COURSERA CAPSTONE PROJECT

WEEK 4 – THE BATTLE OF NEIGHBOURHOODS

Kuala Lumpur, Malaysia vs Tokyo, Japan

Prepared by: Muhammad Najmy Bin Hashim

Github Id: TrufflegarLaw

Introduction

In this project, I would like to answer two questions:

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

For question 1, I will use Kuala Lumpur, Malaysia [3.139003,101.686852] and Tokyo, Japan [35.689487,139.691711]. As we all know, Japan is a highly developed country and most of the countries in Asia including Malaysia idolize Japan and set it as their benchmark. On the other hand, Malaysia is a developing country, despite undergoing rapid economic development over the past five decades. In Kuala Lumpur, we can see that it has a lot of amenities provided as offered by developed country. Therefore, I would like to know is there any similarity between Kuala Lumpur and Tokyo in terms of amenity.

For question 2, I will only focus on Kuala Lumpur, Malaysia. As of now, the number of shopping mall in Malaysia is very high and this is not a good sign. This is because there are many old shopping malls start to close due to cannot compete with the brand-new shopping malls and this will result many vacant buildings in very limited space.

Hypothesis

To answer the first question, how should we define similarity or dissimilarity between two different cities. From my perspective, I always find different city has their own differences as compared to another city. Therefore, to tackle this question, I need to find the nearby venues within the areas/districts of Kuala Lumpur and Tokyo respectively. I will do the clustering model and cluster all of them and see is there any districts in Kuala Lumpur that are clustered together with districts in Tokyo. If all districts are clustered together, this implies that there is no difference between Kuala Lumpur and Tokyo (which is completely untrue and I can prove it to you). My hypothesis for this question is districts of Kuala Lumpur will be separately clustered from districts of Tokyo, Japan. This is because the two countries have two different people, culture and therefore the demand for the area is different.

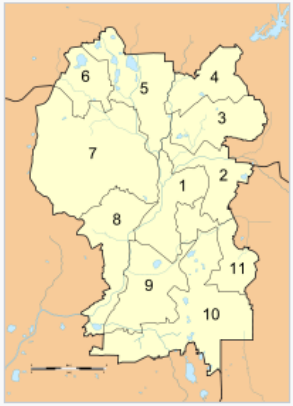
For second question, my hypothesis is that there will be one cluster that has high number of shopping malls and there will be another cluster that has few or zero number of shopping malls. In my point of view, It is better to open it in an area that has few or zero shopping mall. There are two reasons which are to give the area new attraction and to avoid the oversupply of shopping mall in one area. In my point of view, shopping mall is very essential as people can but anything under the sun at one place. Therefore, it always be a demand from the people to have the shopping centre near to their area.

Data

In order to proceed with the analysis, the following data are required:

1) List of districts in Kuala Lumpur, Malaysia. This data can be found in the link here (https://en.wikipedia.org/wiki/Kuala_Lumpur)

Districts [edit]



Districts (divisions) of Kuala Lumpur

Kuala Lumpur's eleven districts, with estimated population and percentage of the total, serve as administrative subdivisions under the authority of the Kuala Lumpur City Hall authority.[89]

1. Bukit Bintang (103,820 - 5.8%)
2. Titiwangsa (198,690 - 11.1%)
3. Setiawangsa (179,000 - 10.0%)
4. Wangsa Maju (227,330 - 12.7%)
5. Batu (91,290 - 5.1%)
6. Kepong (10,740 - 0.6%)
7. Segambut (125,300 - 7%)
8. Lembah Pantai (189,740 - 10.6%)
9. Seputeh (230,910 - 12.9%)
10. Bandar Tun Razak (273,870 - 15.3%)
11. Cheras (159,310 - 8.9%)

2) List of districts in Tokyo, Japan. This data can be found in the link here (https://en.wikipedia.org/wiki/Tokyo)

Special Wards of Tokyo				
	Place Name			Map of the Special Wards
	Rōmaji	Kanji	Color	
1	Adachi	足立区	Red	
2	Arakawa	荒川区	Green	
3	Bunkyo	文京区	Yellow	
4	Chiyoda	千代田区	Orange	
5	Chūō	中央区	Green	
6	Edogawa	江戸川区	Green	
7	Itabashi	板橋区	Yellow	
8	Katsushika	葛飾区	Yellow	
9	Kita	北区	Orange	
10	Kōtō	江東区	Yellow	
11	Meguro	目黒区	Orange	
12	Minato	港区	Yellow	
13	Nakano	中野区	Yellow	
14	Nerima	練馬区	Green	
15	Ōta	大田区	Yellow	
16	Setagaya	世田谷区	Green	
17	Shibuya	渋谷区	Red	
18	Shinagawa	品川区	Green	
19	Shinjuku	新宿区	Green	
20	Suginami	杉並区	Orange	
21	Sumida	墨田区	Orange	
22	Taitō	台東区	Red	
23	Toshima	豊島区	Red	

3) Get the latitude and longitude by using the geocoder library or extract one by one from https://www.latlong.net. In this project, I'd used the latter because the former gave inaccurate latitude and longitude thus it the markers were pointed wrongly on the map that I've created using folium function.

4) Nearby places for each Kuala Lumpur and Tokyo districts. I can get this data by using Foursquare API.

Methodology

1) Question 1: Find the similarity/dissimilarity of two different cities of your choice.

Firstly, we collect the required data from Wikipedia websites as mentioned above. Then, we can acquire their latitudes and longitudes by using the data retrieved from https://www.latlong.net. Next, I will use the Foursquare API to retrieve the nearby venues within each district of Kuala Lumpur and Tokyo. From there, we can find the top 10 most common nearby venues and do the clustering process by using KMeans model from sklearn.cluster library. Finally, we can see is there any district of Kuala Lumpur that is clustered together with Tokyo districts and then we can examine that cluster to find what kind of similarity that they share.

2) Question 2: Where should developer open the new shopping mall in Kuala Lumpur?

Firstly, we collect the required data from Wikipedia websites as mentioned above. Then, we can acquire their latitudes and longitudes by using the data retrieved from https://www.latlong.net. Next, I will use the Foursquare API to retrieve the nearby venues within each district of Kuala Lumpur. From there, we can find the average shopping mall that each district has and then can proceed with clustering process by using KMeans model from sklearn.cluster library. Finally, we can see which cluster has the low or zero average of shopping mall which is the recommended area to open a new one.