Ahmad Zaini Chia
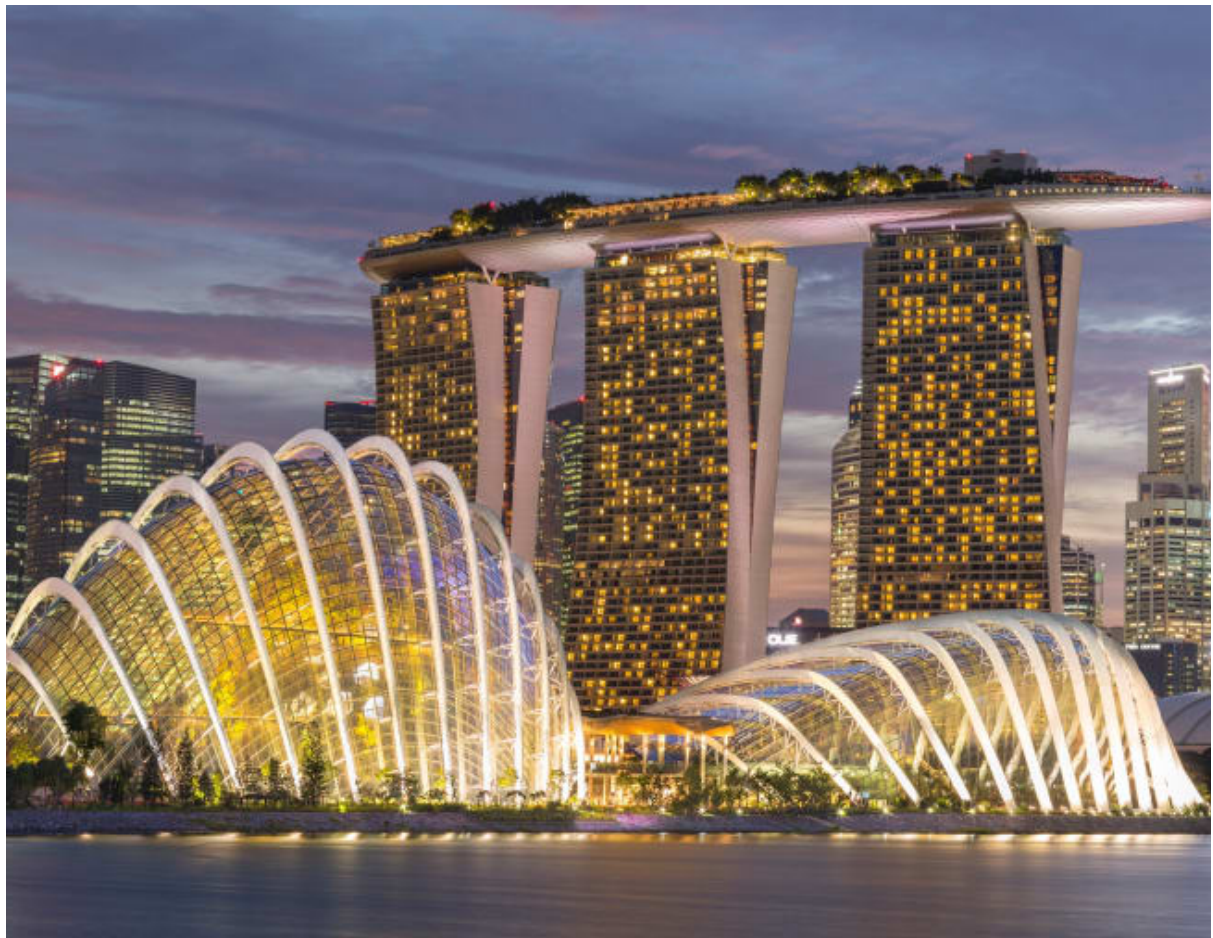IBM Data Science Certificate
Capstone Project
1 December 2019

# Analysing Neighbourhoods in Singapore



## Introduction

Singapore is a small nation-state, and we are one of the most densest countries in the world. In fact, we are the **second-most dense** country in the world.

We are also considered a very developed country, with high standards of education, healthcare, and personal safety. This makes Singapore **the most expensive city** in the world to live in, especially for expats.

