Capstone Project – The Battle of Neighborhoods

Dominic Michael Lee Chao Long

1. Introduction

1.1 Background

Cafes have been a growing trend amongst youth and working class adults for many years now, and the wave does not seem to be dying down. The popularity of cafes is resonant not just in Singapore, but all over the world. The concept of these trendy cafes has been largely attributed to their origins in London and Melbourne, with these spaces providing a comfortable spot for relaxation, some picture perfect-looking foods, and a great cup of coffee.

With this trend comes along many young entrepreneurs looking to cash in on the hype by wanting to set up their own café. However, setting up a café in the current food and beverage(F&B) environment is not as easy as it sounds.

1.2 Objective and Scope

Setting up a café entails several financial risks. It is thus important to identify the factors that may affect eventual success. In the current environment, there is intense competition from existing players in the market that may kill off your business before it even kicks off. Rental prices are also another major issue when it comes to F&B establishments, as rental is known to be the number one killer of businesses in Singapore. Another one will be the difficulty in attracting customers, or the presence of human foot traffic.

The objective of this project is thus to find an optimal location for opening a new café in the city of Singapore. To do so, we would like to define some guiding principles in choosing an optimal location:

1. Locations with few or no existing cafes in the vicinity
2. Locations with lower rental prices
3. Locations with decent human foot traffic

These factors were chosen to ensure a higher chance of success and sustainability in opening a café. The findings of this project will then serve as a guide to the relevant stakeholders interested in carrying out this endeavor.

2. Data

2.1 Data Description

In order to tackle the problem, we would require data to answer the criteria we have chosen.

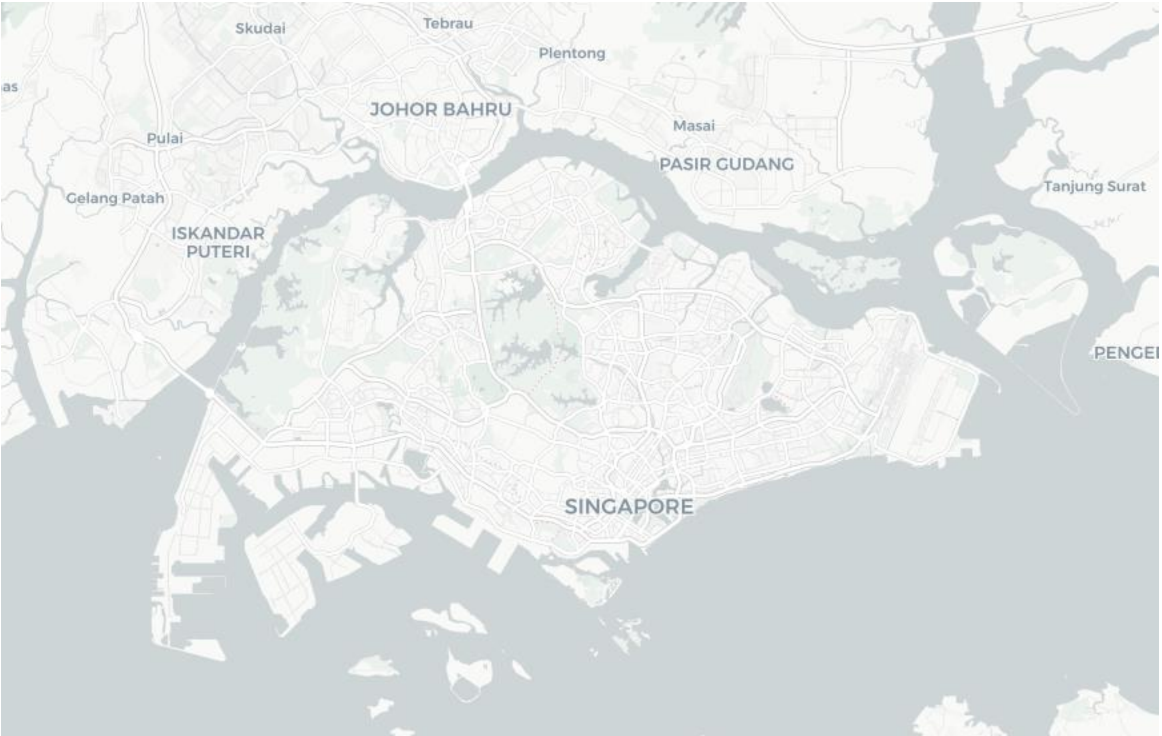
We first obtained a dataset relating to 'Median Rental Prices of Commercial Retail and Office Spaces' [1] from the Urban Redevelopment Authority of Singapore, which shows the median rent per square metre per month of various locations around Singapore to gives us an estimate of rental prices.

