

# Capstone Project IBM

The Battle of Neighborhoods

[https://github.com/TruffleLaw/Coursera\\_Capstone](https://github.com/TruffleLaw/Coursera_Capstone)

# INTRODUCTION

There are two questions  
to be answered

- 1) Is there any similarity between two cities
- 2) If developer would like to open a new shopping mall, where they should open it?

For question No,1, I will use Kuala Lumpur, Malaysia and Tokyo, Japan. We know that Tokyo is a developed country meanwhile Malaysia is still a developing country. Can we find any similarity between these two cities in terms of amenity?

For question No,2, I want to focus on Kuala Lumpur only. My goal in here is to find the location that has less no. of shopping mall to prevent the oversupply







# Hypothesis

## Q1 NO SIMILARITY

Malaysia and Japan are completely two different cities

## Q2 THERE WILL BE A PLACE WITH LOW OR ZERO NO. OF SHOPPING MALLS

Should open the new shopping mall at a location that has a few or zero no of shopping malls.



Kuala Lumpur,  
Malaysia

[3.139003,  
101.686852]



Tokyo, Japan

[35.689487,  
139.691711]



# Methodology

## 1. Web Scraping

Get the neighborhoods list for Kuala Lumpur and Tokyo

## 2. Latitude and Longitude

Get the latitude and longitude for each neighborhood by using Geocoder package.

## 3. Foursquare API

Use Foursquare API to get the nearby venues and then group the data by neighborhood and take the average of frequency of occurrence of each category

## 4. Clustering

Perform clustering on the data by using Kmeans clustering model.

## 5. Data Visualization

Visualize the clusters in a map using Folium package.



A stylized digital eye with a blue and green iris, overlaid with binary code and web development snippets like 'IMG', 'px;', and 'image:'. The background is dark with glowing lines and text, creating a high-tech, digital atmosphere.

# Question 1

Find the similarity or dissimilarity of two different cities of your choice.



# Question 1

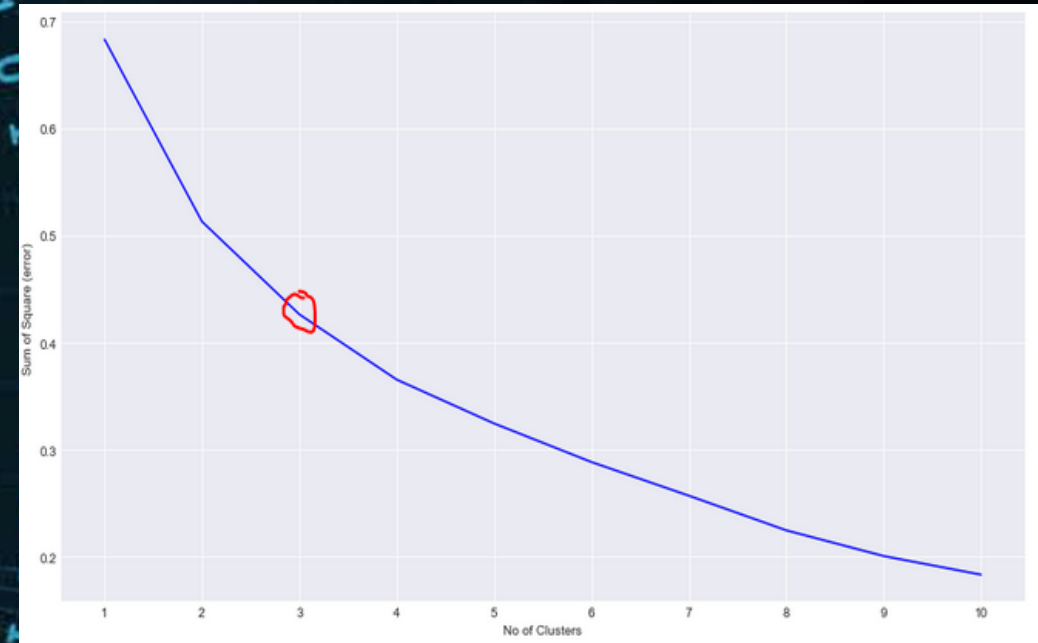
Find the similarity or dissimilarity of two different cities of your choice.



