

# **IBM Applied Data Science Capstone**



## **Building a New Mall In Riyadh, Saudi Arabia (Data Description)**

By:

Marwan Akeel

# Data Types

In this project we need three types of data:

1. Riyadh Neighborhood names. 100 names are used.
2. Neighborhood latitudes and longitudes.
3. Neighborhood avenues (maximum 100 avenues/neighborhood).

## Data Sources

We will download Riyadh neighborhood names from Wikipedia which come within an html code (<https://en.wikipedia.org/wiki/Riyadh>). It is not up to date but still appropriate.

The data has to be extracted using web scraping techniques such as Python request and beautifulsoup package. We will get the latitude and longitude of the Riyadh city using the geocoder package. The neighborhoods latitude and longitude will be collected using (<https://www.coordinatesfinder.com>) and google map. Then, we will use the foursquare API to obtain the avenues of each neighborhood. Foursquare is a famous database that store geographical data containing millions of places from around the world. The data returned by foursquare is the venue's name, district ,tips, location, and others. We are interested only in malls number and the total venues in each neighborhood.

## How we are using the Data

We are going to use the neighborhood names to retrieve the latitude and longitude of each of them. Then sending the latitude and longitude to the foursquare database to get the venues. The more the venues returned the more commercial the neighborhood. We will train the model using the number of malls in the neighborhood and the number of venues we have in that neighborhood.