

APPLIED DATA SCIENCE CAPSTONE

# BATTLE OF THE NEIGHBOURHOOD IN SINGAPORE

GEYLANG VERSUS BUKIT TIMAH

BY NADIAH IBRAHIM

---



## 1. Introduction/Business Problem

According to statistica.com, 58% of Singaporean engage in exercise activity at home during the Covid-19 pandemic in 2020. With the current positive outlook on the pandemic situation due to the discovery of the vaccines, a lot of businesses are taking this opportunity to make a head start by launching their business as soon as they can.

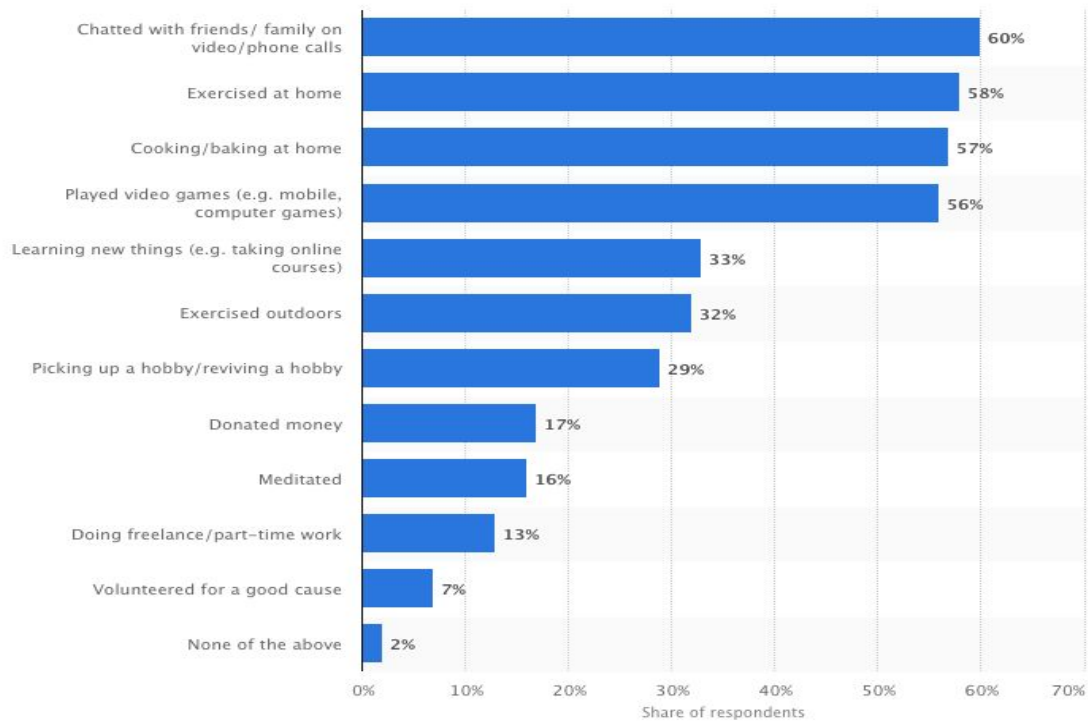
Assuming that the Singaporean community would be keen in continuing the healthy lifestyle, this particular project is looking into the most popular neighbourhood in

---

---

Singapore to set up a new fitness center.

I have decided to narrow down to the two most popular neighbourhood in Singapore that is Bukit Timah and Geylang, which both is centrally located in Singapore.



1

**Activities undertaken by Singaporeans in the previous two weeks during the COVID-19 pandemic as of May 22, 2020**

---

<sup>1</sup> Source:  
<https://www.statista.com/statistics/1119966/singapore-activities-conducted-during-covid-19-pandemic/>

---

## 2. Data Description

Data used in this project was garnered from 3 different sources as follows:

1. Singapore HDB Postal Code Mapper (2018):  
<https://www.kaggle.com/mylee2009/singapore-postal-code-mapper>
2. Singapore Open Data: Location of Gyms@SG: <https://data.gov.sg/dataset/gymssg>
3. Wikipedia: [https://en.wikipedia.org/wiki/Postal\\_codes\\_in\\_Singapore](https://en.wikipedia.org/wiki/Postal_codes_in_Singapore)

### 2.1.1 Singapore HDB Postal Code Mapper

Using data from <https://www.kaggle.com/mylee2009/singapore-postal-code-mapper>, we will get the data on the hotspot location in Singapore which includes the latitude, longitude and the postal code. However, the information on which area these locations are categorized is lacking in these data. Hence, in the next section, I will discussed on the next set of data acquired to supplement this dataset.

