My Coursera Capstone Project

IBM Applied Data Science Professional

Launching a Multipurpose Entertainment Center in Kuala Lumpur, Malaysia

Report By: ***Muhammad Mohsin Javaid***

August 2019

Introduction to The Idea

For many citizens, visiting entertainment centers is the best getaway with family & friends to enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities. Multipurpose entertainment malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the multipurpose entertainment malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more multipurpose entertainment malls to cater to the demand. As a result, there are many entertainment centers in the city of Kuala Lumpur and a considerable more are being built. Opening multipurpose entertainment malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new multipurpose entertainment mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the multipurpose entertainment mall is one of the most important decisions that will determine whether the mall will be a success or a failure. Moreover, growing population in the city of Kuala Lumpur with hike in its inhabitants is making it fairly clear that the saturation point is not even close yet. Therefore, this can be a huge profit-making idea to be executed in near term.

Goals in Sight

The objective of my project is to analyze and select the most feasible location in the vicinity of this city to open a new multipurpose entertainment center. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the questions at hand:

“For an investor, what is the perfect location in city of Kuala Lumpur, Malaysia to open a new multipurpose entertainment center to maximize the footfall, enhance sales and boost profits?”

A recommendation will follow our analysis.

Ultimate beneficiary

This output of this project will be useful to property developers and investors looking to invest in new multipurpose entertainment centers in Kuala Lumpur. The outputs of this data will be beneficial as the rising number of entertainment centers in the city calls for carefully selecting the location to build this center. Data from the National Property Information Centre (NAPIC) released last year showed that an additional 15 per cent will be added to existing mall 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 malls 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 shopping space despite chronic oversupply. These adverse number suggest that an investor conducts a very thorough investment feasibility analysis before jumping in with a hefty investment.

Data

For our problem, we require:

• List of neighborhoods in Kuala Lumpur. This defines the scope of this project which confirms a number of variables in our project including population division across the city and its disparity with other areas.

• Latitude and longitude coordinates of all our sample neighborhoods. This is required to plot the map and to get the venue data. Foursquare will be used for this purpose to lead us to an accurate recommendation.

• Selected venue details, particularly data related to multipurpose entertainment malls. We will use this data to cluster the neighborhoods in our sample.

Data Sources & Methodology Adopted

This Wikipedia page **(https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur)** contains a list of neighborhoods in Kuala Lumpur, with a total of 70 neighborhoods.

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 neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 110+ 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 Entertainment Center 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, the python notebook used, and ultimately the recommendation given.

References

1. https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur

……Continued in Week 5 Assignment