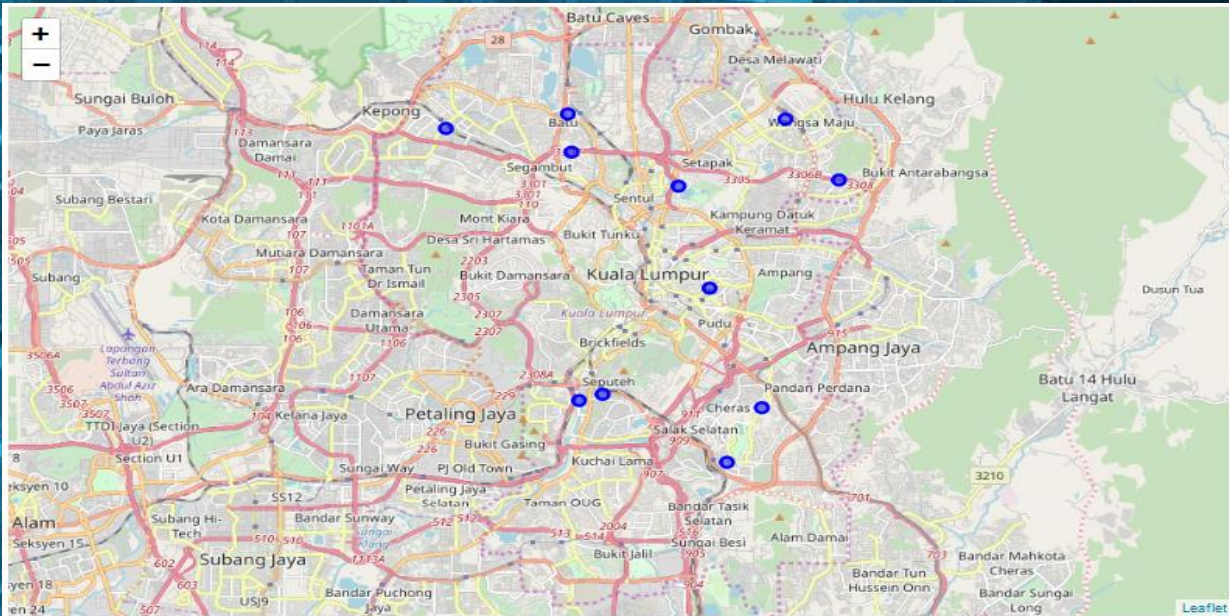
## Cluster0 (red marker) and Cluster1 (green marker)



I'm using the kmeans clustering model with  $n = 3$  as I believe it gives an optimal result.

