**Capstone Project - The Battle of Neighborhoods**
Applied Data Science Capstone by IBM/Coursera
*Prepared by: Michiko De la Cruz*

# Searching for an Optimal Relocation Area in Singapore

## I. Introduction

### Background

Expats from all over the world move to Singapore in large numbers on a yearly basis and at present there are approximately 1.64 million none permanent residents/citizens living in Singapore (September 2020). Singapore is widely regarded as the easiest city in Asia for expats to fit into and it allows foreigners an opportunity to become acquainted with different cultures in a relatively safe and modern environment.

Singapore is an efficient and clean Asian city that retains much of the charm from its British colonial days whilst also offering a cutting edge, well-developed environment. Singapore's position on the southern tip of Malaysia has allowed it to develop into one of the most important trade and finance centers in Asia and today it is the fifth wealthiest country in the world according to GDP. It also has the fourth largest foreign exchange trading center in the world.

Singapore's population of 5.69 million people (September 2020) largely consist of people of Chinese (70%), Malay and Indian backgrounds. The main religions followed in Singapore are Buddhism, Islam, Hinduism, Christianity and Tao.

While Singapore does offer a high standard of living, it is also one of the most expensive cities in the world to live in. Expatriates will find that property prices are very high (whether to rent or buy), alcohol is also taxed at a high rate, and it is also costly to own a car due to the heavy taxes that are imposed on vehicles.

However, due to its authoritarian state, expats will find that food and groceries are quite reasonable and public transportation is efficient and orderly.

### Objective and Scope

As there are still many expats looking to move to Singapore, we would like to define some factors to guide us in choosing an optimal location within the city to move to:

1. Locations near commercial real estate areas (malls, restaurants, groceries, etc.)
2. Locations near transportation hubs (MRT, buses, etc.)
3. Slightly lower rental prices

These factors were chosen to ensure a higher chance of cost-savings and sustainability when moving to Singapore.

