

## **Data Synergists**

**Darimireddy, Sai Raga Mounika  
Polisetty, Yogitha  
Thummu, Anudeep  
Vennela, Venkata Ashwath**

### **Context:**

Understanding global health and demographics is crucial for shaping public health policies and interventions. Life expectancy, a key indicator of population health, is influenced by various factors such as socio-economic status, healthcare infrastructure, and disease prevalence. Our dataset offers a comprehensive look into these variables across different countries and years, providing an opportunity to predict life expectancy.

### **Specific Problem:**

Life expectancy is not solely determined by one factor but is influenced by a myriad of variables including adult mortality rates, disease prevalence, healthcare expenditure, and socio-economic indicators. Identifying the most influential factors and understanding their impact on life expectancy is essential for designing effective public health strategies and interventions.

### **Pitch:**

This project aims to utilize regression analysis to predict life expectancy based on a rich dataset encompassing various health and demographic indicators. By uncovering the key determinants of life expectancy, we can provide valuable insights into population health trends.

### **Data:**

To facilitate our exploration, we will utilize the provided dataset containing information on countries, years, development status, and various health and demographic indicators such as adult mortality, immunization coverage, disease prevalence, GDP, and education levels.

Column Descriptions:

- Country: The name of the country for which the data is recorded.
- Year: The year in which the data was collected, representing the temporal aspect of the dataset.
- Status: The development status of the country, often categorized as "Developed" or "Developing." This indicates the overall economic and social development of the country.
- Life Expectancy: The average number of years a person born in that year can expect to live, serving as an important health indicator.
- Adult Mortality: The probability of dying between the ages of 15 and 60 per 1,000 population.
- Infant Deaths: The number of infant deaths per 1,000 live births.
- Alcohol: The average alcohol consumption in liters per capita.
- Percentage Expenditure: Health expenditure as a percentage of the country's GDP.
- Hepatitis B: Immunization coverage for Hepatitis B in the country.
- Measles: The number of reported cases of measles per 1,000 population.
- BMI: The average Body Mass Index (BMI) of the population.
- Under-Five Deaths: The number of deaths of children under the age of five per 1,000 live births.
- Polio: Immunization coverage for Polio in the country.
- Total Expenditure: The total health expenditure as a percentage of the country's GDP.
- Diphtheria: Immunization coverage for Diphtheria in the country.
- HIV/AIDS: The prevalence of HIV/AIDS as a percentage of the population.
- GDP: Gross Domestic Product, representing the economic output of the country.
- Population: The total population of the country.
- Thinness 1-19 Years: Prevalence of thinness among children and adolescents aged 1-19.
- Thinness 5-9 Years: Prevalence of thinness among children aged 5-9.
- Income Composition of Resources: A composite index reflecting the income distribution and access to resources in the country.
- Schooling: The average number of years of schooling received by the population.

## **Exploratory Data Analysis:**

In this stage, we will analyze the dataset using various techniques including histograms, box plots, scatter plots, and correlation matrices to gain insights into the distribution and relationships among variables. This analysis will provide a deeper understanding of the data before proceeding to model building.

## **Goals:**

Using the dataset, our goals include:

- Identifying the most influential factors affecting life expectancy.
- Developing regression models to predict life expectancy.
- Providing insights into how different variables interact to shape life expectancy outcomes.
- Understanding how socio-economic factors and disease prevalence impact life expectancy trends.

## **Individual Efforts:**

**Mounika** will work with the overall data and identify which variables are significant and correlated to the response variables and determine the life expectancy based on the regression analysis.

**Yogitha** will explore the relationship between education levels and life expectancy and also analyze the dataset to uncover how schooling impacts life expectancy and contributes to developing predictive models.

**Ashwath** will examine the influence of socio-economic factors such as GDP and income composition on life expectancy and conduct regression analysis to understand the relationship between economic indicators and life expectancy outcomes.

**Anudeep** will investigate the role of immunization coverage and disease prevalence in shaping life expectancy and also analyze data related to immunization rates and disease prevalence to identify their impact on life expectancy trends.

## **Conclusion:**

In summary, this project aims to understand how different factors such as health and demographics affect life expectancy. Using statistical analysis and exploring the data, our goal is to offer useful insights to enhance health outcomes globally.