Coursera Capstone Project

# IBM Applied Data Science Capstone

Opening a New Shopping Mall in Kuala Lumpur, Malaysia

## By: Arjun Thottappillil

A view of a city with tall buildings in the background

Description automatically generated

**Introduction**

For many people, shopping malls are locations for them to relax and unwind on the weekends and holidays. Malls are an all-in-one destination for both residents and tourists, by allowing them a multitude of options to spend their time with. For the retail store owners, the central location and large number of potential visitors that most shopping malls attract serves as an ideal distribution channel for them to market their products and services. Property developers can receive a steady stream of rental income from retailer stores who join them. However, the success of a shopping mall is greatly dependent on its location. Some of the problems that can arise from choosing bad locations range from oversupply to intense competition between shopping malls that are in proximity of one another.

**Business Problem**

The main objective of this capstone project is to analyze and then select the ideal locations in the city of Kuala Lumpur, Malaysia in which to build a new shopping mall. Through the use of data science methodology and a variety of machine learning techniques like clustering an k-means, the end goal of this project is to find the best solution to the business question: In the city of Kuala Lumpur, Malaysia, for a property develop seeking to open a new shopping mall, where is the ideal location to build the mall?

**Target Audience of this project**

The main benefiters of this project would be the property developers and investors who are interested in opening or investing in the opening of new shopping malls in the capital city of Malaysia, Kuala Lumpur. In recent years Kuala Lumpur has had issues with oversupply of their shopping malls in one location rather than being more distributed and spread out. This project can serve as a guide to opening malls in more strategic locations to avoid the oversupply issue and increase the profitability of the mall.

**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.