## II. Data

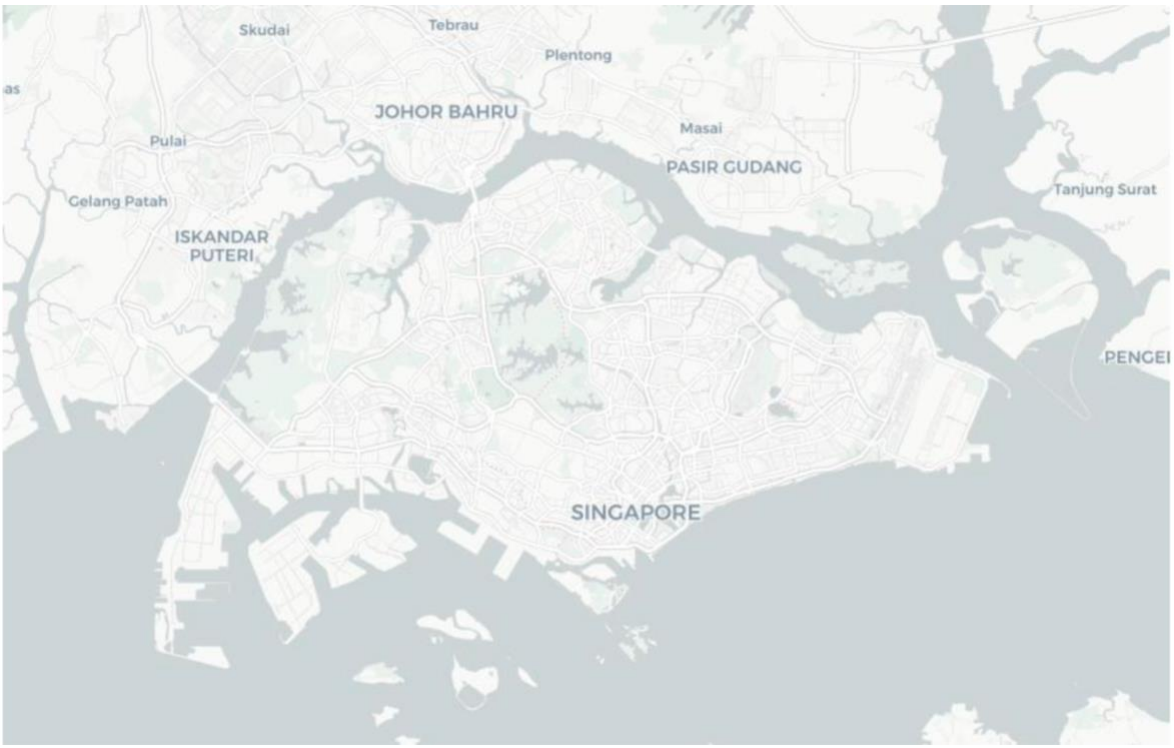
### Data Description

In order to investigate this problem, we will be looking at the following data:

1. We will be using a map of the Region Boundaries depicting the districts of Singapore from the Urban Redevelopment Authority's Master Plan 2014 to plot our data.
2. We will be using a dataset of Median Rentals by Town and Flat Type from the Urban Redevelopment Authority, which shows the median rent per sqm per month to get an estimate of rental prices of different housing building locations around Singapore.
3. We will use the Foursquare API to obtain a list of commercial real estate and transportation hubs in each district around Singapore.

### Data Cleaning

We used Python's Folium library to plot a plain map of Singapore, giving us an idea of the area that we will be tackling in this project.




Map 1: Plain Map of Singapore

We then used a dataset of 'Median Rentals by Town and Flat Type' from the Urban Redevelopment Authority found here: https://data.gov.sg/dataset/median-rent-by-town-and-flat-type.

The dataset contains the median rental prices of different residential properties in Singapore at a per square meter per month basis.

We also cleaned the dataset by removing unnecessary blank rows and dropped all rows with no rental price available to finally obtain the table below consisting of 20 different streets in Singapore.

| | quarter | Street | flat_type | Rent |
|---|---|---|---|---|
| 2 | 2020-Q4 | ANG MO KIO | 3-RM | 1800 |
| 3 | 2020-Q4 | ANG MO KIO | 4-RM | 2100 |
| 4 | 2020-Q4 | ANG MO KIO | 5-RM | 2400 |
| 8 | 2020-Q4 | BEDOK | 3-RM | 1800 |
| 9 | 2020-Q4 | BEDOK | 4-RM | 2100 |

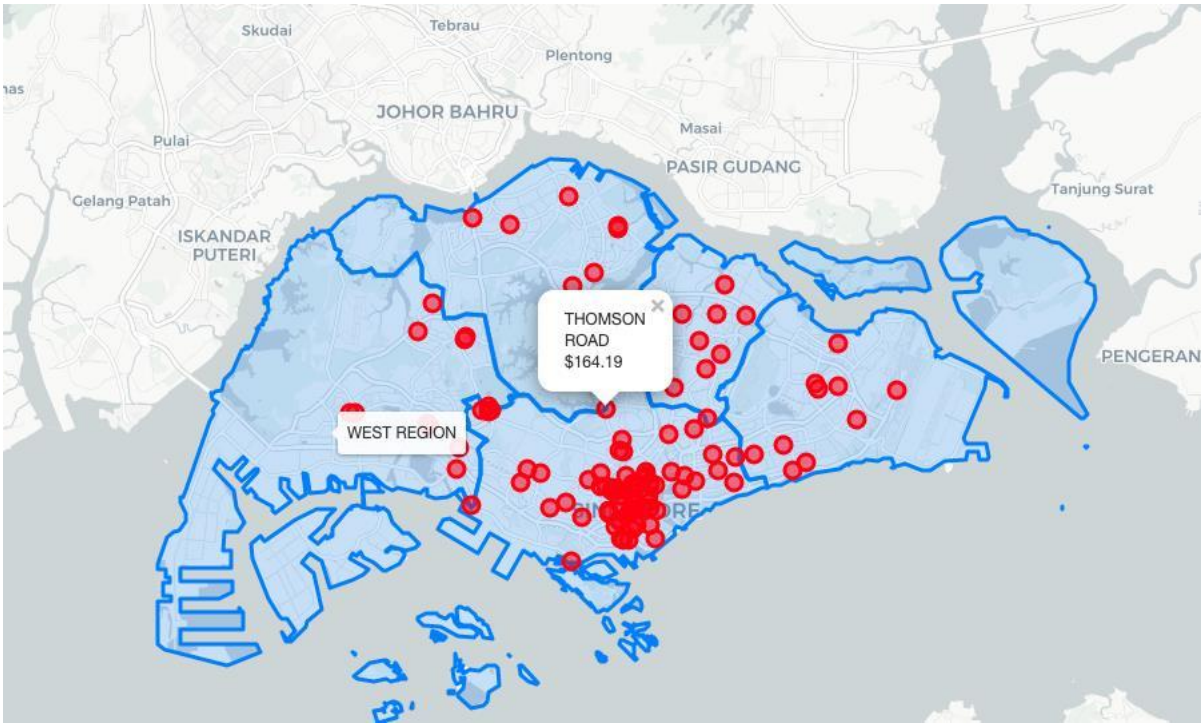
Table 1: Dataframe of Median Rental Prices of Various Streets in Singapore