|   | postal | latitude | longitude  | searchval                                | blk_no | road_name             | building   | address   | postal.1 |
|---|--------|----------|------------|--|--------|-----------------------|------------|---|----------|
| 0 | 398614 | 1.312763 | 103.883519 | # 1 LOFT                                 | 1      | LORONG 24 GEYLANG     | # 1 LOFT   | 1 LORONG 24 GEYLANG # 1 LOFT SINGAPORE 398614   | 398614   |
| 1 | 398721 | 1.312390 | 103.881504 | # 1 SUITES                               | 1      | LORONG 20 GEYLANG     | # 1 SUITES | 1 LORONG 20 GEYLANG # 1 SUITES SINGAPORE 398721 | 398721   |
| 2 | 629875 | 1.309135 | 103.679463 | 1 BENOI ROAD SINGAPORE 629875            | 1      | BENOI ROAD            | NIL        | 1 BENOI ROAD SINGAPORE 629875                   | 629875   |
| 3 | 439731 | 1.305466 | 103.895674 | 1 BOSCOMBE ROAD SINGAPORE 439731         | 1      | BOSCOMBE ROAD         | NIL        | 1 BOSCOMBE ROAD SINGAPORE 439731                | 439731   |
| 4 | 659592 | 1.344619 | 103.749789 | 1 BUKIT BATOK STREET 22 SINGAPORE 659592 | 1      | BUKIT BATOK STREET 22 | NIL        | 1 BUKIT BATOK STREET 22 SINGAPORE 659592        | 659592   |
| 5 | 618292 | 1.314283 | 103.723913 | 1 BUROH LANE SINGAPORE 618292            | 1      | BUROH LANE            | NIL        | 1 BUROH LANE SINGAPORE 618292                   | 618292   |
| 6 | 768069 | 1.437425 | 103.829625 | 1 CANBERRA                               | 7      | CANBERRA DRIVE        | 1 CANBERRA | 7 CANBERRA DRIVE 1 CANBERRA SINGAPORE 768069    | 768069   |
| 7 | 768070 | 1.437019 | 103.828763 | 1 CANBERRA                               | 9      | CANBERRA DRIVE        | 1 CANBERRA | 9 CANBERRA DRIVE 1 CANBERRA SINGAPORE 768070    | 768070   |
| 8 | 768071 | 1.437221 | 103.828939 | 1 CANBERRA                               | 11     | CANBERRA DRIVE        | 1 CANBERRA | 11 CANBERRA DRIVE 1 CANBERRA SINGAPORE 768071   | 768071   |
| 9 | 768073 | 1.437617 | 103.829271 | 1 CANBERRA                               | 15     | CANBERRA DRIVE        | 1 CANBERRA | 15 CANBERRA DRIVE 1 CANBERRA SINGAPORE 768073   | 768073   |

---

## 2.1.2 Wikipedia

Referring to Wikipedia : [https://en.wikipedia.org/wiki/Postal\\_codes\\_in\\_Singapore](https://en.wikipedia.org/wiki/Postal_codes_in_Singapore), we can get the information on the postal codes, the respective area and the general location that corresponds to it as follows:

| Postal district |    | area                   | General location                                  |
|-----------------|----|------------------------|---|
| 0               | 1  | 01, 02, 03, 04, 05, 06 | Raffles Place, Cecil, Marina, People's Park       |
| 1               | 2  | 07, 08                 | Anson, Tanjong Pagar                              |
| 2               | 3  | 14, 15, 16             | Bukit Merah, Queenstown, Tiong Bahru              |
| 3               | 4  | 09, 10                 | Telok Blangah, Harbourfront                       |
| 4               | 5  | 11, 12, 13             | Pasir Panjang, Hong Leong Garden, Clementi New... |
| 5               | 6  | 17                     | High Street, Beach Road (part)                    |
| 6               | 7  | 18, 19                 | Middle Road, Golden Mile                          |
| 7               | 8  | 20, 21                 | Little India, Farrer Park, Jalan Besar, Lavender  |
| 8               | 9  | 22, 23                 | Orchard, Cairnhill, River Valley                  |
| 9               | 10 | 24, 25, 26, 27         | Ardmore, Bukit Timah, Holland Road, Tanglin       |
| 10              | 11 | 28, 29, 30             | Watten Estate, Novena, Thomson                    |
| 11              | 12 | 31, 32, 33             | Balestier, Toa Payoh, Serangoon                   |
| 12              | 13 | 34, 35, 36, 37         | Macpherson, Braddell                              |

Merging the two dataframes, we will get the following dataframe that mapped the hotspot location in Singapore to a district/general location:

|   | postal | latitude | longitude  | searchval                               | blk_no | road_name           | building          | address   | area | index | Postal district | General location         |
|---|--------|----------|------------|---|--------|---------------------|-------------------|---|------|-------|-----------------|--------------------------|
| 0 | 398614 | 1.312763 | 103.883519 | # 1 LOFT                                | 1      | LORONG 24 GEYLANG   | # 1 LOFT          | 1 LORONG 24 GEYLANG # 1 LOFT SINGAPORE 398614     | 39   | 13.0  | 14              | Geylang, Eunos, Aljunied |
| 1 | 398721 | 1.312390 | 103.881504 | # 1 SUITES                              | 1      | LORONG 20 GEYLANG   | # 1 SUITES        | 1 LORONG 20 GEYLANG # 1 SUITES SINGAPORE 398721   | 39   | 13.0  | 14              | Geylang, Eunos, Aljunied |
| 2 | 399870 | 1.309889 | 103.881016 | 1 GUILLEMARD LANE SINGAPORE 399870      | 1      | GUILLEMARD LANE     | NIL               | 1 GUILLEMARD LANE SINGAPORE 399870                | 39   | 13.0  | 14              | Geylang, Eunos, Aljunied |
| 3 | 398146 | 1.313707 | 103.888506 | 11 LORONG 36 GEYLANG SINGAPORE 398146   | 11     | LORONG 36 GEYLANG   | NIL               | 11 LORONG 36 GEYLANG SINGAPORE 398146             | 39   | 13.0  | 14              | Geylang, Eunos, Aljunied |
| 4 | 399268 | 1.310839 | 103.876110 | 11 LORONG 4 GEYLANG SINGAPORE 399268    | 11     | LORONG 4 GEYLANG    | NIL               | 11 LORONG 4 GEYLANG SINGAPORE 399268              | 39   | 13.0  | 14              | Geylang, Eunos, Aljunied |
| 5 | 399911 | 1.309542 | 103.881312 | 16 GUILLEMARD CRESCENT SINGAPORE 399911 | 16     | GUILLEMARD CRESCENT | NIL               | 16 GUILLEMARD CRESCENT SINGAPORE 399911           | 39   | 13.0  | 14              | Geylang, Eunos, Aljunied |
| 6 | 398997 | 1.311348 | 103.878503 | 17 LORONG 12 GEYLANG SINGAPORE 398997   | 17     | LORONG 12 GEYLANG   | NIL               | 17 LORONG 12 GEYLANG SINGAPORE 398997             | 39   | 13.0  | 14              | Geylang, Eunos, Aljunied |
| 7 | 398008 | 1.308466 | 103.878985 | 223 @ MOUNTBATTEN                       | 223    | MOUNTBATTEN ROAD    | 223 @ MOUNTBATTEN | 223 MOUNTBATTEN ROAD 223 @ MOUNTBATTEN SINGAPO... | 39   | 13.0  | 14              | Geylang, Eunos, Aljunied |

With this merged, we now have the information on which area these locations are categorized.

### 2.1.3 Singapore Open Data: Location of Gyms@SG

I have also acquired the data of the existing locations of Fitness Center in Singapore from <https://data.gov.sg/dataset/gymssg>. The data is in the form of GEOJSON. So the first step in handling this data set is to read and parse the data into a data frame. The information needed from this particular dataset is only the location (ie: latitude and longitude) of the fitness center.

---

|     | Latitude         | Longitude        |
|-----|------------------|------------------|
| 0   | 1.27811388348071 | 103.844202079592 |
| 1   | 1.3342195983767  | 103.795898242377 |
| 2   | 1.35528262760188 | 103.850922071159 |
| 3   | 1.30992191079359 | 103.862090963732 |
| 4   | 1.36000966368369 | 103.85957105374  |
| ... | ...              | ...              |
| 154 | 1.2884814659674  | 103.849144620549 |
| 155 | 1.30550611670641 | 103.829984827987 |
| 156 | 1.33983014805774 | 103.706460886826 |
| 157 | 1.29151478613339 | 103.844586298874 |
| 158 | 1.35431556438278 | 103.94501945797  |

159 rows x 2 columns

This dataset showed that there are currently 159 existing Fitness Center available around the city of Singapore.

Using Foursquare and the dataframe above, we can compare any two neighbourhoods in Singapore to determine which neighbourhood will be the most suitable to open a new fitness center.

## 2.2 Data Cleaning

The first step in handling the data is merging the postal code table with the hotspot locations in Singapore.



---

The data acquired from these 2 sources are straightforward and there is very minimal data cleaning needed to be done.

The only major issue with the data are some missing values in some of the columns. To start off the data cleaning, I would first handle the missing data in these columns. The missing values in the column building name is set to Nan as the building name is not a weighted data in this case. Any rows with a missing value in postal code and general location will be dropped as this information is crucial in our data to determine the appropriate location to open the new fitness center.

I also renamed the column general location to Neighbourhood to standardized the variables used throughout this project

Next, I would standardise the data type of postal code and the area code to integer. The data type for latitude and longitude will be left as it is in float format.

The table below shows the summary of the data cleaning done for this dataframe:

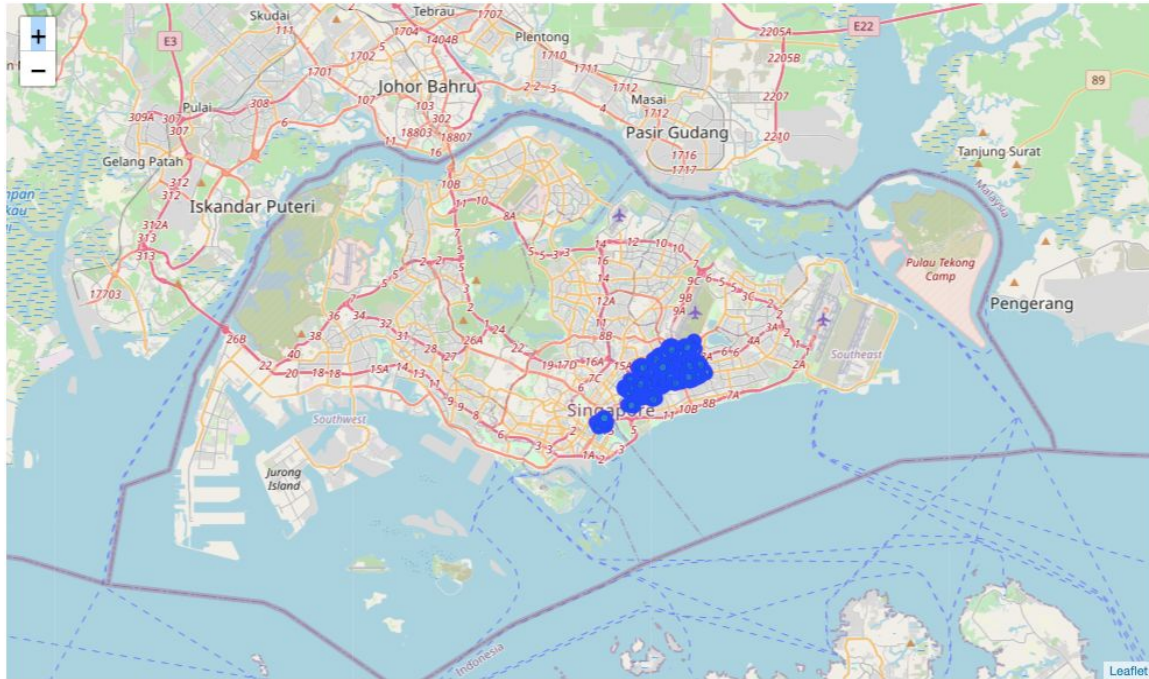
|               |   |
|---------------|---|
| Dropped Rows  | Rows with missing value in postal code and general locations column |
| Column Rename | General location column -> Neighbourhood                            |
| Data Type     | Postal code column (Float -> Int)                                   |

As for the data acquired for the existing Fitness Center in Singapore, the only action applied is reading the JSON file and parsing the latitude and longitude information. This information is then stored in a separate data frame.

## 4. Data Methodology and Results

Using Four Square, we can get the top 10 most common venues by neighbourhood. For example, comparing the two neighbourhoods in Singapore, Geylang versus Bukit Timah.

Foursquare also provides information of which of these locations are most popular/most liked by the community.

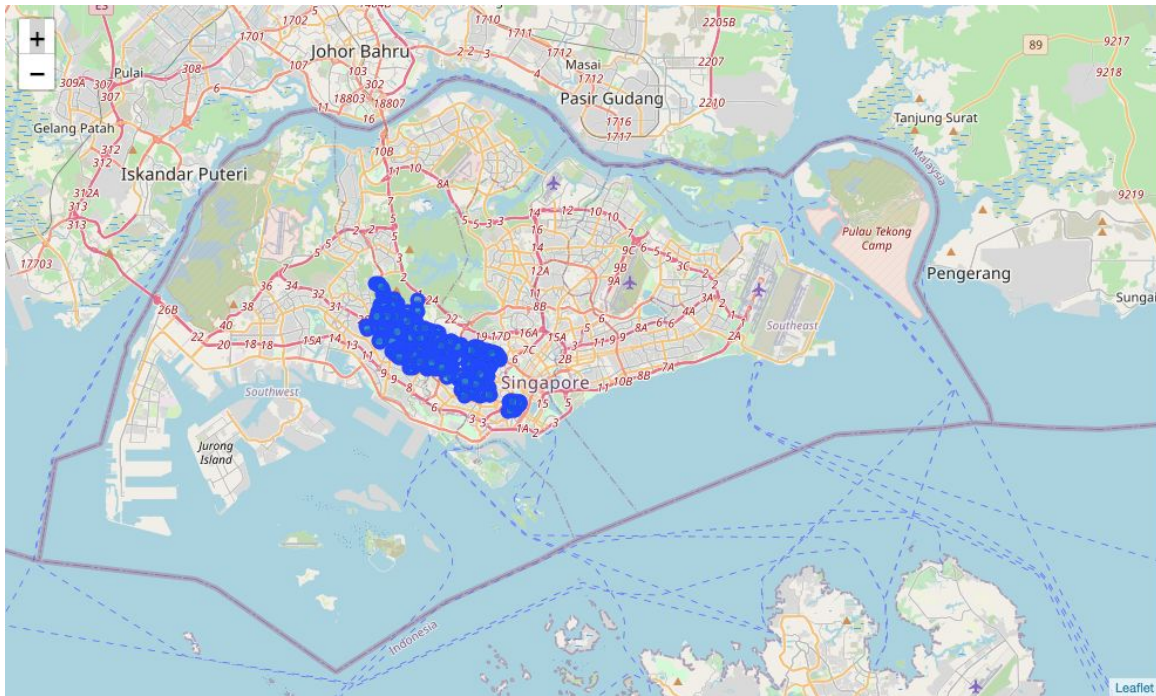


## 4.1 Folium Map

To compare these two neighbourhoods, we must have a rough idea of where they are actually located. Using Folium Map as follows, we can plot the location of the hotspots in Geylang vs Bukit Timah. To ensure that we can explore the data clearly, I created 2 separate Folium Map as follows:

Map Of Hotspot Locations in Geylang,Singapore





Map of Hotspot Locations in Bukit Timah, Singapore

We can see that these two areas are centrally located in Singapore and close to the CBD area as well, making it a very popular location visited by the locals.

