Analysis of Neighborhoods in Penang (Pulau Pinang) state of Malaysia for Starting A Food Business

Capstone Project – The Battle of Neighborhoods (Week 2):

1. **Introduction**
   The aim of this project is to explore cities in Penang state and it's common venues. By analysing the common venues around cities, it can show us the popular type of business.

2. **Business Problem**
   Penang (also known as Pulau Pinang) is famous tourist state in Malaysia. I frequently visit Penang state during long holiday, one of our intention is to have good food. This let me to have a thought about food business in around the city of Penang. I decided to make a clustering analysis on most common venues around different city in Penang where most common venues attract visitors and that will be the source of business for food.

3. **Target Audience**
   In the future, this approach could be used as a service, helping food business entrepreneur to understand which category of food business is preferred in each city of Penang.

4. **Data**
   The following data is required for this project.

   4.1. **Neigborhood Data**
   Neighborhood data such as City and State can be obtain from https://www.weatherdatasource.com/MY-Pulau_Pinang website.

	City	State
0	Bayan Lepas	Penang
1	Bukit Mertajam	Penang
2	Butterworth	Penang
3	Bandaraya George Town	Penang
4	Juru	Penang
5	Kepala Batas	Penang
6	Nibong Tebal	Penang
7	Perai	Penang
9	Sungai Ara	Penang
10	Tanjung Tokong	Penang
11	Tasek Gelugor	Penang

   4.2. **Geographical Coordinates**
   Geographical coordinate of above city list is extracted via Geopy library of Python.

	City	State	Latitude	Longitude
0	Bayan Lepas	Penang	5.29513	100.26
1	Bukit Mertajam	Penang	5.36424	100.461
2	Butterworth	Penang	5.39349	100.366
3	Bandaraya George Town	Penang	5.41457	100.33
4	Juru	Penang	5.31572	100.438
5	Kepala Batas	Penang	5.51438	100.436
6	Nibong Tebal	Penang	5.1701	100.479
7	Perai	Penang	5.38708	100.382
8	Sungai Ara	Penang	5.32088	100.268
9	Tanjung Tokong	Penang	5.45111	100.305
10	Tasek Gelugor	Penang	5.48444	100.49
11	Permatang Kuching	Penang	5.46339	100.381

The coordinate of city is plotted on the map of Penang with Folium library of Python.
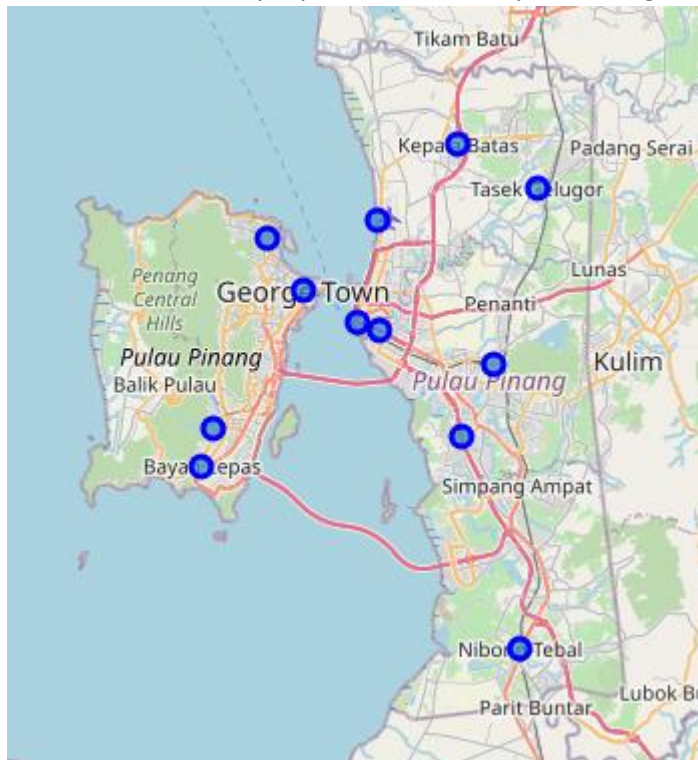
4.3. Venue Data

By using FourSquare API, with "explore" query type, list of 500 common venues within radius of 1KM for each city is extracted. The list contain information such as Venue Category which is useful for business type analysis. Partial screen capture data can be seen from following image.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bayan Lepas	5.29513	100.26	Nasi Dalca Briyani	5.296738	100.259892	Malay Restaurant
1	Bayan Lepas	5.29513	100.26	Bawal Goreng Pokok Cheri	5.299255	100.262345	Malay Restaurant
2	Bayan Lepas	5.29513	100.26	Cargas Cafe	5.294794	100.258380	Malay Restaurant
3	Bayan Lepas	5.29513	100.26	Pasar Ramadhan Bayan Lepas	5.295330	100.262206	Food Truck
4	Bayan Lepas	5.29513	100.26	Fizzy Cafe Mee Udang Ketam	5.296352	100.259004	Asian Restaurant

5. Methodology

5.1. Feature extraction

With FourSquare API, "feature" (business that is rated by users) of each neigborhoods is downloaded into JSON file. A function is created to loop through list of data on JSON file and result assign into new list as following.