However, the dataframe lacked the coordinates required for us to pinpoint and plot the exact locations of these streets. We thus used OpenCage Geocoder to iterate over the rows and obtain their accompanying coordinates.

| | quarter | Street | flat_type | Rent | Latitude | Longitude |
|---|---|---|---|---|---|---|
| 2 | 2020-Q4 | ANG MO KIO | 3-RM | 1800 | 1.370080 | 103.849523 |
| 3 | 2020-Q4 | ANG MO KIO | 4-RM | 2100 | 1.370080 | 103.849523 |
| 4 | 2020-Q4 | ANG MO KIO | 5-RM | 2400 | 1.370080 | 103.849523 |
| 8 | 2020-Q4 | BEDOK | 3-RM | 1800 | 1.323976 | 103.930216 |
| 9 | 2020-Q4 | BEDOK | 4-RM | 2100 | 1.323976 | 103.930216 |

Table 2: Dataframe of Median Rental Prices with Coordinates

We then downloaded the region boundaries map to map out the various districts of Singapore. However, the map is originally in KML format, and in order for it to be used with Folium it needs to be in the GeoJSON format. To do so, we used an external converter to convert it to a GeoJSON file. Using the dataframe and the GeoJSON map, we plotted out the locations to obtain the map below.



Map 2: Map of Singapore with Various Streets and Accompanying Rental Prices

The map shows the various street locations marked out with the red circle markers, which provide the street name and accompanying rental price when clicked on. The region boundaries are also plotted out and shows the respective region name when scrolled over.

We then used the Foursquare API to draw a list of the following in the vicinity of our street locations by searching specifically for each category:

**A. Transportation Hubs**
- MRT Stations
- Bus Stops
- Taxi Stands

**B. Commercial Real Estate**
- Groceries
- Shopping Malls
- Restaurants/Cafés

We set the limit of returned venues to be 50, with an encompassing radius of 500 meters. Foursquare API then returned a list of each item for each street that was iterated, along with their coordinates and we put them into a dataframe.

| Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude |
|---|---|---|---|---|---|---|
| ANG MO KIO | 6 | 6 | 6 | 6 | 6 | 6 |
| BEDOK | 3 | 3 | 3 | 3 | 3 | 3 |
| BISHAN | 9 | 9 | 9 | 9 | 9 | 9 |
| BUKIT BATOK | 12 | 12 | 12 | 12 | 12 | 12 |
| CHOA CHU KANG | 6 | 6 | 6 | 6 | 6 | 6 |
| CLEMENTI | 9 | 9 | 9 | 9 | 9 | 9 |
| HOUGANG | 32 | 32 | 32 | 32 | 32 | 32 |
| JURONG EAST | 12 | 12 | 12 | 12 | 12 | 12 |
| JURONG WEST | 8 | 8 | 8 | 8 | 8 | 8 |
| KALLANG/WHAMPOA | 3 | 3 | 3 | 3 | 3 | 3 |
| PASIR RIS | 3 | 3 | 3 | 3 | 3 | 3 |
| PUNGGOL | 4 | 4 | 4 | 4 | 4 | 4 |
| QUEENSTOWN | 4 | 4 | 4 | 4 | 4 | 4 |
| SEMBAWANG | 9 | 9 | 9 | 9 | 9 | 9 |
| SENGKANG | 9 | 9 | 9 | 9 | 9 | 9 |
| SERANGOON | 3 | 3 | 3 | 3 | 3 | 3 |
| TAMPINES | 8 | 8 | 8 | 8 | 8 | 8 |
| TOA PAYOH | 9 | 9 | 9 | 9 | 9 | 9 |
| WOODLANDS | 12 | 12 | 12 | 12 | 12 | 12 |
| YISHUN | 39 | 39 | 39 | 39 | 39 | 39 |

Table 3: Dataframe of MRT Stations and Their Coordinates

| Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude |
|---|---|---|---|---|---|---|
| ANG MO KIO | 54 | 54 | 54 | 54 | 54 | 54 |
| BEDOK | 24 | 24 | 24 | 24 | 24 | 24 |
| BISHAN | 15 | 15 | 15 | 15 | 15 | 15 |
| BUKIT BATOK | 16 | 16 | 16 | 16 | 16 | 16 |
| BUKIT MERAH | 3 | 3 | 3 | 3 | 3 | 3 |
| BUKIT PANJANG | 24 | 24 | 24 | 24 | 24 | 24 |
| CENTRAL | 4 | 4 | 4 | 4 | 4 | 4 |
| CHOA CHU KANG | 27 | 27 | 27 | 27 | 27 | 27 |
| CLEMENTI | 21 | 21 | 21 | 21 | 21 | 21 |
| GEYLANG | 6 | 6 | 6 | 6 | 6 | 6 |
| HOUGANG | 32 | 32 | 32 | 32 | 32 | 32 |
| JURONG EAST | 45 | 45 | 45 | 45 | 45 | 45 |
| JURONG WEST | 56 | 56 | 56 | 56 | 56 | 56 |
| KALLANG/WHAMPOA | 3 | 3 | 3 | 3 | 3 | 3 |
| MARINE PARADE | 17 | 17 | 17 | 17 | 17 | 17 |
| PASIR RIS | 18 | 18 | 18 | 18 | 18 | 18 |
| PUNGGOL | 6 | 6 | 6 | 6 | 6 | 6 |
| SEMBAWANG | 27 | 27 | 27 | 27 | 27 | 27 |
| SENGKANG | 9 | 9 | 9 | 9 | 9 | 9 |
| SERANGOON | 27 | 27 | 27 | 27 | 27 | 27 |
| TAMPINES | 140 | 140 | 140 | 140 | 140 | 140 |
| TOA PAYOH | 21 | 21 | 21 | 21 | 21 | 21 |
| WOODLANDS | 44 | 44 | 44 | 44 | 44 | 44 |
| YISHUN | 24 | 24 | 24 | 24 | 24 | 24 |

