Opening a new shopping mall in Singapore Final Report

Sissi Lin

August, 2020

1. INTRODUCTION

1.1 Background

This capstone project is made specifically for IBM Data Science Professional Certificate, where a hypothetical scenario is raised. I chose Singapore for this project. Singapore,famously known as the "the red dot" on the world map, is one of the most important financial and transport hubs in Asia. Due to its geographical advantages, Singapore welcomes tourists from all around the world. Each year, Singapore attracts a lot of foreign investment as a result of its location, skilled workforce, low tax rates, and advanced infrastructure.
Singapore has a decent amount of shopping malls, because Singaporeans of all ages like to go to shopping malls for different purposes, some for shopping, some for dining out, some for watching movies... Not only the locals love going to shopping mall, but also the millions of tourists who visit Singapore each year.

1.2 Business Problem

Suppose a shopping mall chain is willing to open a new shopping mall outlet in Singapore to expand their business in South East Asia, however, the stakeholders are not familiar with Singapore. They want to know which location would be the best to suit their needs.

1.3 Target Audience

- The stakeholders who want to invest in the shopping mall.

- Tourists who visit Singapore. Singapore attracted approximately 19.1 million visitors in 2019 with receipts at S$27.1 billion, according to preliminary figures by the Singapore Tourism Board. The top three countries where the tourists are from are China, India and Indonesia.
- Local Singaporean residents. As of June 2019, Singapore's population stood at 5.70 million.

2. DATA

2.1 Data requirements

- List of the planning areas of Singapore
- Coordination (Latitudes, Longitudes) of the planning areas in Singapore
- Top venues of each planning areas

2.2 Data Sources

1. Downloaded the 2019 Singapore planning area dataset from: https://data.gov.sg/dataset/master-plan-2019-planning-area-boundary-no-sea?resource_id=06155d17-a1f2-44e8-926c-768bdba6e621, which contains columns including planning area, region, and latitude, longitude values.

2. Using foursquare API to retrieve venue data, we will focus especially on the category of shopping mall for the planning areas (neighborhoods).

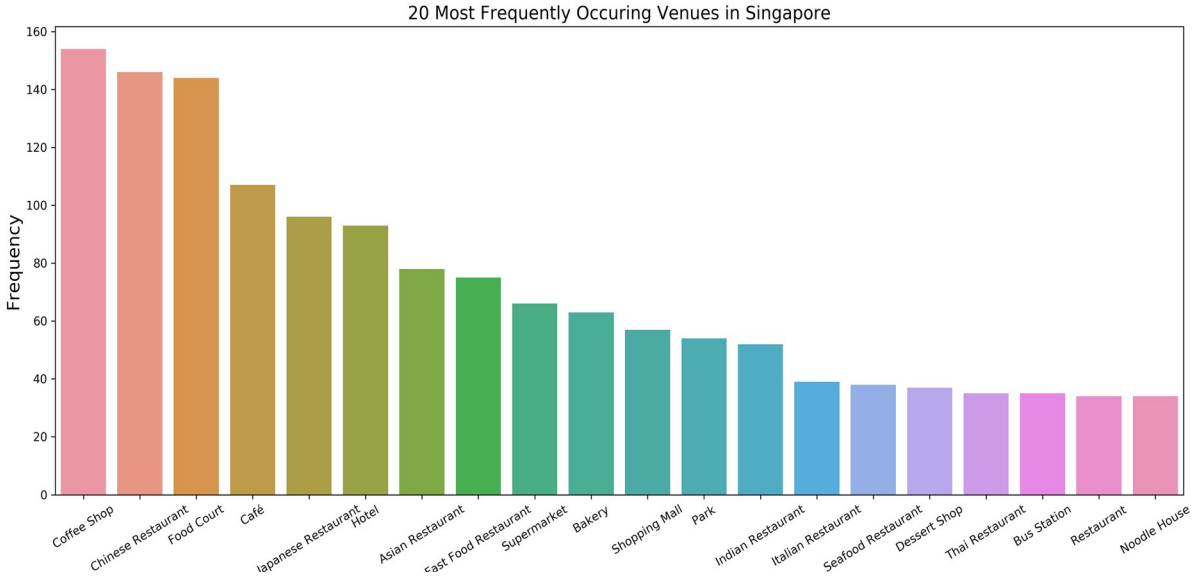
	Longitude	Latitude	Neighborhood	Region
0	103.793357	1.328117	Bukit Timah	Central Region
1	103.801664	1.376076	Central Water Catchment	North Region
2	103.748492	1.387486	Choa Chu Kang	West Region
3	104.049107	1.387936	North-Eastern Islands	North-East Region
4	103.725202	1.362108	Tengah	West Region
5	103.913796	1.406764	Punggol	North-East Region
6	103.698202	1.312923	Boon Lay	West Region
7	103.892283	1.256292	Marina East	Central Region
8	103.667422	1.306725	Pioneer	West Region
9	103.818933	1.457080	Sembawang	North Region
10	103.857281	1.444517	Simpang	North Region

