

Project Report of Data Analysis on Plan of Opening A New Shopping Mall in Kuala Lumpur, Malaysia

Prepared By: Winter Melon
Date: Dec 2020

Introduction Section

For most people, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can go to supermarket for daily supplies, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities. Shopping malls are like a one-stop termination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more shopping malls to cater to the demand. As a result, there are many shopping malls in the city of Kuala Lumpur and many more are being built. Opening shopping malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

Business Problem Section

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

This project is particularly useful for property developers and investors

Assignment on IBM Applied Data Science Capstone

looking to open or invest in new shopping malls in the capital city of Malaysia i.e. Kuala Lumpur. This project is timely as the city is currently suffering from oversupply of shopping malls. 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.

Data Section

To solve the problem, we will need the following data such as list of neighborhoods 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 neighborhoods. This is required in order to plot the map and also to get the venue data. Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

This Wikipedia page with the link showing as below:

(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 beautiful soup 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 neighbourhoods.

Assignment on IBM Applied Data Science Capstone

Foursquare has one of the largest databases 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 Shopping Mall 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.