Table 4: Dataframe of Bus Stops and Their Coordinates

| Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude |
|---|---|---|---|---|---|---|
| ANG MO KIO | 3 | 3 | 3 | 3 | 3 | 3 |
| BUKIT MERAH | 3 | 3 | 3 | 3 | 3 | 3 |
| CENTRAL | 2 | 2 | 2 | 2 | 2 | 2 |
| CLEMENTI | 3 | 3 | 3 | 3 | 3 | 3 |
| JURONG EAST | 6 | 6 | 6 | 6 | 6 | 6 |
| JURONG WEST | 8 | 8 | 8 | 8 | 8 | 8 |
| PUNGGOL | 2 | 2 | 2 | 2 | 2 | 2 |
| SEMBAWANG | 3 | 3 | 3 | 3 | 3 | 3 |
| SENGKANG | 6 | 6 | 6 | 6 | 6 | 6 |
| SERANGOON | 6 | 6 | 6 | 6 | 6 | 6 |
| TAMPINES | 16 | 16 | 16 | 16 | 16 | 16 |
| YISHUN | 3 | 3 | 3 | 3 | 3 | 3 |

Table 5: Dataframe of Taxi Stands and Their Coordinates

| Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude |
|---|---|---|---|---|---|---|
| ANG MO KIO | 15 | 15 | 15 | 15 | 15 | 15 |
| BEDOK | 12 | 12 | 12 | 12 | 12 | 12 |
| BISHAN | 12 | 12 | 12 | 12 | 12 | 12 |
| BUKIT BATOK | 20 | 20 | 20 | 20 | 20 | 20 |
| BUKIT PANJANG | 16 | 16 | 16 | 16 | 16 | 16 |
| CENTRAL | 8 | 8 | 8 | 8 | 8 | 8 |
| CHOA CHU KANG | 9 | 9 | 9 | 9 | 9 | 9 |
| CLEMENTI | 21 | 21 | 21 | 21 | 21 | 21 |
| GEYLANG | 45 | 45 | 45 | 45 | 45 | 45 |
| HOUGANG | 16 | 16 | 16 | 16 | 16 | 16 |
| JURONG EAST | 21 | 21 | 21 | 21 | 21 | 21 |
| JURONG WEST | 8 | 8 | 8 | 8 | 8 | 8 |
| KALLANG/WHAMPOA | 9 | 9 | 9 | 9 | 9 | 9 |
| MARINE PARADE | 9 | 9 | 9 | 9 | 9 | 9 |
| PASIR RIS | 9 | 9 | 9 | 9 | 9 | 9 |
| PUNGGOL | 2 | 2 | 2 | 2 | 2 | 2 |
| QUEENSTOWN | 8 | 8 | 8 | 8 | 8 | 8 |
| SEMBAWANG | 18 | 18 | 18 | 18 | 18 | 18 |
| SENGKANG | 18 | 18 | 18 | 18 | 18 | 18 |
| SERANGOON | 24 | 24 | 24 | 24 | 24 | 24 |
| TAMPINES | 40 | 40 | 40 | 40 | 40 | 40 |
| TOA PAYOH | 27 | 27 | 27 | 27 | 27 | 27 |
| WOODLANDS | 16 | 16 | 16 | 16 | 16 | 16 |
| YISHUN | 42 | 42 | 42 | 42 | 42 | 42 |

