

Kuala Lumpur and Johor Bahru Sightseeing

Mostafa Adel

8, June 2019

1. Introduction

1.1 Background

Malaysia is a country in Southeast Asia, where it is one of the multi-ethnic and multi-cultural countries that being an interest for a huge population. Since independence, Malaysian GDP has grown at an average of 6.5% per annum for almost 50 years. In addition, Kuala Lumpur and Johor Bahru are the two major cities, and both of them become a center of attention for several reasons e.g. education, tourism, job employment, residential, and shopping. Hence, it became a favorable choice globally for the following:

- Kuala Lumpur, officially the Federal Territory of Kuala Lumpur or commonly known as KL, is the national capital and largest city in Malaysia. As the global city of Malaysia, it covers an area of 243 km²(94 sq mi) and has an estimated population of 1.73 million as of 2016. Greater Kuala Lumpur, also known as the Klang Valley, is an urban agglomeration of 7.25 million people as of 2017. It is among the fastest growing metropolitan regions in Southeast Asia, in both population and economic development. (source: https://en.wikipedia.org/wiki/Kuala_Lumpur)

- Johor Bahru, formerly known as Tanjung Puteri or Iskandar Puteri, is the capital of the state of Johor, Malaysia. It is situated along the Straits of Johor at the southern end of Peninsular Malaysia. Johor Bahru has a population of 497,097, while its metropolitan area, with a population of 1,638,219, is the third largest in the country. (source: https://en.wikipedia.org/wiki/Johor_Bahru)

1.2 Business Problem

Since the two cities Kuala Lumpur and Johor Bahru are the top choice for a wide range of people around the world, then I would like to develop a study that facilitate the way forward for selecting the area of interest among these cities.

1.3 Interest

The tourists visiting the two cities would be very interested in such accurate analysis, for which it should save a lot time to determine the top rated venues across the city. In addition, others who do require market survey for advertising purposes might be interested as well.

2. Data

2.1 Data Sources

I will be using dataset published online for both cities from Wikipedia as below:

https://en.wikipedia.org/wiki/Kuala_Lumpur

https://en.wikipedia.org/wiki/Johor_Bahru

2.2 Data cleaning and re-grouping

The retrieved table contains some un-wanted entries and needs some cleanup.

The following tasks will be performed:

- Drop/ignore cells with missing data.
- Use most current data record.
- Fix data types.

