# Identifying the potential locations of setting up a new business in Singapore

Locations are paramount to the success of business especially in business to customers (B2C) models. Singapore is small country and many areas are highly populated. Areas are well planned out and saturated with diverse shops covering different kinds of needs of the population. To identify the suitable location for a new business, it is important to identity competing businesses in the vicinity.

One of the key success factors for B2C businesses is the accessibility of the product (which can be in the form of a physical product or service) to the customers. While location is one of the factors, other factors also include demographics of residents as well as potential building developments around the area.

As such, in this project, I will be using the context of setting up a gym to develop a way of analyzing the densities of gyms in different areas of Singapore to simulate the process of business planning mirroring an entrepreneur who is interested in setting up a gym business

## Data source

To conduct this study, I will be extracting location info from website: http://en.wikipedia.org/wiki/Planning_areas_of_Singpore. Using python request and beautifulsoup packages, I will obtain all areas of Singapore. Finally, to visualize these data, I will use Foursquare API and K means clustering to obtain the venue data which will be overlayed on the map using Folium.

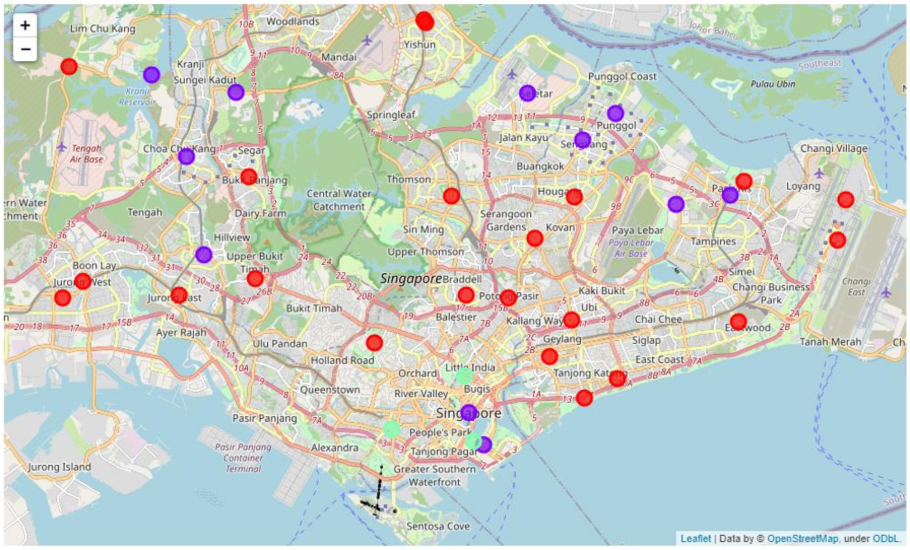
## Methodology

To identify the locations and number of gyms around different areas of Singapore, I first extracted the location data of Singapore. This data set will give the different areas of Singapore which allows visualizing of the number of gyms in the different areas. These data are found in the website: http://en.wikipedia.org/wiki/Planning_areas_of_Singpore. This gives a total of 55 different areas marked by different names and postal codes. Data wrangling was performed to removed irrelevant data which includes areas spanning highly forested areas, military camps and reservoirs as we do not expect these areas to be used for setting of a gym. I obtained the longitude and latitude coordinates of different areas in Singapore using beautifulsoup packages and python request by scraping the website in the above url.

Using Foursquares API, I extracted the venue data for each areas of Singapore. To enable clearer visualization, I extracted the top 100 shops in each area within 5km radius of each of the areas in Singapore and clustered them based on the similarity in category of shops. The clustering gave differing names for gym related categories. To enable accurate counting of the shop number, I standardized all such categories into the category “Gym/Fitness centre” I then obtained the mean frequency of occurrence for “Gym/Fitness centre” for each of the areas of Singapore which will give us the count of the number of gyms in the areas. To observed their density, I performed K means clustering the starting k number for centroid to be 2.

## Results

Out[62]:



- Most numbers of gyms in area = purple
- Moderate numbers of gyms in area = red
- Low numbers of gyms in area = light green

A total of 3 clusters were identified using k means clustering. They are color coded in purple for areas with most numbers of gyms, red for areas with moderate numbers of gyms and light green for areas with least numbers of gyms. Such a visualization can easily tell us where would be potential location to consider. These potential locations are Rochor, Bukit Merah and Straits view.

## Discussion

We are able to easily visualize multiple layers of data and make sense out of it using K means clustering which is a machine learning approach. By performing web scraping and followed by machine learning and representing using data visualization, we are now clear on which of the locations presents better opportunity to set up a gym in Singapore. This approach was showcased to reflect location data and existing densities of different businesses in the specified area (Singapore) and would also be applicable to other forms of businesses as well as location. The developed approach can be of value in order to identify potential location to set up a business in the angle of geographical considerations and neighboring competitors.

