

## **New Shopping Mall in Kuala Lumpur, Malaysia**

### **Introduction**

Malls is great place to relax and enjoy weekends and holidays. As weekend getaway get boring with the traffic to move out and in the city. Multiple activities in the Mall attracts viz. Watching movies at the multiplex, shopping apparels and accessories at the fashion outlets, weekly groceries shopping at the retail store, reliving stress at the game zone and much more.

### **Business case**

Central location and large footfalls is the key for any mall. Property developers are in search of the prime location for a Mall. Objective of the project is to analyse the best location for the Mall in the city of Kuala Lumpur using the data science and machine learning technique we just learnt in the course. To suggest most viable location to the developer as location is the key for success of any mall. Report will be made available to all the Property developers of Kuala Lumpur to focus on the right location in the city which is already facing issues from the oversupply of shopping malls, at the same time to the fashion outlets and other brand stores to help them identify the right place to start their store.

### **Data Requirement**

1. List of neighbourhoods of the city of Kuala Lumpur - To be obtained from Wikipedia using a web-scraping techniques and beautifulsoup packages
2. Latitude and Longitudes coordinates of the neighbourhoods to plot on Maps - To be obtained using the FourSquare API
3. Venue details of the shopping malls to form clusters

### **Sources of data and brief steps followed**

- Wikipedia page for list of neighbourhoods in Kuala Lumpur, with 71 neighbourhoods. Use of web scraping techniques to extract the data from the Wikipedia page.
- Used Python Geocoder to get geographical (latitude and longitude) coordinates of the neighbourhoods of the neighbourhoods.
- Used Foursquare <sup>1</sup>API to get the venue data for those neighbourhoods. Obtained venue details for the Shopping Mall category in order to help us to solve the business problem.
- Used data cleaning, data wrangling, machine learning (K-means clustering) and map visualization (Folium) to solve business problem and present solution.

---

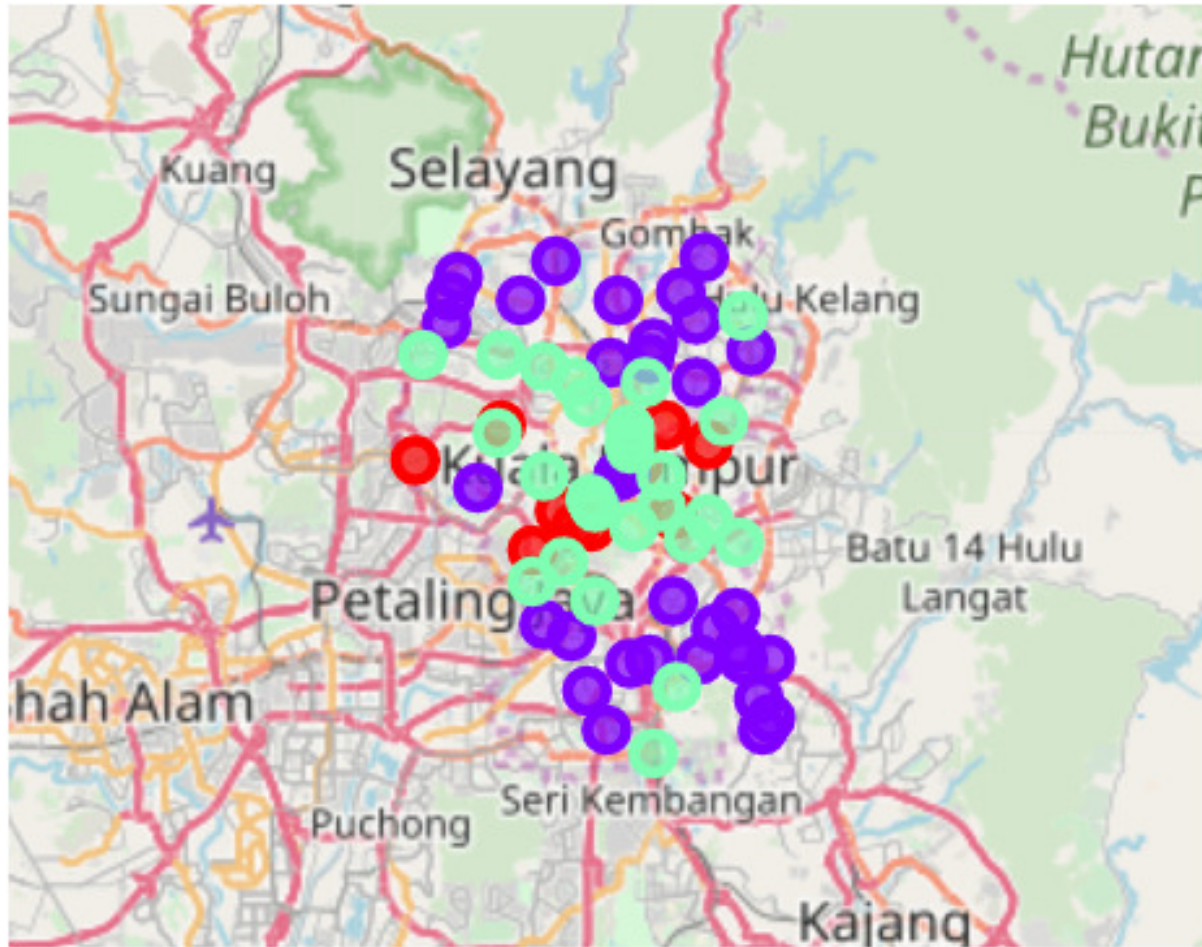
<sup>1</sup> Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers.

## Results

Based on k-means clustering we have divided neighbourhoods into 3 clusters based on the frequency of occurrence for "Shopping Mall":

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

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 green colour.



## Observation

Shopping malls are concentrated in the central area of Kuala Lumpur city, with the highest number in cluster 2 and moderate number in cluster 0. On the other hand, cluster 1 has very low number to totally no shopping mall in the neighborhoods. Its a great opportunity and high potential areas to open new shopping malls as there is very little competition. Meanwhile, shopping malls in cluster 2 are facing intense competition due to oversupply of shopping malls.

Therefore, this project recommends property developers to open new shopping malls in neighborhoods in cluster 1 with near to no competition.

## Limitations

Project is based only on one factor i.e. frequency of occurrence of shopping malls, there are several other factors such as population and income of residents that could influence the location decision of a new shopping mall. Since, this data was not available for the neighbourhood, it was not considered for this project. 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.

## 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 three clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to start a new shopping mall.

The neighbourhoods in cluster 1 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

Category: Suburbs in Kuala Lumpur. *Wikipedia*.

([https://en.wikipedia.org/wiki/Category:Suburbs\\_in\\_Kuala\\_Lumpur](https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur))

Foursquare Developers Documentation. *Foursquare*. (<https://developer.foursquare.com/docs>)