

GYMS IN BOSTON: TO DO OR NOT TO DO

Reha Patel

Problem

- In a city of almost 700,000 people with many more commuting in and out every day, Boston has many different neighborhoods and we will determine *if there is a neighborhood that holds potential for development of a gym.*

Data Used

- I found a list of neighborhoods in a Wikipedia article and web scraped the page to add them to a data frame.
- I used the Foursquare API to gather venue information within a specified limit and radius.
- I manually searched the coordinates of the city of Boston in order to ensure the coordinates were accurate.

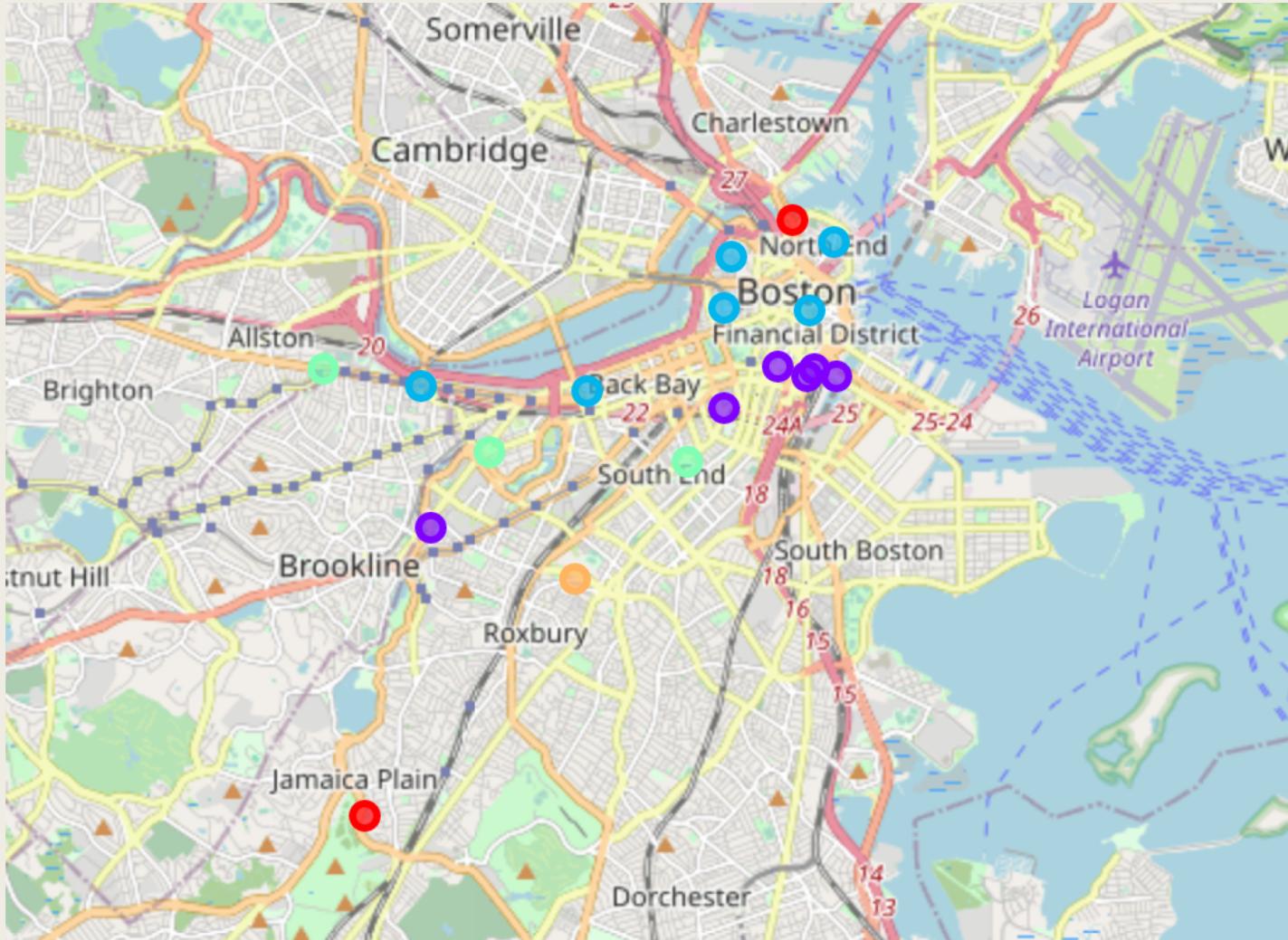
Methodology

- I created numerous data frames, each building off the one before it in order to store data with Neighborhood being the baseline.
- I used *folium* to visualize the Greater Boston Area with the neighborhoods marked and eventually with the neighborhoods marked into their respective clusters.
- I used the Foursquare API to gather a list of 100 venues within a mile radius of the coordinates provided. In a city where walking is common, a mile radius seemed appropriate.
- Almost each neighborhood reached the limit of 100 venues. It was noted that neighborhoods closer to Downtown were more likely to reach the limit.

Observations

- It was noted that the 3 most popular venue categories across neighborhoods were *Coffee Shop*, *Italian Restaurant*, and *Pizza Place*. This fundamentally makes sense given my knowledge of the heavy Italian presence in Boston.
- Gyms were more populous in neighborhoods closer to Downtown, potentially closer to where businesses and businesses are.
- The neighborhoods without gyms within a mile radius are found furthest from Downtown, or in the case of Charlestown, are a small entirely residential neighborhood.

Map with Clustered Neighborhoods



the 5 clusters depicted by 5 different colors

Final Data Frame

- The final data frame clusters each neighborhood into 5 clusters using the k-means algorithm and also depicts the proportion of gyms within each.

	Neighborhood	Gym	Cluster Labels	Latitude	Longitude
5	Charlestown	0.000000	0	42.367771	-71.059016
13	Mattapan	0.000000	0	42.278222	-71.096083
11	Hyde Park	0.000000	0	42.274773	-71.119898
12	Jamaica Plain	0.000000	0	42.305849	-71.119092
18	South Boston	0.030000	1	42.352250	-71.055690
2	Bay Village	0.030000	1	42.348165	-71.068470
6	Chinatown/Leather District	0.030000	1	42.352510	-71.060900
7	Dorchester	0.030000	1	42.351355	-71.052848
14	Mission Hill	0.030000	1	42.335710	-71.109800
9	East Boston	0.030000	1	42.351418	-71.056714
0	Allston	0.010000	2	42.350531	-71.111091
16	Roslindale	0.012987	2	42.281820	-71.137104
15	North End	0.010000	2	42.365490	-71.052970
21	West Roxbury	0.013333	2	42.282201	-71.146000
8	Downtown	0.010000	2	42.358290	-71.056630
3	Beacon Hill	0.010000	2	42.358420	-71.068600
1	Back Bay	0.010000	2	42.349990	-71.087650
20	West End	0.010000	2	42.363940	-71.067390
4	Brighton	0.020000	3	42.352134	-71.124925
19	South End	0.020000	3	42.342560	-71.073580
10	Fenway Kenmore	0.020000	3	42.343550	-71.101570
17	Roxbury	0.040000	4	42.330304	-71.089469

Conclusion

- I would suggest opening a gym in Cluster 0. Cluster 0 is a generally residential area and would potentially find benefits in development.
- I would deter investors from opening any gyms in Clusters 1 and 2. Because these areas are densely populated with venues ranging from bars to restaurants, there are also many gyms and therefore a new gym would face severe competition. One issue the gym would face in this area is convincing gym-goers to switch from their home gym to the new gym and it is ultimately not worth the risk.