*Most Singaporeans eat at Hawker Centres like this one shown in the photo.*



*Majority of Singaporeans also stay in public housing apartments as seen here.*
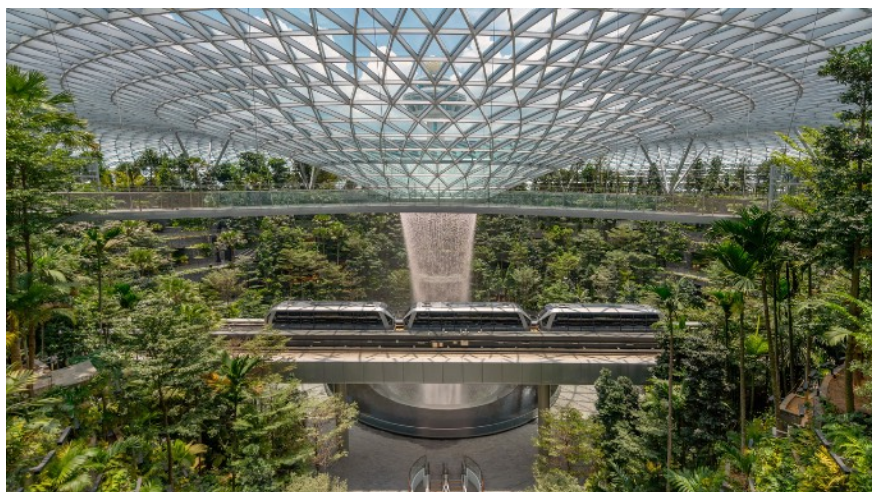


*Singapore prides itself as having the best Airport in the world.*

*Shown here is one of the attractions found at the Airport.*

# Business Problem

Which leads to the business problem: how does one find a place to live in? Of course the most obvious answer is in the central area, but the prices there are prohibitively expensive (unless you are the CEO of Dyson).

And so this is the main question I will attempt to answer for my project: how does one find a place to stay in Singapore? What are the main amenities and facilities in each area, that would suit a prospective buyer?

Given the high cost of buying/renting a house in Singapore, this analysis would be able to help not only expats, but also local Singaporeans like myself who are looking for a place to stay. I hope that I am able to shed some insight on the different neighbourhoods in Singapore and what they might offer to a prospective buyer.

# Data

My main data sources will be Foursquare and Wikipedia.

I will be using Wikipedia mainly to extract a table of the neighbourhoods in Singapore. Singapore is a very tightly controlled nation-state, and the government works hard to ensure that every single plot of land is used as effectively as possible. This leads to the government-created "Planning Areas" for Singapore.

There are 55 such areas, and the purpose/design of each area has been carefully planned and thought out. Of course, not all these areas are meant for housing, and so I will need to carefully filter the data to ensure that I only select the areas which are designated for housing.

I will then use Foursquare to extract data on each of these planning areas. Singapore's size and highly developed transport infrastructure will mean that I will need to carefully tweak the "limit" and "radius" parameters when using the Foursquare API. I will go into more detail in the methodology section.

Using the data extracted, I hope to run clustering and other machine learning techniques to uncover some insights on all the different neighborhoods in Singapore; although we are quite a small nation, our neighborhoods are not as homogenous as one may think.

# Methodology

After the Web-scraping and Data-filtering steps are completed, I managed to get a Dataframe with all the location data I needed:

| | Area | Region | Area (km2) | Population | Density (/km2) | Latitude | Longitude |
|---|---|---|---|---|---|---|---|
| 0 | Sungei Kadut | North | 15.99 | 780 | 53.2 | 1.415 | 103.753 |
| 1 | Orchard | Central | 0.96 | 990 | 960.3 | 1.30527 | 103.833 |
| 2 | Southern Islands | Central | 6.07 | 1800 | 244 | 1.27109 | 103.863 |
| 3 | Changi | East | 40.61 | 2080 | 62.3 | 1.35252 | 103.987 |
| 4 | Mandai | North | 11.77 | 2090 | 180.2 | 1.42353 | 103.803 |
| 5 | Downtown Core | Central | 4.34 | 2510 | 580 | 1.28667 | 103.854 |
| 6 | Singapore River | Central | 0.96 | 2980 | 2842.2 | 1.28918 | 103.845 |
| 7 | Newton | Central | 2.07 | 7640 | 3700 | 1.31298 | 103.839 |
| 8 | River Valley | Central | 1.48 | 9930 | 6230.5 | 1.3084 | 103.886 |
| 9 | Rochor | Central | 1.62 | 13450 | 9034.1 | 1.30398 | 103.853 |