	Neighborhood	Accessories Store	Airport	Airport Lounge	Airport Terminal	American Restaurant	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bank	Bar	Basketball Court	Bed & Breakfas
0	Bayan Lepas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Bayan Lepas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Bayan Lepas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Bayan Lepas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Bayan Lepas	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Feature is normalize by using mean value.

	Neighborhood	Accessories Store	Airport	Airport Lounge	Airport Terminal	American Restaurant	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bank	Bar	Basketball Court	B B
0	Bandaraya George Town	0.000000	0.000000	0.000000	0.000000	0.0	0.01	0.01	0.000000	0.0	0.0	0.07	0.0	0.01	0.0	0.
1	Bayan Lepas	0.021277	0.021277	0.021277	0.021277	0.0	0.00	0.00	0.085106	0.0	0.0	0.00	0.0	0.00	0.0	0.
2	Bukit Mertajam	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.00	0.066667	0.0	0.0	0.00	0.0	0.00	0.0	0.
3	Butterworth	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.00	0.093750	0.0	0.0	0.00	0.0	0.00	0.0	0.
4	Juru	0.000000	0.000000	0.000000	0.000000	0.0	0.00	0.00	0.142857	0.0	0.0	0.00	0.0	0.00	0.0	0.

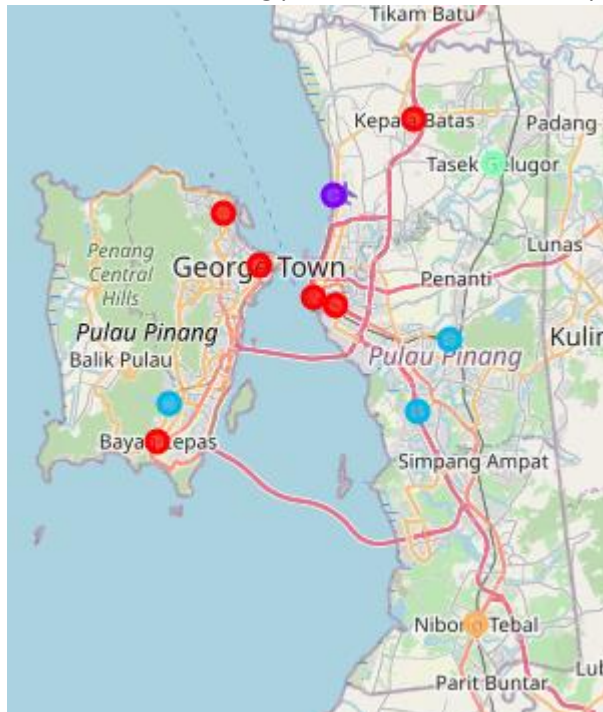With normalized data, only top 20 highest rating venues will be considered.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12t Co Ver
0	Bandaraya George Town	Café	Hotel	Chinese Restaurant	Vegetarian / Vegan Restaurant	Bakery	Food Truck	Dessert Shop	Bed & Breakfast	Halal Restaurant	Wine Bar	Hostel	Ind Res
1	Bayan Lepas	Café	Malay Restaurant	Asian Restaurant	Coffee Shop	Convenience Store	Pharmacy	Food Truck	Indian Restaurant	Thai Restaurant	Pastry Shop	Outdoors & Recreation	Noc Hou
2	Bukit Mertajam	Chinese Restaurant	Malay Restaurant	Food Court	Food Truck	Coffee Shop	Asian Restaurant	Electronics Store	Pizza Place	Market	Department Store	Café	Bus Sta
3	Butterworth	Asian Restaurant	Harbor / Marina	Malay Restaurant	Coffee Shop	Thai Restaurant	Electronics Store	Burger Joint	Business Service	Platform	Pier	Convenience Store	Hoc Bar
4	Juru	Malay Restaurant	Chinese Restaurant	Asian Restaurant	Food Truck	Breakfast Spot	Shipping Store	Hotel	Comfort Food Restaurant	Coffee Shop	Food Court	Food	Par

5.2. Unsuperivised Machine Learning via K-Means

K-Means is applied to normalized data set and it cluster the dataset into 5 clusters

