The Battle of Neighborhoods (Week 2)

Venturing into Coffee Shop Businesses & Venues Data Analysis in Singapore

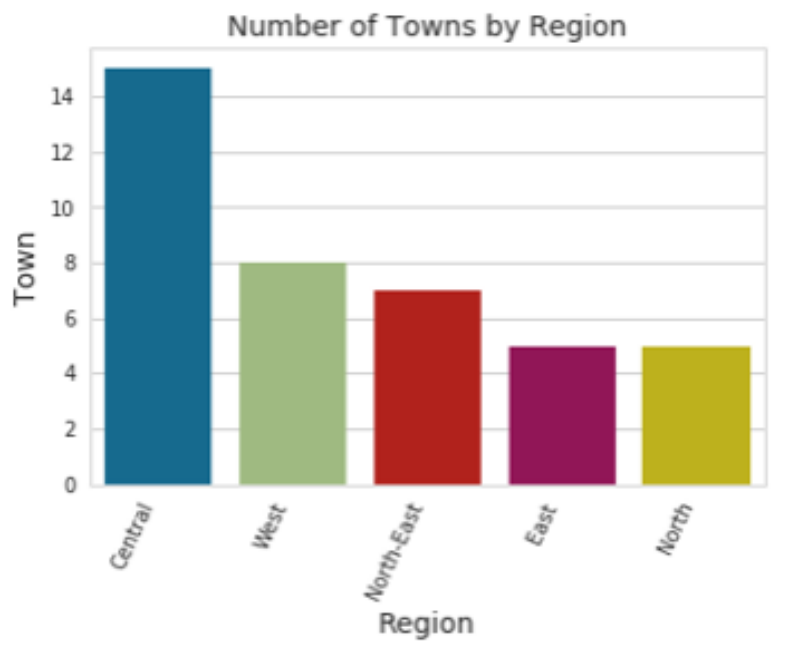
COFFEE SHOP ANALYSIS & RECOMMENDATION

- There are many 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 analyze the following:
  - Coffee shops currently available in the neighborhoods
  - Identify the highest number of coffee shops in the neighborhood
  - Recommend any potential neighborhood to start a coffee shop business.
- Who will be interested in this project?
  - Business owners that are interested in opening a coffee shop
  - Business analyst that are curious about the coffee shop scene in Singapore

DATA PREPARATION & CLEANING

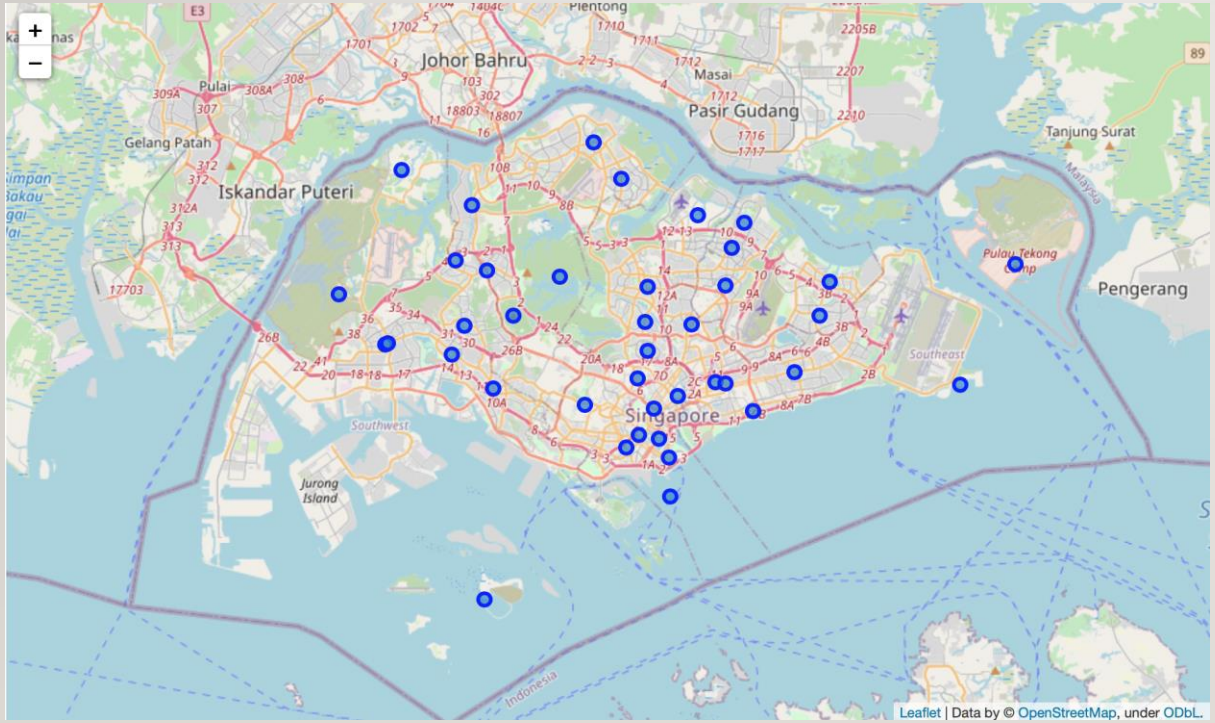
- Scrape the name (Neighborhood) and Region from Wikipedia page: https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore
- Total of 54 rows and 7 features in the raw dataset
- Incorrect coordinates of the neighborhoods were dropped from the dataset
- Cleaned data contains 40 rows and 4 features

