# **+Business Case:**

The Global Health Observatory (GHO) data repository under World Health Organization (WHO) keeps track of the health status as well as many other related factors for all countries. The data-set related to life expectancy, health factors for 193 countries have been collected from the same WHO data repository website and its corresponding economic data was collected from the United Nation website.  It has been observed that in the past 15 years, there has been a huge development in the health sector resulting in improvement of human mortality rates especially in the developing nations in comparison to the past 30 years.

WHO has vast data for various verticals that can help humans to strive and survive under extreme conditions? You have been hired as a Data Analyst to find the hidden key insights that can solve critical questions and provide a solution.

# **Problem Statement:**

Health plays a vital role in deciding Life Expectancy. People all over the world pay a lot to insurance companies to secure the finance if there is a health emergency. It becomes very important to understand which factors directly or indirectly affect life expectancy.

1. What are the predicting variables actually affecting life expectancy?
2. Should a country having a lower life expectancy value (<65) increase its healthcare expenditure to improve its average lifespan?
3. How do Infant and Adult mortality rates affect life expectancy?
4. Does Life Expectancy have a positive or negative correlation with eating habits, lifestyle, exercise, smoking, drinking alcohol, etc?
5. What is the impact of schooling on the lifespan of humans?
6. Does Life Expectancy have a positive or negative relationship with drinking alcohol?
7. Do densely populated countries tend to have a lower life expectancy?
8. What is the impact of Immunization coverage on life Expectancy?

Your role is to draw insights into the countries and their expenses pattern. These solutions can play a critical role regarding human survival. Through the chain of scenarios designed, you will be able to solve problems and follow the process that real Data Analysts perform.

# **Dataset Description:**

**Country Data**

Table showing the metrics of the country with important parameters.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| Country | Categorical | Country name |
| Year | Discrete | Year |
| Status | Categorical | Status of the given country (Either Developing or Developed) |
| Life expectancy | Continuous | in years |
| Adult Mortality | Continuous | Adult Mortality Rates of both sexes (the probability that those who have reached age 15 will die before reaching age 60 (shown per 1,000 persons)) |
| Alcohol | Continuous | Alcohol, recorded per capita (15+) consumption (in liters of pure alcohol) |
| Total expenditure | Continuous | General government expenditure on health as a percentage of total government expenditure (%) |
| GDP | Continuous | Gross Domestic Product per capita (in USD) |
| Population | Continuous | The population of the country |
| thinness 1 - 19 years | Continuous | Prevalence of thinness among children and adolescents for Age 10 to 19 (%) |
| thinness 5 - 9 years | Continuous | Prevalence of thinness among children for Age 5 to 9(%) |
| Income composition of resources (ICOR) | Continuous | Human Development Index in terms of income composition of resources (index ranging from 0 to 1) |
| Schooling | Continuous | Number of years of Schooling(years) |

**Immunization Data**

Table showing diseases, mortality, and immunization coverage factors of the country.

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| Country | Categorical | Country name |
| Year | Discrete | Year |
| Infant deaths | Continuous | It is the death of an infant before his or her first birthday. Number of Infant Deaths per 1000 population |
| Hepatitis B | Continuous | Hepatitis B (HepB) immunization coverage among 1 - year olds (%)  Hepatitis B is an infection of your liver. It’s caused by a virus. |
| Measles | Continuous | Measles  number of reported cases per 1000 population |
| BMI | Continuous | Average Body Mass Index of the entire population |
| Under-five death | Continuous | Number of under-five deaths per 1000 population |
| Polio | Continuous | Polio (Pol3) immunization coverage among 1 - year olds (%) |
| Diphtheria | Continuous | Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1 - year olds (%) |
| HIV/AID | Continuous | Deaths per 1 000 live births HIV/AIDS (0 - 4 years) |

# 

# **Scenario 1:**

You have been given the required dataset by the WHO to help the countries know about their health status, as well as, help individuals to get prepared with the upcoming problems. There are 2 datasets available that are present in CSV format and you have to explore, align and merge it if it is not in a structured format.

## Task 1:

1. Load both the datasets.
2. Explore the shapes and the dataset with .head()
3. Correct data issues (if any)
4. Check missing values in all datasets.
5. Check duplicate values in all datasets.
6. Analyze descriptive statistics.

## Task 2:

1. Merge both the dataset with the unique key (create if not available). Also, remove duplicate columns.
2. Find the continuous and categorical variables.
3. Impute missing value.

# **Scenario 2:**

After the dataset is cleaned and merged we can do univariate, bivariate, and multivariate analysis of all the variables. As life expectancy is the target variable we will analyze all possible correlations with other variables and analyze specific to developing vs developed country to understand what impacts life expectancy.

## Task 1:

1. Let’s start with a Univariate analysis.
2. Let’s do bivariate analysis.
   1. Find the highly correlated variables.
   2. Let’s compare **developed** vs **developing** countries wrt following factors.
      1. Does **GDP per capita** of the country impact life expectancy?
      2. Does the **Productive use of resources(Income composition of resources)** make life expectancy better?
      3. Does **Schooling** impact life expectancy any better?
3. Let’s do multivariate analysis.
   1. Which diseases are correlated with life expectancy?
   2. Does better immunization coverage improve life expectancy?
   3. How do countries' economic conditions affect life expectancy?

## Task 2:

1. Our target variable is **Life expectancy** so let’s start analyzing few important variables wrt it.
   1. Calculate the average life expectancy of all the years and find out Top and Bottom countries.
   2. Rank countries based on their average life expectancy.
   3. Compare a few country’s life expectancies.
   4. Compare life expectancy of **Developed** vs **Developing** country.
2. Let’s take few developing and developed countries and analyze them completely.
3. Let’s look more.

***Action****: Groupby them based on the country and take the mean of the continuous variables. Max for categorical variables.*

* 1. Should a country having a lower life expectancy value(<65) increase its healthcare expenditure to improve its average lifespan?
  2. Does Life Expectancy have a positive or negative correlation with eating habits(BMI), lifestyle, exercise, smoking, drinking alcohol, etc?
  3. Do densely populated countries tend to have a lower life expectancy?

1. Let’s conclude our analysis.

**Coding Task 1:**

Take India, the United States of America, China, the Central African Republic, Bhutan country. Analyze and share your results.

**Coding Task 2:**

After creating a new table using groupby and aggregate on Life Expectancy keeping Index as True with following features on mean.

Life Expectancy, Total expenditure, BMI, Alcohol, Population

* Should a country having a lower life expectancy value(<65) increase its healthcare expenditure to improve its average lifespan?
* Does Life Expectancy have a positive or negative correlation with eating habits(BMI), lifestyle, exercise, smoking, drinking alcohol, etc?
* Do densely populated countries tend to have a lower life expectancy? (Hint : Take 2 densely populated countries like India, and Usa)