Venturing into Coffee Shop Businesses & Venues Data Analysis in Singapore

Submitted by:
Lua Sze Chea

9 September 2020

1 INTRODUCTION

1.1 Background

Like any other nation, Singapore certainly does not lack any of its vibrant coffee scene that is Uniquely Singapore. Fueled by the influx of immigrants in the 19th century, locals open coffee shops to satisfy the growing European working population. Later on, coffee shop owners made an equally delicious brew to the local population by using flannel "socks" to filter the coffee and finishing off with sugar and evaporated milk. The traditional way of coffee making is pocket friendly and among the favorites with locals today.

1.2 Business Problem

The coffee culture scene changes along with time with many new cafes - independent and chain operator marking their ground in all parts of Singapore. This bring in new challenges and competition in the coffee making industry. In this project, we will analyse the coffee shops currently available in the neighborhoods, identify the highest number of coffee shops in the neighborhood and recommend any potential neighborhood to start a coffee shop business.

1.3 Target Audience

- Business owners that are interested in opening a coffee shop
- Business analyst that are curious about the coffee shop scene in Singapore

2 DATA PREPARATION

The data was scraped from the Wikipedia page that features the neighborhood, region, population, area and density of Singapore and read it into a pandas data frame. I cleaned the table and kept only the neighborhood and region column. I used Geopy client to get the coordinates of each neighborhood and created the final frame.

```
df=pd.read_html('https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore')[2]
df.head()
```

	Name (English)	Malay	Chinese	Pinyin	Tamil	Region	Area (km2)	Population[7]	Density (/km2)
0	Ang Mo Kio	NaN	宏茂桥	Hóng mào qiáo	ஆங் மோ கியோ	North-East	13.94	163950	13400
1	Bedok	*	勿洛	Wù luò	பிடோக்	East	21.69	279380	13000
2	Bishan	NaN	碧山	Bì shān	பீஷான்	Central	7.62	88010	12000
3	Boon Lay	NaN	文礼	Wén lǐ	பூன் லே	West	8.23	30	3.6
4	Bukit Batok	*	武吉巴督	Wǔjí bā dū	புக்கிட் பாத்தோக்	West	11.13	153740	14000

	Town	Region	Latitude	Longitude
0	Ang Mo Kio	North-East	1.370080	103.849523
1	Bedok	East	1.323976	103.930216
2	Bishan	Central	1.350986	103.848255
3	Boon Lay	West	1.338550	103.705812
4	Bukit Batok	West	1.349057	103.749591
6	Bukit Panjang	West	1.379149	103.761413
7	Bukit Timah	Central	1.354690	103.776372
8	Central Water Catchment	North	1.375708	103.801743
10	Changi Bay	East	1.316850	104.020649

Using a folium map, I noticed some of the markers are pointed outside Singapore due to the incorrect coordinates provided by the geolocator. As this may affect the venues API result in the later part of the analysis, I removed the affected neighborhoods and this result a total of 40 neighborhoods in the final frame.
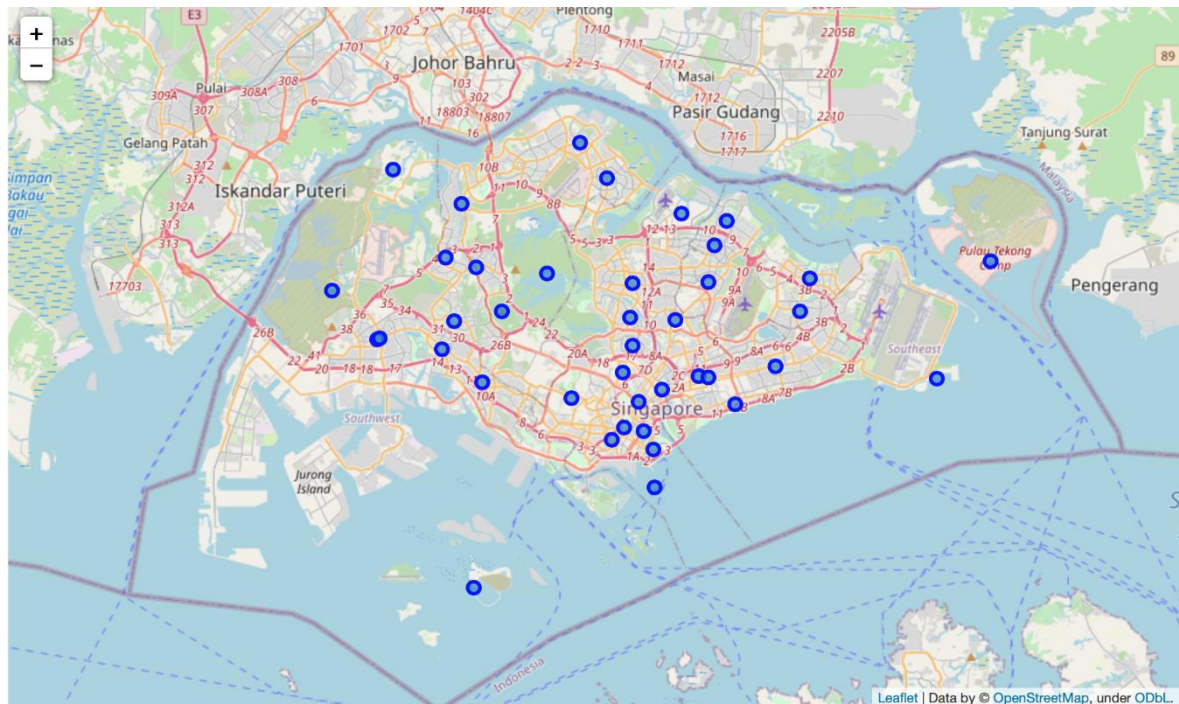
I group the neighborhood by their region and plot a bar chart. The result shows that central and west region has more neighborhoods within their boundary and might form most of the analysis in the later part of the project.


