Determining the Optimal Neighborhood to Build a New Fitness Facility in Boston, Massachusetts

Sayaka Minegishi

April 7, 2021

1. Introduction

Background:
With more jobs requiring people to sit in their office and work using computers all day, an increasing number of individuals are experiencing a sedentary lifestyle.

According to the 2018 Physical Activity Guidelines for Americans, 2nd edition, released by the U.S. Department of Health and Human Services, adults are recommended to do at least 150 minutes a week of moderate-intensity aerobic physical activity, and muscle-strengthening activities of moderate or greater intensity on at least 2 days per week.[1]

However, CDC reports that merely 23.2% of U.S. adults aged 18 and older met the Physical Activity Guidelines for both aerobic and muscle-strengthening activity, based on data from the 2018 National Health Interview Survey.[2] To help maintain the health of Americans, there is an increasing demand to make recreational facilities equally accessible to all.

Researchers from Stanford University School of Medicine have reported that "inactivity rather than overeating could be driving the surge in Americans' obesity"[3]. They have found that while obesity and inactivity has seen a drastic growth from 1988 through 2010, no significant increases were observed in the overall number of calories consumed.[4] It is reasonable to deduce that by reducing the amount of exercise, individuals are burning fewer calories daily, which is leading to obesity.

Problem:
A majority of the Americans are not getting the recommended amount of exercise in their lives[5]. To encourage more individuals to exercise, we need to have more recreational facilities distributed evenly in neighborhoods so that everyone would have equal access to these facilities.

[1] (U.S. Department of Health and Human Services 2018)
[2] ("Exercise or Physical Activity" 2021)
[3] (Bach n.d.)
[4] (Bach n.d.)
[5] ("Exercise or Physical Activity" 2021)

This report aims to find the best neighborhood to open a new gym in Boston by taking into consideration the number of gyms/fitness centers and parks available, as well as the number of restaurants. The optimal location for a new gym would be the neighborhood with a relatively large number of restaurants as compared to gyms or fitness centers, and to a lesser extent, parks.

Interest:
This report is specifically targeted to stakeholders who are considering opening a new gym or a fitness center in Boston, Massachusetts.

2. Data

Data Sources:
This analysis employs data about the neighborhood features in Boston, MA, which is available as a CSV file on https://data.boston.gov/dataset/boston-neighborhoods. This file contains data on all the neighborhoods in Boston, including their neighborhood name and area in square miles. This data is used to find the names of the neighborhoods, which forms the foundational columns for our dataframes.

I utilize the Nominatim library from GeoPy to find the geographical coordinates of each of the neighborhoods.

The geographical coordinates of each neighborhood would be passed to Foursquare API to search for restaurants, gyms and parks in each of the neighborhoods. We would like to place a new gym where there are a relatively large number of restaurants where there would be a greater need or tendency to exercise. At the same time, we would like to construct a gym where there are relatively fewer fitness centers or parks.

Data Cleaning and Feature Selection:
From Boston's neighborhood data (https://data.boston.gov/dataset/boston-neighborhoods), only the neighborhood name and the area of land in square miles seemed relevant to our analysis. Thus I dropped the remaining columns. I saved the remaining data as a pandas dataframe called boston_data. I then created a new dataframe called boston_neighborhoods, which included neighbourhood name, the address of the neighbourhood, area, latitude, and longitude. I filled in the latitude and longitude columns after using GeoPy to find the geographical coordinates. The other columns were filled out by copying the data from the boston_data dataframe.

After obtaining data on the venues in each neighborhood using Foursquare, I constructed data frames to separately store venues in each category of interest: parks, restaurants, and gyms. The restaurants in the analysis included all venues whose venue category contained the word 'restaurant' or 'pizza place'. Similarly, all venues that were categorized as a 'park' included all

venues that contained the word ‘park’ in their venue categories, and all venues that contained the word ‘gym’ were categorized as a ‘gym’. All cells with NaN values were given a value of 0. The three data frames were concatenated into a single dataframe called ‘boston_gymandfood’.

After k-means clustering was performed, a final data frame called ‘boston_table_finalized’ was created to summarize the results. As there were rows in the boston_table_finalized dataframe with NaN values, I dropped these rows from the dataframe. The neighbourhoods that were removed in this process were West Roxbury and Harbor Islands.