3. METHODOLOGY

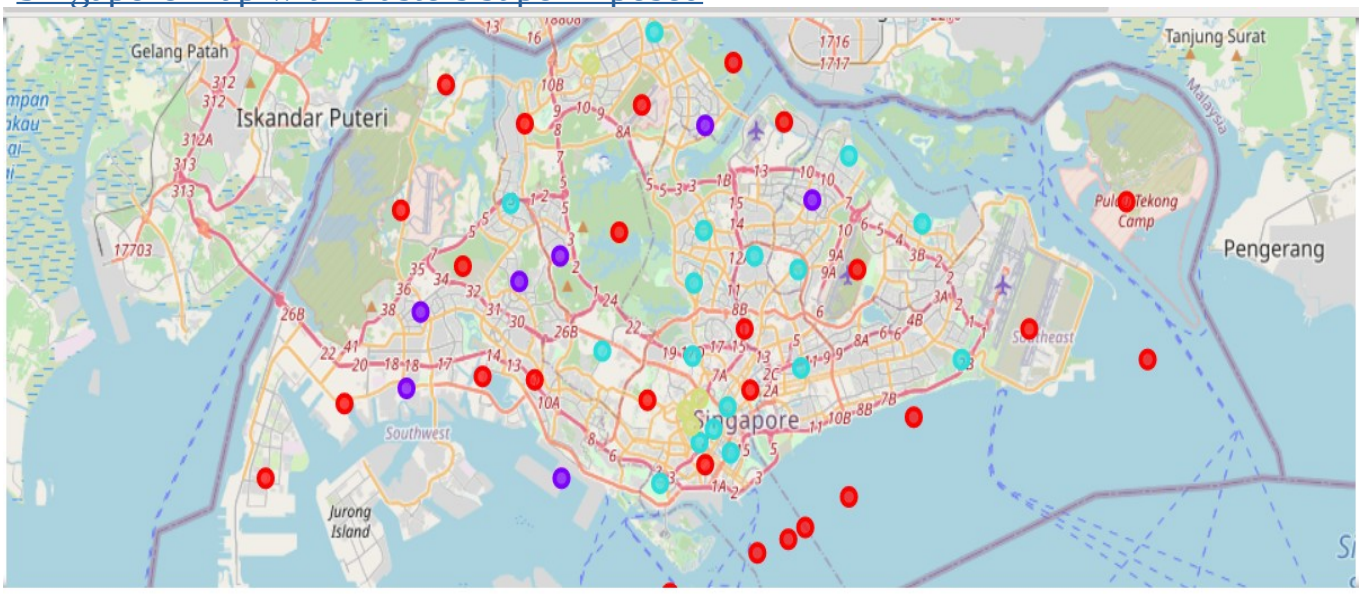
We will use Pandas to pre-process, cleanse the dataset, build a dataframe of the planning areas of Singapore, visualize the neighborhoods on Folium, work with Foursquare API to obtain the venue data in each planning area (neighborhood), then explore and cluster the neighborhoods, we will be using K-Means clustering method to perform. Finally we will pick up the best cluster to recommend to the stakeholders for the best locations to open up a new shopping mall.

Visualization of Singapore neighborhoods by using Folium map

Shopping mall is ranked 11th most popular venue category in Singapore



Singapore map with clusters superimposed



3. RESULTS AND DISCUSSION

First Cluster (Cluster Labels=0)

```
sg_merged.loc[sg_merged['Cluster Labels'] == 0]
```