Number of Towns by Region
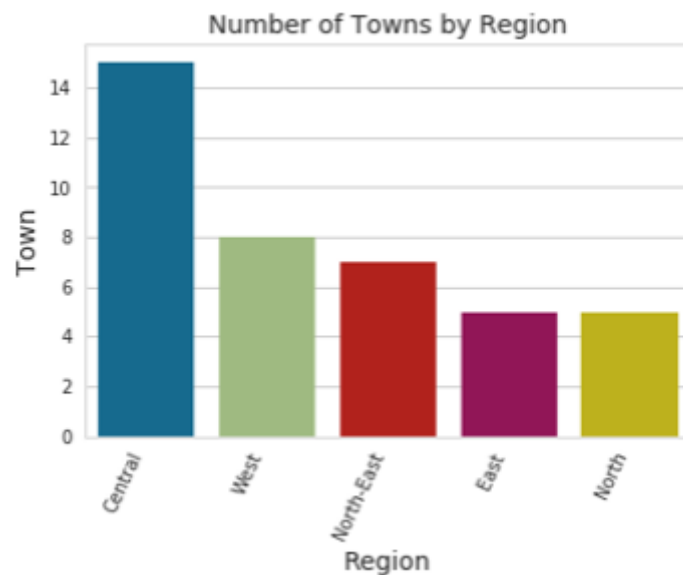
3 DATA EXPLORATION

Using the foursquare API, I retrieved the data limiting to 100 venues and within 500 meter. The API return 1,407 results, featuring the venues and categories of the places within the criteria parameter. I filter the results further to focus only on the 80 coffee shops returned from the API and to further explore all of them in the neighborhood.

There are 26 neighborhoods with coffee shops. I plot the chart to feature the top 10 most number of coffee shops in the neighborhoods. Toa Payoh and Novena top the list, where both are from the central region of Singapore.

