Introduction

The decision to open a new gym or fitness center requires the consideration of several factors to determine the feasibility of the opening. For example, an entrepreneur must consider the right timing, location, services provided, size of gym or fitness center, and branding/marketing campaign. Of those factors the location of the opening must be studied thoroughly to maintain the successfulness of the new business. For example, the new gym is preferably located in a neighborhood with minimum competitors and far away from venues that contradict health wellness such as fast food restaurants. In this project, New York City will be explored for the best place to open a new gym or fitness center. The targeted stakeholders for this project would be business corporation or entrepreneurs interested in open a new gym or fitness center at New York.

Data

To solve the problem, two data sets will be required:

1. New York Neighborhood dataset from the repository of NYU <https://geo.nyu.edu/>

This dataset is needed to explore the neighborhoods segments at New York in addition to the latitude and longitude of each neighborhood.

1. FourSquare API to get the venues in a given borough at New York <https://developer.foursquare.com/>

This dataset will be used as the source for location data.

Methodology

The New York Neighborhood dataset was organized using four columns borough, neighborhood, latitude, and longitude using pandas dataframe. The GeoPy library was used in Python to get the coordinates of New York and the Folium Library was used to visualize the map of New York with neighborhoods (refer to Figure 1).

Next FourSquare API was used to further explore the neighborhoods in New York. After grouping by the neighborhoods, the mean of frequency of occurrences for each category were display. Also, the top 10 most common venues in each neighborhood were displayed. The venues were sorted and put in a new pandas dataframe for further analysis. Finally, the neighborhoods were clustered using k-means clustering with 10 clusters.

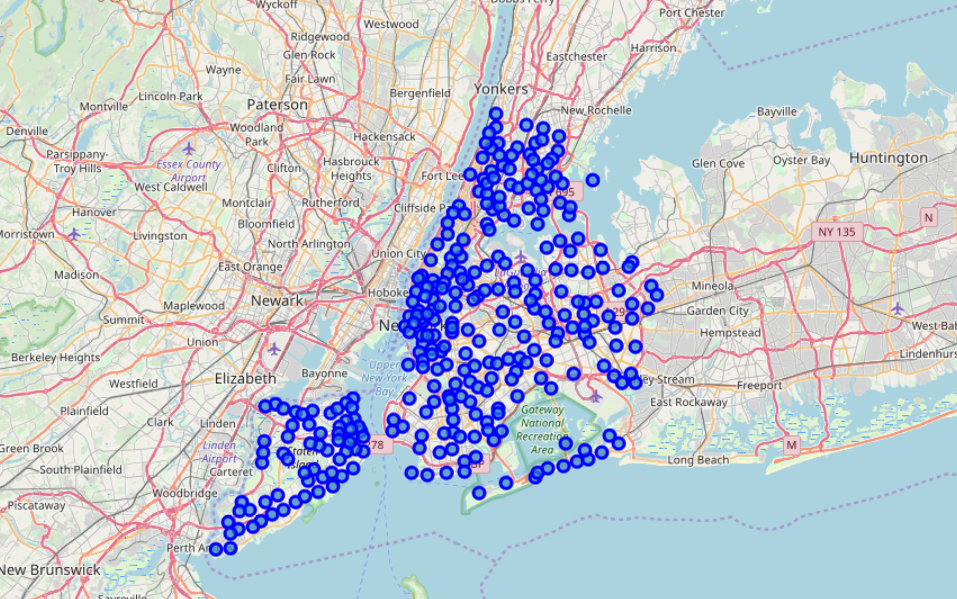


Figure 1. New York Neighborhoods

Results

The original dataset consisted of five boroughs and 306 neighborhoods. When further explored with FourSquare API, New York had 10,412 venues in all neighborhoods. Refer to Figure 2 for a sample screenshot. Neighborhoods were grouped and results of the mean of frequency of occurrences for each category resulted in 300 neighborhoods and 431 different categories. Refer to Figure 3. A sample for a neighborhood with top 10 venues displayed is shown in Figure 4. From the results four terms for venues referring to gym-related activities were found:

Gym

Gym / Fitness Center

Gym Pool

Gymnastics Gym

The results of the k-mean clustering displayed the neighborhood in each cluster along with the top 10 venues in order. Refer to figure 5.

