📝 📄 **Life Expectancy Analysis — Final Report**

Project Title:

Life Expectancy Analysis

Submitted by:

Mohini Pasi
Data Science Batch 054
Dcodetech Training Center

1️⃣ Introduction

Life expectancy is a crucial indicator of the health and development status of a country's population.
This project aims to analyze how various factors — such as economic conditions, social status, immunization coverage, and lifestyle habits — affect life expectancy across different countries.

The dataset used for this analysis was collected from the **World Health Organization (WHO)** and the **United Nations**, covering data from **193 countries** for the years **2000 to 2015**.

2️⃣ Problem Statement

In the past, many studies ignored the impact of immunization programs and human development index factors on life expectancy.
This project solves that gap by combining multiple socio-economic and health-related variables and building a regression model to identify the most significant predictors of life expectancy.

3️⃣ Objectives

- To identify which factors significantly influence life expectancy.

- To analyze whether increased healthcare expenditure improves lifespan.

- To study the impact of infant and adult mortality rates.

- To examine correlations with alcohol consumption, BMI, and lifestyle.

- To find out how education (schooling) affects lifespan.

- To provide recommendations for countries with low life expectancy.

4️⃣ Dataset Details

Description	Details
Source	WHO, United Nations
Countries	193
Years Covered	2000 – 2015
Total Records	2,938 rows
Total Columns	22 features

Key Columns:

- Life_Expectancy, Adult Mortality, infant deaths, Alcohol, BMI, Polio, Diphtheria, Hepatitis B, GDP, Schooling, Population, percentage expenditure, etc.

5️⃣ Tools & Technologies

- **Languages:** Python, SQL

- **Libraries:** Pandas, NumPy, Seaborn, Matplotlib, Scikit-Learn, Statsmodels

- **IDE:** VS Code, Jupyter Notebook

- **Other:** Excel for data cleaning

6️⃣ Methodology

Step 1: Data Cleaning

- Checked for missing values.

- Handled missing data using mean imputation for numeric columns.

- Removed unreliable rows for countries with insufficient data.

Step 2: Exploratory Data Analysis (EDA)

- Correlation matrix to study relationships between variables.

- Scatter plots for GDP vs Life Expectancy, Schooling vs Life Expectancy.

- Distribution plots to inspect variable spread.

Step 3: Model Building

- Applied Multiple Linear Regression using Scikit-Learn and Statsmodels.

- Trained the model with selected features:

    - GDP, Schooling, Alcohol, BMI, infant deaths, Polio, Hepatitis B

- Evaluated performance using R^2 Score, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

7️⃣ Results

- **Correlation Findings:**
  GDP, Schooling, Immunization (Polio, Hepatitis B) show strong positive correlation with Life Expectancy.
  Higher Alcohol consumption and higher mortality rates show negative impact.

- **Model Performance:**
  The regression model explains life expectancy variations well with good R^2 value and acceptable error margins.

8️⃣ Insights

- **Economic Development:** Countries with higher GDP per capita generally have higher life expectancy.

- **Education:** More schooling years are strongly linked to improved lifespan.

- **Immunization:** Vaccinations for Polio and Hepatitis B significantly increase average life expectancy.

- **Healthcare Spending:** Percentage expenditure shows that countries spending more on healthcare tend to have better outcomes.

9️⃣ Conclusion

This project demonstrates that improving GDP, increasing educational opportunities, and expanding immunization programs can boost life expectancy in developing countries. Policymakers can use these insights to plan health budgets, awareness campaigns, and public health interventions.

1️⃣0️⃣ References

- WHO Global Health Observatory (GHO) Data Repository

- United Nations Economic Data

- Guided under Dcodetech Training Center, Data Science Batch 054

1️⃣1️⃣ Acknowledgement

I thank **Dcodetech Training Center**, my mentors, and teammates for their valuable support and guidance in completing this project.

✅ **Thank You!**

Submitted by: Mohini Pasi
Data Science Batch 054
Dcodetech Training Center