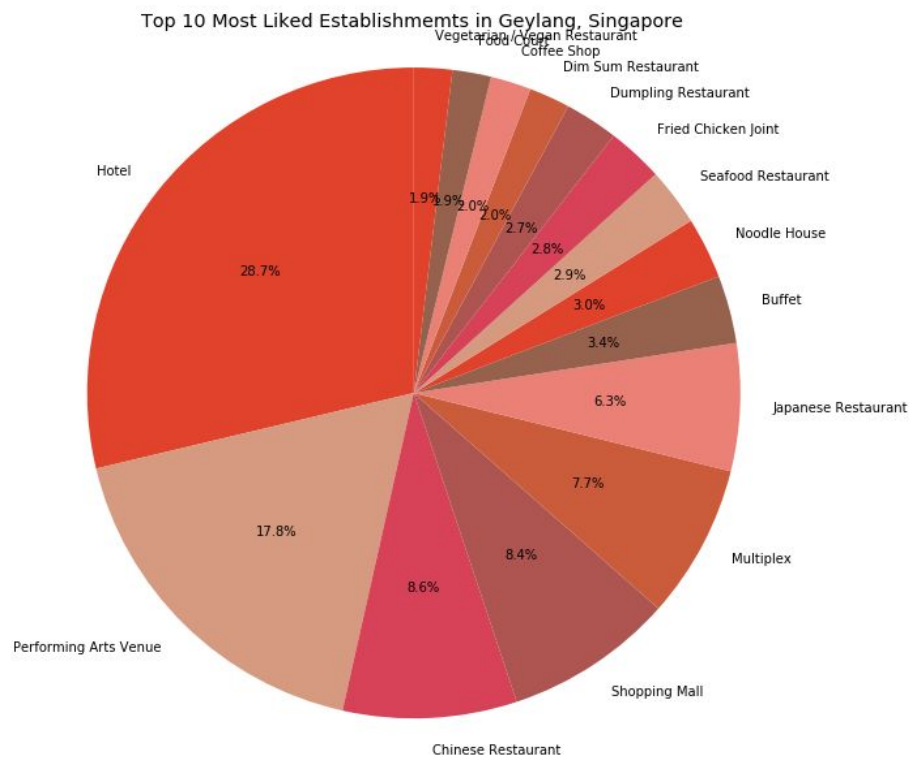
## 4.2 Foursquare

Using Foursquare to collect all information on the popular establishments surrounding these locations, I managed to get the information on how close the existing Fitness Center are located from these hotspot locations.

In the district of Geylang, there are 34 popular venues and 210 nearby venues return by Foursquare.

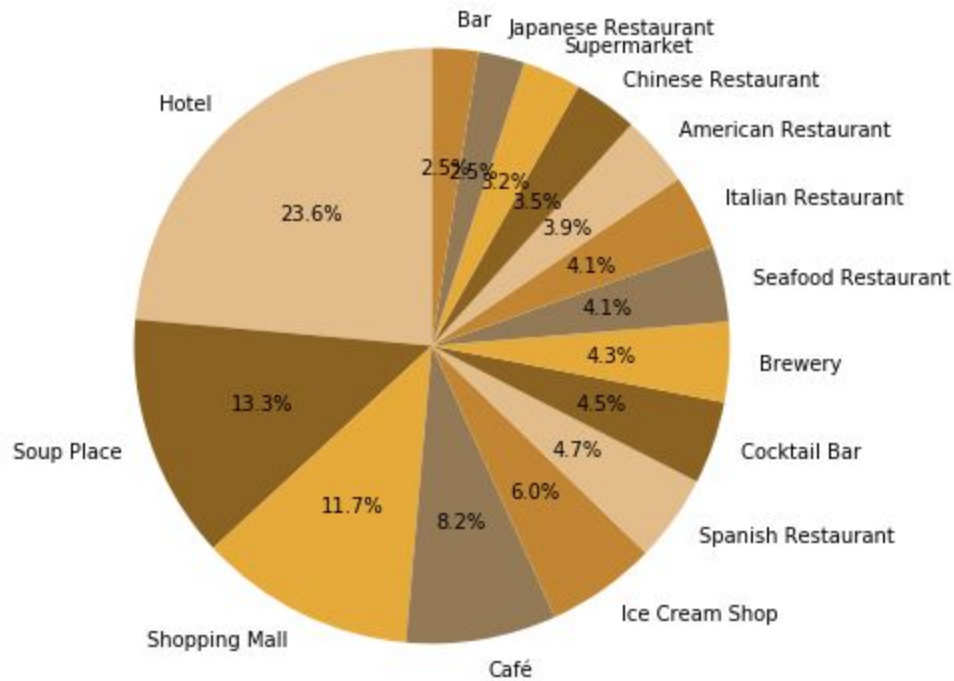
In the case of Bukit Timah, 12 popular venues were return with 300 nearby venues.

Using pie charts, we can model the information on most popular hotspots in the area.



