The Battle of Neighbourhoods week-IV Report

Project Title: Analyzing and Recommending the best locations for opening hotel in the city of Kuala Lumpur, Malaysia.

Introduction:

For many tourists, booking good hotel is very tedious job for holidays. There are many websites available for hotel booking. The Hotels should provide good residential facility, nearer to airport and railways, easy access to transportation services and nearer to tourists spots. These provide the hotels to provide their services in a better way. Property developers are also taking advantage of this trend to build more hotels to cater to the demand. As a result, there are hotels in the city of Kuala Lumpur and many more are being built. Opening hotels allows property developers to earn consistent rental income. Of course, as with any business decision, opening a hotel requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the hotel is one of the most important decisions that will determine whether the hotel will be a success or a failure.

Business Problem:

The objective of this capstone project is to analyse and select the best locations in the city of Kuala Lumpur, Malaysia to hotel. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Kuala Lumpur, Malaysia, if a property developer is looking to open a hotel, where would you recommend that they open it?

Target Audience of this project:

This project is particularly useful to property developers and investors looking to open or invest in hotel in the capital city of Malaysia i.e. Kuala Lumpur. This project is timely as the city is currently suffering from scarcity of good hotels. Data from the National Property Information Centre (NAPIC) released last year showed that an additional 15 per cent will be added to existing hotel space, and the agency predicted that total occupancy may dip below 86 per cent. The local newspaper The Malay Mail also reported in March last year that the true occupancy rates in Hotel may be as low as 40 per cent in some areas, quoting a Financial Times (FT) article cataloguing the country's continued obsession with building more hotel space .

Data

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

> List of neighbourhoods in Kuala Lumpur. This defines the scope of this project which is confined to the city of Kuala Lumpur, the capital city of the country of Malaysia in South East Asia.

> Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

> Venue data, particularly data related to Hotel. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them:

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur) contains a list of neighbourhoods in Kuala Lumpur, with a total of 70 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful soup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package

which will give us the latitude and longitude coordinates of the neighbourhoods.

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers.

Foursquare API will provide many categories of the venue data, we are particularly interested in the hotel category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.