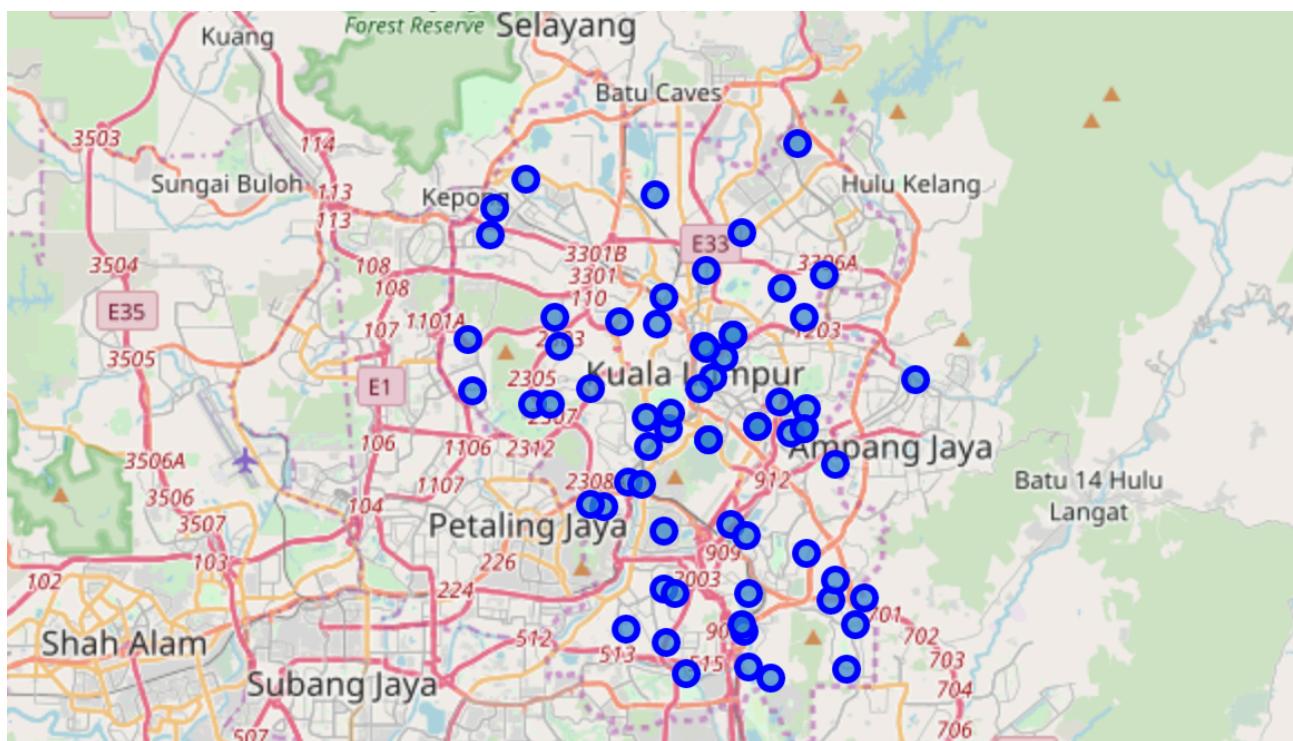
3. Methodology

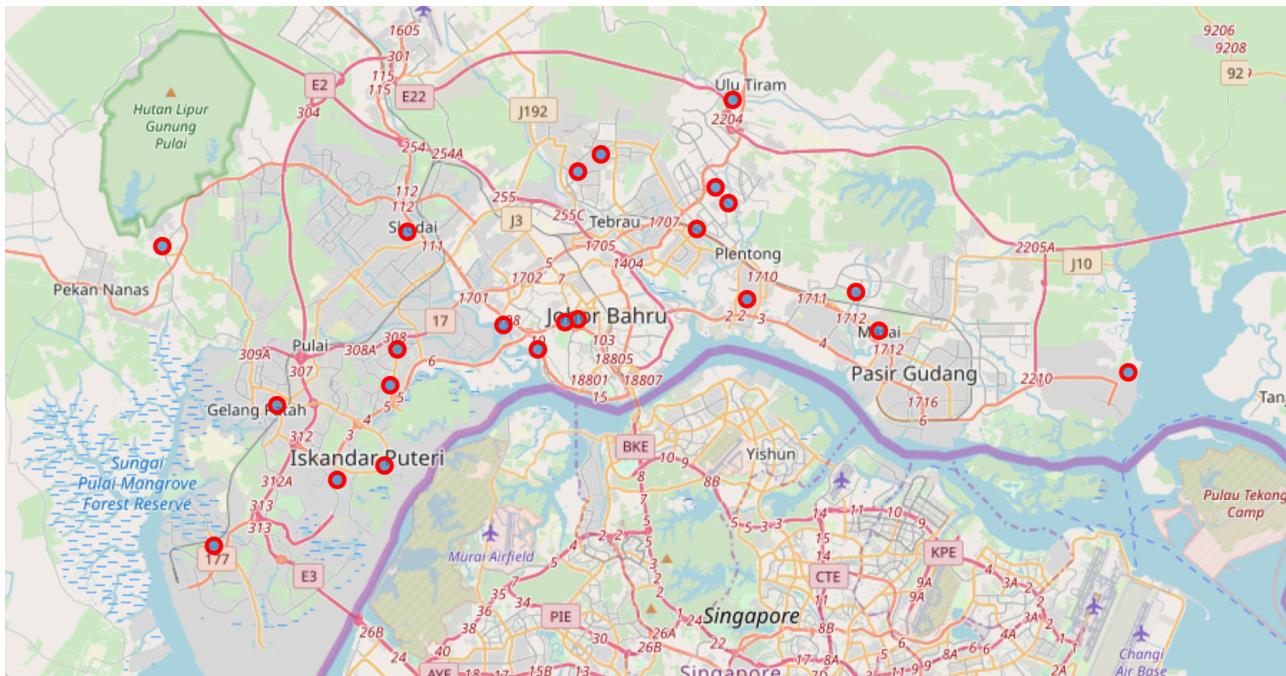
In this project, we will be picking the area of interest among the cities for the tourists. Hence, in order to do the same, we *have got the coordinates of each area/district using Google API*.