Top 10 Most Number of Coffee Shops by Town
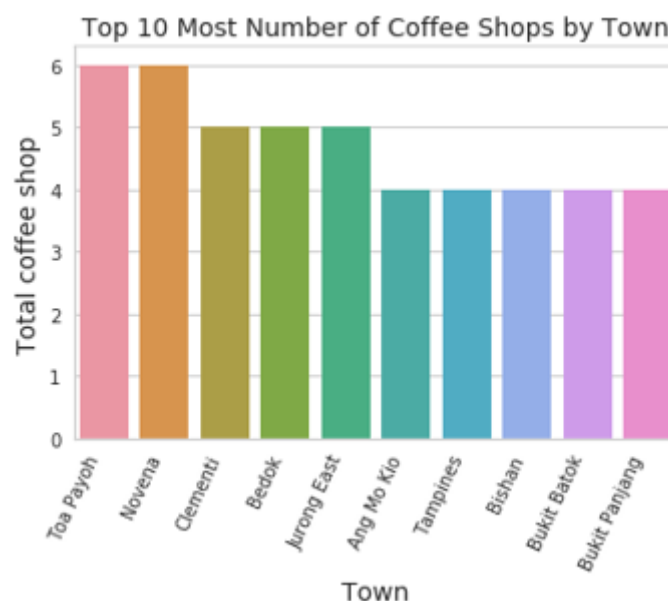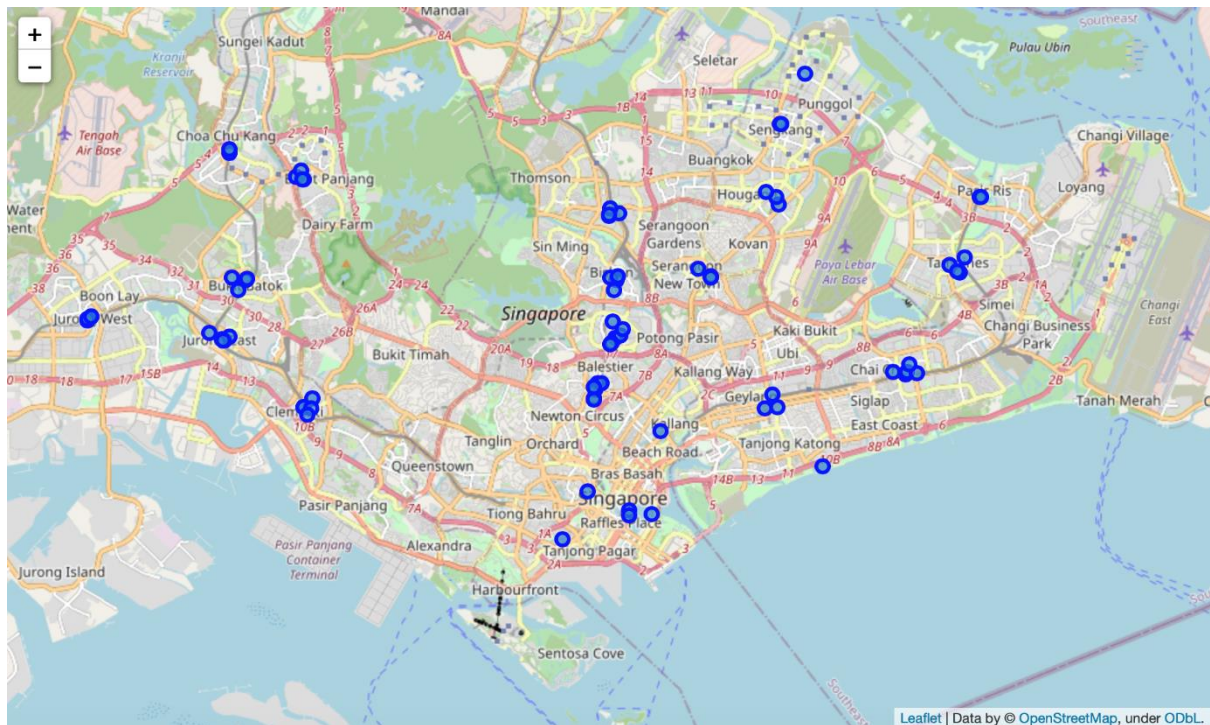
An illustration showing the density of the coffee shops clustered within the neighborhood

With the venue places return from the FourSquare API, I analyze and calculate the mean of the frequency of the venue category in each neighborhood.

|	Zoo Exhibit	Accessories Store	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	...	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

5 rows × 205 columns

```
#group rows by neighborhood and by taking the mean of the frequency of occurrence of each category
coffee_grouped = coffee_onehot.groupby('Neighborhood').mean().reset_index()
coffee_grouped
```

	Neighborhood	Zoo Exhibit	Accessories Store	African Restaurant	American Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	...	Trail	Train Station	Vegetarian / Vegan Restaurant	
0	Ang Mo Kio	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.024390	...	0.000000	0.000000	0.000000	0.0
1	Bedok	0.0	0.000000	0.000000	0.017544	0.000000	0.000000	0.00	0.000000	0.052632	...	0.000000	0.000000	0.017544	0.0
2	Bishan	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.046512	...	0.000000	0.000000	0.000000	0.0
3	Boon Lay	0.0	0.000000	0.000000	0.014286	0.000000	0.000000	0.00	0.000000	0.085714	...	0.000000	0.000000	0.000000	0.0
4	Bukit Batok	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.0
5	Bukit Panjang	0.0	0.000000	0.000000	0.025000	0.000000	0.000000	0.00	0.000000	0.050000	...	0.000000	0.000000	0.000000	0.0
6	Bukit Timah	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.583333	0.000000	0.000000	0.0
7	Changi Bay	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.0
8	Choa Chu Kang	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.045455	...	0.000000	0.000000	0.000000	0.0