Table 6: Dataframe of Groceries and Their Coordinates

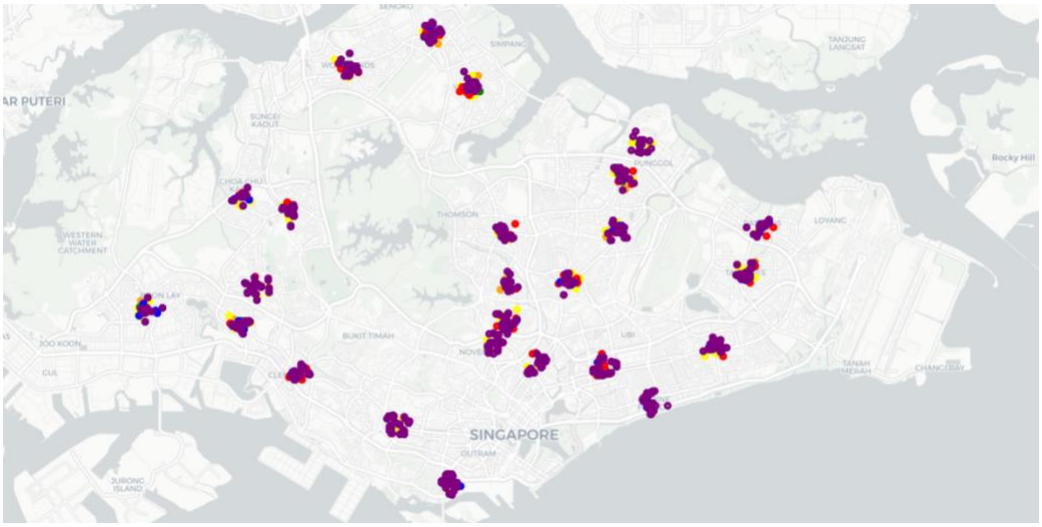
| Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude |
|---|---|---|---|---|---|---|
| ANG MO KIO | 54 | 54 | 54 | 54 | 54 | 54 |
| BEDOK | 24 | 24 | 24 | 24 | 24 | 24 |
| BISHAN | 15 | 15 | 15 | 15 | 15 | 15 |
| BUKIT BATOK | 16 | 16 | 16 | 16 | 16 | 16 |
| BUKIT MERAH | 3 | 3 | 3 | 3 | 3 | 3 |
| BUKIT PANJANG | 24 | 24 | 24 | 24 | 24 | 24 |
| CENTRAL | 4 | 4 | 4 | 4 | 4 | 4 |
| CHOA CHU KANG | 27 | 27 | 27 | 27 | 27 | 27 |
| CLEMENTI | 21 | 21 | 21 | 21 | 21 | 21 |
| GEYLANG | 6 | 6 | 6 | 6 | 6 | 6 |
| HOUGANG | 32 | 32 | 32 | 32 | 32 | 32 |
| JURONG EAST | 45 | 45 | 45 | 45 | 45 | 45 |
| JURONG WEST | 56 | 56 | 56 | 56 | 56 | 56 |
| KALLANG/WHAMPOA | 3 | 3 | 3 | 3 | 3 | 3 |
| MARINE PARADE | 17 | 17 | 17 | 17 | 17 | 17 |
| PASIR RIS | 18 | 18 | 18 | 18 | 18 | 18 |
| PUNGGOL | 6 | 6 | 6 | 6 | 6 | 6 |
| SEMBAWANG | 27 | 27 | 27 | 27 | 27 | 27 |
| SENGKANG | 9 | 9 | 9 | 9 | 9 | 9 |
| SERANGOON | 27 | 27 | 27 | 27 | 27 | 27 |
| TAMPINES | 140 | 140 | 140 | 140 | 140 | 140 |
| TOA PAYOH | 21 | 21 | 21 | 21 | 21 | 21 |
| WOODLANDS | 44 | 44 | 44 | 44 | 44 | 44 |
| YISHUN | 24 | 24 | 24 | 24 | 24 | 24 |

