Coursera Capstone Project

- Project Overview -

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

Introduction
Kuala Lumpur has been fighting with congested streets and a constant lack of free parking space especially in its commercial districts. On the one hand citizens need to park their cars while they are at work, on the other hand malls are extremely popular for spending free time. Due to its geographical location the weather is extremely hot and humid. Therefore people prefer to stay inside to enjoy air-condition. There are many malls and office buildings that keep the city vibrant and similar new objects are on their way. It is therefore important to keep up with the need of parking space and build the required infrastructure.

Business Case
I want to find out, which locations in Kuala Lumpur (KL) are best for building a new car park. Using the techniques taught during the other courses of the IBM Data Science program, I will use FourSquare and various machine learning 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 time in their cars looking for a parking space.

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 parking spaces.
3. Geographical data about existing car parks, so we can see where supply and demand are not matching up.

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 mall, office space and car park data for these neighborhoods.

I will then use machine learning techniques like clustering and visualization tools like Folium to provide an answer to the question above.