Determining the optimum location to set up a gym in Manhattan

IBM Capstone Project

Business Problem

The problem I have chosen to analyse involves the decision regarding where to open a gym in Manhattan, New York City , USA. There are several factors which will play a role in determining this optimal location. The gym needs to be centrally located and needs to be easily accessible from all parts of the city.

The gym should also have complementary venues near it. For example, the presence of a spa nearby will encourage customers to purchase gym membership of said gym. The gym should also not exist in a place where several gyms already exist which would generate unnecessary competition. Clusters will be generated to assess each neighbourhood.

I will attempt to generate a list of neighbourhoods most suitable which will result in maximum revenue for the stakeholder wishing to set up the gym.

DATA

Based on the factors I discussed above, the following data will be required:

- The distance of each neighbourhood from the centre of the city.
- The venues nearby which are complementary to the gym.
- The number of gyms that exist in each neighbourhood.

The neighbourhoods have been identified from the source: https://cocl.us/new_york_dataset. All the features will be extracted from this data source. The other data will be extracted using the Foursquare API. The location of each neighbourhood will be obtained from the GeoPy GeoCoder package in Python.

Data frame

[10]:

	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
...	...	...	...	...
301	Manhattan	Hudson Yards	40.756658	-74.000111
302	Queens	Hammels	40.587338	-73.805530
303	Queens	Bayswater	40.611322	-73.765968
304	Queens	Queensbridge	40.756091	-73.945631
305	Staten Island	Fox Hills	40.617311	-74.081740

306 rows × 4 columns

Map of Manhattan containing the relevant neighborhoods

Analysis using FOURSQUARE API

[38]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Carnegie Hill	Coffee Shop	Yoga Studio	Bakery	Gym / Fitness Center	Gym
1	Central Harlem	Cosmetics Shop	African Restaurant	Seafood Restaurant	Bar	French Restaurant
2	Clinton	Theater	Coffee Shop	Gym / Fitness Center	Wine Shop	Gym
3	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Latin American Restaurant	Deli / Bodega
4	Hamilton Heights	Pizza Place	Café	Coffee Shop	Deli / Bodega	Mexican Restaurant
5	Hudson Yards	Hotel	American Restaurant	Gym / Fitness Center	Café	Italian Restaurant
6	Lenox Hill	Italian Restaurant	Pizza Place	Coffee Shop	Café	Cocktail Bar
7	Lincoln Square	Plaza	Café	Concert Hall	Gym / Fitness Center	Performing Arts Venue
8	Manhattan Valley	Yoga Studio	Bar	Pizza Place	Coffee Shop	Mexican Restaurant
9	Manhattanville	Seafood Restaurant	Coffee Shop	Italian Restaurant	Park	Mexican Restaurant
10	Midtown	Coffee Shop	Hotel	Clothing Store	Theater	Spa
11	Midtown South	Korean Restaurant	Hotel	Japanese Restaurant	Burger Joint	Dessert Shop
12	Morningside Heights	Park	American Restaurant	Coffee Shop	Bookstore	Burger Joint
13	Murray Hill	Sandwich Place	Coffee Shop	Hotel	Pizza Place	Gym / Fitness Center
14	Roosevelt Island	Park	Plaza	Japanese Restaurant	Soccer Field	Farmers Market
15	Sutton Place	Italian Restaurant	Gym / Fitness Center	Park	Coffee Shop	Gym
16	Tudor City	Café	Park	Mexican Restaurant	Deli / Bodega	Pizza Place
17	Turtle Bay	Italian Restaurant	Coffee Shop	Park	Deli / Bodega	Wine Bar
18	Upper East Side	Italian Restaurant	Juice Bar	Bakery	Gym / Fitness Center	Exhibit
19	Upper West Side	Italian Restaurant	Bakery	Wine Bar	Coffee Shop	Mediterranean Restaurant
20	Yorkville	Coffee Shop	Italian Restaurant	Gym	Deli / Bodega	Bar

KNN CLUSTERS ON THE MAP OF MANHATTAN

INFORMATION ABOUT THE CLUSTERS

Cluster 1- It represents the neighbourhoods that are at a greater distance from the centre of Manhattan. Each neighbourhood as at least 1 gym. I would not recommend setting up a gym here as it is far from Central Park. These neighbourhoods are marked in **red**.