	Neighborhood	State	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Mo Comm Venue
0	Bayan Lepas	Penang	5.29513	100.26	0	Café	Malay Restaurant	Asian Restaurant	Coffee Shop	Convenience Store	Pharmacy	Food Truck	Indian Restaurant	Thai Restau
1	Bukit Mertajam	Penang	5.36424	100.461	2	Chinese Restaurant	Malay Restaurant	Food Court	Food Truck	Coffee Shop	Asian Restaurant	Electronics Store	Pizza Place	Market
2	Butterworth	Penang	5.39349	100.366	0	Asian Restaurant	Harbor / Marina	Malay Restaurant	Coffee Shop	Thai Restaurant	Electronics Store	Burger Joint	Business Service	Platfor
3	Bandaraya George Town	Penang	5.41457	100.33	0	Café	Hotel	Chinese Restaurant	Vegetarian / Vegan Restaurant	Bakery	Food Truck	Dessert Shop	Bed & Breakfast	Halal Restau
4	Juru	Penang	5.31572	100.438	2	Malay Restaurant	Chinese Restaurant	Asian Restaurant	Food Truck	Breakfast Spot	Shipping Store	Hotel	Comfort Food Restaurant	Coffee Shop

The cluster then being plotted with the same map.

6. Results

First cluster

	State	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
0	Penang	Bayan Lepas	Café	Malay Restaurant	Asian Restaurant	Coffee Shop	Convenience Store	Pharmacy	Food Truck	Indian Restaurant	Thai Restaurant	Pastry Shop	Outdoors Recreatic
2	Penang	Butterworth	Asian Restaurant	Harbor / Marina	Malay Restaurant	Coffee Shop	Thai Restaurant	Electronics Store	Burger Joint	Business Service	Platform	Pier	Convenie Store
3	Penang	Bandaraya George Town	Café	Hotel	Chinese Restaurant	Vegetarian / Vegan Restaurant	Bakery	Food Truck	Dessert Shop	Bed & Breakfast	Halal Restaurant	Wine Bar	Hostel
5	Penang	Kepala Batas	Asian Restaurant	Mobile Phone Shop	Breakfast Spot	Indian Restaurant	Burger Joint	Café	Boutique	Convenience Store	Malay Restaurant	Chinese Restaurant	Fast Food Restaura
7	Penang	Perai	Malay Restaurant	Diner	Coffee Shop	Asian Restaurant	Chinese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Furniture / Home Store	Food Truck	Food Court	Fast Food Restaura
9	Penang	Tanjung Tokong	Japanese Restaurant	Coffee Shop	Korean Restaurant	Café	Malay Restaurant	Chinese Restaurant	Asian Restaurant	Food Truck	Convenience Store	Fast Food Restaurant	Departme Store

Second cluster

	State	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue	
11	Penang	Permatang Kuching	Golf Course	Malay Restaurant	Diner	Asian Restaurant	Water Park	Soccer Field	Night Market	Breakfast Spot	Flea Market	Fast Food Restaurant	Farmers Market	Yoga Studio	

Third cluster

	State	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
1	Penang	Bukit Mertajam	Chinese Restaurant	Malay Restaurant	Food Court	Food Truck	Coffee Shop	Asian Restaurant	Electronics Store	Pizza Place	Market	Department Store	Café
4	Penang	Juru	Malay Restaurant	Chinese Restaurant	Asian Restaurant	Food Truck	Breakfast Spot	Shipping Store	Hotel	Comfort Food Restaurant	Coffee Shop	Food Court	Food
8	Penang	Sungai Ara	Malay Restaurant	Asian Restaurant	Coffee Shop	Chinese Restaurant	Convenience Store	Seafood Restaurant	Bakery	Farmers Market	Restaurant	Resort	Yoga Studio

Fourth cluster

	State	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
10	Penang	Tasek Gelugor	Asian Restaurant	Flea Market	Department Store	Pharmacy	Restaurant	Bistro	Food Truck	Basketball Court	Malay Restaurant	Clothing Store	Train Station

Fifth cluster

	State	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue
6	Penang	Nibong Tebal	Noodle House	Seafood Restaurant	Food Court	Café	Chinese Restaurant	Coffee Shop	Bakery	Asian Restaurant	Fast Food Restaurant	Restaurant	Sandwich Place	Flea Market

7. Discussion

The analysis is showing that restaurant are the most common business in each city. Cluster 0 represent the major city in Penang, they are more 'famous' compared to other city. This could be a good spot for food truck business. Cluster 1 is near to military based and it is not really common venues to visit by tourist. Cluster 2 and 3 is considered as outskirt and not quite a tourist hotspot where local restaurant is among the top venues where local frequently visit. Cluster 4 is far from center of Penang and it is close to fishing village, hence seafood restaurant is famous in that area.

8. Conclusion

In this project, neighborhood of Penang had been successfully analyse. It provide an idea on what business is popular in the above mentioned city. As tourist state, those venues will be heart of attraction and with crowd of tourist, food will be one of the necessity and it present opportunity opening restaurant. Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every city zone.

9. Link Jupyter Notebook in Github

https://github.com/edmundcham/Capstone-Project---The-Battle-of-Neighborhoods-Week-1-/tree/master