Coursera Capstone Project

- Project Overview -

Course	IBM Data Science Professional Certificate
Project	Where to build a train station in Kuala Lumpur, Malaysia?
Author	Christoph Yew
Date	July 2020

Introduction
Kuala Lumpur is a major hub in Malaysia. Thousands of people come to the city every day for leisure and business activities. In order to offer a convenient way to reach the major shopping malls and office spaces in town when coming from another city, Kuala Lumpur needs to have a train station that is located very centrally.

Business Case
I want to find out, which location in Kuala Lumpur (KL) are best for building a new train station. Using the techniques taught during the other courses of the IBM Data Science program, I will use FourSquare and various techniques to provide an answer to the question above.

An answer to question is especially interesting for city management and real estate developers who want to solve the problem that people are spending too much finding their ways to the major venues in town after they arrived via train.

Required Data
1. Latitude and longitude data of KL, including its neighborhoods. So we can plot the neighborhoods to a map.
2. Geographical data about existing shopping malls and office towers so we know where there is high demand for a train station.

Data Sources
A list of the neighborhoods in KL can be found on Wikipedia (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur). After web-scraping the list, I will get the location data of the neighborhoods using a Python package like e.g. GeoCoder.

After that I will use FourSquare to get the shopping and, office space data for these neighborhoods.

I will then calculate the ideal location and use visualization tools like Folium to provide an answer to the question above.

Methodology
I first scraped the list of all neighborhoods in Kuala Lumpur and then got the location data for them. Next I used FourSquare to get a list of the major shopping venues and office spaces located in these neighborhoods as well as their location data. Based on the location data I then found the best location for a train station using the average of all venues I found.

Results
I found that the best location for a train station based on the distribution of shopping venues and office spaces is located at:

Best Train Station	3.150112	101.700595

Discussion
In this project I realized that malls and office spaces are mostly located around the city centre of Kuala Lumpur. We can see from the visualization that there are hardly any such venues located in more remote areas. Based on this I was able to find a fantastic location for building a train station or other major transportation hub averaging out the geo data for major shopping and office spaces.

Conclusion
This project can only work as a first step to locate the best location for a train station in Kuala Lumpur. Besides the points modeled, it would be also good to check where existing traffic hubs are as well as consider the size of the shopping malls and office spaces to weight them properly.