WHO Life Expectancy – Initial Results and Code

Student Name :- Himit Patel

Student Number :- 501344641

Supervisor Name:- Dr. Ceni Babaoglu

Date:- November 5 ,2025

1. Data Analysis

- I created a new Google Colab notebook and loaded the WHO Life Expectancy CSV. After cleaning the column names, the dataset shows about 2,775 rows and 22 columns. Each row is a country–year record, and the main target is Life expectancy (years).

- I reviewed the basic statistics for all numeric columns. The average life expectancy across countries is around 70 years, with values ranging from about 40 to 85 years. Other indicators such as GDP and Schooling also vary widely between countries.

- A year-wise record check shows data available from 2000 to 2014, with about 183 country observations each year, confirming that the dataset is balanced across years.

2 Data Preparation

- I checked for missing values in the dataset. Some columns such as Hepatitis B and GDP had gaps. To keep the analysis simple and consistent, I replaced missing numeric values with their median, and missing categorical values with their most frequent category (mode). After this step, there were no missing values remaining in the dataset.

3. Model Evaluation

- I trained a simple Linear Regression model to predict Life expectancy using six predictor variables (GDP, Schooling, BMI, Alcohol, Adult Mortality, and infant deaths). I used an 80/20 train–test split. The model achieved a Mean Absolute Error of about 5 years and an R^2 score of roughly 0.75, showing that these basic factors explain most of the variation in life expectancy across countries.

4. Code Documentation

- All code and analysis files are available in my GitHub repository:
  https://github.com/Himitpatel1/WHO-Life-Expectancy-Project

- The repository includes the Google Colab notebook, PDF export, dataset, and a README file that explains how to reproduce the results.

- The code is commented for clarity, and all steps can be executed from start to finish without errors.

5. References

- World Health Organization. (2015). WHO Life Expectancy Data.

- Kaggle. (n.d.). Life Expectancy (WHO) Dataset. Retrieved from https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who

- Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

- McKinney, W. (2017). Python for Data Analysis. O'Reilly Media.

- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90–95.

- The pandas development team. (2024). pandas Documentation. Retrieved from https://pandas.pydata.org