Locations for Sports Stores in Brooklyn, NY

Business Problem

Goal of this project is to find the best location/neighborhood for new sports equipment shops in Brooklyn, NY. The criteria for a good neighborhood are that it should be located in a neighborhood with a high density of gyms, sports stadiums, baseball fields, and all sports-related venues indeed, in order to attract the highest possible amount of customers. However, the customer also wants to know about neighborhood classifications in terms of what types of sports venues are, so he can open up the right type of sports shop in the right neighborhood (i.e. a personal fitness store in a neighborhood with a high amount of gyms, or a team sport store where many people are witnessed playing basketball, soccer, etc.).

Data and Procedure

The data provided is of geographical nature (coordinates) regarding the neighborhoods and also venue-related (categories, locations, etc.). First, I will retrieve the geographical data, store in a data frame and reduce it to the relevant Brooklyn content. Then, I will use the folium package to visualize the relevant neighborhoods in Brooklyn. The venue data will be analyzed regarding the discussed target venues and assessed in terms of the highest amount of sports venues. Finally, the neighborhoods will be classified into clusters that appropriately represent the types of sports venues, so one can make recommendations as to which type of stores should be opened in which neighborhoods.

Data Wrangling

The geographical data is converted into a DataFrame (see extract in Figure 1) and visualized by the folium package displayed in Figure 2. The blue dots represent the corresponding neighborhoods. Next, the webpage Foursquare.com is utilized to extract venue data. The relevant data can be stored in a DataFrame, as shown by Figure 3. Of special importance is the 'venue category', which will be used to categorize the neighborhoods.

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

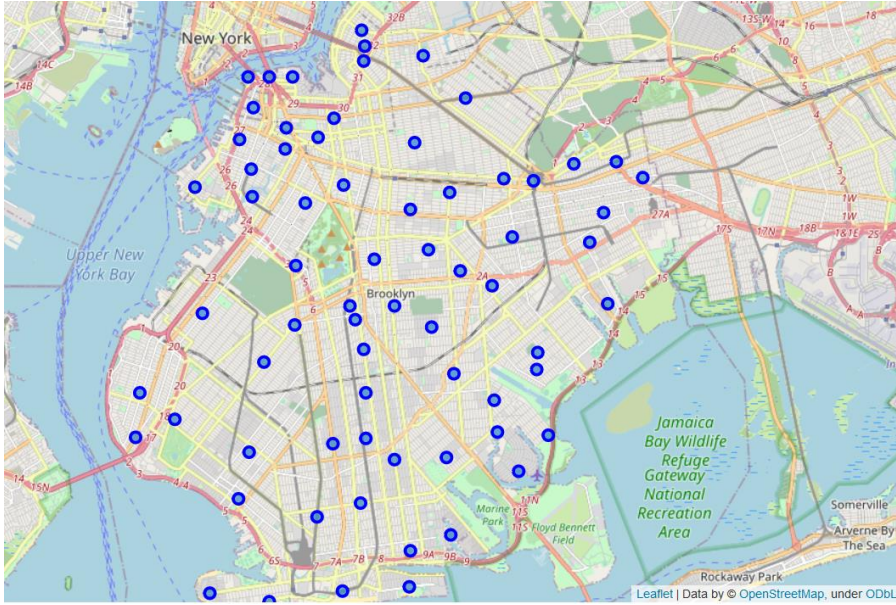Figure 1: extract oft he geographical data from Brooklyn, NY

Figure 2: visualization of the geographical data in Brooklyn, NY

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Leo's Casa Calamari	40.624200	-74.030931	Pizza Place
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	The Bookmark Shoppe	40.624577	-74.030562	Bookstore

Figure 3: DataFrame of venue data from the Brooklyn neighborhoods

The venue categories are filtered are filtered for a relation to sports or outdoor activities such as yoga, gym, basketball, pool, skating, as well as park areas. This information can once again be visualized together with the respective neighborhoods, as displayed in Figure 4.