3. Methodology

Exploratory Data Analysis:
The project analyzed a total of 23 neighborhoods in Boston. When Foursquare API was used to obtain the venues within 500 meters radius of the geographical coordinates of each of the neighborhoods, I used the ‘.shape’ function to check the number of venues returned, which was 1100. After I have separated the results into 3 data frames separated by the category of the venue, I used the same method to find that there were 29 parks, 345 restaurants, and 31 gyms in Boston’s neighbourhoods altogether. The final data frame containing information about the frequency of the venues for each neighborhood shows that restaurants are the most common venue out of the three venue types in all neighborhoods except in Roxbury, where gym is the most common.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
9	East Boston	Restaurant	Park	Gym
16	Mission Hill	Restaurant	Park	Gym
15	Mattapan	Restaurant	Park	Gym
14	Longwood	Restaurant	Park	Gym
12	Jamaica Plain	Restaurant	Park	Gym
20	South Boston	Restaurant	Park	Gym
21	South Boston Waterfront	Restaurant	Park	Gym
18	Roslindale	Restaurant	Park	Gym
5	Charlestown	Restaurant	Park	Gym
3	Beacon Hill	Restaurant	Park	Gym
17	North End	Restaurant	Park	Gym
0	Allston	Restaurant	Gym	Park
10	Fenway	Restaurant	Gym	Park
2	Bay Village	Restaurant	Park	Gym
4	Brighton	Restaurant	Park	Gym
6	Chinatown	Restaurant	Gym	Park
22	South End	Restaurant	Park	Gym
1	Back Bay	Restaurant	Gym	Park
13	Leather District	Restaurant	Gym	Park
11	Hyde Park	Restaurant	Gym	Park
8	Downtown	Restaurant	Park	Gym
23	West End	Restaurant	Gym	Park
7	Dorchester	Restaurant	Gym	Park
19	Roxbury	Gym	Restaurant	Park

Table 1: The three types of venues compared (Gym, Restaurant and Park) ranked in order of frequency in each Boston neighborhood.

Data Analysis by Clustering:

I have conducted k-means clustering, a machine learning algorithm, to cluster Boston's neighborhoods into four separate clusters.

The neighborhoods were color coded based on their cluster labels and were superimposed on the map of Boston, as shown below.

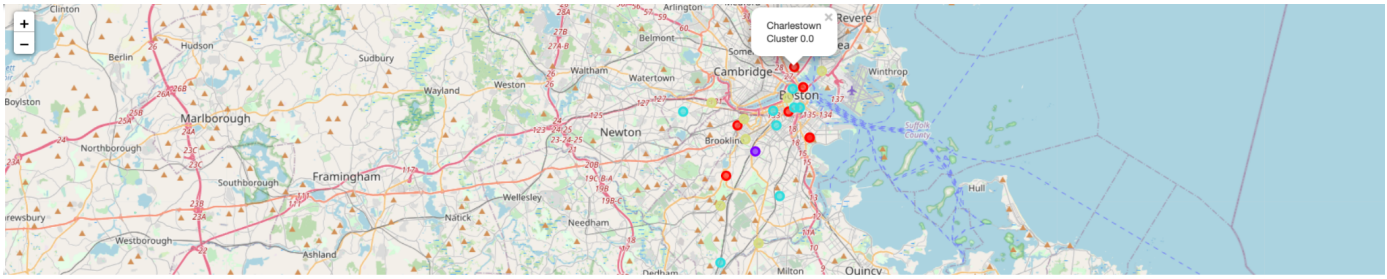

Figure 1: Map of Boston showing the neighborhoods as points, color-coded based on their cluster labels. Red represents Cluster 0, purple represents Cluster 1, light blue represents Cluster 2, and yellow represents Cluster 3 neighborhoods.

The neighborhoods in each cluster were stored in separate data frames, together with data showing the percentage frequency of each type of venue (restaurant, gym or park) relative to each other. These data are represented below.

	Address	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Gym	Park	Restaurant
1	Jamaica Plain, Massachusetts	0.0	Restaurant	Park	Gym	0.000000	0.250000	0.750000
3	Longwood, Massachusetts	0.0	Restaurant	Park	Gym	0.000000	0.200000	0.800000
4	Bay Village, Massachusetts	0.0	Restaurant	Park	Gym	0.066667	0.133333	0.800000
7	North End, Massachusetts	0.0	Restaurant	Park	Gym	0.022222	0.111111	0.866667
12	Charlestown, Massachusetts	0.0	Restaurant	Park	Gym	0.000000	0.222222	0.777778
21	South Boston Waterfront, Massachusetts	0.0	Restaurant	Park	Gym	0.000000	0.153846	0.846154
22	South Boston, Massachusetts	0.0	Restaurant	Park	Gym	0.000000	0.153846	0.846154

Table 2: Neighborhoods in Cluster 0, displaying the percentage frequency of each type of venue.