Next, we obtained a map of region boundaries [2] of Singapore from the Urban Redevelopment Authority's Master Plan 2014. This gives us the various districts of Singapore, which will aid in visualisation and clustering analysis later on.

In order to identify existing cafes, we used Foursquare API [3] to draw a list of existing cafés and their locations around Singapore.

2.2 Data Cleaning

To begin, we used Python's Folium library [4] 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 looked at loading our first dataset of median rental prices. Unfortunately, the dataset is in a PDF document. We thus used an external converter to convert it to csv format to make it easier for Python to read. We also cleaned the dataset by removing unnecessary blank rows and dropping the column for office rental prices as it is not in our scope of interest. We then dropped all rows with no rental price available to finally obtain the table below consisting of 109 different streets in Singapore.

	Street	Rent
0	AIRPORT BOULEVARD	270.63
2	ALEXANDRA ROAD	102.91
3	ANG MO KIO AVENUE 3	303.79
4	ANSON ROAD	75.36
5	BALESTIER ROAD	76.67

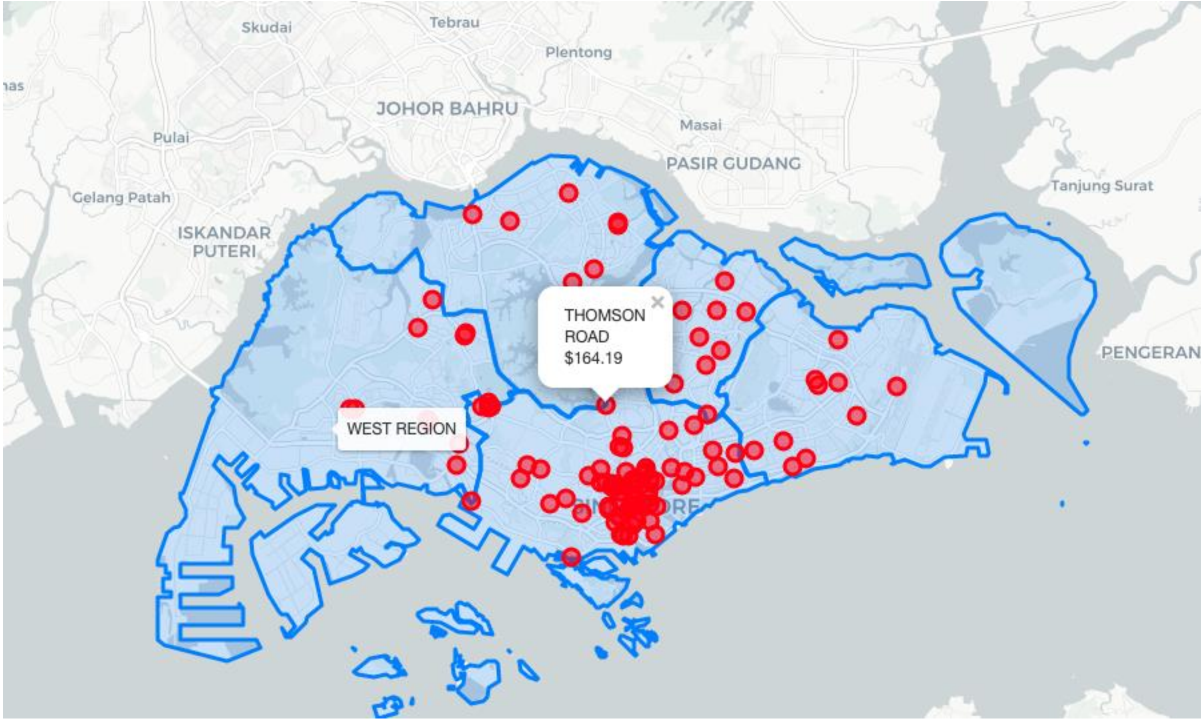
Table 1: Dataframe of Median Retail Rental Prices of Various Streets in Singapore

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.