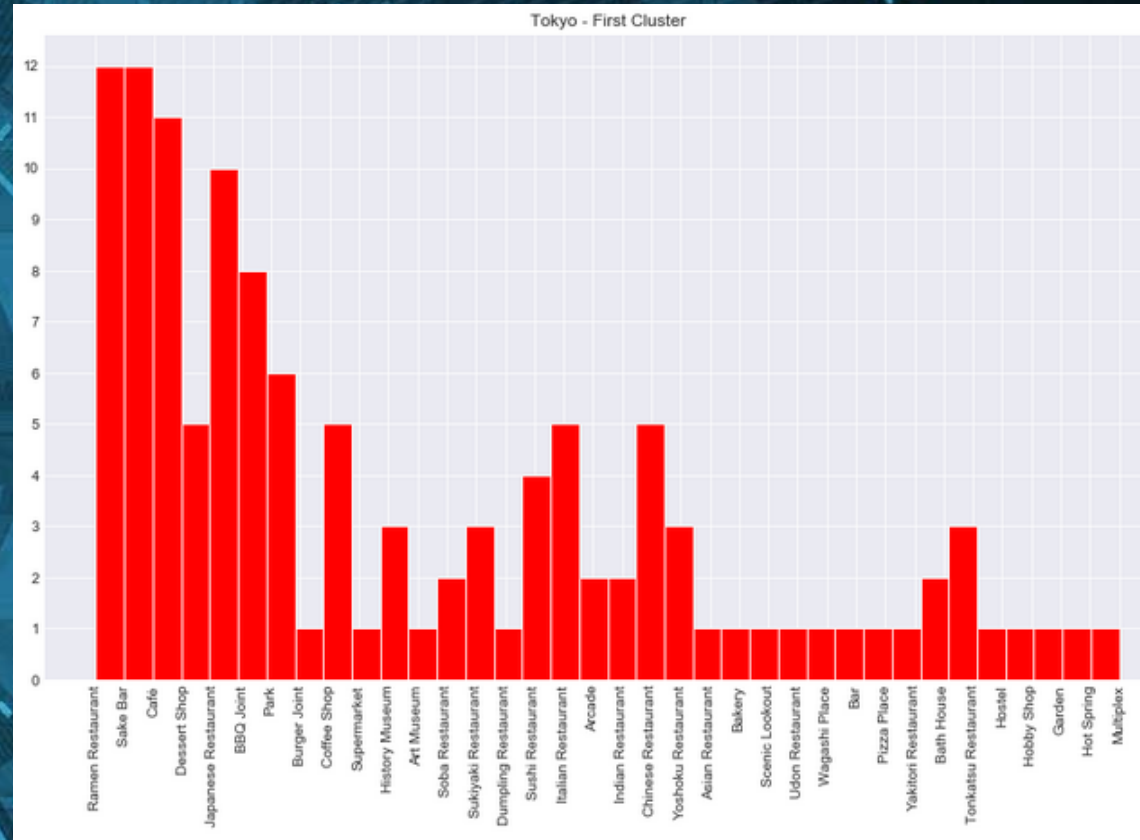


## Cluster2 (blue marker)





District	Latitude	Longitude	Cluster
Adachi	35.7748	139.805	0
Arakawa	35.7361	139.783	0
Edogawa	35.707	139.868	0
Itabashi	35.7511	139.709	0
Katsushika	35.7435	139.847	0
Kita	35.7528	139.734	0
Nerima	35.7357	139.652	0
Suginami	35.6997	139.636	0
Sumida	35.7107	139.802	0
Taitō	35.7126	139.78	0
Toshima	35.7362	139.714	0
Ōta	35.5614	139.716	0