Table 7: Dataframe of Shopping Malls and Their Coordinates

| Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude |
|---|---|---|---|---|---|---|
| ANG MO KIO | 150 | 150 | 150 | 150 | 150 | 150 |
| BEDOK | 150 | 150 | 150 | 150 | 150 | 150 |
| BISHAN | 150 | 150 | 150 | 150 | 150 | 150 |
| BUKIT BATOK | 200 | 200 | 200 | 200 | 200 | 200 |
| BUKIT MERAH | 144 | 144 | 144 | 144 | 144 | 144 |
| BUKIT PANJANG | 200 | 200 | 200 | 200 | 200 | 200 |
| CENTRAL | 98 | 98 | 98 | 98 | 98 | 98 |
| CHOA CHU KANG | 150 | 150 | 150 | 150 | 150 | 150 |
| CLEMENTI | 150 | 150 | 150 | 150 | 150 | 150 |
| GEYLANG | 150 | 150 | 150 | 150 | 150 | 150 |
| HOUGANG | 200 | 200 | 200 | 200 | 200 | 200 |
| JURONG EAST | 150 | 150 | 150 | 150 | 150 | 150 |
| JURONG WEST | 200 | 200 | 200 | 200 | 200 | 200 |
| KALLANG/WHAMPOA | 150 | 150 | 150 | 150 | 150 | 150 |
| MARINE PARADE | 50 | 50 | 50 | 50 | 50 | 50 |
| PASIR RIS | 150 | 150 | 150 | 150 | 150 | 150 |
| PUNGGOL | 100 | 100 | 100 | 100 | 100 | 100 |
| QUEENSTOWN | 200 | 200 | 200 | 200 | 200 | 200 |
| SEMBAWANG | 147 | 147 | 147 | 147 | 147 | 147 |
| SENGKANG | 150 | 150 | 150 | 150 | 150 | 150 |
| SERANGOON | 150 | 150 | 150 | 150 | 150 | 150 |
| TAMPINES | 200 | 200 | 200 | 200 | 200 | 200 |
| TOA PAYOH | 150 | 150 | 150 | 150 | 150 | 150 |
| WOODLANDS | 200 | 200 | 200 | 200 | 200 | 200 |
| YISHUN | 150 | 150 | 150 | 150 | 150 | 150 |

Table 8: Dataframe of Restaurants/Cafes and Their Coordinates

We also grouped them according to their respective street location to get a brief idea of how many transportation hubs and commercial real estate areas are there per street.




Map 3: Map of Singapore with Various Transport Hubs and Commercial Real Estate Per Area

Each category is indicated by a different color as indicated below:

**A. Transportation Hubs**
- MRT Stations -Orange
- Bus Stops -Yellow
- Taxi Stands -Green

**B. Commercial Real Estate**
- Groceries -Red
- Shopping Malls -Blue
- Restaurants/Cafés -Purple

## III. Methodology

### Clustering

In order to better analyze our data, we will use k-means clustering method to create a model of 5 clusters of cafes. The decision to have 5 clusters is to match the number of districts in Singapore, and the resulting clusters should be similarly mapped to the districts. K-means clustering was chosen due to its ease of use and efficiency for medium to large datasets. We used the coordinates of our cafes to fit our clustering model, and generated the resulting dataframe of accompanying clusters below.

| | Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude | Cluster |
|---|---|---|---|---|---|---|---|---|
| 0 | ANG MO KIO | 1.37008 | 103.849523 | 1800 | Koi Café Express | 1.369342 | 103.849820 | 2 |
| 1 | ANG MO KIO | 1.37008 | 103.849523 | 1800 | A&W | 1.369541 | 103.849043 | 2 |
| 2 | ANG MO KIO | 1.37008 | 103.849523 | 1800 | Phoon Huat & Co (Pte) Ltd | 1.368318 | 103.851639 | 2 |
| 3 | ANG MO KIO | 1.37008 | 103.849523 | 1800 | Malaysia Boleh! | 1.369441 | 103.849066 | 2 |
| 4 | ANG MO KIO | 1.37008 | 103.849523 | 1800 | BreadTalk / Toast Box | 1.369177 | 103.848874 | 2 |

Table 9: Dataframe of Transport Hubs and Commercial Real Estate Areas and Their Clusters

| Cluster | Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude |
|---|---|---|---|---|---|---|---|
| 0 | 950 | 950 | 950 | 950 | 950 | 950 | 950 |
| 1 | 1050 | 1050 | 1050 | 1050 | 1050 | 1050 | 1050 |
| 2 | 1048 | 1048 | 1048 | 1048 | 1048 | 1048 | 1048 |
| 3 | 497 | 497 | 497 | 497 | 497 | 497 | 497 |
| 4 | 344 | 344 | 344 | 344 | 344 | 344 | 344 |

Table 10: Dataframe of Count of Transport Hubs and Commercial Real Estate Areas in Each Cluster

Cluster 2 has the highest number of transportation hubs and commercial real estate areas because it is not situated in the central business district, meaning that there is more space for land redevelopment at lower rental rates.