NO. OF REGIONS & NEIGHBORHOODS



- Neighborhoods were grouped accordingly to their region
- Central and west region might form most of the analysis in the later part of the project.
- Total 40 neighborhoods – Central (15), West (8), North-East (7), East(5), North (5)

FOLIUM MAP



- Using the folium package, we created the map with 40 neighborhoods with their latitude and longitude

EXPLORING COFFEE SHOPS DATA (1)

- Using foursquare API to retrieve the data limiting to 100 venues and 500 meter.
- 80 coffee shops returned from the API

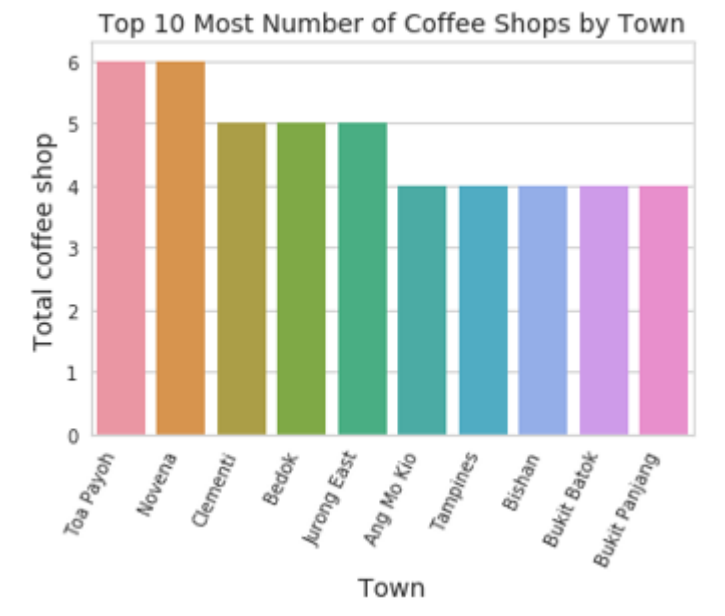
```
print ("Venue Category: Coffee Shop: ", coffee_shop.shape)
coffee_shop.head(10)
```

```
Venue Category: Coffee Shop:  (80, 7)
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Ang Mo Kio	1.370080	103.849523	ST31 Coffee Shop	1.367478	103.848334	Coffee Shop
2	Ang Mo Kio	1.370080	103.849523	Coffee Shop, Block 422, Ang Mo Kio Ave 3	1.368119	103.851093	Coffee Shop
3	Ang Mo Kio	1.370080	103.849523	Starbucks	1.369230	103.848683	Coffee Shop
4	Ang Mo Kio	1.370080	103.849523	AMK 347 Food House	1.367602	103.848372	Coffee Shop
5	Bedok	1.323976	103.930216	Ya Kun Kaya Toast 亞坤	1.324095	103.929198	Coffee Shop
6	Bedok	1.323976	103.930216	Starbucks	1.324760	103.929644	Coffee Shop
7	Bedok	1.323976	103.930216	Starbucks	1.324568	103.932518	Coffee Shop
8	Bedok	1.323976	103.930216	Mr Teh Tarik Eating House	1.326869	103.930212	Coffee Shop
9	Bedok	1.323976	103.930216	金福海鲜菜馆 @ Chai Chee Ave	1.324798	103.925924	Coffee Shop
10	Bishan	1.350986	103.848255	Starbucks	1.349849	103.850415	Coffee Shop