---

### Top 10 Most Liked Establishments in Bukit Timah, Singapore

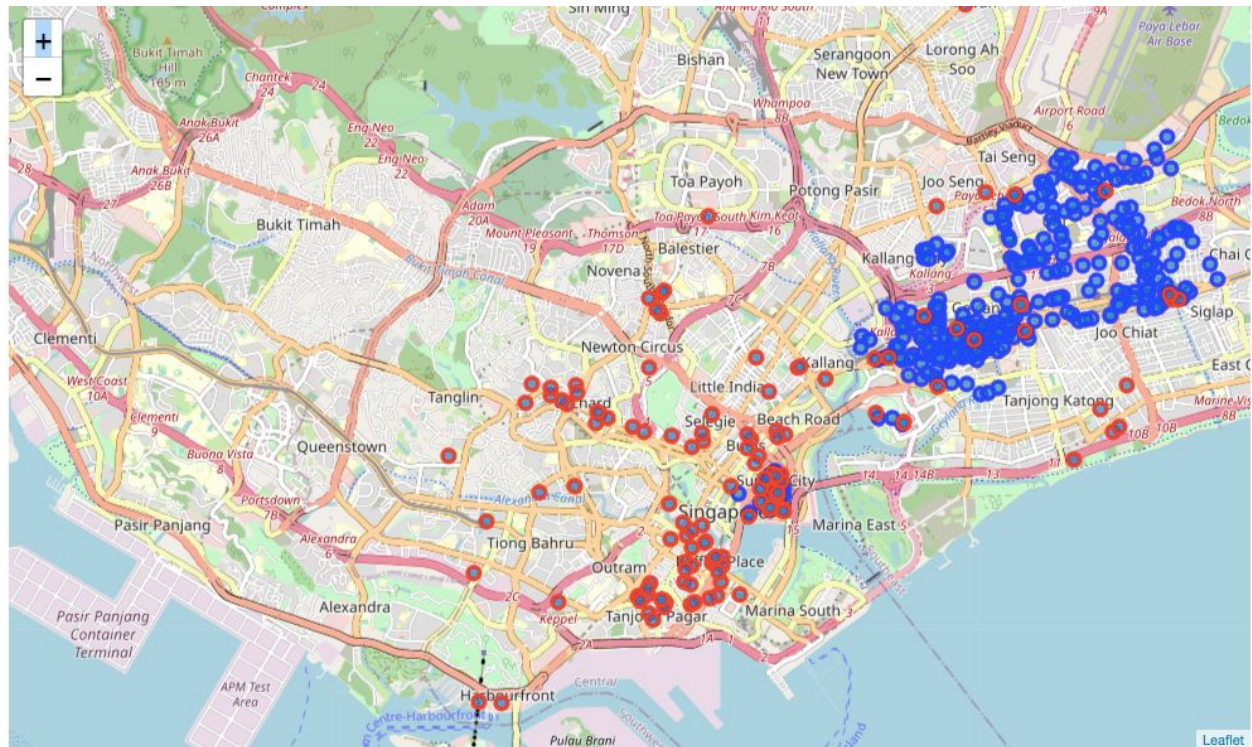


From these two pie charts, we can tell that in the district of Geylang, the most popular location is Hotel at 28.7%, Performing Arts at 17.8%, followed by various of restaurants.

For the Bukit Timah district, Hotel is the most popular location as well with 23.6%, followed by Soup place at 13.3%, Shopping Mall at 11.7% and restaurants and cafe for the rest. We can see from the 2 pie charts that the Bukit Timah district has a more even distribution of popularity compared to the Geylang district.

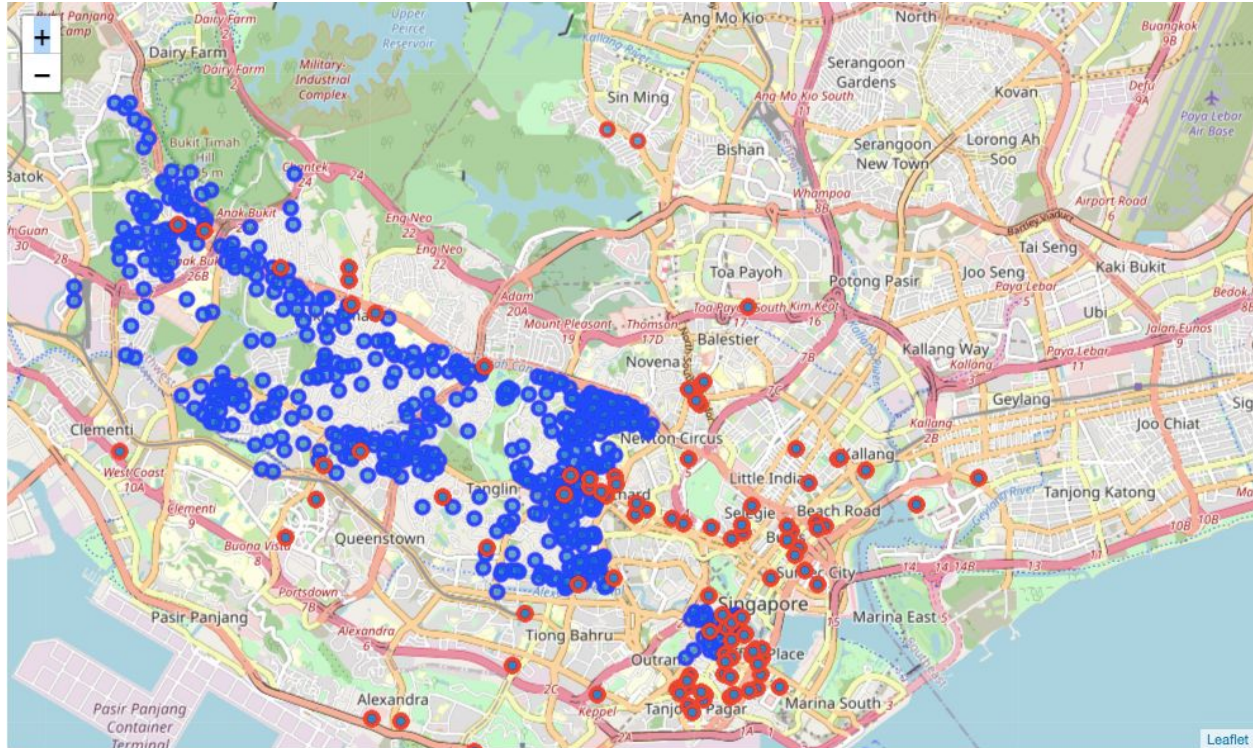
Using simple mathematical calculations, I figured out if any of these locations is nearby the existing Fitness Center defining the nearby location as any location within the radius of 5km of the hotspots

