

Capstone Project - The Battle of Neighborhoods

from Ho Hin Wing

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 dissimilarity 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/ringohohw/Capstone/blob/master/FinalProject/JB_disrict.csv

https://github.com/ringohohw/Capstone/blob/master/FinalProject/KL_disrict.csv

Another aspect of this project is the Foursquare data. I beleive 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.