

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, but 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 mall in an area.

At 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.