Visualising the data helps a lot, and is a good technique to check on the accuracy of the data. Here is a map of all the planning areas we will be analysing:

**Foursquare Data**

The next step is to retrieve data on venues located nearby each planning area. Setting the limit of 45 venues per planning area, 1287 venues were retrieved. The venues spanned a total of 213 categories.

**Clustering**

In order to run the clustering algorithm, there were 2 main factors that we need to consider:

1. **Limit** (of the number of venues per location)
2. **Radius** (of venues from location)

After running the clustering algorithm a few times, I've found that the results made most sense when I set the Limit to 45 venues per location, and the Radius to 750 metres.

As I mentioned earlier, Singapore has a very highly developed public transport infrastructure, so 750 metres is considered to be quite near as it would be very easy to travel that distance even by public transport.

I set the limit to 45 because there were quite a few planning areas located very close to each other, so setting it higher than that would cause a lot of overlap in the venue data.

After running the clustering algorithm, the planning areas were clustered as such:

# Results & Discussion

As we can see, there are 2 main clusters, 1 small cluster, and two outliers. We will visualise the clusters to better understand the clustering results.

**Central Cluster**

As we can see from the map above, there is a cluster of areas all in the central/south location. This is the central area of Singapore, and is the most developed and cosmopolitan area. As expected, we see more affluent venues like Restaurants, Hotels and Café's. If we look at the Word-Cloud generated from the venues, we can also see that there are many Cosmopolitan Dining venues available as well, from Italian, to Japanese, and even Vegan food (something not easily found in other areas of Singapore).



**"Heartland" Cluster**

The next biggest cluster are what we call the "Heartlands" - areas where most of the working class people stay in. Here we see venues that are more accessible, like Food Courts, Supermarkets, and Asian/Seafood Restaurants.

**Red Dots**: *"Central" Cluster*          **Purple Dots**: *"Heartland" Cluster*

It is worth noting that Changi Airport, located at the Far east side of Singapore, is also clustered as "Central". That is probably due to the fact that Singapore prides itself as having a World-Class Airport, and has built many amenities and attractions around the Airport Area.

## Outlier Clusters

There are three areas which are all assigned their own cluster. The first one is <u>Punggol</u>, and the venue data for Punggol is shown below:

| Area | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Punggol | 4 | Fast Food Restaurant | Sandwich Place | Café | Dessert Shop | Gym / Fitness Center | Supermarket | Chinese Restaurant | Japanese Restaurant | High School | Diner |

We can see that the venues here typically fit those of a heartland neighborhood: fastfood restaurant, supermarket, etc. However, Punggol is a newly developing housing area, and as such the venues and amenities are quite sparse within that area. If we look closer at the venue data, Punggol only has 10 venues within a 750 metre radius of the town center (at least from Foursquare data), and it is an outlier in that sense.

The next outlier is Bukit Timah, and the venue data is as shown:

| Area | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| Bukit Timah | 2 | Trail | Mountain | Rest Area | Scenic Lookout | Nature Preserve | Lake | Tourist Information Center | Bike Trail | Park | Yoga Studio |

Bukit Timah is home to Singapore's largest nature reserve, and we can see that from the venue data around the location.

The last outlier is the Southern Islands:

| Area | Cluster Labels | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|------|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| Southern Islands | 3 | Cruise | Boat or Ferry | Pier | Cafeteria | BBQ Joint | Metro Station | Mexican Restaurant | Bus Line | Government Building | History Museum |

The Southern Islands are small islands, with Sentosa being the biggest one. All the Islands are tourist spots, and we can see that in the venue data as well. Sentosa has many high end resorts, spas, as well as a host of other tourist attractions.

