Feasibility Analysis: Opening a Shopping mall in Kuala Lumpur

Introduction:

Shopping malls have started to become increasingly popular in the 21st century. Especially when people go for leisurely travel, shopping is an absolute necessity for most. Also, at present, the shopping malls offer a variety of activities other than shopping itself. One of the most popular places for shopping malls in East Asia is none other than Kuala Lumpur. The city is trendy among tourists from different parts of the world and is one of the favourite places for shopaholics. Starting from the huge shopping malls like Times Square and Surya KLCC to the small stalls of Petaling Street, Kuala Lumpur is full of places to shop. Although opening a shopping mall in Kuala Lumpur can be beneficial, the margin of profit will differ depending on the area. The realtors must undergo serious consideration before deciding to build a shopping mall on a respective location.

The objective of this project is to find out the best potential locations to build a shopping mall in the city of Kuala Lumpur. Using Data Science Methodologies, this project aims to find out if one is to make a Shopping Mall in Kuala Lumpur, which would be the best location to do so.

Data:

The data used for this project are as following:

- List of suburbs in Kuala Lumpur. This data will help in narrowing down the target areas where the investors/realtors can build shopping malls.
- Latitude and Longitude of those suburbs. This data is necessary to find out if there are any shopping malls in these areas. This data will help in plotting the map and provide us with the venue data.
- Data of the venues. Specifically, data related to shopping malls. Using this data, we can perform clustering to find out the amount of shopping malls in an area.

First, we will get the list of suburbs in Kuala Lumpur from this link: https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur. Then, by using web scraping, the data will be extracted from the Wikipedia page. After getting the data, we will get the geographical location of the listed places by utilizing Python Geocoder Package. Next, we will use the Foursquare API to find out the venue data. It will provide us with many categories, but since we are only interested in data related to shopping malls, we can use only that specific category and leave out the rest.

Methodology:

At first, we need to retrieve the list of neighbourhoods located in the city of Kuala Lumpur. To accomplish this task, we used the Wikipedia page located in the following link: https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur. To extract the data, we will then use the Web Scraping technique using Python requests and beautifulsoup packages. At this stage, we will get a table containing the names of the neighbourhoods. Then we will use the Python Geocoder Package to find out the coordinates of the areas in the form of longitude and latitude. After gathering this data, we will populate the data into a Pandas DataFrame and visualize the locations in a map using the Folium package. Then, we will use the Foursquare API to generate the top 100 venues within a radius of 2000 meters. We will make the API calls by passing the coordinates through a loop. Foursquare will return the venue data in JSON format, and we will extract the venue name, venue category, venue latitude, and longitude. With the data, we can check how

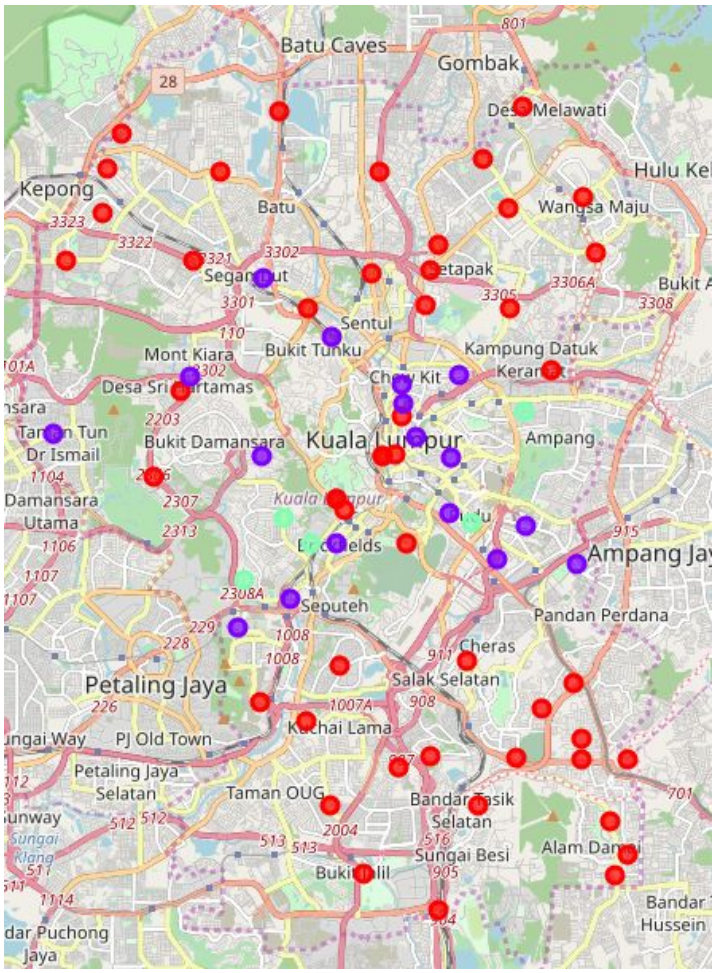
many venues were returned for each neighbourhood and examine how many unique categories we can curate from all the returned venues. Then, we will analyze each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analyzing the “Shopping Mall” data, we will filter the “Shopping Mall” as the venue category for the neighbourhoods. Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. 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 a higher concentration of shopping malls while neighbourhoods have a 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 how frequently “Shopping Mall” occurred:

- **Cluster 0:** Neighbourhoods with a low number to no existence of shopping malls
- **Cluster 1:** Neighbourhoods with a moderate number of shopping malls
- **Cluster 2:** Neighbourhoods with a high concentration of shopping malls

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



Discussion

As per the findings of the Results section, most of the shopping malls are concentrated in the central area of Kuala Lumpur city, with the highest number in cluster 2 and moderate number in cluster 1. On the other hand, cluster 0 has a meagre number to no shopping mall in the neighbourhoods which represents a great opportunity and high potential areas to open new shopping malls as it is very little to no competition from existing malls. Meanwhile, shopping malls in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of shopping malls. From another perspective, this also shows that the oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still having very few shopping malls. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighbourhoods in cluster 0 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 1 with moderate competition. Lastly, property developers should avoid neighbourhoods in cluster 2 which already have a high concentration of shopping malls and suffering from intense competition.

Limitations and Future Research

In our research, the only factor that we considered is how frequent shopping malls occur in an area. There are also other factors (i.e. population, resident's income, location type etc.) that might influence the decision of building a shopping mall greatly that are not available to such extent. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new shopping mall. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of a paid account to bypass these limitations and obtain more results.

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 lastly providing recommendations to the relevant stakeholders, i.e. property developers and investors regarding the best locations to open a new shopping mall. To answer the business question raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 0 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 in high potential locations while avoiding congested areas in their decisions to open a new shopping mall.