion that the borough of Manhattan has the highest number of Gyms and Fitness Centers in the city of New York. The data is also split up based on the neighborhoods of Manhattan. The Manhattan neighborhoods containing Gyms and Fitness Centers are plotted on a map of New York for visualization. This project can be further improved by collecting the ratings and tips of the Gyms and Fitness Centers and cluster the top rated ones. By doing so, we can filter the best neighborhoods for the settlements of athletes, sportspersons and gym rats. On an overview, I can conclude that this project serves as a solution for those who are looking to settle in neighborhoods in the New York City with good facilities of Gyms and Fitness Centers.