**Introduction**

The dataset used can be found in Kaggle:

<https://www.kaggle.com/tomasdrietomsky/life-expectancy-gdp-per-capita-food-security>

The dataset chosen for this case study is GDP & Life Expectancy. The data was retrieved from an opensource dataset found in Kaggle, and it is made-up with two main excel sheets. The first is about life expectancy recorded over a yearly course between 2000 – 2015. A total of 264 countries will be available for analysis use, providing 2905 total rows. The second will be about the GDP of each country between 1960-2019. Each country’s life expectancy is different than another due to health development, economic stability and many more factors.

**Business Case Scenario (I created my own case):**

The business case of this study is to investigate the physical health of the population in developing and developed countries over the years, and whether GDP played a role in equation.

Attributes such as GDP, life expectancy, adult mortality, infant deaths are all promising attributes to analyze and find interesting predictions. Two countries have been chosen to narrow the focus of the study and make it more impactful and specific. The countries are first divided into two sections:

**Business Questions:**

* What is the future direction of GDP in the stated countries?
* What is the future direction for life expectancy, and some of the other variable in the stated countries?
* Which attributes are affected by the high/low GDP of the stated countries?

**The way to a successful forecast goes through various steps:**

• Cleaning the data.

• Analyzing, visualizing & understanding the data.

• Perform Forecasting and predictive methods.

• Provide Recommendations based on results.

**I. Data and Preprocessing (cleaning, missing values, recoding)**

Cleaning and checking for missing data were done in Excel.

The columns of years between 1960-1999 and 2016-2019 will also be removed, due to having only the years between 2000-2015 in the Life Expectancy sheet. Due to the lack of documentation, the following attributes were removed from both sheets:

* Schooling
* Percentage Expenditure
* Total expenditure
* Income composition of resources
* Hepatitis
* Country Code
* Indicator Name
* Indicator Code

#Libraries that will be used for this project

library(readxl)

library(readr)

library(fpp3)

library(ggplot2)

library(tsibble)

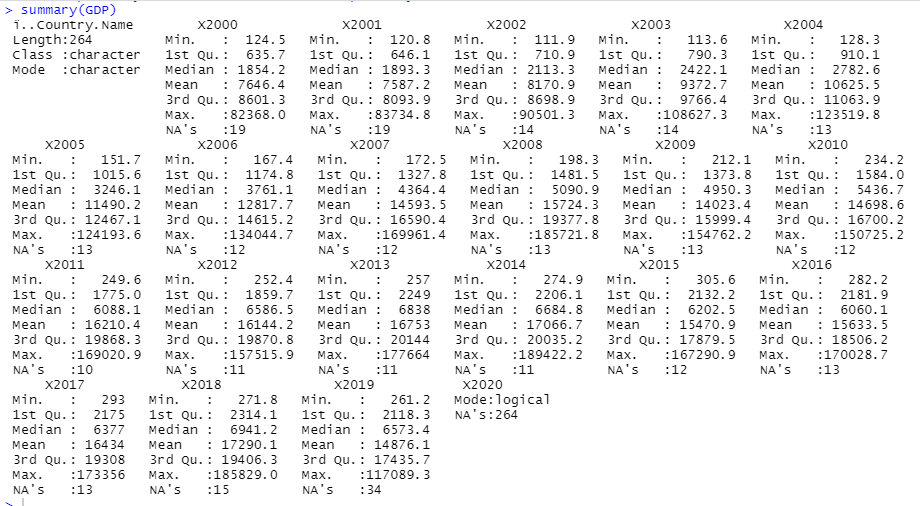
library(feasts)

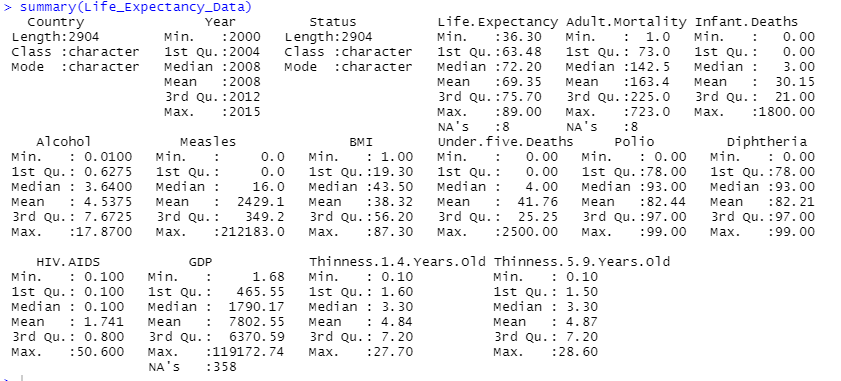
# II. Visualization and Insights

#Information about the dataset

summary(Life\_Expectancy\_Data)

view(Life\_Expectancy\_Data)





# III. Machine Learning - Making Predictions[¶](https://www.kaggle.com/code/andradaolteanu/model-and-visualize-mental-health-in-tech#III.-Machine-Learning---Making-Predictions)

# Conclusions

