

## Introduction

With the aging population growing in numbers in Singapore, many people are focusing more on health in order to achieve healthy aging. This perception has influenced many people from all range of ages to participate actively in physical exercises. As a result, there are increasing number of gyms or fitness centers that are being set up across all parts of Singapore. Accessibility is one of the most important factors in which people chooses their gym/ fitness center. Hence, understanding the locations of gyms and their respective densities represent the first step in deciding the where should the next gym/ fitness center be such that it could best cater to consumer needs.

### Business problem:

Therefore, the specific business problem that this study will attempt to address is: Where will be the best location to set up a gym/fitness center in Singapore now? This information will be especially useful for investors and developers, who are the main target audience of this study, to decide on the next location for setting up fitness centers or deciding the best location for conducting fitness classes.

### Data

The following data will be used to solve the above problem. 1) All 40 areas of Singapore will be analyzed for the numbers of enrichment centers. The list of the 40 areas can be obtained from the Wikipedia page ([https://en.wikipedia.org/wiki/Planning\\_Areas\\_of\\_Singapore](https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore)). Using python request and beautiful Soup packages, data to these areas will be extracted. Geographical coordinates will be obtained via python geocoder package. 2) Venue data for each of these areas will be extracted using the Foursquare API. These data will then be used for machine learning by K-Means clustering and each clusters will be overlapped on the map using folium to visualize the density of gyms across each area.

## Methodology

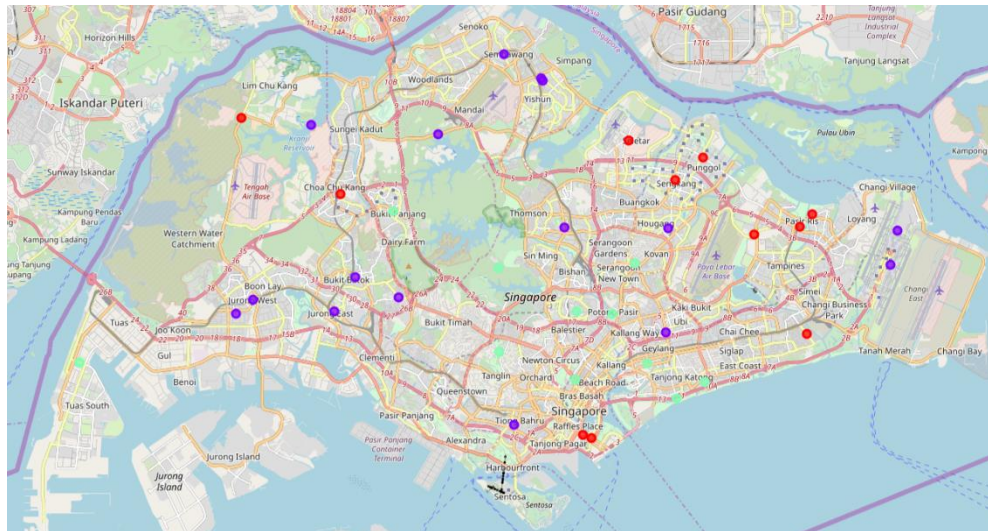
As I aim to identify the different locations and densities of gym/fitness centers across different parts of Singapore, the first step is to segment Singapore into areas in which venue data can be obtain. For segmenting Singapore, I obtain the list of planning areas of Singapore from Wikipedia ([https://en.wikipedia.org/wiki/Planning\\_Areas\\_of\\_Singapore](https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore)). A total of 55 areas were listed on the website. However, there were some areas that were geographically important for Singapore planning but are irrelevant for identifying gyms/fitness center locations such as nature reserves or reservoirs. Hence the list of Singapore planning areas was obtained by web scraping via python request and beautiful soup packages. The geographical longitude and latitude coordinates of each area were obtained Geocoder package. The information was converted into a pandas data frame after which planning areas that were irrelevant for identifying gym / fitness center locations were removed from the data. The resultant areas of interest were visualized in the Singapore map using Folium packages.

Venue data of each of these areas were then extracted from the Foursquares API. A total of 100 venues over a distance range of 5000m for each area were extracted. The extracted data were processed by grouping together the top 100 venues in each area by each category. Examination of the unique entries in the venue categories showed a variability in venue curation such that some entries were tagged as gym and some tagged as gym/fitness center. Hence the data were cleaned up such that "Gym" entries are replaced with "Gym/Fitness Center" in order to obtain a more complete data set for subsequent

analysis. After data cleaning, the mean frequency of occurrence for each category for each area were calculated. K-Means clustering were then performed to identify which the density of gyms/fitness center in each area. The starting k number of centroids were selected to be 2.

