Report – Open new shopping mall in KualaLumpur

Introduction –

There are many shopping malls in the city of Kuala Lumpur and many more are being built currently. Opening shopping malls allows Property Developers to earn consistent rental income. Opening a new shopping mall requires serious consideration of many aspects 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 –

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 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 new shopping mall, where would you recommend that they should open it?

Target audience for this project –

This project is particularly useful to Property Developers and Real Estate Investors looking to open or invest in new shopping malls in Kuala Lumpur Malaysia.

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 Kuala Lumpur city. 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 shopping malls. We will later use this data to perform clustering on the neighbourhoods.

Sources of data and How to extract them –

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur) contains a list of neighbourhoods in Kuala Lumpur.

Methodology –

Firstly, we need to get the list of neighbourhoods in the city of Kuala Lumpur. This list is available at the Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur). We will do web scraping using Python requests and beautifulsoup packages to extract the list of neighbourhoods data. We then need to get geographical coordinates in the form of latitude and longitude in order to use Foursquare API. We will use Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package allowing us to perform a check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Kuala Lumpur. Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters.

We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. We can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Lastly, we will perform clustering on the data by using k-means clustering. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for “Shopping Mall”. The results will allow us to identify which neighbourhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

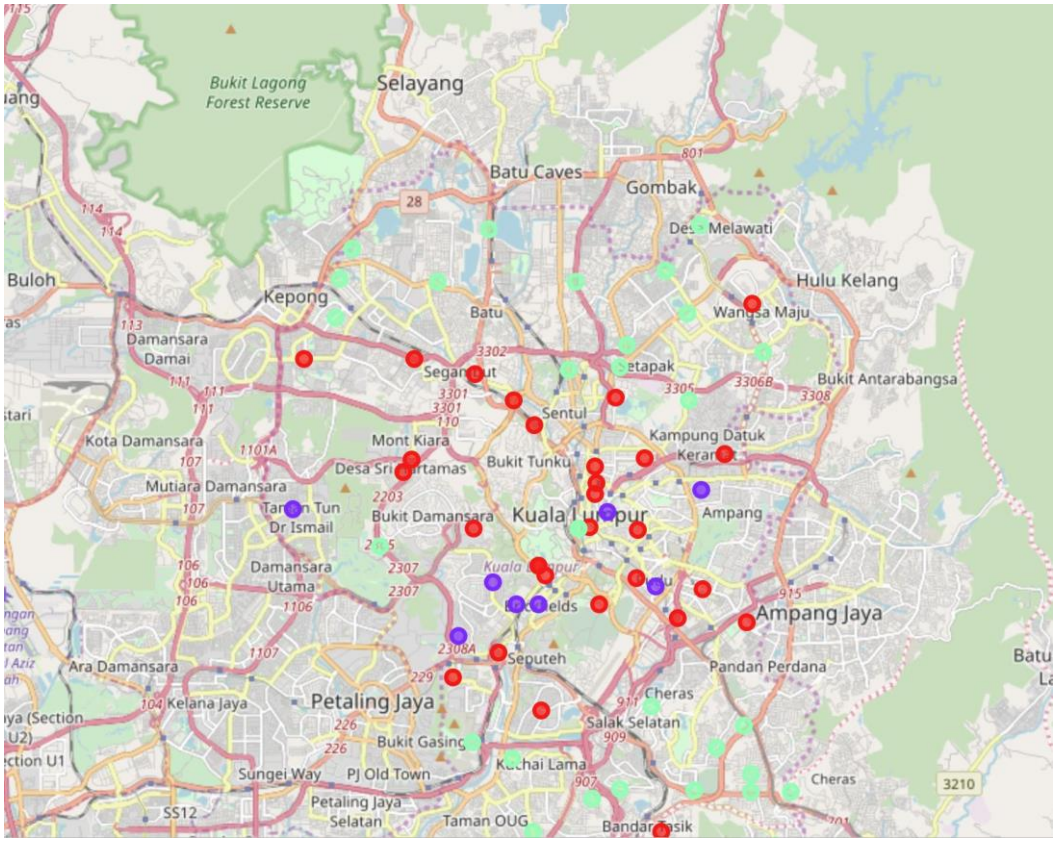
Results –

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Shopping Mall”:

Cluster 2 has no shopping malls. Cluster 1 has highest shopping malls. Cluster 0 has moderate number of shopping malls.

Cluster 2 is the best neighbourhood cluster for opening a new shopping mall as there is very little to no competition from existing malls. Cluster 0 neighbourhoods can also be considered as next alternative to open new shopping malls. It is advised for Property Developers not to open new shopping malls in Cluster 1 neighbourhoods because of very high competition.

The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour.



Discussion -

As observations noted from the map in the Results section, most of the shopping malls are concentrated in the central area of Kuala Lumpur city, with the highest number in Cluster 1 and moderate number in Cluster 0. On the other hand, Cluster 2 has very low number to no shopping mall in the neighbourhoods. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls. Meanwhile, shopping malls in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of shopping malls. From another perspective, the results also show that the oversupply of shopping malls mostly happened

in the central area of the city, with the suburb area still have very few shopping malls. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighbourhoods in Cluster 2 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighbourhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighbourhoods in Cluster 1 which already have high concentration of shopping malls and suffering from intense competition.

Conclusion -

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and providing recommendations to the relevant stakeholders i.e. Property Developers and Real Estate Investors regarding the best locations to open a new shopping mall. To answer the business question that was raised in the introduction section, the answer proposed by this project is:

The neighbourhoods in cluster 2 are the most preferred locations to open a new shopping mall. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall.

References -

Wikipedia Page -

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