	Street	Rent	Latitude	Longitude
0	AIRPORT BOULEVARD	270.63	1.351973	103.986400
2	ALEXANDRA ROAD	102.91	1.291695	103.808030
3	ANG MO KIO AVENUE 3	303.79	1.369119	103.850404
4	ANSON ROAD	75.36	1.274970	103.845806
5	BALESTIER ROAD	76.67	1.326651	103.844867

Table 2: Dataframe of Median Retail Rental Prices with Coordinates

Next, we 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 moused over.

Next, we used Foursquare API to draw a list of existing cafes in the vicinity of our street locations by searching specifically for the 'Cafés' category. We set the limit of returned venues to be 50, with an encompassing radius of 500 metres. Foursquare API then returns a list of cafes for each street that was iterated, along with their coordinates and we put them into a dataframe consisting 3,304 entries below.

	Street	Street_Latitude	Street_Longitude	Rent	Venue	Venue_Latitude	Venue_Longitude
0	AIRPORT BOULEVARD	1.351973	103.9864	270.63	Pacific Coffee Company Changi Airport	1.353395	103.985315
1	AIRPORT BOULEVARD	1.351973	103.9864	270.63	Yan's Cafe	1.350228	103.984512
2	AIRPORT BOULEVARD	1.351973	103.9864	270.63	Heavenly Wang	1.356417	103.987323
3	AIRPORT BOULEVARD	1.351973	103.9864	270.63	The Coffee Bean & Tea Leaf	1.355613	103.985587
4	AIRPORT BOULEVARD	1.351973	103.9864	270.63	Paris Baguette Café	1.356248	103.988332

Table 3: Dataframe of Existing Cafes and Their Coordinates

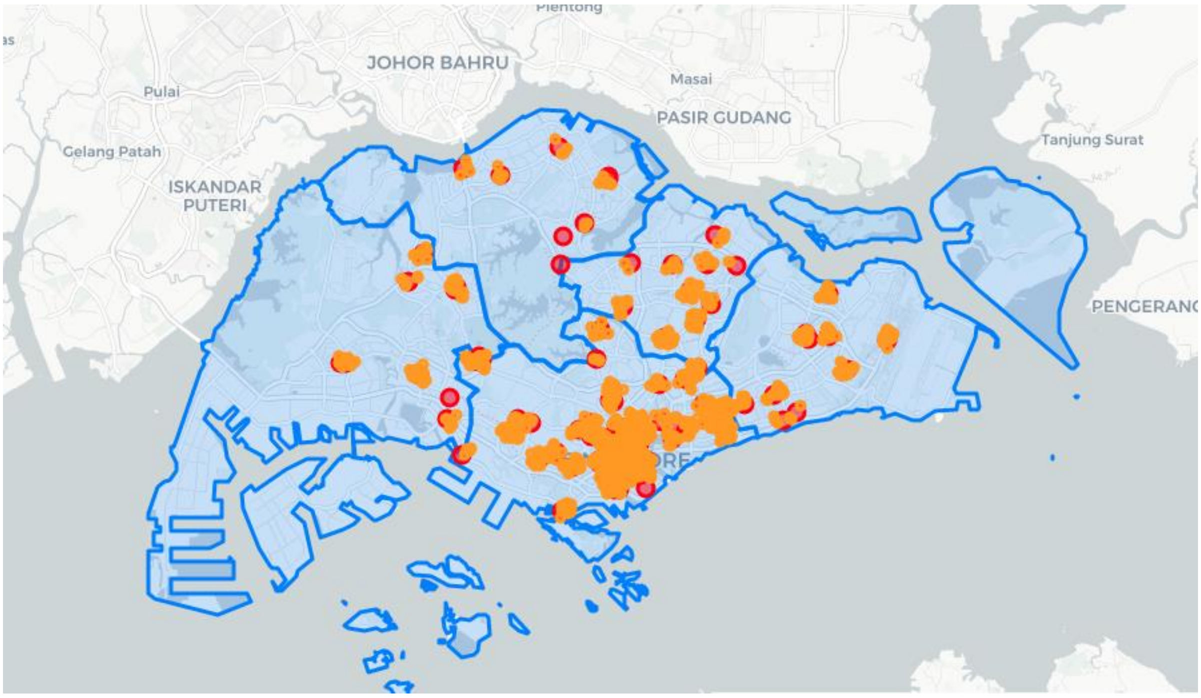
We also grouped them according to their respective street location to get a brief idea of how many cafes are there per street.

