

# Investigate a Dataset Project

## Dataset:

For this project, I choose Gapminder World datasets, I picked 2 XLSX files to analyze ([indicator\\_t 25-54 labour to pop percent.xlsx](#), [indicator\\_f 25-54 labour to pop percent.xlsx](#)), then I convert them to CSV files.

## Questions Posed:

1. What are the indicators of Total Labour (%) and Female Labour (%) in the age between 25-54 in the years 2003-2007 in all countries?
2. Are there trends that can be observed between Total Labour Indicator (%) and Female Labour Indicator (%) in the age between 25-54 in the years 2003-2007 in Egypt and Singapore?

## Description:

- First of all, I import the packages and libraries such as: Pandas, Numpy and Matplotlib.
- Then, I specified the data\_frames to work on.
- After that, I started to wrangle and clean the data.
- So, I started to answer the questions (posed above) by using some functions like `.merge()` to merge the data\_frames, `.loc()` to specify rows and columns, `.describe()` to return the statistical summary and `.plot()` to visualize the returns.

## Wrangling:

For wrangling date, I used some function such as `.head()` to return the first five rows of the data\_frame, `.info()` to return all the information about the data\_frame like data type, null values, `.shape()` to return the dimension of the data\_frame, `.duplicated()` to return the duplicated values, and `.isnull()` to return the null values.

## Summary Statistics:

By using the function `.describe()`, the summary statistics for Labour Indicator and Female Indicator in World Countries, and it returned: Count of rows, Mean of each column, Standard Deviation, Min values, First Quartile (25%), Median (50%), Third Quartile (75%) and Max values.

**Plots:**

I used two kinds of plots to display the data like Scatter Plot and Line Plot.

\*\* I used N/A external resources, only the resources provided by Udacity.