## Results

K-means clustering of the dataset identified three clusters of neighborhood based on mean frequency of occurrence. Singapore areas belonging to Cluster 0 had the most number of gyms/fitness center, Cluster 1 had moderate number of gyms/fitness center and Cluster 2 has no gym/fitness center. These results are represented in the Singapore map as follows:



Cluster 0 = red

Cluster 1 = Purple

Cluster 2 = Light green

## Discussion

Using K-means clustering, a machine learning approach, the density of gyms in different areas of Singapore can be identified. As Cluster 0 represent the areas with the greatest number of gyms, while Cluster 2 represents areas without gyms, it is recommended, based on the number of gyms in the location, to set up gyms in areas belonging to cluster 2. This recommendation, however, falls short of the consideration of the density of people in each area. It is logical to assume that areas with more people has greater business opportunities for setting up of gyms. Nonetheless, this geographical analysis of gym numbers offers valuable information for investors or gym owners in deciding where they would prefer to set up the next gym.

### Limitation

A limitation of this study is that the venue information used in this exercise is clearly incomplete. Two possible reasons contribute to this limitation. The first being that we could only extract the top 100 venue data from the Foursquare API. This is a very small subset considering the densely populated areas in Singapore. I have attempted to get a more complete data by expanding the radius from each area during data extraction from Foursquare. However, only the top 100 data could be retrieved. In light of this limitation, an improvement could be made by further adding location markers in each area such that 100 venue data from each marker could provide a good coverage of venues. The second limitation

is that the venue data from Foursquare were obtained by consumers. Therefore, it is not possible to assess the completeness of data.

## Conclusion

In conclusion, I would recommend areas belonging to Cluster 2 to be the ideal location for setting up the next gym/fitness center. It is also important to recognize that more information is required to come to an accurate recommendation. A potential future study could hence be, to identify if other categories of shops affects the patronization of people to gyms.

## References

Singapore planning List: [https://en.wikipedia.org/wiki/Planning\\_Areas\\_of\\_Singapore](https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore)

Foursquare API: <https://developer.foursquares.com>

## Appendix

### Cluster 0

	SG_Area	Gym / Fitness Center	Cluster Labels	Latitude	Longitude
26	Seletar	0.030000	0	1.410000	103.874170
19	Marina South	0.030000	0	1.278570	103.857620
22	Pasir Ris	0.030000	0	1.371940	103.949940
24	Punggol	0.020000	0	1.402460	103.906960
10	Downtown Core	0.030000	0	1.377160	103.955530
9	Choa Chu Kang	0.020000	0	1.386160	103.746180
16	Lim Chu Kang	0.030303	0	1.419670	103.702320
31	Straits View	0.030000	0	1.278660	103.853590
33	Tampines	0.030000	0	1.366190	103.829480
1	Bedok	0.020000	0	1.324250	103.952970
28	Sengkang	0.020000	0	1.362460	103.894590

### Cluster 1

	SG_Area	Gym / Fitness Center	Cluster Labels	Latitude	Longitude
27	Sembawang	0.010000	1	1.44794	103.818910
17	Mandai	0.010000	1	1.41259	103.789687
30	Simpang	0.010000	1	1.43722	103.835280
32	Sungei Kadut	0.012987	1	1.41667	103.733330
23	Paya Lebar	0.010000	1	1.32503	103.896480
0	Ang Mo Kio	0.010000	1	1.37161	103.845480
37	Yishun	0.010000	1	1.43621	103.835620
7	Changi	0.010000	1	1.36514	103.990960
14	Jurong West	0.010000	1	1.33949	103.707390
13	Jurong East	0.010000	1	1.33437	103.743670
12	Hougang	0.010000	1	1.37124	103.891620
2	Boon Lay Planning Area	0.010000	1	1.33333	103.700000
3	Bukit Batok	0.010000	1	1.34952	103.752770
8	Changi Bay	0.010000	1	1.36996	103.993110
4	Bukit Merah	0.010000	1	1.28417	103.823060
6	Bukit Timah	0.010000	1	1.34041	103.772210

### Cluster 2

	SG_Area	Gym / Fitness Center	Cluster Labels	Latitude	Longitude
35	Toa Payoh	0.0	2	1.334480	103.851080
34	Tanglin	0.0	2	1.316670	103.816670
5	Bukit Panjang	0.0	2	1.378770	103.769770
36	Tuas	0.0	2	1.311820	103.630520
15	Kallang	0.0	2	1.333330	103.866670
25	Rochor	0.0	2	1.304130	103.850290
11	Geylang	0.0	2	1.311470	103.882180
21	Museum Planning Area	0.0	2	1.353444	103.816509
20	Marine Parade	0.0	2	1.303060	103.907780
29	Serangoon	0.0	2	1.355540	103.876600
18	Marina East	0.0	2	1.295790	103.895440