Coursera Capstone

IBM Applied Data Science Capstone

Building A Park in Kuala Lumpur, Malaysia

By: Nizam Asmari

Oktober 2019

Introduction

As countries aiming to reach a developed country, lot of projects have been made in order to meet the requirement. Developing the residential area, infrastructures as well as the improving of towns and cities management were made mainly in big cities. Therefore, in recent years, we can see a lot of concrete jungles in big cities like in Kuala Lumpur. As a result, it helps increasing the carbon emission rate. Furthermore, it also leads to a physical and mental health problem to cities resident as there are not many places to socialize or to getaway from the crowded and hectic life. Building a park is a way to solve this matter and it also line with Malaysia government who is currently study on low-carbon footprint in an effort to create more green spaces and linkages in Kuala Lumpur. However, it is very difficult to find a suitable place to build it because it requires serious consideration and a lot more complicated than it seems. Particularly, the location of the park is one of the important decisions that will determine its function towards the environment and citizens.

Business Problem

The objective of this capstone project is to analyze and explore the best and suitable place to build a park in Kuala Lumpur. Using the data science methodology and machine learning techniques such as clustering, in order to solve the business question; If your government, developers or any related NGOs what to build a park in Kuala Lumpur metropolitan city, where will you recommend it?

Target Audience of the project

This project is very useful for the environmentalist NGOs, developers and government who want to solve the environment problem in Kuala Lumpur. This project indeed is taking time, as Kuala Lumpur is crowded and bustled with concretes and bricks. Thus, the target audience of this project includes developers, NGOs, constructors and anyone who deals with environment.

Data

To solve this problem, we will need the following data:

1. The list of Kuala Lumpur neigbourhood. This define the project which is confined the city of Kuala Lumpur, the capital city of Malaysia.
2. The longitude and latitude of Kuala Lumpur. This is necessary in order to plot the map and to get the venues data.
3. Venues data, particularly the one who related for building a park. We will use this data to perform clustering on the neighbourhood.

Sources of data and methods to extract them

(https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur)

Refer the above Wikipedia link, it is the list of Kuala Lumpur neighborhood. We will be using the link and scrap the webpage by using BeautifulSoup Package and Request Package in order to obtain the necessary data for this project. Then, using the Python Geocoder Package in order to get the longitudes and latitudes of each neighbourhood data.

After that, we will use Foursquare Application Programming Interface (API) to get the venues data of each neighbourhood. Foursquare has one of the largest database, with 105 millions places and 150 thousands developers. Foursquare will provide many categories on each of the venues data; particularly, we are interested in park, shopping mall and residential area in order to find the best solution in this matter. This project will make use of many data science skills; from web scrapping, API, data wrangling, data develop and data evaluation to machine learning and map visualization. In the next section, we will discuss about the methodology of the project.