Street	Street_Latitude	Street_Longitude	Rent	Venue	Venue_Latitude	Venue_Longitude
AIRPORT BOULEVARD	27	27	27	27	27	27
ALEXANDRA ROAD	22	22	22	22	22	22
ANG MO KIO AVENUE 3	23	23	23	23	23	23
ANSON ROAD	50	50	50	50	50	50
BALESTIER ROAD	15	15	15	15	15	15
...	...	...	...	...	...	...
WOODLANDS ROAD	4	4	4	4	4	4
WOODLANDS SQUARE	18	18	18	18	18	18
YIO CHU KANG ROAD	2	2	2	2	2	2
YISHUN AVENUE 2	2	2	2	2	2	2
YISHUN AVENUE 9	7	7	7	7	7	7

Table 4: Count of Number of Cafes for Each Street

We can see that Anson Road returned the highest number of 50 cafes in its vicinity, whilst Yio Chu Kang Road and Yishun Avenue 2 returned the lowest of 2 in their vicinity.

Based on this dataframe, we plotted out the cafes onto a map alongside their street locations to give us a better picture.



Map 3: Map of Existing Cafes and Accompanying Streets

3. Methodology

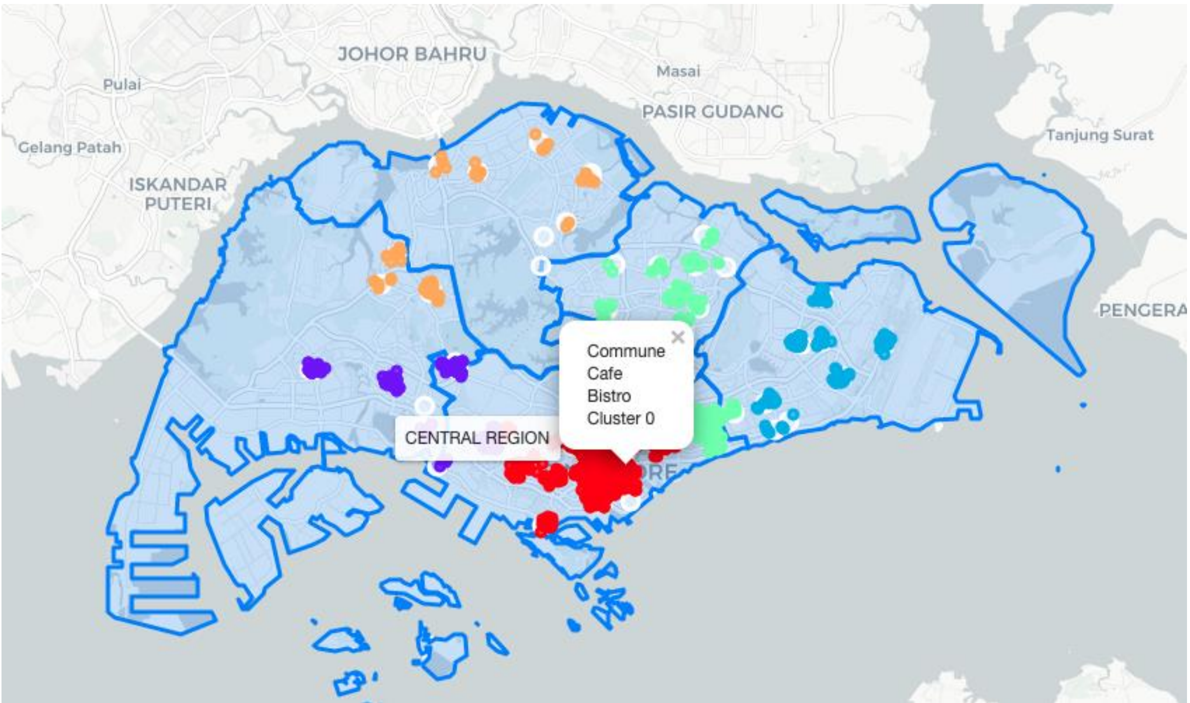
3.1 Clustering

In order to better analyse 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	AIRPORT BOULEVARD	1.351973	103.9864	270.63	Pacific Coffee Company Changi Airport	1.353395	103.985315	2
1	AIRPORT BOULEVARD	1.351973	103.9864	270.63	Yan's Cafe	1.350228	103.984512	2
2	AIRPORT BOULEVARD	1.351973	103.9864	270.63	Heavenly Wang	1.356417	103.987323	2
3	AIRPORT BOULEVARD	1.351973	103.9864	270.63	The Coffee Bean & Tea Leaf	1.355613	103.985587	2
4	AIRPORT BOULEVARD	1.351973	103.9864	270.63	Paris Baguette Café	1.356248	103.988332	2

Table 5: Dataframe of Cafes with Their Clusters

To better visualise this, we plot out the cafes in their clusters onto a map. The clusters are differentiated by colour, and the points show their cluster number when clicked on.



Map 4: Map of Cafes in Their Clusters

