

Capstone Project – The Battle of Neighborhoods

2. Data

2.1 Data Description

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

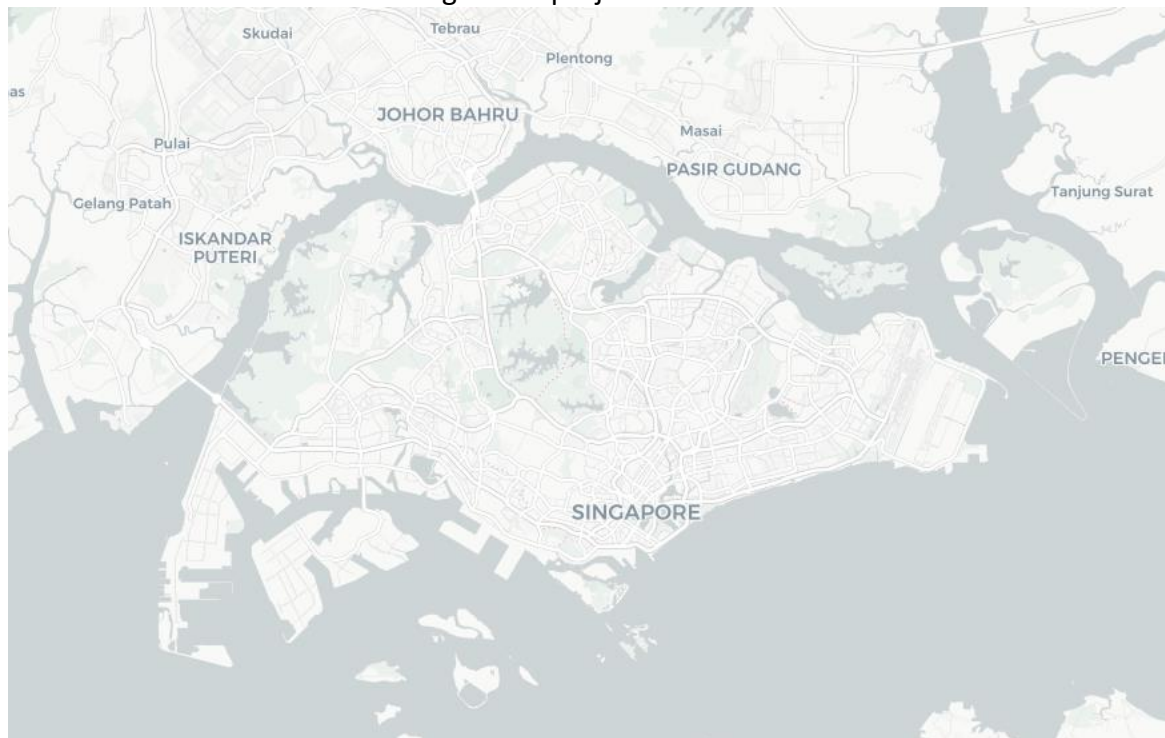
We first obtained a dataset relating to [‘Median Rental Prices of Commercial Retail and Office Spaces’](#) from the Urban Redevelopment Authority of Singapore, which shows the median rent per square metre per month of various locations around Singapore to give us an estimate of rental prices.

Next, we obtained [a map of region boundaries](#) 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](#) to draw a list of existing cafés and their locations around Singapore.

2.2 Data Cleaning

To begin, 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 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