As I am only interested in the coffee shops, I filter the data frame to show only the neighborhood and its corresponding coffee shop result.
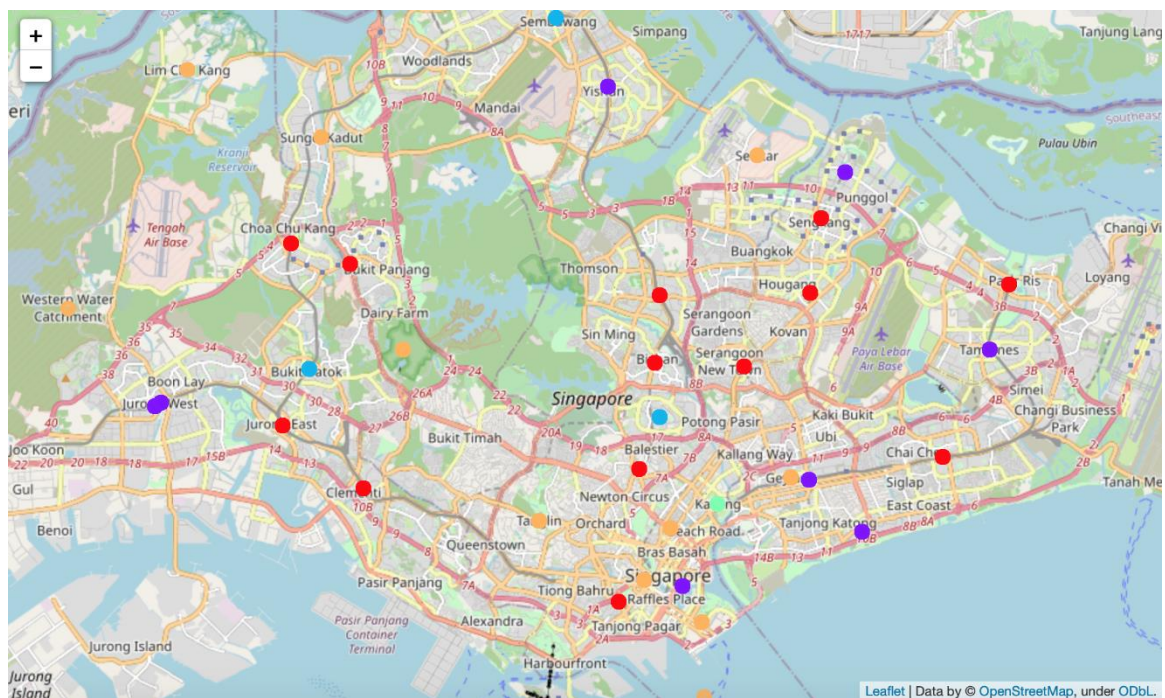
	Neighborhood	Coffee Shop
0	Ang Mo Kio	0.097561
1	Bedok	0.087719
2	Bishan	0.093023
3	Boon Lay	0.042857
4	Bukit Batok	0.166667
5	Bukit Panjang	0.100000
6	Bukit Timah	0.000000
7	Changi Bay	0.000000
8	Choa Chu Kang	0.090909
9	Clementi	0.084746
10	Downtown Core	0.030000
11	Geylang	0.000000
12	Hougang	0.076923
13	Jurong East	0.068493
14	Jurong West	0.046875
15	Kallang	0.250000
16	Lim Chu Kang	0.000000
17	Marina South	0.000000

Now, I cluster the coffee shops in a number of groups using K-Means, an unsupervised method to group each data point based on their feature similarities. I used the elbow method to find the best "K". From the result, we can tell that the best K is 5.