From the map above, we can see that the clusters are decently mapped to the different regions of Singapore. This will help in narrowing down the regions to look into.

We also do a count to see the number of cafes located in each cluster.

Cluster	Street	Street_Latitude	Street_Longitude	Rent	Venue	Venue_Latitude	Venue_Longitude
0	2250	2250	2250	2250	2250	2250	2250
1	336	336	336	336	336	336	336
2	245	245	245	245	245	245	245
3	374	374	374	374	374	374	374
4	99	99	99	99	99	99	99

Table 6: Count of Number of Cafes in Each Cluster

Looking at the count, Cluster 0 has the highest number of cafes, which is expected since it is located in the Central Region. Whereas Cluster 2 and 4 have the lowest number of cafes, hence they may be suitable for our objective.

3.2 Statistical Analysis

From our previous count, we wanted to obtain the average, maximum, and minimum rent of each cluster for analysis.

Cluster	Street	Street_Latitude	Street_Longitude	Rent	Venue	Venue_Latitude	Venue_Longitude	Avg_Rent	Max_Rent	Min_Rent
0	2250	2250	2250	2250	2250	2250	2250	122.839813	515.29	57.81
1	336	336	336	336	336	336	336	146.555565	263.65	35.71
2	245	245	245	245	245	245	245	253.336939	409.56	67.71
3	374	374	374	374	374	374	374	154.317112	328.46	35.65
4	99	99	99	99	99	99	99	230.626465	293.70	90.91

Table 7: Dataframe with Cluster Rental Statistics

From the table above, we observed some surprising results:

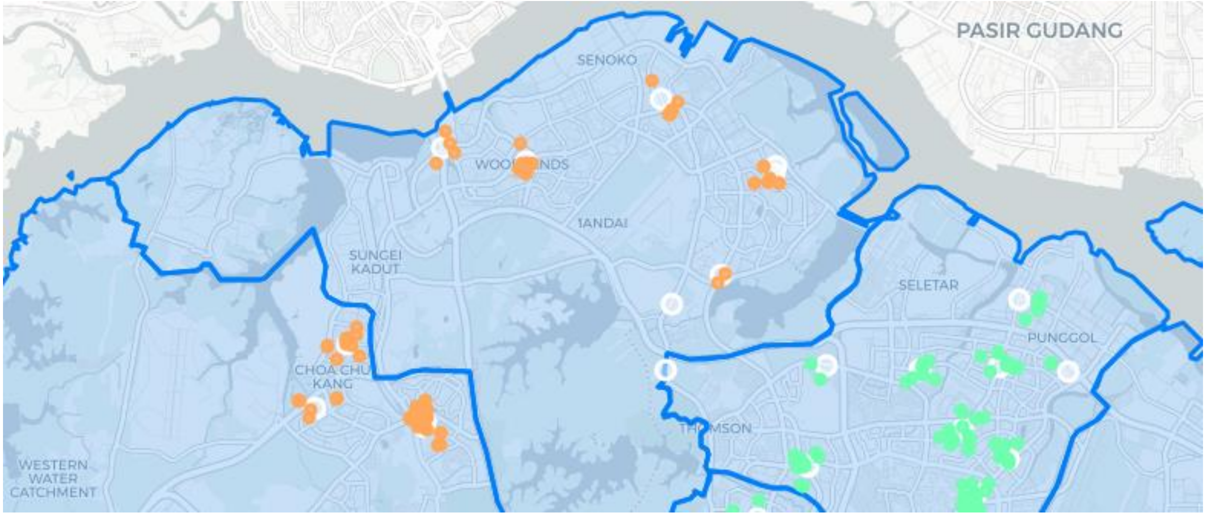
1. The most populated cluster has the lowest average rental.
2. However, it also consists of the highest rental.
3. The highest average rental is in the 2nd least populated cluster.
4. The least populated cluster has the highest minimum rental.