Methodology, Data Exploration and Analyzation

The DataFrame with the venue data is onehot-encoded in order to perform a clustering algorithm. This is necessary, because the categorical variables (venue categories) must be handled in terms of numeric values for a successful clustering operation. The neighborhoods with the highest number of sports venues can be extracted and are shown in Figure 5

Figure 4: visualization of the sports-related venues (green) and neighborhoods in Brooklyn, NY

	Neighborhood
Brooklyn Heights	15
Dumbo	13
South Side	9
North Side	9
Fulton Ferry	8

Figure 5: Neighborhoods with highest amount of sports venues in Brooklyn, NY

The data is clustered by a kmeans methodology to reveal similarities in the neighborhoods regarding the sports venues. An analysis of the error of the method for different cluster numbers by the elbow-method showed no distinct value for the cluster number, however a value of k = 3 appears reasonable (see Figure 6).

The data was subsequently merged with the calculated most common venues per neighborhood. Figure 7 shows this, but it also reveals that some neighborhoods could not be classified, because they do not contain any sports venues. This was remedied by deleting these irrelevant neighborhoods.
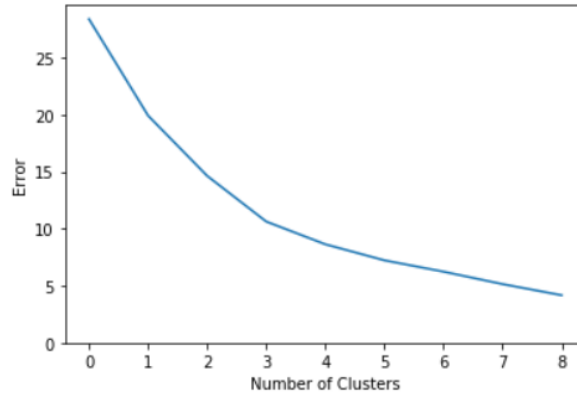
Figure 6: Error of the clustering algorithm as a function of the number of clusters used

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Brooklyn	Bay Ridge	40.625801	-74.030621	0.0	Yoga Studio	Park	Gym / Fitness Center
1	Brooklyn	Bensonhurst	40.611009	-73.995180	1.0	Park	Yoga Studio	Gym
2	Brooklyn	Sunset Park	40.645103	-74.010316	2.0	Gym	Yoga Studio	Basketball Court
3	Brooklyn	Greenpoint	40.730201	-73.954241	0.0	Yoga Studio	Gymnastics Gym	Gym
4	Brooklyn	Gravesend	40.595260	-73.973471	2.0	Gym	Yoga Studio	Basketball Court
...	...	...	...	...	...	...	...	...
65	Brooklyn	Dumbo	40.703176	-73.988753	0.0	Park	Yoga Studio	Gym
66	Brooklyn	Homecrest	40.598525	-73.959185	NaN	NaN	NaN	NaN
67	Brooklyn	Highland Park	40.681999	-73.890346	0.0	Gym / Fitness Center	Park	Tennis Court
68	Brooklyn	Madison	40.609378	-73.948415	NaN	NaN	NaN	NaN
69	Brooklyn	Erasmus	40.646926	-73.948177	0.0	Yoga Studio	Gym / Fitness Center	Surf Spot

Figure 7: venue data with calculated cluster labels and determined most common venues

Interpretation of the Results

The analysis is visualized in Figure 8. The 53 relevant neighborhoods were clustered into a large group 0 (33 neighborhoods), a medium sized group 1 (14 neighborhoods), and a small group 2 (6 neighborhoods). According to the information provided in Figure 9, the three clusters can be explained by the following preferences regarding sports venues:

