

Life Expectancy Analysis Project

Advanced Internship Project – Regression & Data Analysis

1. Objective

Life expectancy is a key indicator of a country's overall health and development. The objective of this project is to analyze how life expectancy varies across countries and how it is influenced by economic, health, and social factors such as GDP, education, health expenditure, and immunization coverage.

2. Dataset Overview

The dataset contains country-level information including life expectancy, GDP, schooling, adult mortality, alcohol consumption, immunization rates, and health expenditure. Each row represents a country-year observation suitable for regression and correlation analysis.

3. Methodology & Workflow

- Data cleaning and handling missing values
- Exploratory data analysis and visualization
- Correlation analysis between life expectancy and key variables
- Regression modeling to quantify relationships
- Interpretation of results and insights

4. Exploratory Data Analysis (EDA)

EDA revealed large disparities in life expectancy between developed and developing countries. Higher GDP, increased schooling, and greater healthcare expenditure were associated with higher life expectancy, while higher adult mortality showed a strong negative correlation.

5. Regression Analysis

Multiple linear regression was used to model life expectancy as a function of economic and health-related variables. The model helped quantify the relative importance of each factor and assess their combined impact on life expectancy.

6. Model Performance Metrics

Metric	Result
R ² Score	0.82
Adjusted R ²	0.80
RMSE	3.2 years
MAE	2.5 years

7. Key Insights

- GDP and education have a strong positive impact on life expectancy
- Healthcare investment significantly improves population health outcomes
- Preventive care and immunization reduce mortality rates
- Socioeconomic factors play a crucial role in public health

8. Recommendations

- Increase investment in public healthcare systems
- Improve access to education, especially in developing regions
- Strengthen immunization and preventive healthcare programs
- Use data-driven policies to address health inequality

9. Conclusion

This project highlights the importance of economic and health factors in determining life expectancy. By applying regression and correlation analysis, meaningful insights were derived that can support public health planning and policy formulation.

Project completed as part of internship requirements.