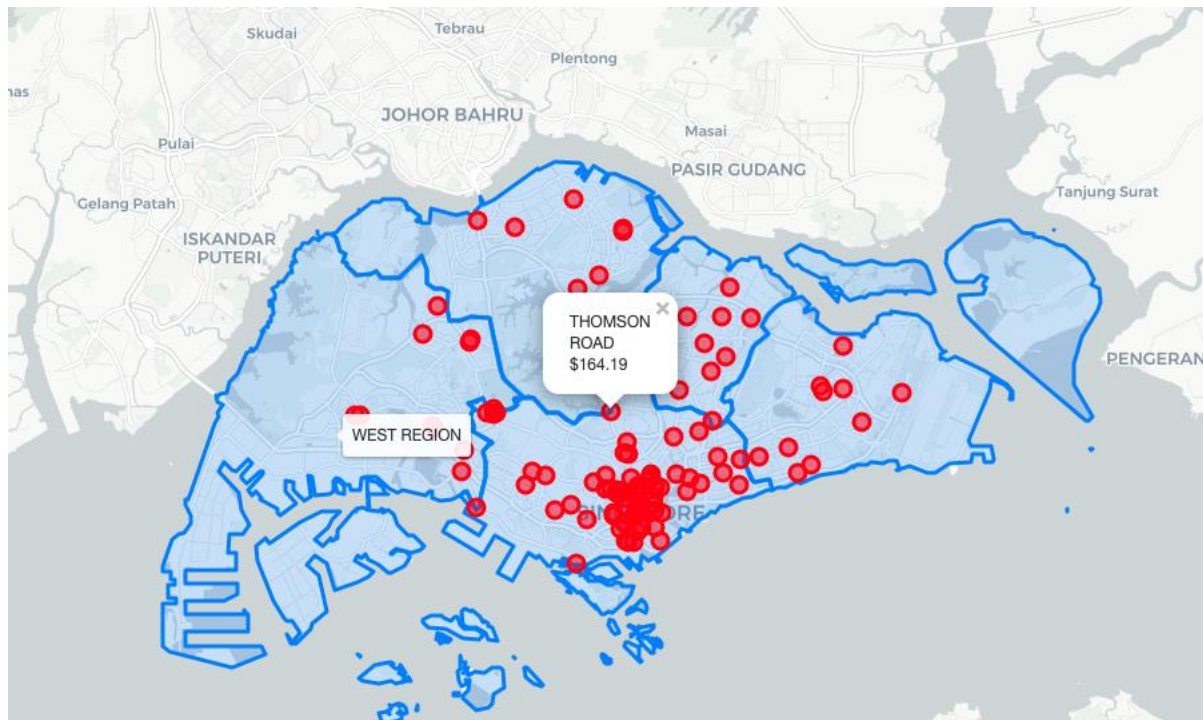
Table 1: Dataframe of Median Retail Rental Prices of Various Streets in Singapore

However, the dataframe lacked the coordinates 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

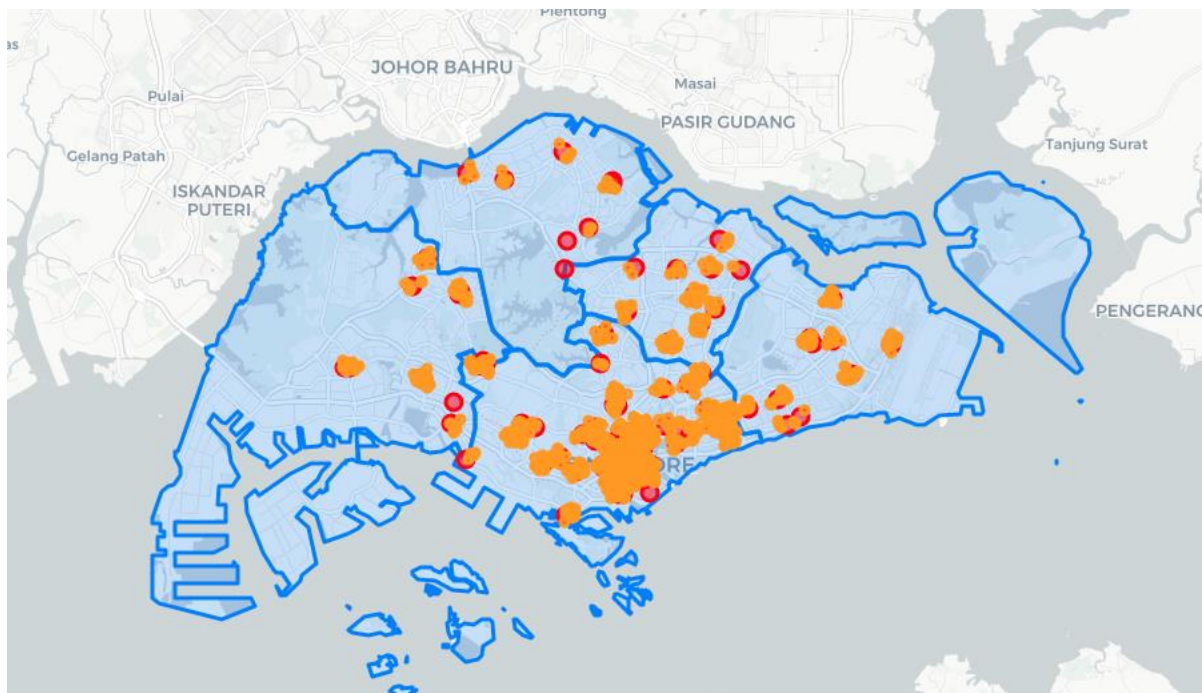
We also grouped them according to their respective street location to get a brief idea of how many cafes are there per street.

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