Coursera Capstone

IBM Data Science Professional Certificate

Opening a new shopping mall in Kualalampur, Malaysia.

By, Meka Aditya

December,2019

Introduction

For many shoppers, visiting shopping malls is a great way to relax themselves on weekends and holidays. They can do grocery shopping, dine at restaurants, watch movies and perform many more activities. Shopping malls are like a one stop destination for all these activities. For retailers, the central location and the large number of crowds provides a great distribution channel for marketing their products and services. Property developers are also taking advantage of this trend and are trying to open a greater number of shopping malls to cater the increasing demand. As a result, there are many shopping malls in the city of Kualalampur and many more are being built. Opening shopping malls provides a stable source of rental income for the property developers. Of course, opening a shopping mall is not as simple as it looks because of its numerous business considerations before opening it. Particularly, it is the location of the shopping mall that will play a major role in determining whether the mall will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in Kualalampur, Malaysia to

open a shopping mall. 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 Kualalampur, Malaysia, if a property developer is looking to open a new shopping mall, where should you recommend that they open it?

Target audience of this project

This is particularly useful for property developers and investors who are looking to open or invest in new shopping malls in the city of Malaysia i.e. Kualalampur. This project is timely as the city is suffering from an oversupply of shopping malls.

Data

To solve the problem, we need the following data:

1.List of neighbourhoods in Kualalampur. This defines the scope of the project which is confined to the city of Kualalampur, Malaysia.

2.Latitude and longitude coordinates of neighbourhoods. This is required to plot the map and also get the venue data.

3.Venue data, particularly data related to the shopping malls. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

This Wikipedia page consists of a list of neighbourhoods in Kualalampur (https://en.wikipedia.org/wiki/Category:Suburbs_in_K uala_Lumpur) with a total of 71 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighbourhoods using python geocoder package which will give us the latitudes and longitudes of the neighbourhoods.

After that, we will use Foursquare API to get the venue data of these neighbourhoods. Foursquare has one of the largest databases of more than 105 million places and is used by over 125,000 developers. Foursquare will provide may categories of venue data, we are particularly interested in shopping mall category which will help us to solve our problem. This project will make use of many data science skills such as 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