## Analyze cluster0

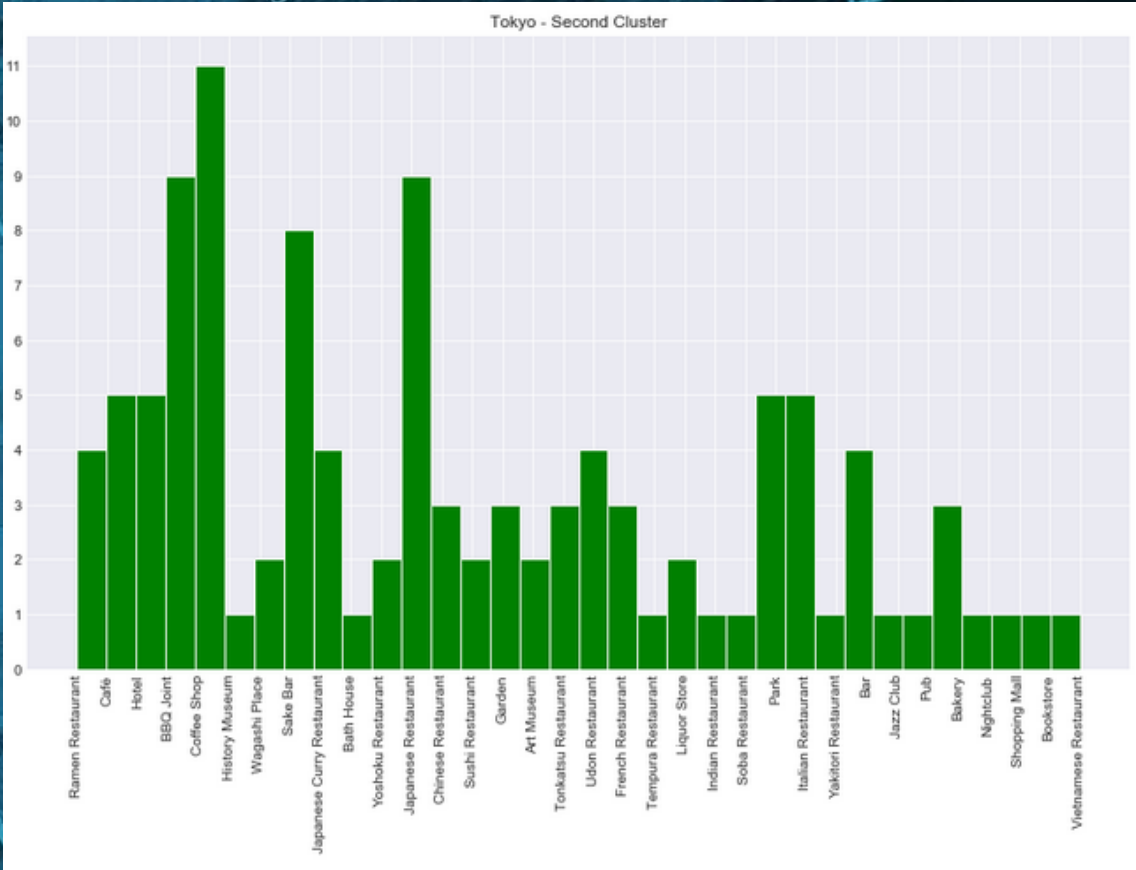
Cluster0 consist of some of Tokyo districts. No Kuala Lumpur districts are found.

The top 5 most common venues:

- 1) Ramen Restaurant
- 2) Sake Bar
- 3) Café
- 4) BBQ Joint
- 5) Japanese Restaurant



	District	Latitude	Longitude	Cluster
0	Bunkyo	35.7076	139.752	1
1	Chiyoda	35.6939	139.754	1
2	Chūō	35.6706	139.772	1
3	Kōtō	35.6728	139.817	1
4	Meguro	35.6415	139.698	1
5	Minato	35.658	139.752	1
6	Nakano	35.7073	139.664	1
7	Setagaya	35.6465	139.653	1
8	Shibuya	35.6618	139.704	1
9	Shinagawa	35.6092	139.73	1
0	Shinjuku	35.6938	139.703	1