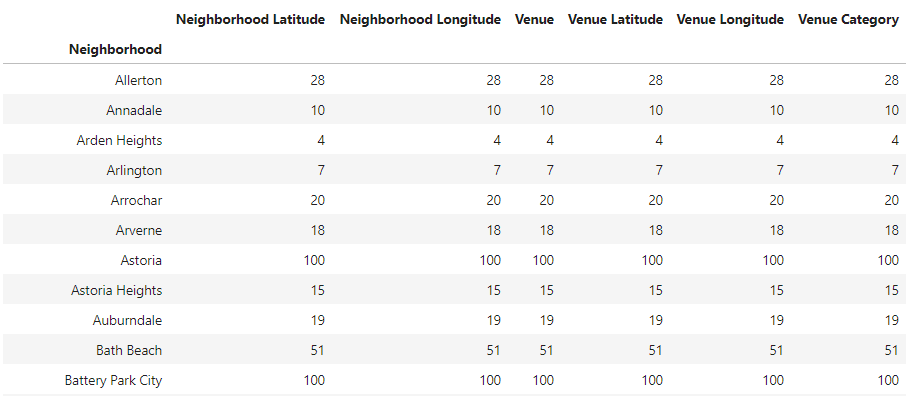


Figure 2. Sample of New York Neighborhood Venues

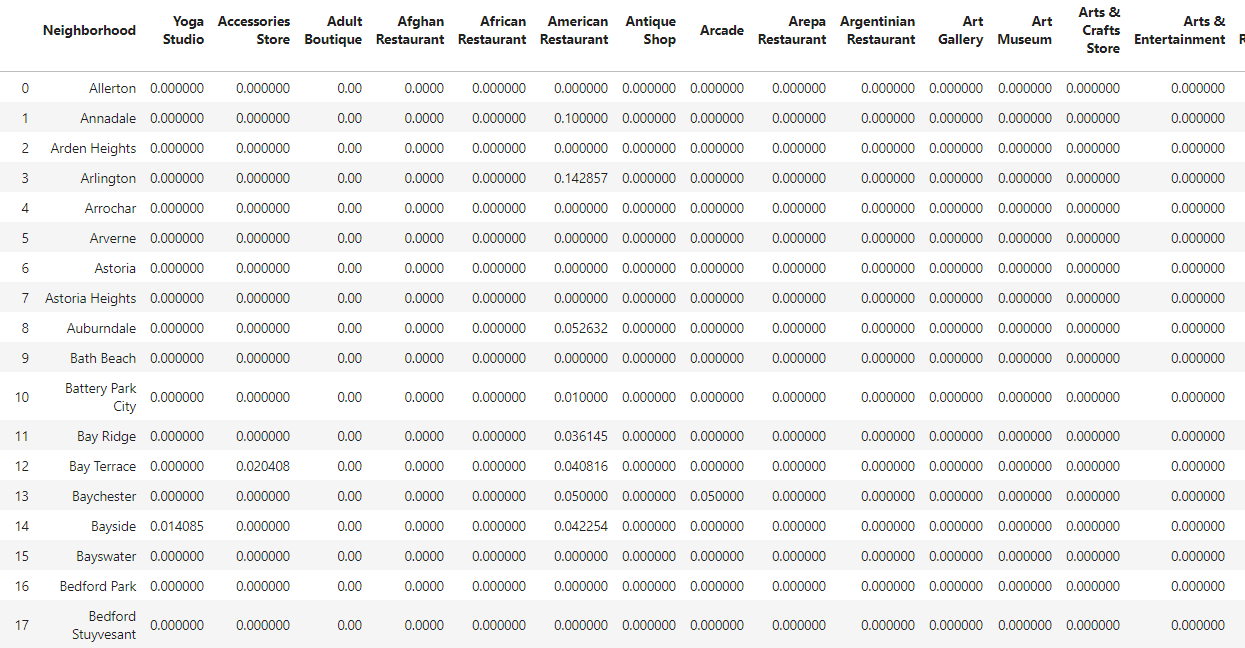


Figure 3. Mean of frequency of occurrences for each category in Neighborhoods

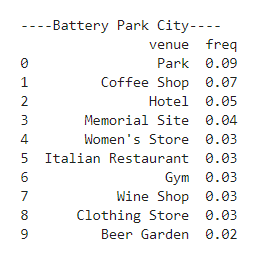


Figure 4. A Neighborhood with top 10 venues



Figure 5. Sample Result of k-means clustering (Cluster 1)

Discussion

Based on the results from the above analysis, it is recommended to open a gym or fitness centers on neighborhoods such as Todt Hill where the top 10 venues suggest there is no gym or fitness center. Moreover, there are three pet stores and no junk food restaurants. This makes it a recommended neighborhood. Refer to Figure 6.

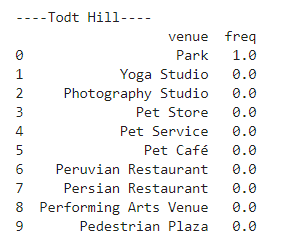


Figure 6. Recommended Neighborhood for a Gym

Conclusion

In this project, the location opening a new gym or fitness center was discussed. The New York City dataset was be explored for the best place to open a new gym or fitness center. Based of the results of the analysis several neighborhoods were found as good candidates for locations to open a gym or fitness center. Those neighborhoods had no gyms listed as the top 10 venues, nor any junk food within a proximity distance.