

Coursera course – IBM Data Science Professional Certificate

Capstone Project -

Opening a shopping mall in Kuala Lumpur, Malaysia

Introduction:

For most of us, visiting shopping malls is a great way to relax and enjoy ourselves during weekends. We do our grocery shopping, dine at restaurants, shop at the fashion outlets, watch movies and do so many more things. Shopping malls are a one-stop destination for all type of shoppers. For retailers, the central location and the large crowd provides a great distribution channel to market their products and services. Property developers also take 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 in Malaysia. Shopping malls paves way for property developers to earn rental income consistently. Opening a new shopping mall requires a serious consideration as with any business decision and it is very complicated. Mostly importantly and 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:

The objective of this project is to analyze and select the best neighbourhood or location in Kuala Lumpur, Malaysia to open a new shopping mall. Using data science techniques and machine learning algorithm like clustering, we aim to provide a solution to this Business problem. That is, where would we recommend opening a new shopping mall in Kuala Lumpur, Malaysia?

Target Audience:

The answer to this problem would greatly benefit property developers and investors who would love to open or invest in new shopping malls in Kuala Lumpur.

Data:

For this project, we would need the following data.

The list of neighbourhoods in Kuala Lumpur.

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

Venue data related to shopping malls. This data will be used to perform clustering on neighbourhoods.

Sources of data and how to get them

This Wikipedia page https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur has the list of neighbourhoods in Kuala Lumpur, about 70 neighbours in total. Web scraping will be used to extract the data from the Wikipedia page by using the Python requests and BeautifulSoup packages. Then, the geographical coordinates of the neighbours – the latitudes and longitudes will be got using the python Geocoder package. Then, the venue data for the neighbourhoods will be got by using the Foursquare API.