Distortion Score Elbow for KMeans Clustering
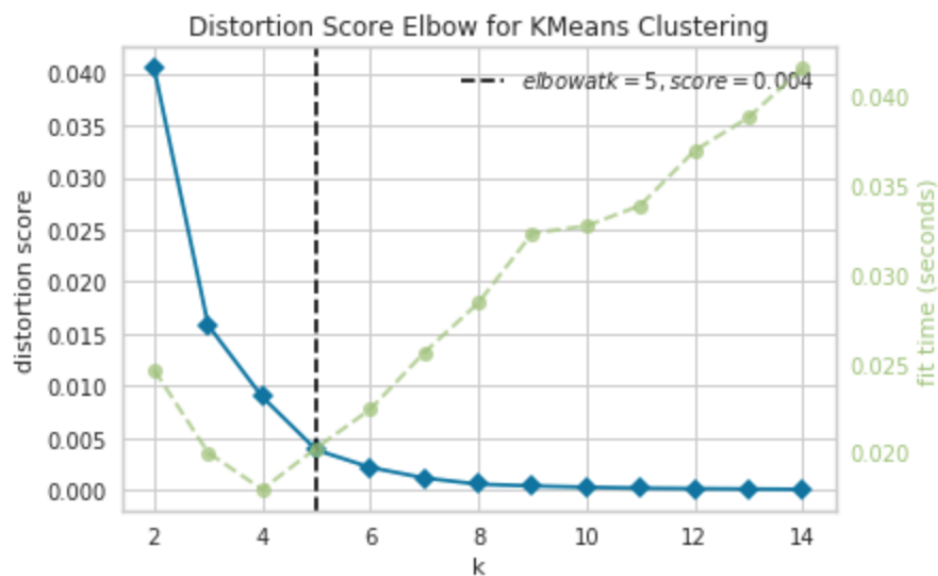
I created a new data frame that merge the cluster labels with the geographical locations and create a folium map to visualize the 5 clusters as follow:

	Neighborhood	Coffee Shop	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ang Mo Kio	0.097561	0	1.370080	103.849523	Old Chang Kee	1.369094	103.848389	Snack Place
27	Sengkang	0.064516	0	1.391924	103.895491	Sengkang Sculpture Park	1.394549	103.896630	Sculpture Garden
27	Sengkang	0.064516	0	1.391924	103.895491	Cold Storage	1.392342	103.895037	Supermarket
27	Sengkang	0.064516	0	1.391924	103.895491	LeNu 樂牛	1.392415	103.895052	Noodle House
27	Sengkang	0.064516	0	1.391924	103.895491	Châteraisé	1.392245	103.895079	Bakery
27	Sengkang	0.064516	0	1.391924	103.895491	Ya Kun Family Cafe	1.391903	103.894913	Café
27	Sengkang	0.064516	0	1.391924	103.895491	Starbucks	1.392367	103.895018	Coffee Shop
27	Sengkang	0.064516	0	1.391924	103.895491	Maki-san	1.392311	103.894969	Sushi Restaurant
27	Sengkang	0.064516	0	1.391924	103.895491	Canton Paradise	1.392299	103.895128	Chinese Restaurant
12	Hougang	0.076923	0	1.370801	103.892544	Phoon Huat & Co	1.371222	103.894695	Kitchen Supply Store

4 RESULTS AND DISCUSSION

Cluster 1 and 3 shows a good mix of local independent and multi-chain coffee shops within the same neighborhood.

Cluster 2 is comparable, but has fewer coffee shops within the same neighborhood and lesser traditional and local coffee shops.

- Other than coffee shops, the surrounding area are competitive with many restaurants, food court and café available at any corner. The neighborhoods are mostly mature and commercialize.
- Starbucks outlet is available in each neighborhood. It serves westernize food and beverages. Ya Kun Kaya Toast tops the list next and serves local traditional beverages and toast. Both are coffeeshop chains.

Cluster 4 only comprises one neighborhood and the reason I could derive is due to having a unique venue category – hostel, which is not a common category in other neighborhoods. It could also be due to the fact that the neighborhood does not have a Starbucks or a Ya Kun Kaya Toast like the rest. As such, it form its own cluster.

Cluster 5 groups all the neighborhoods without a coffee shop in the vicinity. However, there are other restaurants and cafes. These group of neighborhoods are potential areas to start a coffee shop business and worth to analyze further, to survey the ground and its demographics.

5 CONCLUSION

Although some neighborhoods are already saturated with coffee shops, it does not mean zero potential to start a coffee shop business. There are other perspective to take into account such as the business model, the target audience, the objective and mission of the business that is unique, appealing and stands out differently from its competitors.

Similarly, to venture a coffee shop business in a neighborhood without a coffee shop also require market research and on the ground walk to evaluate the pros and cons.

This project can be improved with the correction of the coordinates from the geolocator. Further studies can be continued by incorporating population in the neighborhood or the popularity of the competitive coffee shops nearby.

Ultimately, this project only provides a fundamental analysis of the current coffee business scene in Singapore. A deeper analysis may also be required before business analyst or business owners take their own judgement and decisions further.