# ***My Coursera Capstone Project***

**IBM Applied Data Science Professional**

## Launching a Multipurpose Entertainment Center in Kuala Lumpur, Malaysia

Report By: ***Muhammad Mohsin Javaid***

***August 2019***

# Introduction to The Idea

For many citizens, visiting entertainment centers is the best getaway with family & friends to enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and perform many more activities. Multipurpose entertainment malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the multipurpose entertainment malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more multipurpose entertainment malls to cater to the demand. As a result, there are many entertainment centers in the city of Kuala Lumpur and a considerable more are being built. Opening multipurpose entertainment malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new multipurpose entertainment mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the multipurpose entertainment mall is one of the most important decisions that will determine whether the mall will be a success or a failure. Moreover, growing population in the city of Kuala Lumpur with hike in its inhabitants is making it fairly clear that the saturation point is not even close yet. Therefore, this can be a huge profit-making idea to be executed in near term.

## Goals in Sight

The objective of my project is to analyze and select the most feasible location in the vicinity of this city to open a new multipurpose entertainment center. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the questions at hand:

“For an investor, what is the perfect location in city of Kuala Lumpur, Malaysia to open a new multipurpose entertainment center to maximize the footfall, enhance sales and boost profits?”

A recommendation will follow our analysis.

## Ultimate beneficiary

This output of this project will be useful to property developers and investors looking to invest in new multipurpose entertainment centers in Kuala Lumpur. The outputs of this data will be beneficial as the rising number of entertainment centers in the city calls for carefully selecting the location to build this center. Data from the National Property Information Centre (NAPIC) released last year showed that an additional 15 per cent will be added to existing mall space, and the agency predicted that total occupancy may dip below 86 per cent. The local newspaper The Malay Mail also reported in March last year that the true occupancy rates in malls may be as low as 40 per cent in some areas, quoting a Financial Times (FT) article cataloguing the country's continued obsession with building more shopping space despite chronic oversupply. These adverse number suggest that an investor conducts a very thorough investment feasibility analysis before jumping in with a hefty investment.

## Data

For our problem, we require:

• List of neighborhoods in Kuala Lumpur. This defines the scope of this project which confirms a number of variables in our project including population division across the city and its disparity with other areas.

• Latitude and longitude coordinates of all our sample neighborhoods. This is required to plot the map and to get the venue data. Foursquare will be used for this purpose to lead us to an accurate recommendation.

• Selected venue details, particularly data related to multipurpose entertainment malls. We will use this data to cluster the neighborhoods in our sample.

## Data Sources & Methodology Adopted

This Wikipedia page **(https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur)** contains a list of neighborhoods in Kuala Lumpur, with a total of 70 neighborhoods.

We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 110+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Entertainment Center category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used, the python notebook used, and ultimately the recommendation given.

## Methodology

Firstly, we need to get the list of neighborhoods in the city of Kuala Lumpur. Fortunately, the list is available in 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 neighborhoods data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful 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 data frame and then visualize the neighborhoods in a map using Folium package. This allows us to perform a sanity 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 need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare

secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighborhoods 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. With the data, we can check how many venues were returned for each neighborhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighborhood by grouping the rows by neighborhood 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 venue category for the neighborhoods.