	Address	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Gym	Park	Restaurant
8	Roxbury, Massachusetts	1.0	Gym	Restaurant	Park	0.5	0.25	0.25

Table 3: Neighborhoods in Cluster 1, displaying the percentage frequency of each type of venue.

	Address	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Gym	Park	Restaurant
5	Leather District, Massachusetts	2.0	Restaurant	Gym	Park	0.139535	0.023256	0.837209
6	Chinatown, Massachusetts	2.0	Restaurant	Gym	Park	0.090909	0.045455	0.863636
9	South End, Massachusetts	2.0	Restaurant	Park	Gym	0.105263	0.105263	0.789474
10	Back Bay, Massachusetts	2.0	Restaurant	Gym	Park	0.111111	0.027778	0.861111
13	West End, Massachusetts	2.0	Restaurant	Gym	Park	0.200000	0.040000	0.760000
15	Downtown, Massachusetts	2.0	Restaurant	Park	Gym	0.166667	0.166667	0.666667
17	Brighton, Massachusetts	2.0	Restaurant	Park	Gym	0.083333	0.083333	0.833333
18	Hyde Park, Massachusetts	2.0	Restaurant	Gym	Park	0.166667	0.000000	0.833333
20	Dorchester, Massachusetts	2.0	Restaurant	Gym	Park	0.250000	0.000000	0.750000

Table 4: Neighborhoods in Cluster 2, displaying the percentage frequency of each type of venue.

	Address	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Gym	Park	Restaurant
0	Roslindale, Massachusetts	3.0	Restaurant	Park	Gym	0.000000	0.000000	1.000000
2	Mission Hill, Massachusetts	3.0	Restaurant	Park	Gym	0.000000	0.100000	0.900000
11	East Boston, Massachusetts	3.0	Restaurant	Park	Gym	0.000000	0.071429	0.928571
14	Beacon Hill, Massachusetts	3.0	Restaurant	Park	Gym	0.000000	0.062500	0.937500
16	Fenway, Massachusetts	3.0	Restaurant	Gym	Park	0.066667	0.000000	0.933333
19	Mattapan, Massachusetts	3.0	Restaurant	Park	Gym	0.000000	0.000000	1.000000
23	Allston, Massachusetts	3.0	Restaurant	Gym	Park	0.027027	0.000000	0.972973

Table 5: Neighborhoods in Cluster 3, displaying the percentage frequency of each type of venue.

For each cluster, I have created a series of bar charts to illustrate the percentage frequency of each type of venue in each Boston neighborhood.

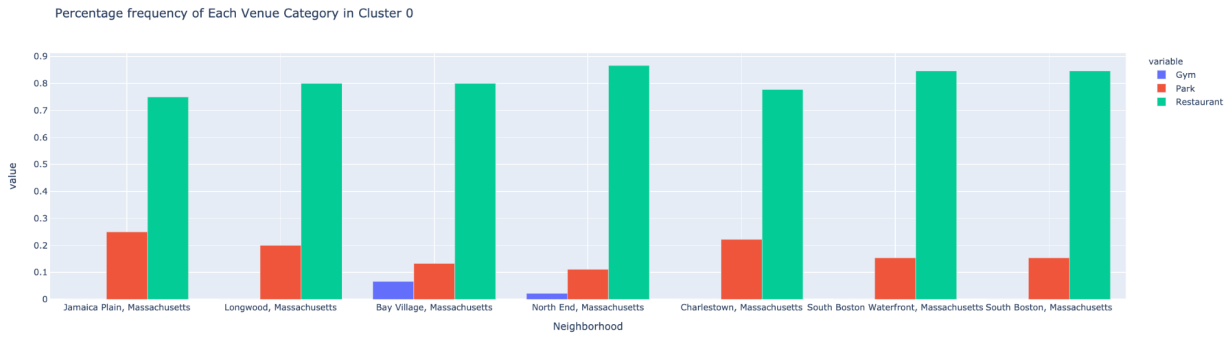

Figure 2: Bar chart comparing the percentage frequency of each venue category in Cluster 0