	Postcode	District		Area	Latitude	Longitude
0	52100	Kepong		Jinjang	3.211033	101.642303
1	52100	Kepong	Taman Bukit Maluri	3.201923	101.632259	
2	51200	Segambut	Bandar Menjalara	3.193871	101.630880	
3	51200	Segambut	Bukit Kiara	3.142163	101.644358	
4	51200	Segambut	Bukit Tunku	3.166521	101.682767	
5	51200	Segambut	Damansara	3.142145	101.649912	
6	51200	Segambut	Damansara Town Centre	3.146779	101.662265	
7	51200	Segambut	Jalan Duta	3.167529	101.670687	
8	51200	Segambut	Kampung Kasipillay	3.174557	101.684333	
9	51200	Segambut	Kampung Sungai Penchala	3.162039	101.624515	

	Postcode	District	Area	Latitude	Longitude
0	80000	Johor Bahru	Johor Bahru	1.492659	103.741359
1	81100	Johor Bahru	Bandar Dato' Onn	1.563273	103.741075
2	80200	Johor Bahru	Danga Bay	1.478378	103.722255
3	81100	Johor Bahru	Johor Jaya	1.535573	103.797820
4	81100	Johor Bahru	Desa Jaya	1.556072	103.807075
5	81100	Johor Bahru	Ehsan Jaya	1.548463	103.813092
6	80350	Johor Bahru	Larkin	1.491506	103.734709
7	81200	Johor Bahru	Tampoi	1.489990	103.705399
8	81100	Johor Bahru	Tebrau	1.571448	103.752088
9	81800	Johor Bahru	Ulu Tiram	1.597369	103.815095

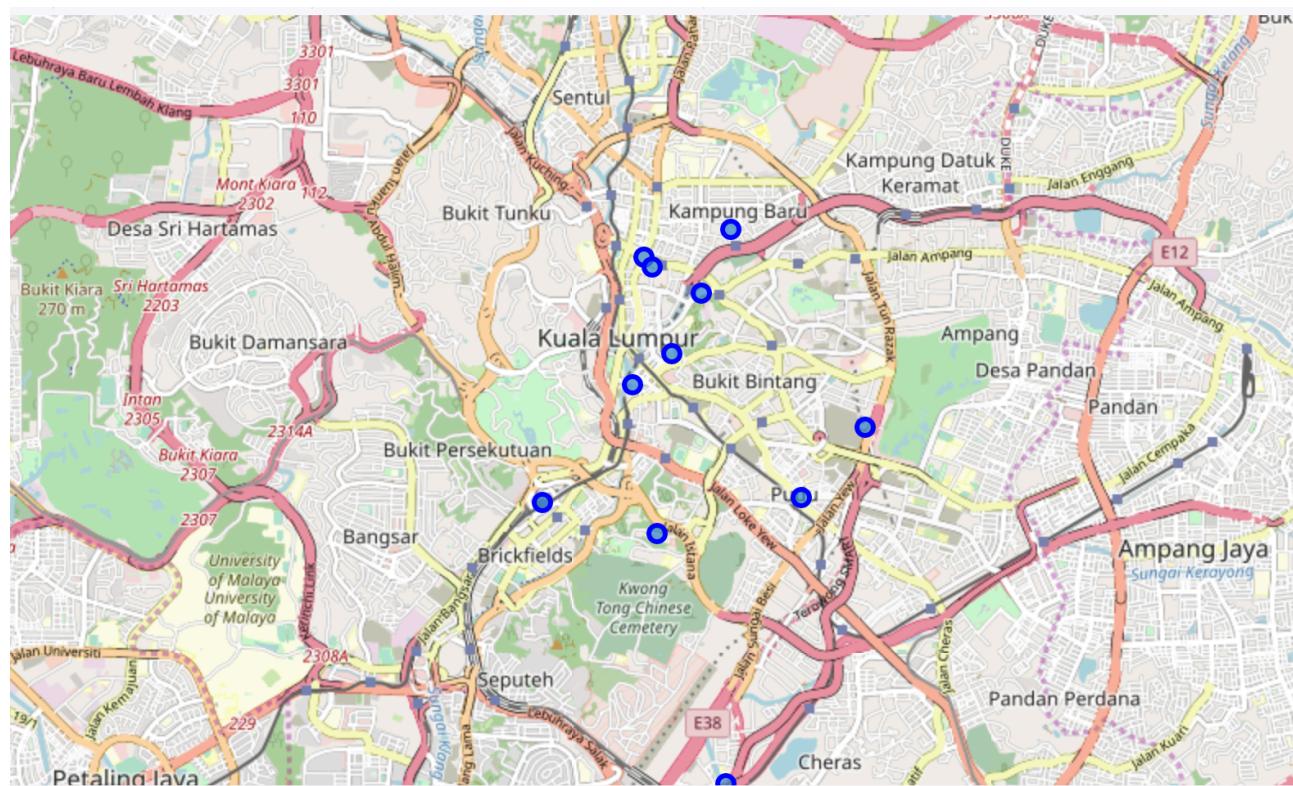
Then, we have created a map for the acquainted points/areas of Kuala Lumpur and Johor Bahru as below.