EXPLORING COFFEE SHOPS DATA (2)

- From the resulting bar chart, we display the top 10 neighborhood with the most coffee shops. Toa Payoh and Novena (Central region) tops the chart with 6 coffee shops.



1. Toa Payoh (Central Region)
2. Novena (Central Region)
3. Clementi (West Region)
4. Bedok (East Region)
5. Jurong East (West Region)
6. Ang Mo Kio (Central Region)
7. Tampines (East Region)
8. Bishan (Central Region)
9. Bukit Batok (West Region)
10. Bukit Panjang (West Region)

EXPLORING COFFEE SHOPS DATA (3)

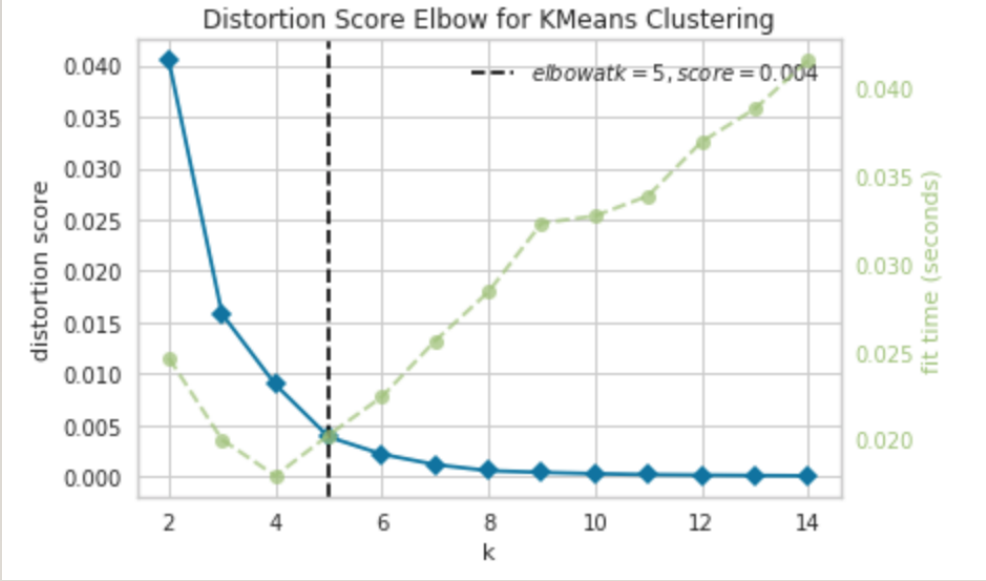
	Neighborhood	Coffee Shop
0	Ang Mo Kio	0.097561
1	Bedok	0.087719
2	Bishan	0.093023
3	Boon Lay	0.042857
4	Bukit Batok	0.166667
5	Bukit Panjang	0.100000
6	Bukit Timah	0.000000
7	Changi Bay	0.000000
8	Choa Chu Kang	0.090909
9	Clementi	0.084746
10	Downtown Core	0.030000
11	Geylang	0.000000
12	Hougang	0.076923
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

Based on the results, we can tell which neighborhood has coffee shops in the nearby area

And which neighborhood has zero presence for coffee shops in the nearby area

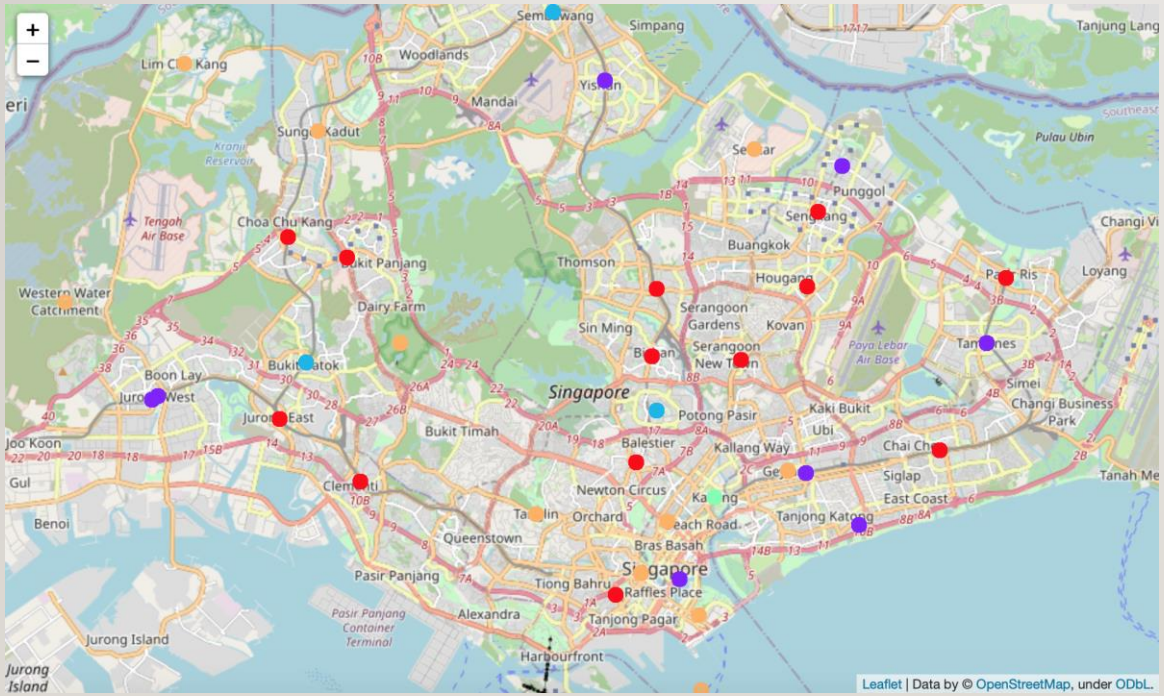
K-MEANS CLUSTERING

- K-Means is an unsupervised method to group each data point based on their feature similarities. We use the elbow method to find the best “K”.
- From the result, we can tell that the best K is 5



FOLIUM MAP - CLUSTER

- We create a folium map to visualize the 5 clusters as follow



RESULTS (1)

- Cluster 1 and Cluster 2 – Top 2 clusters with the most coffee shops

551 rows × 10 columns

