

IBM Applied Data Science Capstone



Opening a Gym in New York City

Salin Dalabehera

Coursera Capstone Project

Date: 21 May 2020

Business Problem

- Location of the gym is one of the crucial factor to avoid competition in neighborhood, because of facilities and people living nearby
- The objective of this project to find the best suitable neighborhood with in borough of the New York City to open gym.
- Business question will be what will be the recommended place to open a gym in New York City?

Data & Source

- Data:

- List of neighborhood of the New York City
- Coordinates of the neighborhoods
- Nearby venues of the neighborhoods

- Source:

- Wikipedia (https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City)
- Python Geopy to get the latitude and longitude
- Foursquare API to get the nearby venues of the neighborhoods

Methodology

In this method we perform data extracting , data cleaning, data validation and apply machine leaning to get the results

- Use web scrapping (requests and beautiful soup) to get neighborhood from wikipedia
- Get latitude and longitude using python Geopy.geocoder
- Get the nearby venues using foursquare API
- Filter data based on the gym category
- Apply machine leaning technique “KMeans clustering”
- Visualize results using Folium

Results

Cluster into 3 category. If we look into Bronx

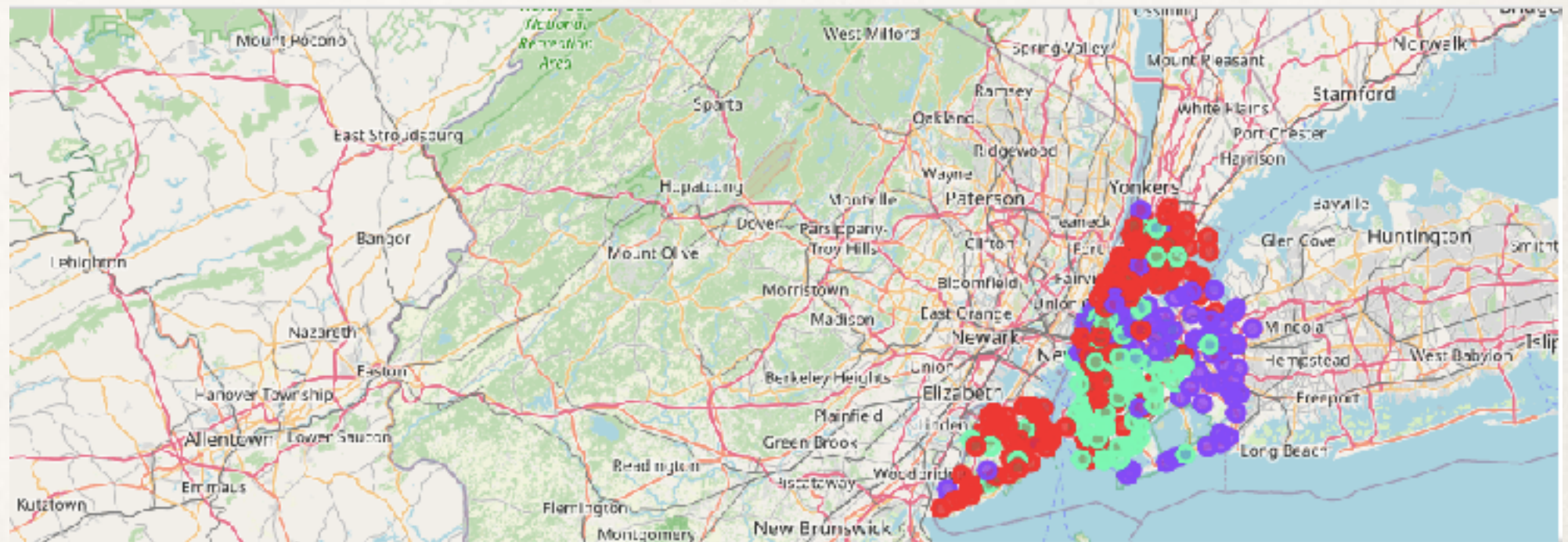
- Cluster 0
 - Neighborhoods with low frequency of gym
- Cluster 1
 - Neighborhoods with high frequency of gym
- Cluster 2
 - Neighborhoods with moderate frequency of gym

Borough	Neighborhood		Gym	Latitude	Longitude
	cluster				
Bronx	0	47	47	47	47
	1	3	3	3	3
	2	11	11	11	11
Brooklyn	0	32	32	32	32
	1	1	1	1	1
	2	57	57	57	57
Manhattan	0	31	31	31	31
	1	9	9	9	9
	2	12	12	12	12
Queens	0	6	6	6	6
	1	63	63	63	63
	2	16	16	16	16
Staten Island	0	42	42	42	42
	1	4	4	4	4
	2	5	5	5	5

Visualization

Using folium we can visualize neighborhoods with assign clustering.

- Cluster 0 : Red
- Cluster 1: Blue
- Cluster 2: Green



Disussion

To answer the business question, below are the observation

- Cluster 0 has very less number of gyms, which will be the best neighborhood to open gym with almost none competitions
- Cluster 1 has moderate number of gyms, which can be also a potential neighborhood to open a gym, due to less competition and already existing interested gym goer.
- Cluster 2 has high number of gyms, which can be avoided due to high competition.

Conclusion

- To the business question, Cluster with 0/1 are the potential neighborhood to open a gym.
- While considering the place, we can include population to finalize the answer.
- This model can be useful for the investor to open a gym in New York City to utilize their capital in correct place