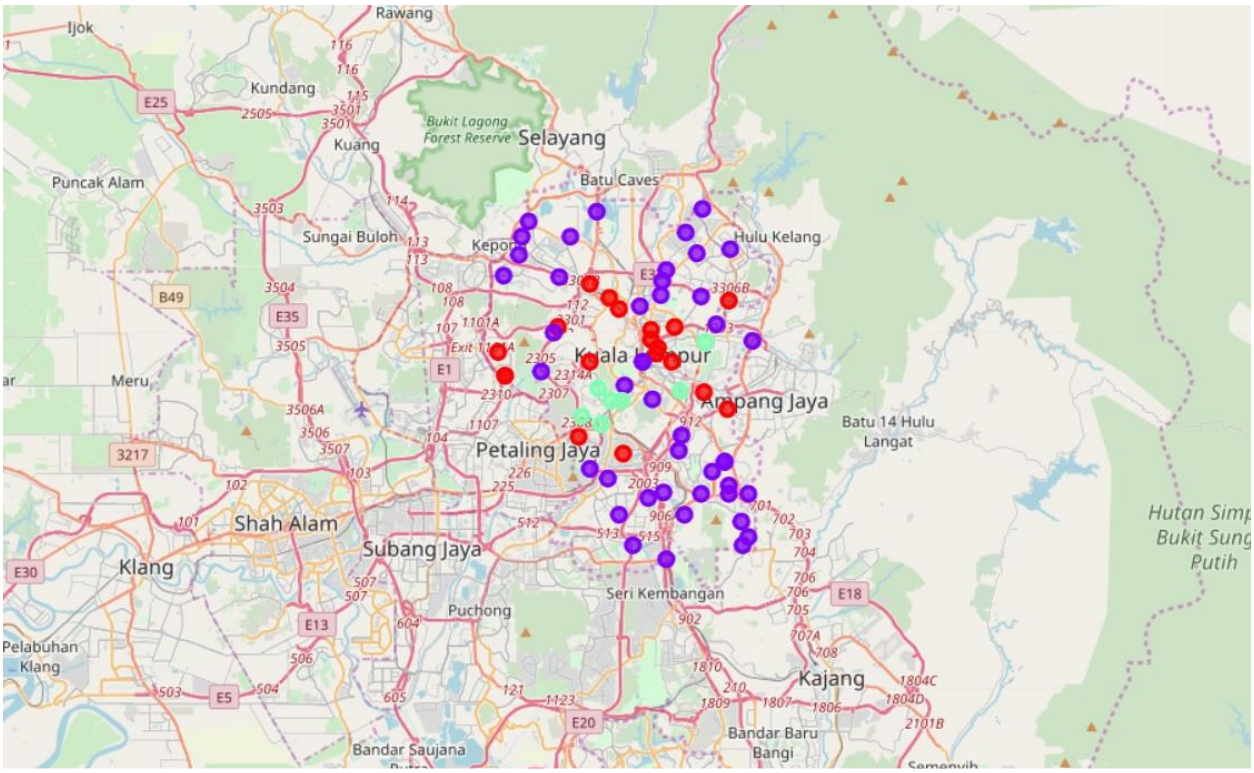
Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighborhoods into 3 clusters based on their frequency of occurrence for "Shopping Mall". The results will allow us to identify which neighborhoods have higher concentration of shopping malls while which neighborhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighborhoods, it will help us to answer the question as to which neighborhoods are most suitable to open new shopping malls.

## Results

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

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

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




## 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 2 and moderate number in cluster 0. On the other hand, cluster 1 has very low number to no shopping mall in the neighborhoods. 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 2 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 neighborhoods in cluster 1 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in cluster 0 with moderate competition. Lastly, property developers are advised to avoid neighborhoods in cluster 2 which already have high concentration of shopping malls and suffering from intense competition.

## Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of shopping malls, there are other factors such as population and income of residents that could influence the location decision of a new shopping mall. However, to the best knowledge of this researcher such data are not available to the neighborhood level required by this project. 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 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 that was raised in the introduction section, the answer proposed by this project is: The neighborhoods 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**


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

2. Foursquare Developers Documentation. *Foursquare*. Retrieved from https://developer.foursquare.com/docs

3. Malay Mail. (2018, March 14). Malls facing meltdown as glut continues. *Malay Mail*. Retrieved from https://www.malaymail.com/s/1597735/malls-facing-meltdown-as-glut-continues

4. Tan, H. H. (2018, April 5). An oversupply of retail space in Malaysia. *StarProperty.my*. Retrieved from http://www.starproperty.my/index.php/articles/property-news/an-oversupply-of-retail-space-in-malaysia/