IBM DATA SCIENCE

COURSERA CAPSTONE

PROJECT:

**PROJECT TOPIC** -->

**OPENING A NEW SHOPPING MALL IN KUALA LUMPUR, MALAYSIA**

BY🡪 SHIVANI GUPTA

 **Malaysia**

**INTRODUCTION:**

1. It is very beneficial for a business owner to set up a store in a shopping mall because shopping centres are sought-after shopping destinations, which are usually located in prime and easily accessible locations. Therefore, it is beneficial for retail store owners to rent shop space in a mall.
2. Shopping malls have products from competing producers available under one roof. So, making it easier to compare and make purchases.
3. In a shopping mall you are never left hunting for somewhere to eat. Shopping malls are filled with eating options to suit any budget, like restaurants, diners, and food courts.

**PROBLEM STATEMENT:**

The main objective of this project is to find the best place in Kuala Lumpur, Malaysia for opening a new shopping mall.

In this project we have used data science techniques and machine learning algorithm such as clustering to find solution to the above problem.

**DATA:**

To solve the problem, we need the following data:

• List of neighbourhoods in Kuala Lumpur. This defines the scope of this project which is confined to the city of Kuala Lumpur, the capital city of the country of Malaysia in South East Asia.

• Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

**Sources of data and methods to extract them:**

**This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs\_in\_Kuala\_Lumpur) contains a list of neighbourhoods in Kuala Lumpur, with a total of 70 neighbourhoods. 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 neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.**

**After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest databases of 105+ 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 Shopping Mall 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.**

**Thankyou!**