- 0: high amount of yoga studios and good amount of gyms (red)
- 1: Outdoor and Park neighborhoods with a good amount of yoga studios (purple)
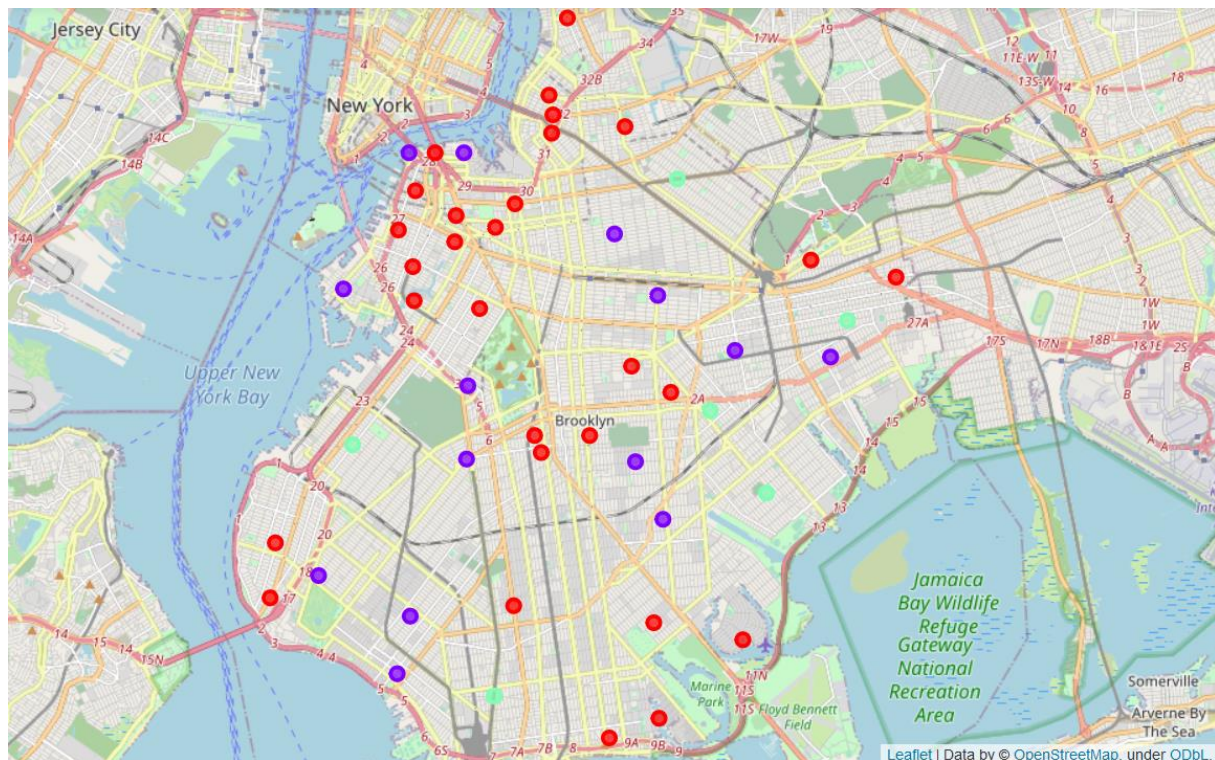- 2: high concentration of gyms with a good amount of yoga studios (green)

Figure 8: visualization of the relevant neighborhoods and their corresponding clusters

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Bay Ridge	Yoga Studio	Park	Gym / Fitness Center
3	Greenpoint	Yoga Studio	Gymnastics Gym	Gym
5	Brighton Beach	Beach	Yoga Studio	Gym
6	Sheepshead Bay	Yoga Studio	Gym	Basketball Court
15	Williamsburg	Yoga Studio	Gym	Park

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	Bensonhurst	Park	Yoga Studio	Gym
10	East Flatbush	Park	Yoga Studio	Gym
11	Kensington	Park	Yoga Studio	Gym
12	Windsor Terrace	Park	Yoga Studio	Gym
14	Brownsville	Park	Yoga Studio	Gym

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
2	Sunset Park	Gym	Yoga Studio	Basketball Court
4	Gravesend	Gym	Yoga Studio	Basketball Court
16	Bushwick	Gym	Yoga Studio	Basketball Court
26	East New York	Gym	Yoga Studio	Basketball Court
28	Canarsie	Gym	Yoga Studio	Basketball Court
57	Remsen Village	Gym	Yoga Studio	Basketball Court

Figure 9: extract of 5 neighborhoods in cluster 0 (top), cluster 1 (middle) and cluster 2 (bottom)

Conclusion

The neighborhoods in Brooklyn with the highest amount of sports venues are Brooklyn Heights, Dumbo, North Side, South Side and Fulton Ferry (Figure 10). Sports equipment shops will most likely attract the most attention if opened in one of these neighborhoods. Furthermore, they are all very close to Manhattan, which will most likely attract a large number of customers from there as well.

