

Applied Data Science Capston Assignment

Capston Project: The Battle of Neighborhoods

Report

Pedro García Iturrade

April 27th

1. Description of the problem:

The city of Manhattan is a very large city with very different neighborhoods inside.

The Project explores this neighborhoods in order to decide which of them are good places to open a Sports Clothing and Goods store.

The analysis will be centered in the number of sport venues such as gyms and sports clubs and on the other side the number of shops like the one to open. The criteria will be to choose the neighborhood with more sport venues and less shops in order to maximize the clients, reducing the competition with other stores.

Target: sport shops chain owners or anybody who wants to open a sports clothing and goods shop in Manhattan.

2. Description of the data:

Data sources:

- Geographical coordinates of the neighborhoods in New York city from https://cocl.us/new_york_dataset
- Venues location and category from Foursquare using requests to get a Json file.

Data cleaning and feature selection:

- Geographical coordinates:
After getting the dataset of New York and adding it to a dataframe it was filtered obtaining only Manhattan Neighborhoods information. The features of interest where the name of the neighborhoods and their coordinates.
- Neighborhood Venues
With the coordinates of the neighborhoods a request was sent to Foursquare to get a Json file with the venues. Afterwards, the Json file was transformed into a DataFrame with the necessary data, the venues for each neighborhood. The DataFrame was then filtered keeping the "Venues Categories" related to sports and sports shops.

3. Methodology

The information obtained after cleaning the data was kept in three dataframes as follows:

- Neighborhoods and coordinates

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

- Neighborhoods and Sport Venues

	Baseball Field	Boxing Gym	Climbing Gym	Gym	Gym / Fitness Center	Soccer Field	Sports Bar	Sports Club	Tennis Court	Tennis Stadium	Neighborhood
0	0	0	0	2	0	0	0	0	0	0	Battery Park City
1	0	0	0	2	3	0	0	0	0	0	Carnegie Hill
2	0	0	0	0	1	0	0	0	0	0	Central Harlem
3	0	0	0	0	0	0	0	0	0	0	Chelsea
4	0	0	0	0	0	0	0	0	0	0	Chinatown

- Neighborhoods and Sport Shops

	Shopping Mall	Sporting Goods Shop	Neighborhood
0	1	0	Battery Park City
1	0	0	Carnegie Hill
2	0	0	Central Harlem
3	0	0	Chelsea
4	0	0	Chinatown

This information was grouped by neighborhoods and then totalized, obtaining two new features per neighborhood:

- Total Sport Venues
- Total Sport Shops

	Neighborhood	Total Sport Venues	Total Sport Shops
0	Battery Park City	2	1
1	Carnegie Hill	5	0
2	Central Harlem	1	0
3	Chelsea	0	0
4	Chinatown	0	0

Then a point sistem was thought in order to classify the neighborhoods.

$$NBH_{points} = \frac{Total\ Sport\ Venues}{Total\ Sport\ Shops + 1}$$

This points simbolize the distribution of posible customers of the venues in the the exsiting shops plus the new shop.