3.3 Region Analysis

One of the factors for consideration was locations with few or no existing cafes in the vicinity. Thus, based on the 2 least populated clusters, we decided to focus on the North and East Regions, where Clusters 2 and 4 are located.

3.3.1 North Region

We first selected a street from our dataset based in the North Region, then created a map focused on that area.



Map 5: Map of North Region

We assessed the area above, and found that locations with few or no cafes in the vicinity would include Woodlands Road, Sembawang Drive, Sembawang Road, and Yishun Avenue 2.

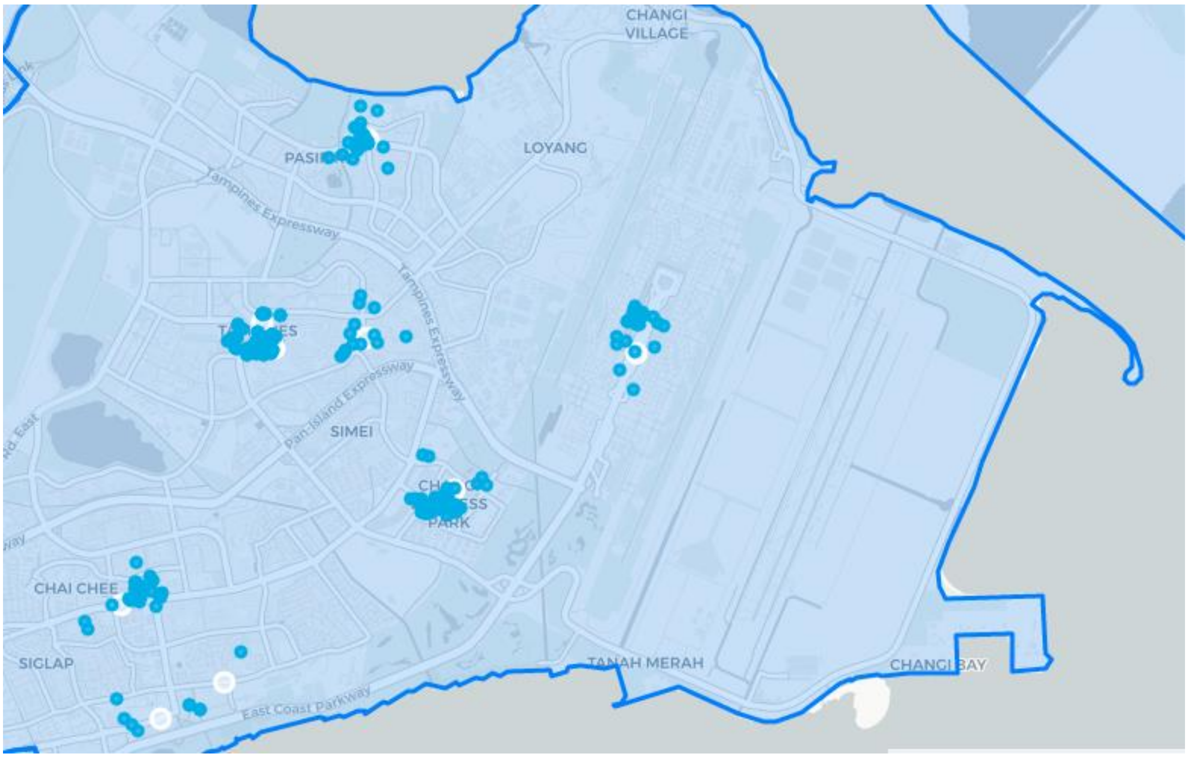
Woodlands Road has the lowest rental at $98.16 amongst these 4 locations, whilst Yishun Avenue 2 has the highest at $292.59. The average rental for Cluster 4 is $230.62.

