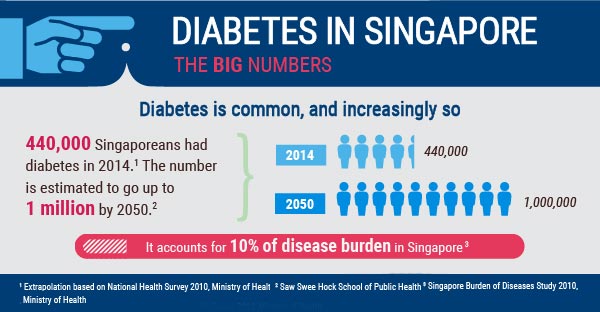
Healthy Café Location Selection in Singapore

IMB Data Science Capstone Project – The Battles of Neighbourhoods

# **Introduction**

1. **Background**

About 400,000 Singaporeans are diabetes and one in three has lifetime risk of developing the disease. And if nothing is done, the number of diabetes under age 70 is expected to rise to 670,000 by 2030 and one million by 2050.



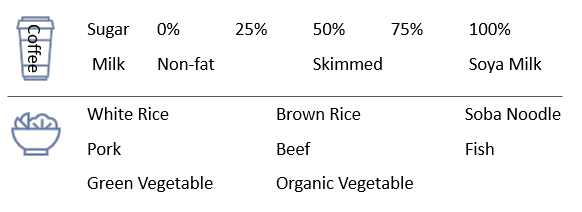
Singapore government adopts a multi-pronged strategy to encourage Singaporeans to eat healthily and reduce their sugar intake from foods and drinks.

Coffee is one of the common sugar intakes for Singaporean. We are going to open a café to server coffee with less sugar and food with balanced nutrition.

1. **Healthy Café Concept**

**Target customers**: Office clericals who care about healthy lifestyle but don’t have time to prepare food for lunch or rush for dinner due to overtime.

**Food Service**: Coffee with sugar level choices and different types of milk (e.g. non-fat, skimmed milk, soya milk, etc). Food with nutrition label, and different size options to match and mix.



1. **Objective**

To find out suitable locations in Singapore

* 1. Near business area
  2. Ares with high density of gyms (To filter those areas passed by more people with healthy concept)
  3. Distribution of restaurants (types and density) – to understand competitor distribution

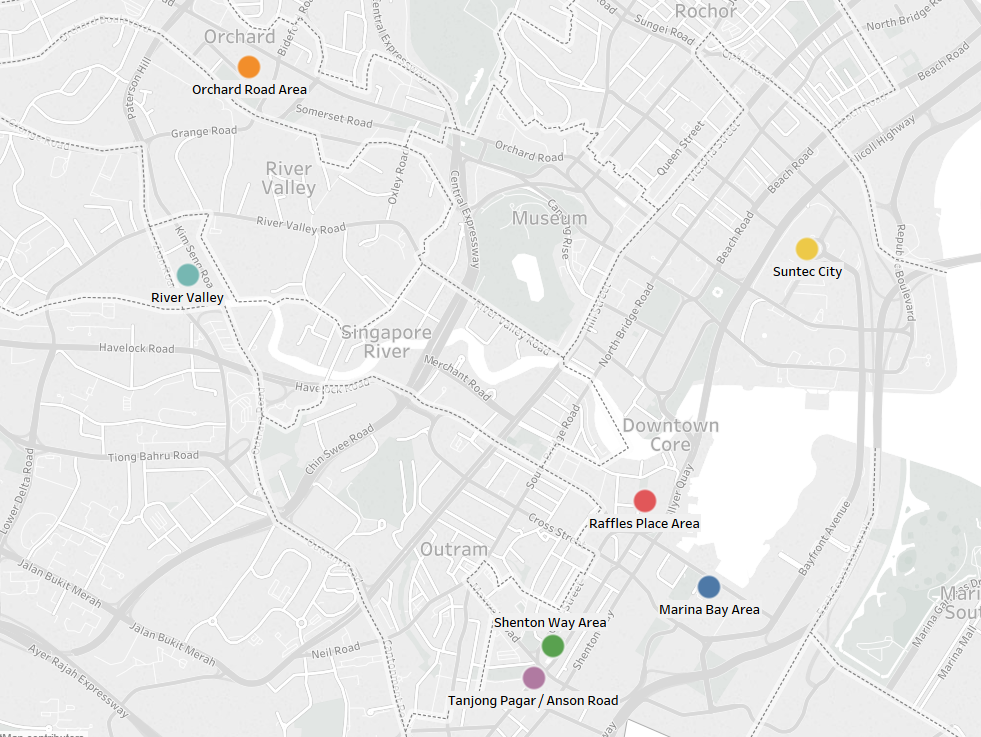
# **Data Acquisition**

1. Target business area in Singapore: <https://www.corporateservicessingapore.com/7-popular-business-locations-singapore/>

So, we can get below starting points

* 1. Raffles Place Area
  2. Marina Bay Area
  3. Tanjong Pagar / Anson Road
  4. Orchard Road Area
  5. Shenton Way Area
  6. River Valley
  7. Suntec City

1. Get geocodes of above areas via Google Map



1. Collect location data from Foursquare website: <https://developer.foursquare.com/>
   1. Location of gyms
   2. Location and category of restaurants
2. Based on the categories of venues, we can find out some abnormal venues under wrong category. So, we need to clean data before starting to do data analysis.