# Conclusion & Recommendations



Contrary to what I mentioned at the start of this report, it turns out that Singapore neighborhoods are quite homogenous. And it does seem that if you want to have the benefits of the central area, you would have to pay the high prices of staying there.

That said, there are certain areas that can bring you to the popular Downtown area within a few minutes, thanks to a new subway line called the "Downtown Line":
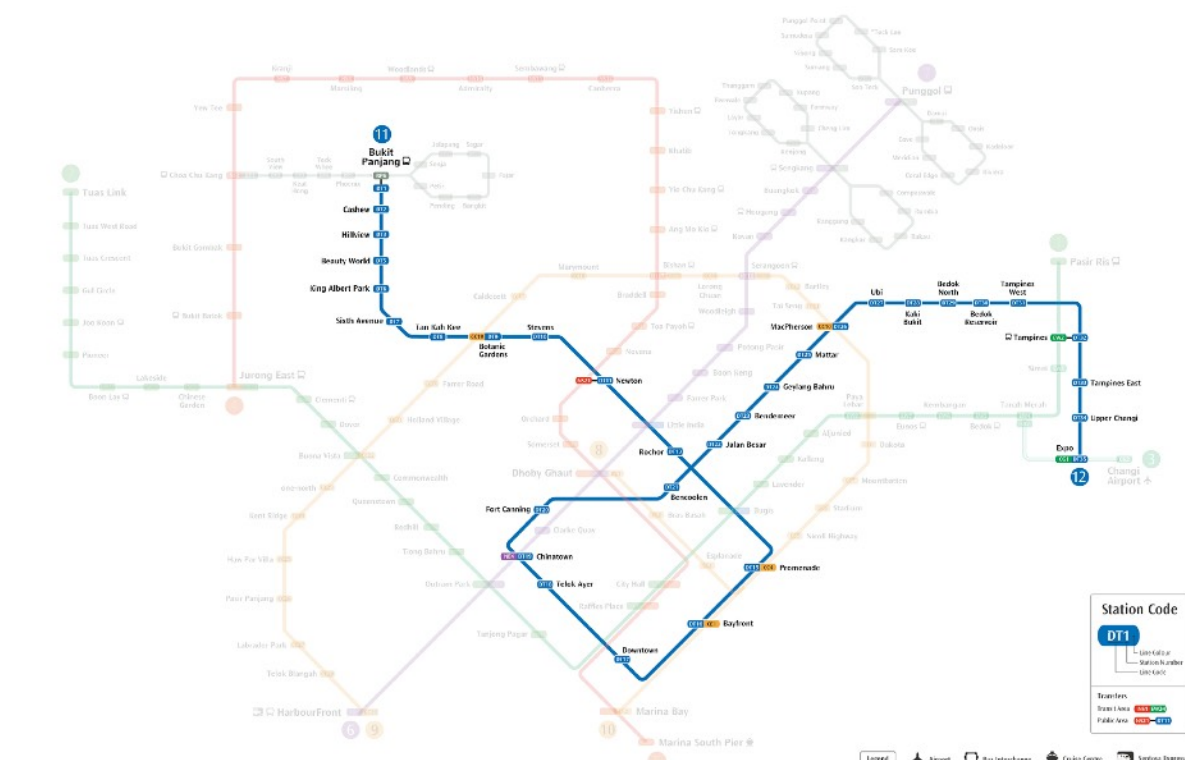
We can also see from the subway map above that it is generally quite easy to get anywhere within Singapore in 30-40 mins (unless you're travelling from one end of the Island to another).

And so if you are able to find a place close enough to the Downtown line, you can hop on a train (which comes every 2 to 7 minutes) and be in the Central/Downtown area within minutes.

As we can also see from our analysis of the "Heartland" clusters, all of them have very important nearby venues like Supermarkets, Food courts, and you may even be able to find a couple of restaurants that you like.



**Further Analysis & Improvements**

An improvement which I could have made is to analyse travelling data between locations. For example, how long does it take to travel from one location to another, and how would that affect the process of choosing a place to live in? Singapore's transport infrastructure, while not perfect, is always expanding and improving. There will definitely be a lot of data we can look at in that area, and analysing it can be useful to many people.