However, upon deeper analysis, these 2 locations may not be as ideal. Woodlands Road is located very near the Causeway bridge to Malaysia, with mostly vehicle traffic going to/fro Malaysia and has little foot traffic. Whereas Yishun Avenue 2 commands a high rental, while being on the fringe of a housing estate, resulting in little human traffic.

Sembawang Drive and Sembawang Road may be better options, as they have below average rental prices ($187.80 and $139.51 respectively) for the region while being near or in the middle of some housing estates.

3.3.2 East Region

Similar to before, we selected a street from our dataset based in the East Region, and plotted the map below.



Map 6: Map of East Region

The East Region was easier to analyse as majority of the locations are already filled with cafes nearby, thus making them less ideal.

The only 2 viable locations that meet our criteria are Marina Parade Road and Upper East Coast Road.

Upper East Coast Road might be a great choice, as it has a low rental of $81.08, and it situated in the middle of a private housing estate, which may be the right market to target for a café.

On the other hand, Marine Parade Road has an above average rental at $281.60, but is also in the middle of a private housing estate and is near a neighbourhood town center, which might mean high human foot traffic.

This concludes our analysis.

4. Results

To reiterate, we are focusing on the following criteria:

1. Locations with few or no existing cafes in the vicinity
2. Locations with lower rental prices
3. Locations with decent human foot traffic

Our analysis shows that the North and East Regions of Singapore may be ideal locations to set up a new café, with 4 streets (Sembawang Drive, Sembawang Road, Upper East Coast Road, and Marine Parade Road) turning up as potential candidates that fulfill the criteria we have set out.

The highest concentration of 2,250 cafes was found to be in the Central Region with Cluster 0, whereas the lowest concentration of 99 cafes was found to be in the North Region with Cluster 4. The highest rental of $515.29 was found in Cluster 0 in the Central Region, whilst the lowest rental of $35.65 was found in Cluster 3 in the North-East Region. The highest average rental of $253.33 was found in Cluster 2 in the North Region, whilst the lowest average rental of $122.83 was found in Cluster 0 in the Central Region.

5. Discussions

Despite having 4 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 metre per month, and not the mean rental prices. Having the mean rental prices may give us a better indication.

At the same time, we are dealing with retail property space, which comes in all shapes and sizes. They may be a shophouse, or a stall in a shopping mall, or in an office building. Thus despite having the prices in the per square metre per month format, it may not give a complete indication as some spaces may be much larger in size compared to others., resulting in a much higher overall quantum per month despite having a reasonable per square metre price. It will be better if we had data of the average retail space size or we had a standard space model to do a comparison.

One interesting part during the Foursquare API café density acquisition phase, is that we considered an additional search on the 'Coffee Shops' category to generate more cafes for our data, as coffee shops and cafes are synonymous in most countries. However in Singapore, this search instead turned up many traditional coffee shops aka. *kopitiams*, which are in fact hawker centers selling food. These venues are not congruent with our objective thus we decided to exclude them. This also brings about the issue of our data being dependent on how a venue is categorized in Foursquare. If a venue is wrongly categorized, it may end up as a false positive in our data.

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.

6. Conclusion

The objective of this project was to identify potential candidate locations to aid stakeholders in setting up a new café in Singapore. With a dataset we obtained several locations around Singapore with their accompanying rental price. We then used Foursquare API to obtain and calculate café 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 2 viable regions in Singapore along with 4 potential candidate locations 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 business location 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 café in Singapore.

References

[1] Urban Redevelopment Authority, *Median Rental Prices of Commercial Retail and Office Spaces*, 2020.
[2] Urban Redevelopment Authority, *Master Plan 2014 Region Boundary (No Sea)*, 2014.
[3] Foursquare, Foursquare Developers, 2020.
[4] Folium, *Folium 0.11.0 Documentation, 2020.*