# Business Proposal

# Opening a new café in Kuala Lumpur

## IBM Applied Data Science Capstone

By: Ahmad Syahmi Adnan
5th August 2020

## Introduction & Background Study

Coffee is the most popular beverage in the world, with more than 400 billion cups consumed each year. More than 450 million cups of coffee are consumed in the United States every day. It's the world's 2nd largest traded commodity. Coffee is consumed in great quantities, making it the most beloved beverage after water. Its worth is over $100 billion worldwide.

I am a coffee lover and have a plan to open my own café or coffee chop business in Kuala Lumpur. So the purpose of this project is to identify the best neighbourhood to open the coffee shop since there are tonnes of café in Kuala Lumpur.

## Problem Statement

The objective of this capstone project is to analyse and select the best locations in the city of Kuala Lumpur, Malaysia to open a new café. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: Which area/neighbourhood in Kuala Lumpur is the best to open a new café?

## The Location

Kuala Lumpur is the capital city of Malaysia. It is the largest city in Malaysia, covering an area of 243 $km^2$ (94 sq mi) with an estimated population of 1.96 million. Kuala Lumpur is the cultural, financial and economic centre of Malaysia and it is among the fastest growing metropolitan regions in Southeast Asia, in both population and economic develop

## Data Collection & Data Source

**Data that are needed for this project is:**

- List of neighbourhood in Kuala Lumpur.
    - Web scrapping from (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur)
- Venue data in each neighbourhood to know what kind of venue by category in each area and to extract venue related to café and coffee house.
    - From Foursquare API
- Coordinate of neighbourhoods in Kuala Lumpur to get the venue data from Foursquare API and to plot it on the map.
    - Using Geocoder

## Libraries

**Libraries Which are used to Develop the Project:**

- Pandas: For creating and manipulating dataframes.
- Folium: Python visualization library would be used to visualize the neighbourhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: For importing k-means clustering.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation and XML stores data in plain text format.
- Geocoder: To retrieve Location Data.
- Beautiful Soup and Requests: To scrap and library to handle http requests.
- Matplotlib: Python Plotting Module.