Coursera Capstone Project

# IBM Applied Data Science Capstone

Opening a New Shopping Mall in Kuala Lumpur, Malaysia

## By: Arjun Thottappillil

**Data**

**To find an optimal solution to this issue, we will be using the following data:**

* A list of all the neighborhoods in Kuala Lumpur. This will help use determine the scope of the project which will be restricted to the capital city of Malaysia in South East Asia, Kuala Lumpur.
* Latitude and Longitude coordinates from these neighborhoods. For creating a plotted map and to get the venue data.
* Venue data, specifically data retaining to shopping malls. We will make use of this data to perform clustering on the neighborhoods.

**Data Sources and Tools**

Wikipedia provides the list of all 71 neighborhoods in Kuala Lumpur, Malaysia. (<https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur>)

The Python Geocoder package provides the means to calculate the latitude and longitude for all neighborhoods.

The Foursquare API provides the venue data for these neighborhoods. Foursquare is one of the largest databases of over 105 million locations and is used by over 150,000 developers. While the Foursquare API will give us all the venues near the neighborhoods, we will focus on the Shopping Mall category for our use case.

We use k-means and clustering to categorize the frequency of the venues and then filter them for only the shopping malls nearby and then record it for the next step.

Folium is used for the final map visualization of the data.