# Life Expectancy Analysis

## Table of Contents



## Introduction

Life expectancy is one of the most important indicators of a nation's health and development. This project explores how various socio-economic and health factors contribute to life expectancy around the world. By analyzing this data, we aim to uncover patterns and derive meaningful insights to help improve global health strategies.

## Problem Statement

**Objective:**

To investigate the factors affecting life expectancy and understand the correlations between life expectancy and features like GDP, education, immunization, and mortality rates. The goal is to:

- Identify key drivers of high/low life expectancy
- Visualize global trends
- Provide insights that can support policy-making and public health efforts

## Dataset Description

**Source:** WHO (via Kaggle)
**Link:** Life Expectancy Dataset

**Key Features Include:**

- Country

- Year
- Status (Developed / Developing)
- Life expectancy
- Adult mortality
- Infant deaths
- Alcohol consumption
- Percentage expenditure on health
- Hepatitis B immunization
- Measles reported cases
- BMI
- Under-five deaths
- GDP
- Schooling
- HIV/AIDS death rate

## Tools and Technologies Used

 **Programming Language:** Python

 **IDE:** Jupyter Notebook

 **Libraries:**

- Pandas (data handling)
- NumPy (numerical computations)
- Matplotlib & Seaborn (visualization)
- Scikit-learn

## Data Cleaning and Preprocessing

Steps involved:

- Handled missing values using mean/median or forward-fill where appropriate

- Dropped irrelevant or redundant columns
- Converted data types (e.g., year to datetime)
- Checked and corrected inconsistent data
- Encoded categorical features like 'Status'

## Exploratory Data Analysis (EDA)

Key EDA activities:

- Distribution plots of life expectancy across countries
- Heatmaps to explore correlation between features
- Time-series trends of life expectancy over years
- Box plots by income level (Developed vs Developing)
- Scatter plots to observe relationship with GDP, Schooling, and Alcohol

## Feature Engineering

Although the project is analysis-focused, some feature engineering included:

- Calculating health expenditure per capita
- Creating a normalized health index
- Transforming skewed variables (log transform for GDP, etc.)

## Modeling

This project focused on analysis and visualization, not predictive modeling. However:

- Correlation matrix was generated to identify highly correlated features
- Optional regression models (Linear Regression, Random Forest) can be used to predict life expectancy using selected features

## Results and Interpretation

- **Positive correlations:**
    - Life expectancy increases with GDP, Schooling, and Health Expenditure
    - Immunization and BMI also positively affect life expectancy
- **Negative correlations:**
    - Infant Mortality and HIV/AIDS are strongly inversely correlated with life expectancy

- Developing countries typically have lower life expectancy due to lower investment in health and education

Visualizations clearly show how consistent investment in education, economic development, and healthcare contributes to longer lives.

## Limitations and Future Scope

**Limitations:**

- Missing data in several countries for key years
- No detailed regional data (urban vs rural)
- Does not consider climate, pollution, or conflict-related factors

**Future Scope:**

- Integrate other datasets (climate, health insurance access, etc.)
- Predictive modeling to forecast life expectancy
- Dashboard creation for real-time policy simulation

## Conclusion

This analysis provides an insightful look into the global disparities and influencers of life expectancy. Education, economic stability, and healthcare access are the most crucial drivers. These findings can assist global organizations and governments in targeting areas of improvement for better health outcomes worldwide.