Four of the mentioned top five neighborhoods fall into the category 0 (high amount of yoga studios and good amount of gyms (red)). An equipment store related to these categories would be preferable to attract this clientele. Cycling equipment could be a promising merchandize to sell in North and South Side, since many cycling studios are located here as well.

Fulton Ferry is characterized as an Outdoor and Park neighborhood. Sports stores that sell park equipment (i.e. frisbees, rollerblades) are the recommendation for this area.

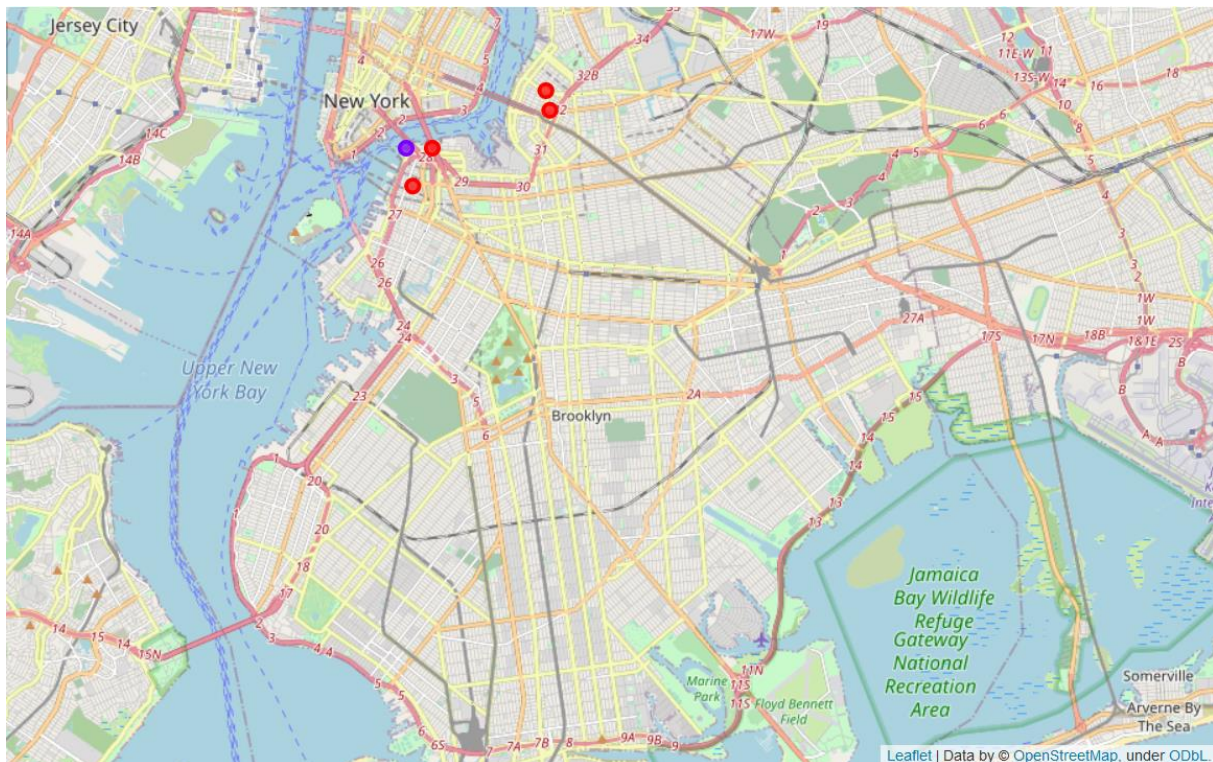|	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
18	Brooklyn	Brooklyn Heights	40.695864	-73.993782	0.0	Yoga Studio	Park	Gym
50	Brooklyn	North Side	40.714823	-73.958809	0.0	Yoga Studio	Cycle Studio	Gym
51	Brooklyn	South Side	40.710861	-73.958001	0.0	Yoga Studio	Cycle Studio	Park
61	Brooklyn	Fulton Ferry	40.703281	-73.995508	1.0	Park	Beach	Yoga Studio
65	Brooklyn	Dumbo	40.703176	-73.988753	0.0	Park	Yoga Studio	Gym

Figure 10: five places with the highest amount of sports venues in Brooklyn, NY