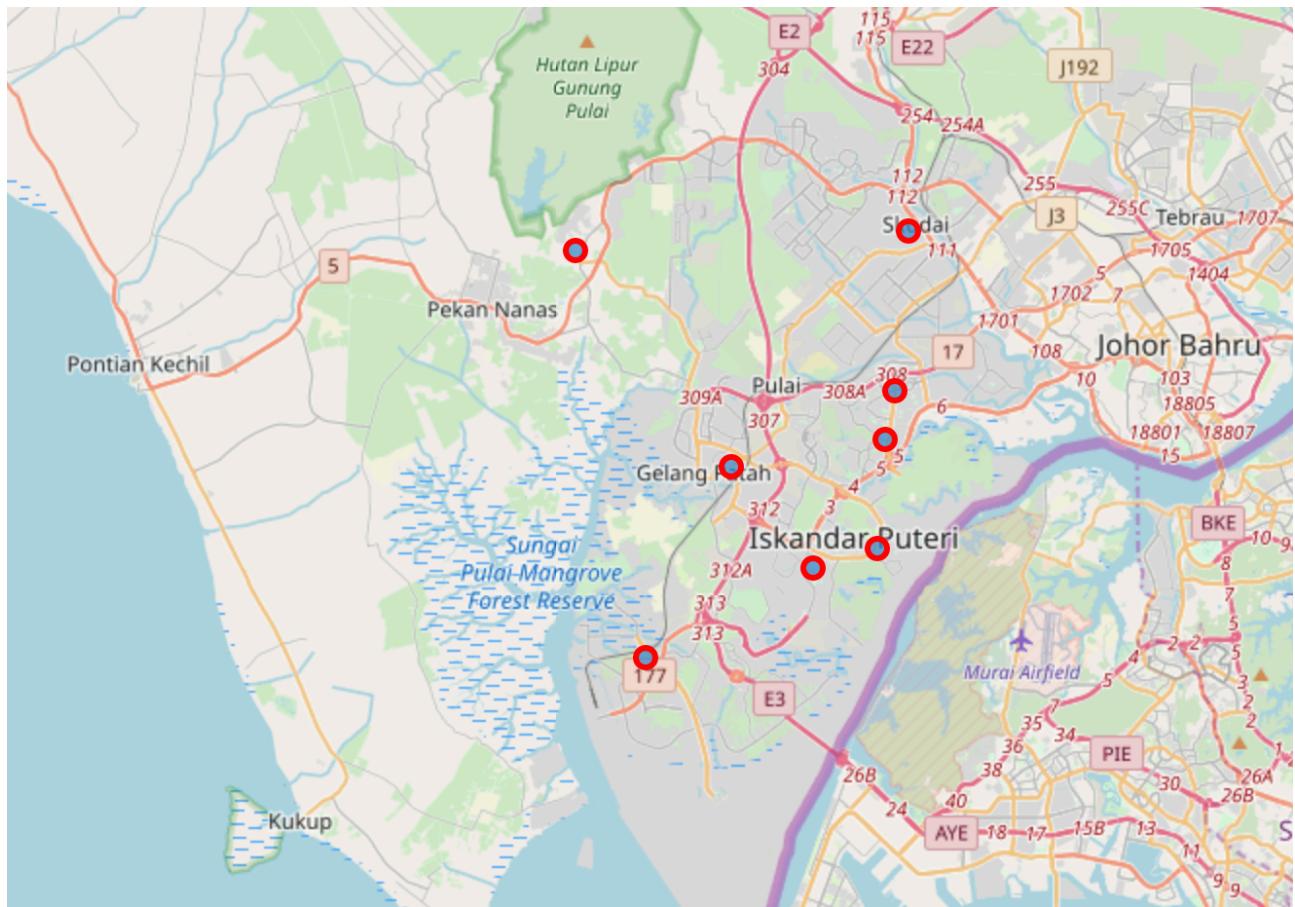




Post the analysis made, it's obvious that the highest area for Kuala Lumpur is *Bukit Bintang* and for Johor Bahru is *Iskandar Puteri*.

So, we will start exploring these areas in deep details to understand geographical distribution among such places. First of all, we have pulled their maps respectively.





Moreover, we will be using Foursquare location data in order to classify and segment the areas of these cities using the prioritizing list provided by Foursquare. Also, we will be spotting the light on the similarities and differences in between both of cities, and defining each area based on its cluster e.g. education, residential, and shopping.

For instance, these are the venues details through each area.

Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	
0	KL Sentral	3.134339	101.686337	Hilton Kuala Lumpur	3.135405	101.685690	Hotel
1	KL Sentral	3.134339	101.686337	Family Mart	3.132960	101.687480	Convenience Store
2	KL Sentral	3.134339	101.686337	Hilton Executive Lounge	3.135923	101.685782	Hotel Bar
3	KL Sentral	3.134339	101.686337	Aloft Kuala Lumpur Sentral	3.132767	101.686094	Hotel
4	KL Sentral	3.134339	101.686337	Aloft Hotel Grand Ballroom	3.133410	101.686029	Hotel

	Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bukit Indah	1.478332	103.654542	Subway	1.480537	103.656018	Sandwich Place
1	Bukit Indah	1.478332	103.654542	TGV Cinemas	1.482024	103.656241	Multiplex
2	Bukit Indah	1.478332	103.654542	Starbucks	1.481445	103.656229	Coffee Shop
3	Bukit Indah	1.478332	103.654542	Thai Relax Massage	1.479739	103.657446	Spa
4	Bukit Indah	1.478332	103.654542	Soul Thai Restaurant	1.480937	103.656256	Thai Restaurant

We will use `explore` function to identify the highly rated venues category within each neighbour.