## Analyze cluster1

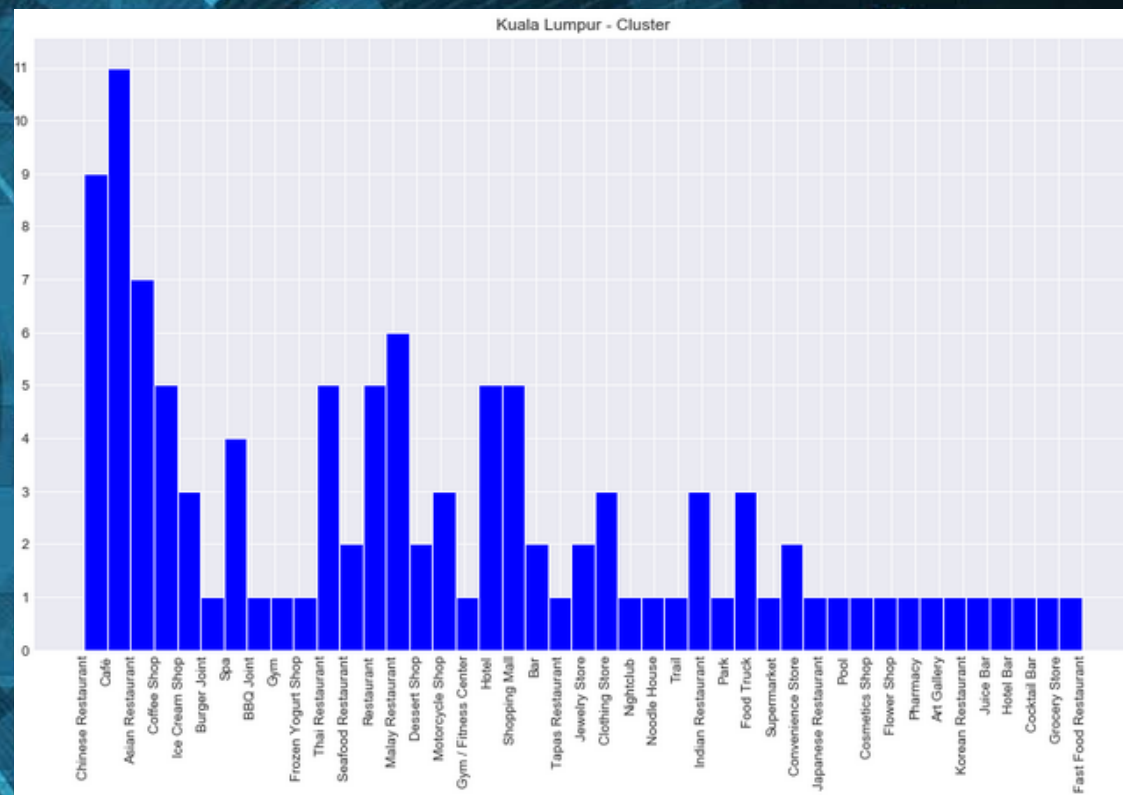
Cluster1 consist of another part of Tokyo districts. No Kuala Lumpur districts are found.

The top 5 most common venues:

- 1) Coffee Shop
- 2) BBQ Joint
- 3) Japanese Restaurant
- 4) Sake Bar
- 5) Hotel/Italian Restaurant



District	Latitude	Longitude	Cluster
Bandar Tun Razak	3.08863	101.716	2
Batu	3.20467	101.672	2
Bukit Bintang	3.14679	101.711	2
Cheras	3.10684	101.726	2
Kepong	3.19967	101.639	2
Lembah Pantai	3.1094	101.675	2
Segambut	3.19209	101.673	2
Seputeh	3.11138	101.682	2
Setiawangsa	3.18253	101.747	2
Titiwangsa	3.18067	101.703	2
Wangsa Maju	3.20299	101.732	2



## Analyze cluster2

Cluster2 consist of all Kuala Lumpur districts.

The top 5 most common venues:

- 1) Café
- 2) Chinese Restaurant
- 3) Asian Restaurant
- 4) Malay Restaurant
- 5) Hotel/Shopping Mall