```
#identify the number of occurence for each type of venue category
cluster1['Venue Category'].value_counts()[0:10]
```

```
Coffee Shop              46
Food Court               27
Chinese Restaurant       26
Fast Food Restaurant     23
Asian Restaurant         21
Café                     20
Japanese Restaurant      20
Supermarket              16
Sandwich Place           16
Shopping Mall            16
Name: Venue Category, dtype: int64
```

530 rows × 10 columns

```
#identify the number of occurence for each type of venue category
cluster2['Venue Category'].value_counts()[0:10]
```

```
Japanese Restaurant      28
Chinese Restaurant       24
Fast Food Restaurant     24
Asian Restaurant         21
Coffee Shop              20
Café                     19
Food Court               16
Bakery                   15
Shopping Mall            14
Supermarket              12
Name: Venue Category, dtype: int64
```

**Cluster 3 has 12 coffee shops*

RESULTS (2)

- Cluster 4 and 5 – Bottom 2 clusters with the least numbers of coffee shops

```
#identify the number of occurence for each type of venue category
cluster4['Venue Category'].value_counts()[0:10]
```

```
Coffee Shop    1
Hostel         1
Restaurant     1
Supermarket    1
Name: Venue Category, dtype: int64
```

*Cluster 5 only has 1 coffee shop and is not displayed in the top 10 frequent occurrence of venue categories

232 rows × 10 columns

```
#identify the number of occurence for each type of venue category
cluster5['Venue Category'].value_counts()[0:10]
```

```
Indian Restaurant                11
Café                             10
Vegetarian / Vegan Restaurant     9
Chinese Restaurant                9
Trail                             9
Japanese Restaurant               9
Hotel                             8
Food Court                        8
Nightclub                         7
Noodle House                      6
Name: Venue Category, dtype: int64
```

DISCUSSION: CLUSTER 1 - 3

- Cluster 1 and 3 shows a good mix of local independent and multi-chain coffee shops
- Cluster 2 is comparable, but has fewer coffee shops and local coffee shops.
- Competitive area with many food and beverages businesses
- Starbucks outlet is available in each neighborhood. Ya Kun Kaya Toast comes after next and both are coffeeshop chains.

DISCUSSION: CLUSTER 4 & 5

- Cluster 4 only comprises one neighborhood and could be due to the following:
  - Unique venue category – hostel, which is not a common category in other neighborhoods
  - Absence of Starbucks or a Ya Kun Kaya Toast. As such, it forms its own cluster
- Cluster 5 groups all the neighborhoods without a coffee shop in the vicinity. However, there are other restaurants and cafes.
  - Potential areas to start a coffee shop business
  - Further analysis can be conducted to survey the ground and its demographics

CONCLUSION (1)

- 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.

CONCLUSION (2)

- This project can be improved by correcting 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 be required before business analyst or business owners take their own judgement and decisions further.