THE BATTLE OF NEIGHBORHOOD

Index :

This project discuss about the two cities in Malaysia about all of their Specialty to start a new business in the cities and also discuss the local and Foreign communities.

INTRODUCTION:

Kuala Lumpur and Johor Bahru are two major cities in Malaysia. Both cities become a center of attention for residential, job employment, tourism, education, shopping and sports activity. Both cities are well known in Malaysia, and become the top choice for local and foreign communities.

Brief information about both cities:
- Kuala Lumpur: is the national capital of Malaysia as well as its largest city. The only global city in 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 South-East 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)

OBJECTIVES:

In this project, we will study in details the area classification using Foursquare data and machine learning segmentation and clustering. The aim of this project is to segment areas of Kuala Lumpur and Johor Bahru based on the most common places captured from Foursquare.

Using segmentation and clustering, we hope we can determine:
1. the similarity or dissimilarirty of both cities
2. classification of area located inside the city whether it is residential, tourism places, or others

DATA:

The data acquired from wikipedia pages and restructure to csv file for easier manipulation and reading. Both files uploaded to my github for references. Link to the files are:

- https://github.com/zaephaer/CapstoneProject/blob/master/JB_disrict.csv
- https://github.com/zaephaer/CapstoneProject/blob/master/KL_disrict.csv

Another aspect to consider for this project is the Foursquare data. I believe that the data as good as provided, meaning although we are using Foursquare data for segmentation and clustering, the amount and accuracy of data captured can't 100% determine correct classification in real world.