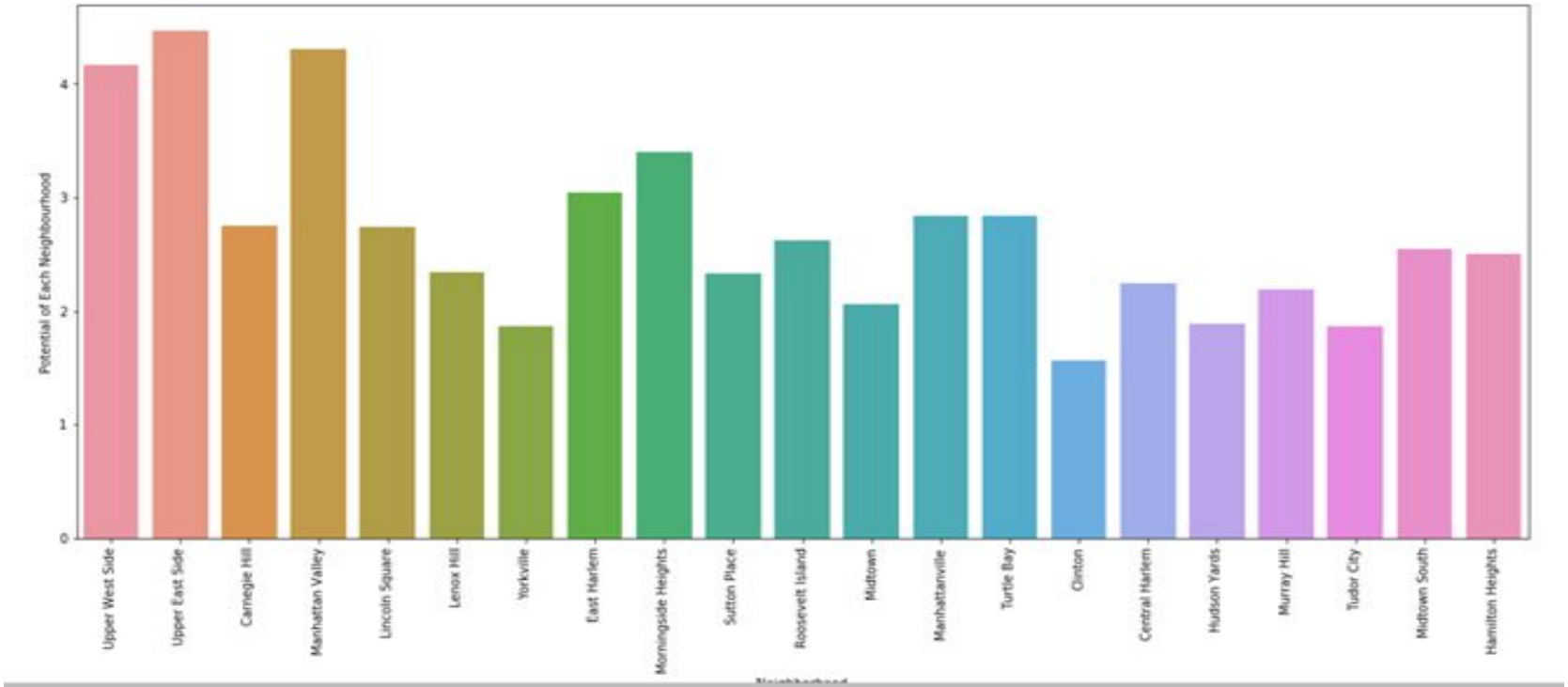
Cluster 2 - These are neighbourhoods with plenty of gyms already. There will be fierce competition here and that may impact revenue. These points are marked in **purple**.

Cluster 3 - These are the points closest to Central Park and are greatly untapped. These neighbourhoods offer great potential. I would recommend these neighbourhoods. However, one must keep in mind that these neighbourhoods are expensive and there will be significant overhead costs to setting up a gym here. These points are marked in light blue.

Cluster 4 - These are points that are slightly further away from the centre. However, no gym exists in the neighbourhood of these vicinities so it may result in high revenue. The cost of setting up the gym might also be much lower than cluster 2. They are marked in light green.

Cluster 5- This cluster shows points that are somewhat between the outskirts and centre of Manhattan. There exist gyms already which may offer some competition. However, that number is far lesser than that of cluster 1. They can also be promising as prices will be lower. These are shown in orange.

BAR GRAPH OF POTENTIAL VS NEIGHBOURHOOD



CONCLUSION

The purpose of this project was to identify a suitable location for stakeholders looking to set up a gym. Through the application of clustering and Foursquare API, I have identified the best neighbourhoods for this project. They are

- · Manhattan Valley
- · Upper East Side
- · Manhattanville
- · Turtle Bay

Manhattan Valley would require higher capital but would generate the maximum revenue. It is the closest to the centre and has complementary businesses. Upper East Side is a good prospect as well but will be even more expensive. Manhattanville and Turtle Bay are ideal prospects for stakeholders who would not wish to play the excessive charges of location. The competition in this region is higher as most gym owners would prefer this region.