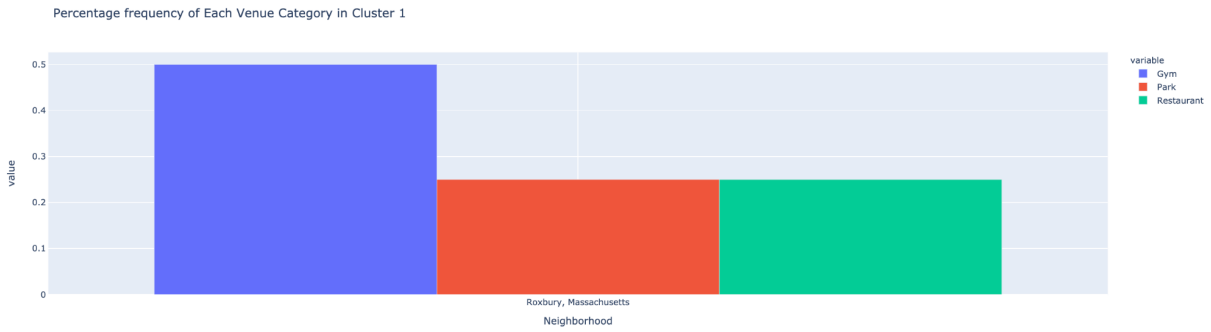

Figure 3: Bar chart comparing the percentage frequency of each venue category in Cluster 1

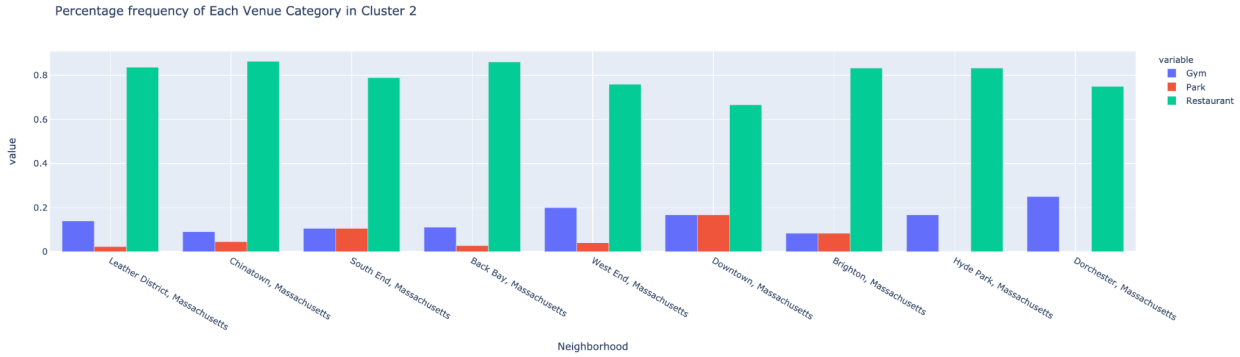

Figure 4: Bar chart comparing the percentage frequency of each venue category in Cluster 2

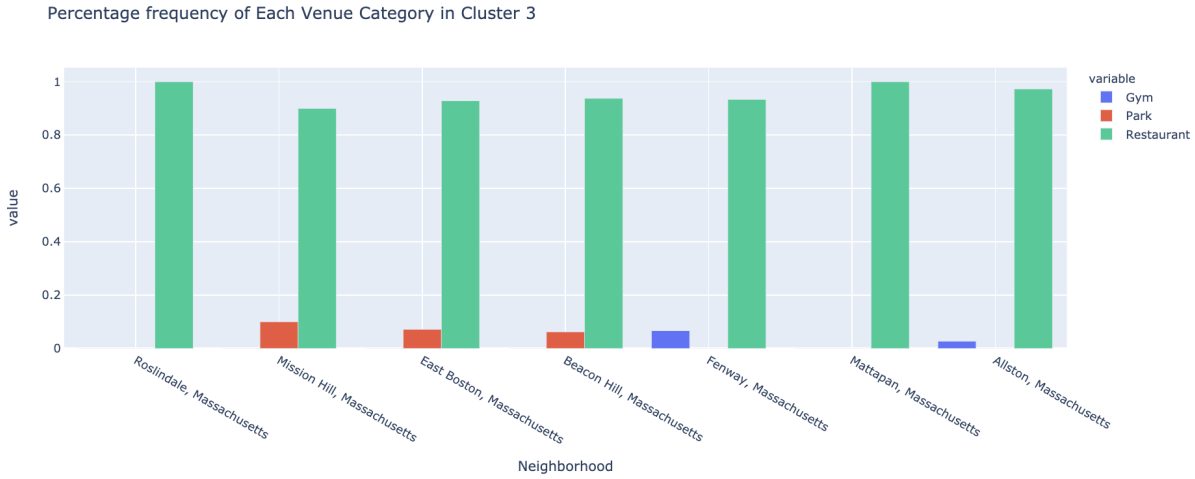

Figure 5: Bar chart comparing the percentage frequency of each venue category in Cluster 3