Using Folium Map again, I mapped out the Fitness Center that located nearby the most liked locations in these 2 districts as follows:



Existing Fitness Center (in red dots) that located nearby favorite hotspot in Geylang





Existing Fitness Center (in red dots) that located nearby favorite hotspot in Bukit Timah

There are 116 Existing Fitness Center located nearby the hotspot in Geylang versus 111 in Bukit Timah.

### 4.3 K-mean clustering

Using **K-mean clustering** we can categorize the hotspot locations to further analyze the potential of opening a new fitness center around these 2 districts.

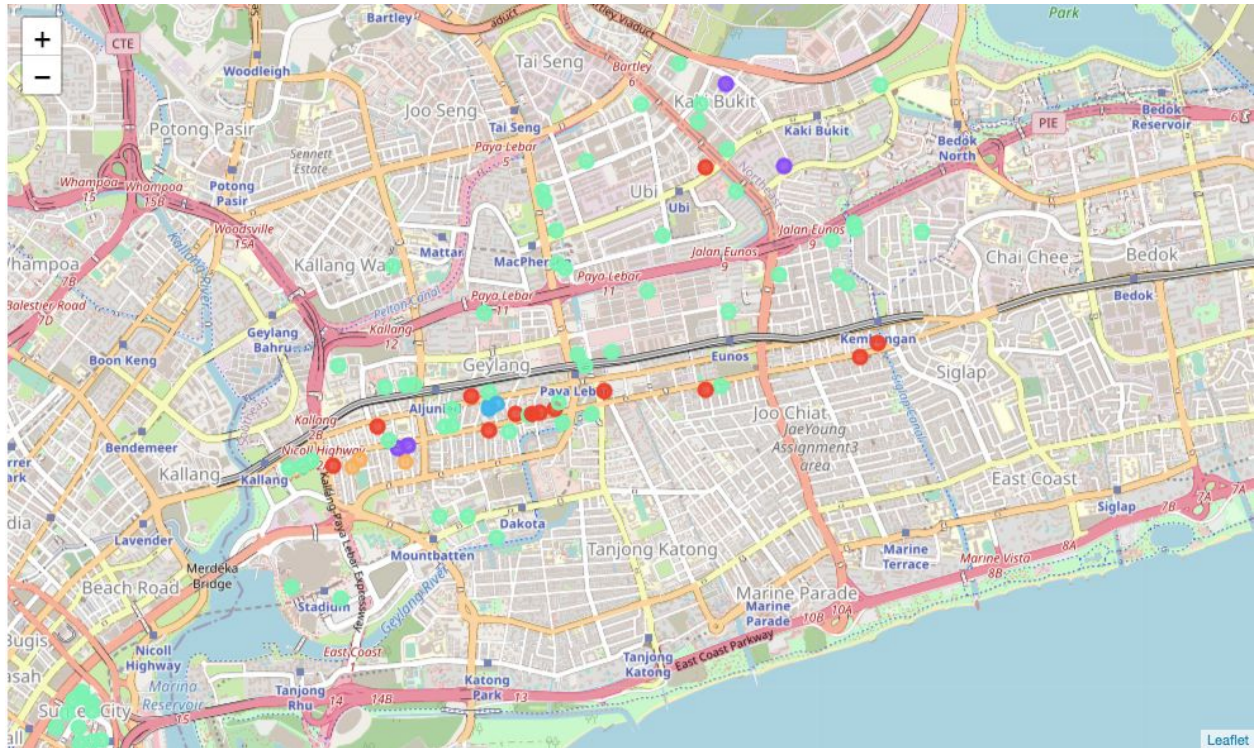
I have decided to set the value of  $k$  to  $k = 5$ .