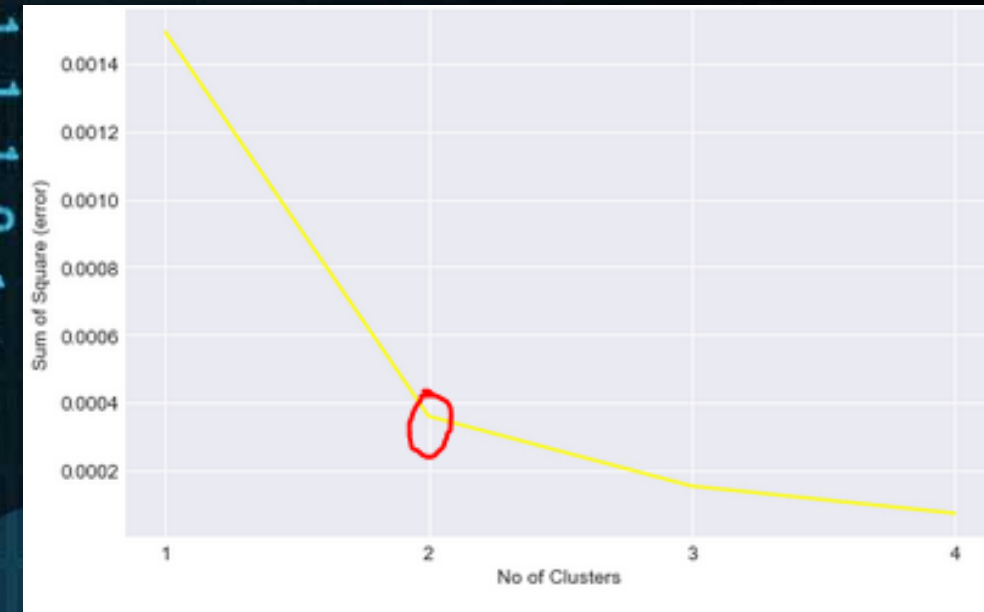
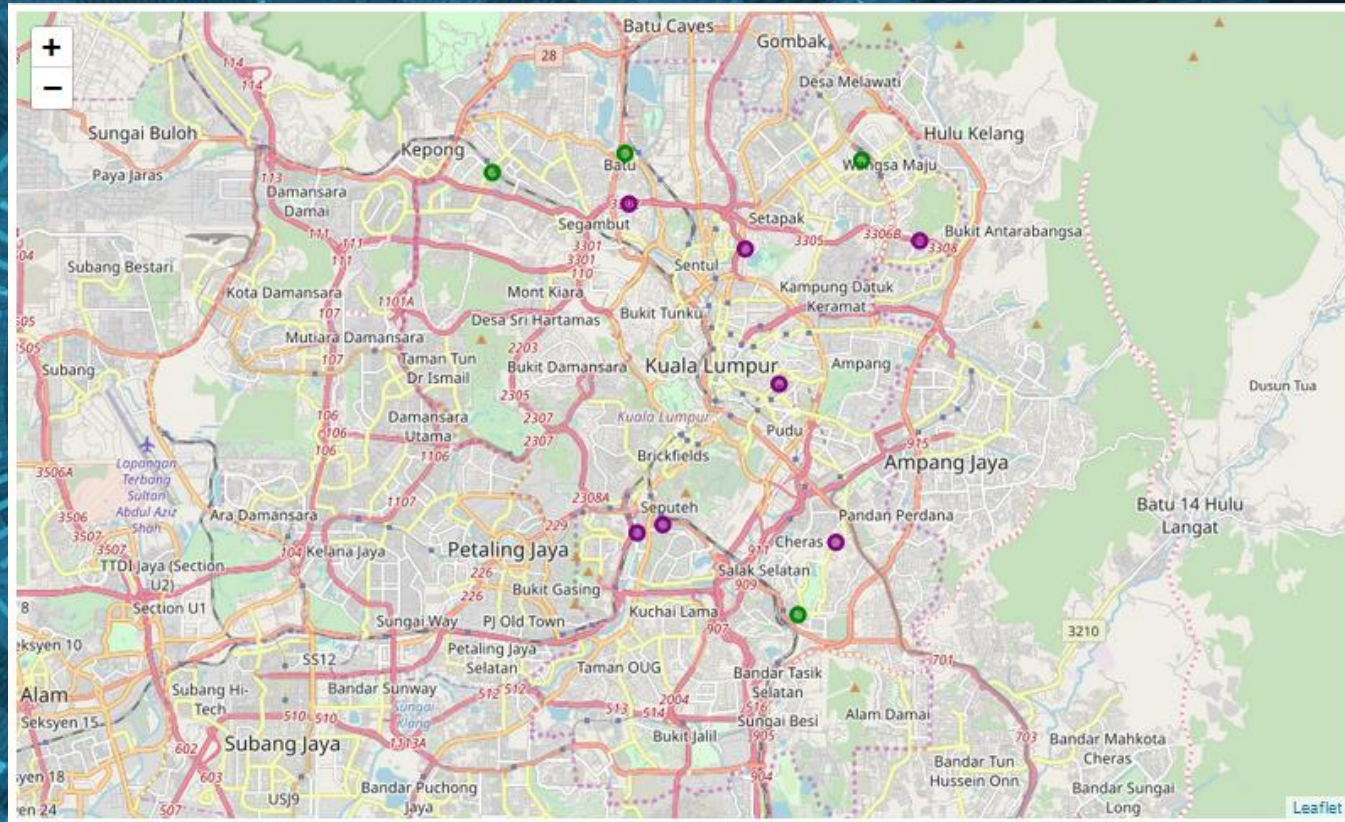
## Question 2

