

# **Coursera IBM Data Science Capstone Project**

## **Opening New Malay Restaurant in Kuala Lumpur, Malaysia**

### **Data Description:**

The data set that I have used for solving the problem is:

- A complete list of neighborhoods in Kuala Lumpur, Malaysia. Source of the data is Wikipedia.org
- Geographical coordinates (latitude and longitude) of those neighborhoods. Source of the data will be FourSquare.
- FourSquare provided Venue data which is related to Malay restaurants. We will use this data to perform clustering on the neighborhoods.

### **Data Sources**

This wikipedia page [https://en.wikipedia.org/wiki/Kuala\\_Lumpur](https://en.wikipedia.org/wiki/Kuala_Lumpur) contains a list of neighborhoods in Kuala Lumpur, with a total of 11 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 neighbors. After that, we will use the Foursquare API to get the venue data for those neighborhoods.

Now that we know the data we need, we use the Foursquare API to get the venue data for the environment. Foursquare is one of the largest databases with 105+ million places and is used by more than 125,000 developers. Foursquare provides many categories of venue data, and what I use here is the Malay restaurants venue data. This is a project that will take advantage of many data science skills, from web scraping (Wikipedia), working with APIs (Foursquare), data cleaning, data disputing, 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 we performed and the machine learning techniques used.