For the Geylang District, the establishments are divided into the following clusters:

1. Cluster 1: Ethnic Restaurants - Chinese, Philipines Cuisine
2. Cluster 2: Daily needs supply - Convenience store, Deli, Flea markets
3. Cluster 3: Arts & Dining - Arts museum, Vegan restaurant, Dessert shop

4. Cluster 4: Health, Beauty & Dining - Basketball court, playground, sports bar, frozen yogurt
5. Cluster 5: Hotels & dining - Hotel, Vegan restaurant, fast food restaurant

From this data, we can tell that cluster 4 has a more diversified business.



Geylang District by cluster

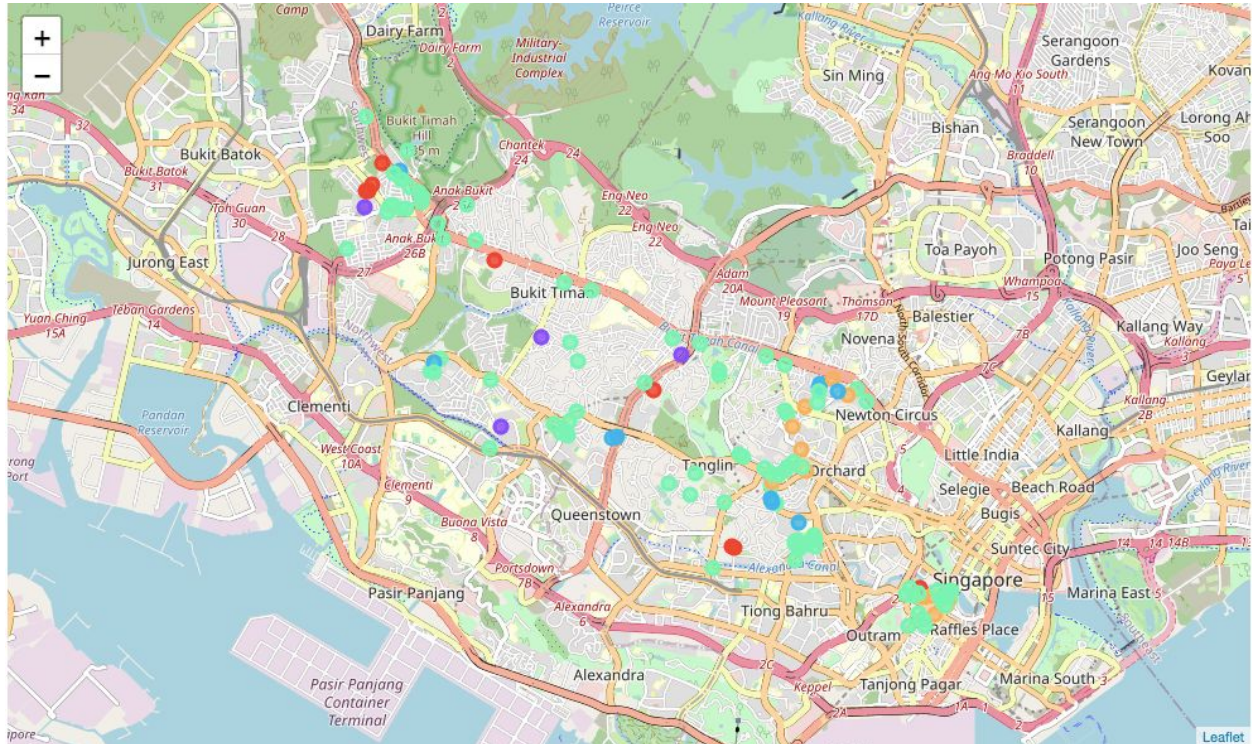
For the Bukit Timah District, the establishments are divided into the following clusters:

1. Cluster 1: Health & Dining - Gym, dance studio, fast food restaurants
2. Cluster 2: Health, Dining & Transport - Bus station, wine shop, dance studio
3. Cluster 3: Health, Beauty & Dining - Wine shop, Coffee shop, Cosmetic shop, Dance studio
4. Cluster 4: Dining, Entertainment & Health - Cafe, American restaurant, bar, bakery, dance studio, escape room, steakhouse, burger joint



- 
- Cluster 5: Hotel, Dining, Entertainment & Health - Hotel, gym, cupcake shop, dance studio, nightclub, doctor's office

For this district we can see that all the clusters are quite diversified with multiple business categories.



Bukit Timah district by cluster

---

## 5. Discussion

From the data, we can see that Bukit Timah district is more diversified. There is a variety of business ranging from Health, Entertainment and Dining establishments around this area.

If we look into the data of the existing fitness center, we can tell that most of the fitness center are actually located in between these 2 districts.

The number of the existing business center are located nearer to the most popular establishments in the Geylang district is higher than the most popular establishments in Bukit Timah district by 5. It is a very small difference thus we can say that it is almost equally the same.

## 6.Result

We can conclude that both areas have a great potential for a new fitness center. However, due to the more diversified business establishments in the Bukit Timah district, I would recommend to focus on this district if anyone were to open a new fitness center. A diversified business establishment reflects more human traffic and a better accessibilities to the area.