However, the eventual decision as to where is the best place to set up any business requires also considerations in terms of population densities, traffic flow of people and future development of the area of neighboring areas. These are other layers of information that is not shown in this report. By considering all layers of information, a sound decision can thus be made.

## Conclusion

This exercise showcased the value of applying data science methodology to aid in decision making of setting up a gym in Singapore. While the data reflected here clearly indicates that light green clusters are the areas with the highest potential in terms of geographical considerations, many other layers of information has to be in placed in order to increase the chances of a success gym set up.

## References

Singapore planning list: [http://en.wikipedia.org/wiki/Planning_areas_of_Singpore.](http://en.wikipedia.org/wiki/Planning_areas_of_Singpore.)

Foursquares API: http://www.developer.foursquares.com

## Appendix:

Cluster 0

Out[59]:

| | SG_Area | Gym / Fitness Center | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | Ang Mo Kio | 0.000000 | 0 | 1.37161 | 103.84546 |
| 35 | Toa Payoh | 0.010000 | 0 | 1.33448 | 103.85108 |
| 34 | Tanglin | 0.010000 | 0 | 1.31667 | 103.81667 |
| 30 | Simpang | 0.010000 | 0 | 1.43722 | 103.83528 |
| 29 | Serangoon | 0.010000 | 0 | 1.35554 | 103.87660 |
| 27 | Sembawang | 0.010000 | 0 | 1.44794 | 103.81891 |
| 23 | Paya Lebar | 0.000000 | 0 | 1.32503 | 103.89049 |
| 20 | Marine Parade | 0.000000 | 0 | 1.30306 | 103.90778 |
| 36 | Tuas | 0.000000 | 0 | 1.31182 | 103.63052 |
| 16 | Lim Chu Kang | 0.000000 | 0 | 1.41967 | 103.70232 |
| 15 | Kallang | 0.010000 | 0 | 1.33333 | 103.86667 |
| 14 | Jurong West | 0.010000 | 0 | 1.33949 | 103.70739 |
| 13 | Jurong East | 0.010000 | 0 | 1.33437 | 103.74367 |
| 18 | Marina East | 0.000000 | 0 | 1.29579 | 103.89544 |
| 11 | Geylang | 0.000000 | 0 | 1.31147 | 103.88218 |
| 1 | Bedok | 0.010000 | 0 | 1.32425 | 103.95297 |
| 2 | Boon Lay Planning Area | 0.010000 | 0 | 1.33333 | 103.70000 |
| 5 | Bukit Panjang | 0.000000 | 0 | 1.37877 | 103.76977 |
| 12 | Hougang | 0.010000 | 0 | 1.37124 | 103.89162 |
| 6 | Bukit Timah | 0.010000 | 0 | 1.34041 | 103.77221 |
| 37 | Yishun | 0.010000 | 0 | 1.43621 | 103.83582 |
| 8 | Changi Bay | 0.010000 | 0 | 1.36998 | 103.99307 |
| 7 | Changi | 0.010000 | 0 | 1.35514 | 103.99006 |
| 10 | Downtown Core | 0.013333 | 0 | 1.37691 | 103.95504 |

Cluster 1

Out[60]:

| | SG_Area | Gym / Fitness Center | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|
| 21 | Museum Planning Area | 0.030000 | 1 | 1.290410 | 103.852110 |
| 22 | Pasir Ris | 0.030000 | 1 | 1.371940 | 103.949940 |
| 19 | Marina South | 0.030000 | 1 | 1.278570 | 103.857620 |
| 24 | Punggol | 0.020000 | 1 | 1.402460 | 103.906860 |
| 26 | Seletar | 0.030000 | 1 | 1.410000 | 103.874170 |
| 17 | Mandai | 0.020000 | 1 | 1.410356 | 103.764728 |
| 28 | Sengkang | 0.020000 | 1 | 1.392440 | 103.894700 |
| 3 | Bukit Batok | 0.020000 | 1 | 1.349520 | 103.752770 |
| 32 | Sungei Kadut | 0.023529 | 1 | 1.416670 | 103.733330 |
| 33 | Tampines | 0.030000 | 1 | 1.368190 | 103.929480 |
| 9 | Choa Chu Kang | 0.020000 | 1 | 1.386160 | 103.746180 |

Cluster 2

Out[61]:

| | SG_Area | Gym / Fitness Center | Cluster Labels | Latitude | Longitude |
|---|---|---|---|---|---|
| 4 | Bukit Merah | 0.06 | 2 | 1.284170 | 103.823060 |
| 31 | Straits View | 0.06 | 2 | 1.279864 | 103.853593 |
| 25 | Rochor | 0.05 | 2 | 1.304130 | 103.850290 |