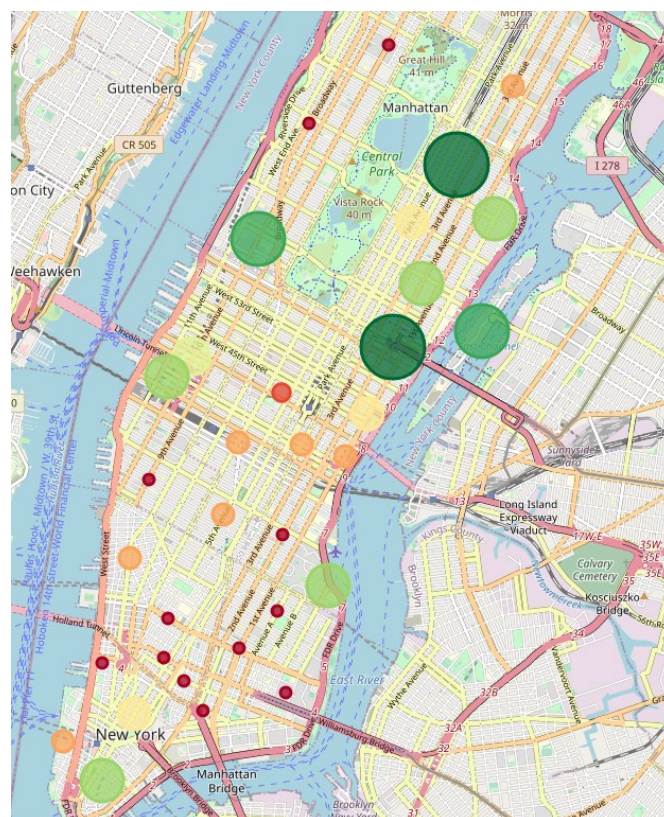
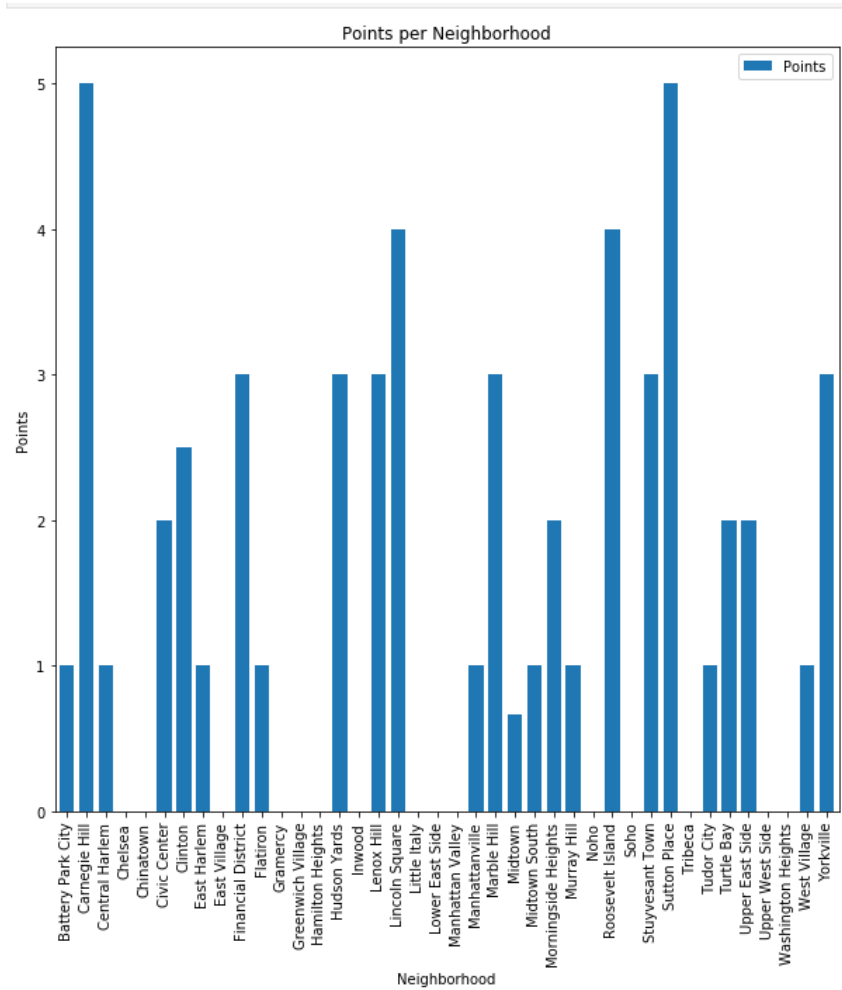
So a new feature “Points” with this operation was added.

	Neighborhood	Total Sport Venues	Total Sport Shops	Points
0	Battery Park City	2	1	1.0
1	Carnegie Hill	5	0	5.0
2	Central Harlem	1	0	1.0
3	Chelsea	0	0	0.0
4	Chinatown	0	0	0.0

Finally the neighborhoods were clustered according to their points to see which of them were mor suitable according to the criteria.

4. Results

To visualize the results, a bar chart showing the points for each neighborhood was plotted and then a map showing the location of the neighborhoods and their conviennene colored from green to red.



5. Conclusion and Observations

According to the criteria proposed, Carnegie Hill and Sutton Place are the two most suitable neighborhoods to open a Sports Clothing and Goods Store. In order to choose one of them or perhaps another with also a good score such as Lincoln Square or Roosevelt Island, there are other features that can be taken into account like the cost of renting or buying the place for example.

However, this may be a good approach in order to decide where to open the store in the city of Manhattan.