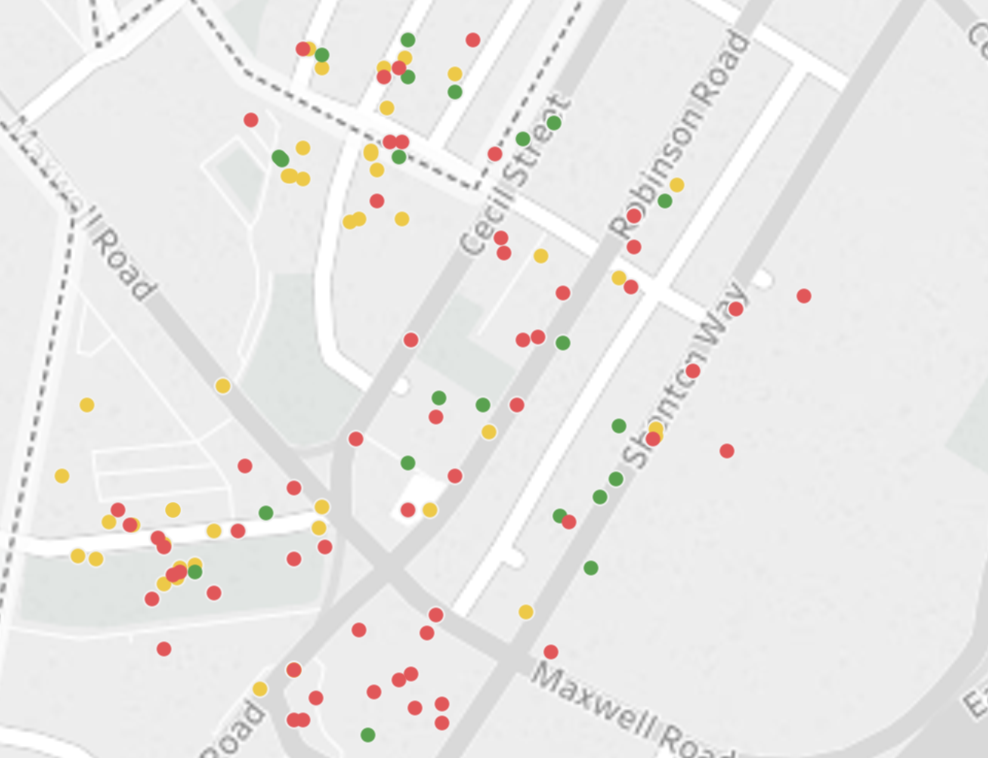
# **Data Analysis**

1. Get the density of gyms, café and other types of restaurants

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Area | Gym | Café | Other Restaurant | Competitive Index |
| Shenton Way Area | 22 | 57 | 43 | 357 |
| Tanjong Pagar / Anson Road | 16 | 45 | 29 | 372 |
| Raffles Place Area | 17 | 55 | 19 | 379 |
| River Valley | 4 | 9 | 32 | 625 |
| Orchard Road Area | 10 | 46 | 68 | 800 |
| Marina Bay Area | 4 | 20 | 27 | 838 |
| Suntec City | 10 | 56 | 59 | 855 |

To have a quantitative concept, we computed **Competitive Index** = (Count of Café + Count of other restaurant) / (Count of gyms) \* 100. It shows the level of competitors in each location. The lowest the better. So, we will choose **Shenton Way**.

1. Explore Shenton way and select actual location

Let’s based on location, count of nearby café and count of nearby other restaurants to classify gyms in Shenton Way area. It’s not very easy to identify which road is more suitable for out café. So, let’s classify by KNN method. Below is the result.

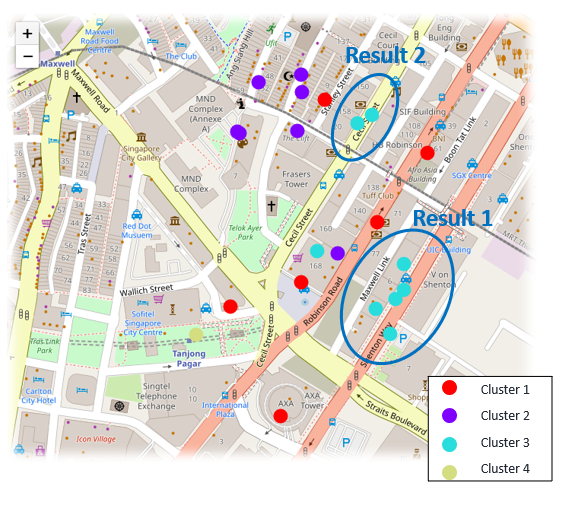
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Count of locations | Avg. Count of Gyms Nearby  (Main Target Customers) | Avg. Count of Café Nearby  (Direct Competitors) | Avg. Count of Other Restaurants Nearby  (Relatively Indirect Competitors) |
| 1 | 6 | 1.2 | 3.3 | 1.3 |
| 2 | 7 | 2.4 | 2.3 | 4.7 |
| 3 | 8 | 2.4 | 1.0 | 0.8 |
| 4 | 1 | 1.0 | 7.0 | 9.0 |

Based on the table, we can convert it into below format,

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Avg. Count of Gyms Nearby  (Main Target Customers) | Avg. Count of Café Nearby  (Direct Competitors) | Avg. Count of Other Restaurants Nearby  (Relatively Indirect Competitors) |
| 1 | Medium | Medium | Low |
| 2 | High | Medium | High |
| 3 | **High** | **Low** | **Low** |
| 4 | Low | High | Very High |

So, now it’s obviously cluster 3 is the ideal location.

Let’s find it in map.



Let’s start to search any available rental in these two places.

# **Conclusion:**

We can find out the two places highlighted in map are good choices. We can do further on-site investigation and make the final decision.