# The Battle of Neighborhood¶

## 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.

## Objective¶

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.

To start, let's get and look at the data. I've already downloaded it, so let's read it (from local drive) and load it to dataframe:

## Methodology¶

In this project, I will use the basic methodology as taught in Week 3 lab.

- Above, we have done convert addresses into their equivalent latitude and longitude values.
- Then we will use the Foursquare API to explore neighborhoods in both cities, Kuala Lumpur and Johor Bahru
- After that, explore function to get the most common venue categories in each neighborhood,
- and then use this feature to group the neighborhoods into clusters

K-means clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighborhoods in Kuala Lumpur and Johor Bahru and their emerging clusters.

Based on dataframe analysis above, we found out that Bukit Bintang area in Kuala Lumpur and Johor Bahru area in Johor Bahru are both have the highest number of area within it those district.

# Discussion¶

Based on cluster for each cities above, we believe that classification for each cluster can be done better with calculation of venues categories (most common) in each cities. we assumed each cluster as follow:

- Cluster 1: Kuala Lumpur: Tourism
- Cluster 2: Kuala Lumpur: Residental
- Cluster 3: Kuala Lumpur: Mix
- Cluster 1: Johor Bahru: Residental
- Cluster 2: Johor Bahru: Tourism
- Cluster 3: Johor Bahru: Sport

# Conclusion¶

Using Foursquare API, we can captured data of common places all around the world. Using it, we refer back to our main objectives, which is to determine;

- the similarity or dissimilarirty of both cities
- classification of area located inside the city whether it is residential, tourism places, or others

In conclusion, both cities Kuala Lumpur and Johor Bahru are the center of attraction among Malaysian. However, to declare both cities are similar or dissimilar base on common venues visited is quite difficult. Both cities is similar in some venues also dissimilar in certain venues.