----Bukit Nanas----

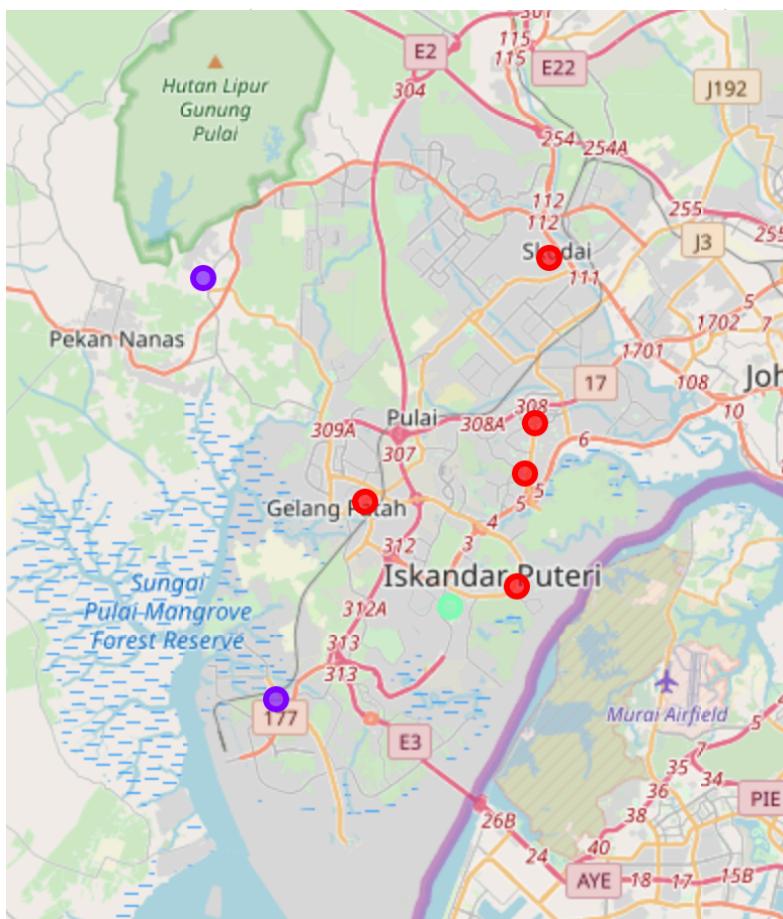
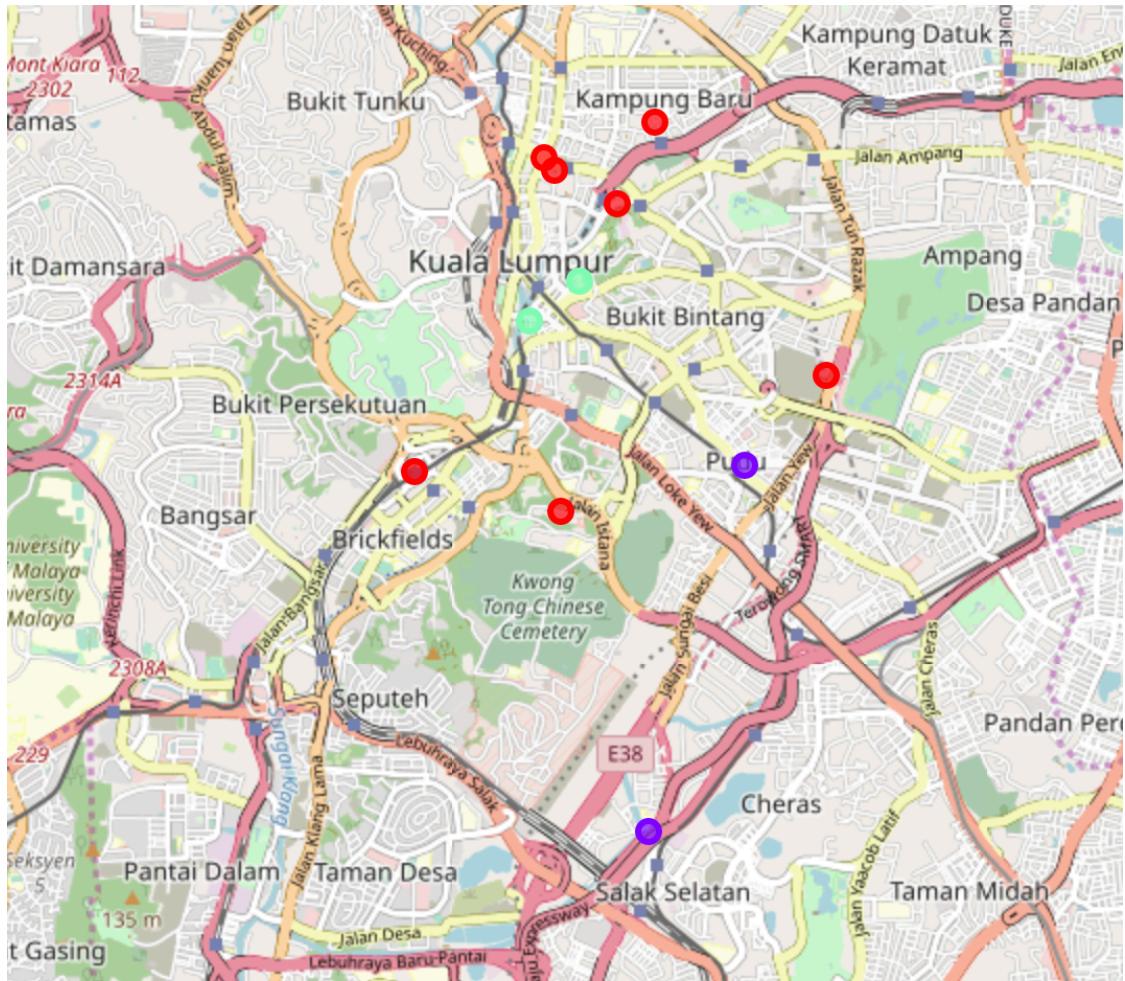
	venue	freq
0	Indian Restaurant	0.14
1	Café	0.09
2	Malay Restaurant	0.07
3	Coffee Shop	0.05
4	Adult Boutique	0.02