4. Results

In this analysis, we applied the k-means clustering technique to separate Boston neighborhoods into 4 distinct clusters, based on the similarities in the proportions of gym, parks and restaurants present in each neighborhood. Examination of the clusters revealed that Cluster 0 was composed solely of neighborhoods whose most common venue was 'Restaurant', followed by 'Park' and 'Gym'. More than half of the neighborhoods in this cluster had very few or no gyms in the neighborhood, while the proportions of restaurants were consistently very high, constituting at least 75% of the facilities compared. Neighborhoods in Cluster 2 and Cluster 3 all shared 'Restaurant' category as the most common venue type; however, unlike in Cluster 0, 'Park' and 'Gym' switched between being the 2nd most common venue and the 3rd most common venue. Upon examination of the proportions for each venue category, it becomes apparent that all of the neighborhoods in Cluster 2 contain a fitness center or a gym, while this is not so for Cluster 3. Unlike in Cluster 0 where all neighborhoods had at least one park despite having no gyms, there are two neighborhoods in Cluster 3 (namely Roslindale and Mattapan) where there are no gyms or parks visible in the data, and restaurants comprise 100% of the venues compared. Cluster 1, which comprises only 1 neighborhood (Roxbury), is unique from all other clusters in that the gym is the most common venue, followed by restaurants and parks.

5. Discussion

The results of this analysis guide us to prioritize the neighborhoods in Clusters 0 and 3 as potential locations to build a new fitness facility such as a gym. In particular, neighborhoods in Cluster 3 have the highest proportion of restaurants compared to the proportions of gyms or parks, as compared to the neighborhoods in Cluster 0. This is exemplified by how restaurants in all of the Cluster 3 neighborhoods comprise over 92% of the venues compared, while the proportion of restaurants never exceeds 90% in Cluster 0 neighborhoods. Thus, stakeholders should especially consider neighborhoods in Cluster 3. Having made this recommendation, it must be noted that this analysis has not compared all possible factors that would make a neighborhood suitable for building a new gym. To further narrow down our decisions on the best venues, a deeper analysis of the neighborhoods is recommended, including the demographics of the neighborhoods and property prices.

6. Conclusion

An analysis using k-means clustering was conducted to determine which Boston neighborhood should be prioritized when stakeholders are considering the next location to build a gym. Upon examination of our clusters, it became evident that Cluster 3 consistently had the highest proportion of restaurants and a very small proportion of gyms and parks combined. Thus,

neighborhoods that are included in Cluster 3 should be prioritized when considering the location to build a new fitness facility. These neighborhoods are: Roslindale, Mission Hill, East Boston, Beacon Hill, Fenway, Mattapan, and Allston. South Boston and South Boston Waterfront should be especially considered, as these neighborhoods had the highest proportion of restaurants and the lowest proportion of gyms among all the neighborhoods in this cluster.

7. Bibliography

Alkson, Alex, and Lin, Polong. 2020. "Segmenting and Clustering Neighborhoods in New York City." IBM Corporation. November 26, 2020. https://labs.cognitiveclass.ai/tools/jupyterlab/lab/tree/labs/DS0701EN/DS0701EN-3-3-2-Neighborhoods-New-York-py-v1.0.ipynb?lti=true.

"American Heart Association Recommendations for Physical Activity in Adults and Kids." n.d. Accessed April 6, 2021. https://www.heart.org/en/healthy-living/fitness/fitness-basics/aha-recs-for-physical-activity-in-adults.

Bach, Becky. n.d. "Lack of Exercise, Not Diet, Linked to Rise in Obesity, Stanford Research Shows." Accessed April 7, 2021. https://med.stanford.edu/news/all-news/2014/07/lack-of-exercise--not-diet--linked-to-rise-in-obesity--stanford-.html.

"Exercise or Physical Activity." 2021. March 1, 2021. https://www.cdc.gov/nchs/fastats/exercise.htm.

U.S. Department of Health and Human Services. 2018. *Physical Activity Guidelines for Americans, 2nd Edition*. U.S. Department of Health and Human Services.