	Neighborhood	Shopping Mall	Cluster Labels	Longitude	Latitude	Region
18	Kallang	0.0	0	103.865107	1.312327	Central Region
32	Pioneer	0.0	0	103.667422	1.306725	West Region
31	Paya Lebar	0.0	0	103.917381	1.360554	East Region
41	Simpang	0.0	0	103.857281	1.444517	North Region
29	Outram	0.0	0	103.843701	1.281626	Central Region
43	Southern Islands	0.0	0	103.826155	1.229460	Central Region
26	North-Eastern Islands	0.0	0	104.049107	1.387936	North-East Region
44	Straits View	0.0	0	103.868795	1.246133	Central Region
45	Sungei Kadut	0.0	0	103.755742	1.419650	North Region
23	Marine Parade	0.0	0	103.913258	1.268724	Central Region
22	Marina South	0.0	0	103.884283	1.251698	Central Region
21	Marina East	0.0	0	103.892283	1.256292	Central Region
20	Mandai	0.0	0	103.812815	1.427104	North Region
19	Lim Chu Kang	0.0	0	103.716885	1.435699	North Region
16	Jurong East	0.0	0	103.734817	1.317268	West Region
37	Seletar	0.0	0	103.881743	1.420722	North-East Region
47	Tanglin	0.0	0	103.815114	1.307610	Central Region
48	Tengah	0.0	0	103.725202	1.362108	West Region
12	Clementi	0.0	0	103.760437	1.316511	West Region
49	Toa Payoh	0.0	0	103.862479	1.336457	Central Region
10	Changi Bay	0.0	0	104.058846	1.324230	East Region
9	Changi	0.0	0	104.001643	1.337079	East Region
8	Central Water Catchment	0.0	0	103.801664	1.376076	North Region
50	Tuas	0.0	0	103.628906	1.276268	West Region
1	Bedok	0.0	0	103.944936	1.301108	East Region
51	Western Islands	0.0	0	103.727146	1.224018	West Region
52	Western Water Catchment	0.0	0	103.694919	1.384956	West Region

Second Cluster (Cluster Labels=1)

```
sg_merged.loc[sg_merged['Cluster Labels'] == 1]
```

	Neighborhood	Shopping Mall	Cluster Labels	Longitude	Latitude	Region
39	Sengkang	0.029851	1	103.895550	1.388638	North-East Region
34	Queenstown	0.043478	1	103.773753	1.276400	Central Region
3	Boon Lay	0.038462	1	103.698202	1.312923	West Region
6	Bukit Panjang	0.032258	1	103.773045	1.366217	West Region
17	Jurong West	0.030928	1	103.704815	1.343955	West Region
4	Bukit Batok	0.036145	1	103.752601	1.356029	West Region
54	Yishun	0.030000	1	103.843395	1.419031	North Region

Third Cluster (Cluster Labels=2)

```
sg_merged.loc[sg_merged['Cluster Labels'] == 2]
```

	Neighborhood	Shopping Mall	Cluster Labels	Longitude	Latitude	Region
42	Singapore River	0.020000	2	103.840498	1.290871	Central Region
46	Tampines	0.010417	2	103.968145	1.324290	East Region
40	Serangoon	0.010000	2	103.867606	1.366010	North-East Region
38	Sembawang	0.023256	2	103.818933	1.457080	North Region
0	Ang Mo Kio	0.010000	2	103.842565	1.376729	North-East Region
33	Punggol	0.012346	2	103.913796	1.406764	North-East Region
30	Pasir Ris	0.018868	2	103.949546	1.379571	East Region
24	Museum	0.020000	2	103.847505	1.295972	Central Region
15	Hougang	0.010000	2	103.888800	1.360946	North-East Region
14	Geylang	0.010000	2	103.889852	1.321163	Central Region
13	Downtown Core	0.010000	2	103.856088	1.286587	Central Region
11	Choa Chu Kang	0.012048	2	103.748492	1.387486	West Region
7	Bukit Timah	0.022472	2	103.793357	1.328117	Central Region
5	Bukit Merah	0.010000	2	103.821630	1.274457	Central Region
2	Bishan	0.010000	2	103.837734	1.355160	Central Region
36	Rochor	0.010000	2	103.854284	1.304995	Central Region
27	Novena	0.010000	2	103.837228	1.326091	Central Region

Fourth Cluster (Cluster Labels=3)

```
sg_merged.loc[sg_merged['Cluster Labels'] == 3]
```

	Neighborhood	Shopping Mall	Cluster Labels	Longitude	Latitude	Region
35	River Valley	0.07	3	103.836371	1.297884	Central Region
28	Orchard	0.06	3	103.834065	1.304012	Central Region
53	Woodlands	0.05	3	103.787925	1.443618	North Region
25	Newton	0.05	3	103.840985	1.308506	Central Region

1. First cluster (Cluster label 0) has no shopping malls, as we can see, the neighborhoods are mostly in suburb areas and distant islands.

2. Second cluster (Cluster label 1) has an above-average number of shopping malls, mostly in West Region.

3. Third cluster (Cluster label 2) shows a limited amount of shopping malls, located mostly in Central Region.

4. Fourth cluster (Cluster label 3) has the highest amount of shopping malls. It's easy to identify most places are in central region.

4. CONCLUSION

The neighborhoods in the second cluster such as Bukit Batok, Jurong West, Queenstown would be the most ideal location for opening a new shopping mall. The areas in the third cluster could be considered too. Those areas have a large number of HDB flats and private condos. However, admittedly, there are certain limits in this analysis, as we don’t have sufficient data to make a more accurate prediction, such as population density and rental prices in each planning area.