----Bukit Petaling----

	venue	freq
0	Malay Restaurant	0.29
1	Breakfast Spot	0.06
2	Travel & Transport	0.06
3	Food Court	0.06
4	Museum	0.06

Area	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	
0	Bukit Nanas	Indian Restaurant	Café	Malay Restaurant	Coffee Shop	Zoo	General Travel	Road	Noodle House	Monument / Landmark	Latin American Restaurant
1	Bukit Petaling	Malay Restaurant	Convenience Store	Food Court	Park	Museum	Seafood Restaurant	Falafel Restaurant	Building	Breakfast Spot	Outlet Store
2	Chow Kit	Malay Restaurant	Hotel	Asian Restaurant	Chinese Restaurant	Coffee Shop	Food Court	Bakery	Soup Place	Indian Restaurant	Dim Sum Restaurant
3	Dang Wangi	Hotel	Malay Restaurant	Soup Place	Spa	Bar	French Restaurant	Sandwich Place	Restaurant	Café	Chinese Restaurant
4	KL City Centre	Indian Restaurant	Hotel	Asian Restaurant	Chinese Restaurant	Coffee Shop	Food Truck	Café	Malay Restaurant	Convenience Store	Restaurant

Based on such ratings we will start grouping and clustering them using K-means, and you can find underneath a map drawing for illustrating the clusters among each city.



4.Results

We can observe from the results of Kuala Lumpur city the following points:

- . First cluster has the most common venue 'Malay Restaurant'
- . Second cluster has the most common venue 'Chinese Restaurant'
- . Third cluster has the most common venue 'Indian Restaurant'

We can observe from the results of Johor Bahru city the following points:

- . First cluster has the most common venue 'Construction & Landscaping'
- . Second cluster has the most common venue 'Asian Restaurant'
- . Third cluster has the most common venue 'Chinese Restaurant'

5.Discussion

. Kuala Lumpur city first and third clusters has more residential locations where we can find several spots for restaurants and hotel

- . The second cluster has more tourism locations including mixture of nightlife venues.

. Johor Bahru city The second and third clusters has more residential locations where we can find several spots for restaurants and cafe

- . The first cluster has more tourism venues such as park and construction venue.

Using Foursquare API, we have collected an undeniable amount of venue recommendations in Malaysia cities. Sourcing from the venue recommendations from FourSquare has its limitation, and the list of venues is not exhaustive list of all the available venues are the area.

This project would help the stakeholders take a better decision on choosing the best neighborhood through identifying the better place to settle.

Further enhancement for the research would be analyzing the correlation between the districts and areas based on the venues.

6.Conclusion

The analysis offer a valuable information for travelers, researchers, and many other people whom might find an interest in allocating the venues across both cities Kuala Lumpur and Iskandar. I have tried to do my utmost efforts to analyze the data using machine learning.

It's obvious that illustrating the similarities between both cities would require further studies, in case we need to recommend which city is having attractive locations based on a specific characteristic. However, we can now identify the best neighborhood through on points of interests, which is the aim of our project.