## Statistical Analysis

From the previous table, we wanted to obtain the average, maximum, and minimum rental rates for each cluster.

| Cluster | Street | Street_Latitude | Street_Longitude | Rent | Venue | Venue_Latitude | Venue_Longitude | Avg_Rent | Max_Rent | Min_Rent |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 950 | 950 | 950 | 950 | 950 | 950 | 950 | 2077.894737 | 2400 | 1750 |
| 1 | 1050 | 1050 | 1050 | 1050 | 1050 | 1050 | 1050 | 2045.238095 | 2600 | 1550 |
| 2 | 1048 | 1048 | 1048 | 1048 | 1048 | 1048 | 1048 | 2223.377863 | 2700 | 1800 |
| 3 | 497 | 497 | 497 | 497 | 497 | 497 | 497 | 1897.283702 | 2150 | 1550 |
| 4 | 344 | 344 | 344 | 344 | 344 | 344 | 344 | 2347.965116 | 2850 | 1500 |

Table 11: Dataframe with Cluster Rental Statistics

The table above shows the following results:

1. The most populated clusters (1-2) have the second to the lowest average rental. These areas were mostly outside the city center, in more residential areas of Singapore.
2. The least populated cluster (4) has the highest average rental as these are located close to the Central Business District.

Considering locations with more transportation hubs and commercial real estate areas in the vicinity is one of our objectives, we will focus on looking at the North and East Regions, where Clusters 2 and 4 are located.

This concludes our analysis.

## IV. Results and Discussion

To reiterate, we are focusing on the following factors:

1. Locations that are near transportation hubs, like MRT Stations, Bus Stops, and Taxi Stands.
2. Locations that are near commercial real estate areas such as, Groceries, Shopping Centers, Restaurants and Cafes.
3. Locations with mid-level rental rates.

Our analysis shows that the East and West Regions of Singapore may be ideal locations to relocate to, with 3 areas (Tampines, Jurong East, Jurong West) turning up as potential candidates that fulfill the above criteria.

The highest concentration of transport hubs and commercial real estate were found to be in the East Region with Cluster 1, whereas the lowest concentration was found to be in the North Region with Cluster 4. The highest average rental was found in Cluster 4 in the Central Region, whilst the lowest average rental was found in Cluster 3 in the North-East Region.

Despite having 3 potential candidates, there is much room for improvement to draw better conclusions.

For a start, the data we have obtained for rental prices and locations does not contain all the streets in Singapore. In addition, we dropped several streets due to their lack of retail rental information. The data we have is thus a small sample size of Singapore, and not representative of the entire country.

Next, the rental prices are the median rental per square meter per month, and not the mean rental prices. Having the mean rental prices may give us a better indication.

Another factor we did not take into account for this exercise is the relocator's objective for moving such as, job opportunities, family-building, which may veer the objective towards either looking for an area closer to their office or children's schools.

Lastly, we used the k-means clustering method which is efficient, but may not give the best results. The clustering centers for k-means are randomly selected, thus they may give different results each time it is initiated. If we could fix the clustering centers to pinpoint and separate the different regions of Singapore, we might be able to obtain more accurate results. It might also be a good idea to experiment with other clustering methods.

## V. Conclusion

The objective of this project was to identify potential candidate locations to aid stakeholders in relocating or moving within Singapore. With the dataset, we were able to obtain several locations around Singapore with their accompanying rental price. We then used Foursquare API to obtain and calculate transportation and commercial real estate density distribution for our locations. We then used k-means clustering to generate 5 clusters of interest, and proceeded to perform critical analysis on them.

Through our analysis, we identified 3 viable regions in Singapore that have fulfilled the criteria we set out at the beginning of the project. In reality, there are many more factors that affect the decision of selecting a relocation area that this project has not been able to take into consideration. In future iterations of this project, we hope to be able to obtain more data, and improve the findings to serve as a guide for all potential stakeholders to set up a successful and sustainable home in Singapore.