Salin Dalabehera
IBM Applied Data Science Capstone
May 16, 2020

Opening of Gym in New York City

Coursera Capstone Project

Introduction:

Lifestyle behaviors are the most important factors in short-term and long-term well being. Developing healthy lifestyle habits will have a dramatic effect on overall wellness. Components of a healthy lifestyle include: physical activity, balanced and adequate nutrition, rest, recovery and stress management. Daily physical activity is one of the key components in developing and leading healthy lifestyle. Physical activity improve overall quality of life. It helps boost energy, assists with weight management and improves self-esteem.

"Health is wealth". This is the new mantra of everybody in metro cities now-a-days. Good health can enjoy the pleasure of life. People are more focus toward healthy lifestyle and for that they are willing to spend money for it. Instead of buying equipment for exercise in home, they normally prefer gym in nearby area.

The fitness industry has exploded in recents years. In the U.S., total revenues reached 3.1 billion U.S. dollars. There are 2111 health clubs in New York and increasing every year. In addition, many of the gym and fitness centers employ very little to no staff, which makes startup costs and barriers to entry low. With the proper mix of skills, training and commitment, starting a gym or fitness center can prove a successful business move.

Business Problem:

The objective of this project is to analyze the best location to open a gym/fitness center in New York city. New York has 5 boroughs i.e. Bronx, Queens, Manhattan, Brooklyn and Staten Island. By using the data science methodology and machine learning, we will find the best location to open a gym in any of the borough.

Audiences:

This project can be useful for the investors/new entrepreneur to open a gym in New York area. They can see the ares with high, moderate and low fitness centers and include in their analysis or decision.

Data:

For this project, we need following data to build our model:

* List of neighborhoods of New York city: List of neighborhoods of all the 5 boroughs
* Location coordinates of neighborhood: Latitude and Longitude of the neighborhoods
* Venues of the neighborhoods: Venues data near to the neighborhood, particularly Gym or fitness centers

Sources:

One of the wikipedia page (https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City) contains the neighborhoods of all 5 boroughs. We will use the python modules "requests" and "BeautifulSoup" to extract the data and save it to a pandas dataframe.

For latitude and longitude coordinates of each neighborhood, we will use python "geopy" package.

After that, we will use Foursquare developer api to extract venues of each neighborhood of New York city. Foursquare has the largest database with 150+millions places. Foursquare api provides many venues with category, which will help us to extract Gym from the result set.

In this project, we will use many tools and methods like web scrapping, Geo coordinates package, API(Foursquare), data cleaning, data wrangling, machine learning (Kmean clustering) and Folium to visualize the locations in map.