Where should developer open the new shopping mall in Kuala Lumpur?



I'm using the kmeans clustering model with  $n = 2$  as I believe it gives an optimal result.

There are two clusters formed which are in green and purple as shown in the figure on the left.



## Cluster0 (green marker)

	District	Shopping Mall	Latitude	Longitude	Cluster
0	Bandar Tun Razak	0.00	3.08863	101.716	0
1	Batu	0.01	3.20467	101.672	0
2	Kepong	0.01	3.19987	101.639	0
3	Wangsa Maju	0.01	3.20299	101.732	0

## Cluster1 (purple marker)

	District	Shopping Mall	Latitude	Longitude	Cluster
0	Bukit Bintang	0.03	3.14679	101.711	1
1	Cheras	0.03	3.10684	101.726	1
2	Lembah Pantai	0.03	3.1094	101.675	1
3	Segambut	0.02	3.19209	101.673	1
4	Seputeh	0.04	3.11138	101.682	1
5	Setiawangsa	0.02	3.18253	101.747	1
6	Titiwangsa	0.03	3.18067	101.703	1

## Analysis on each Cluster

As we can see, the Cluster0 has 'Shopping Mall' value of less and equal to 0.01.

Meanwhile, Cluster1 has the value or more than 0.01.





### Question 1

From the result that I have obtained, we can see that cluster0 and cluster1 consist of Tokyo districts only and on the other hand, cluster2 only consist of Kuala Lumpur districts. Cluster0 and cluster1 have Japanese Restaurant and Sake Bar in their top 5 most common venues near the area. For Kuala Lumpur city, the most demand category is Chinese Restaurant and then followed by Café. This shows that the demand from the people who live within the area plays an important role to shape the city. As for Japanese, they love to eat Japanese/local food and enjoy drinking sake after the office hour. Meanwhile, Malaysian loves to eat rice and also the main dish in all Chinese restaurants. Apart from that, malaysian also love to spend their time having a discussion/meeting in a café.

### Question 2

From the result that I have obtained, I found that almost all districts of Kuala Lumpur have their own shopping malls. In this project, I use the radius of 5 km and to have so many shopping malls within this small radius indicates that there is oversupply in shopping mall. To build more shopping mall has its own pros and cons. The pros are can attract more people and therefore can encourage them to spend more and this will contribute to positive growth in GDP of a country. The cons are will be many 'dead' shopping malls or abandoned buildings that tear down the landscape of the area. However, there is a good location to open a new shopping mall which is at Bandar Tun Razak as the number of existing shopping mall is zero at the moment. This will give a new attraction to the place. In my point of view, shopping mall is very essential to the people as this make their life easier as they can buy everything under one roof. Therefore, there is always a demand for shopping mall and it is good to open it at a location that has less or zero number of shopping malls nearby.

# Discussion



## In a nutshell

My earlier hypotheses are true for question 1 and 2. Kuala Lumpur and Tokyo are completely two different cities and we cannot find the similarity based on amenity. This is because Malaysia and Japan have different people and culture that shape the city to be different from one another.

In Kuala Lumpur, we can cluster the districts into two which are 'high' and 'low' in terms of number of shopping malls. The best location to open a new shopping mall is in cluster0 or to be precise at Bandar Tun Razak as it has zero shopping mall at the moment.

What I can conclude?

# Conclusion





# THANK YOU

Capstone Project