

DATA ANALYSIS PROJECT

Anantha Krishnan K

PROJECT INTRODUCTION

Objective : Analyze factors affecting life expectancy across countries.

Dataset : WHO life expectancy data from 2000-2015.

Goals:

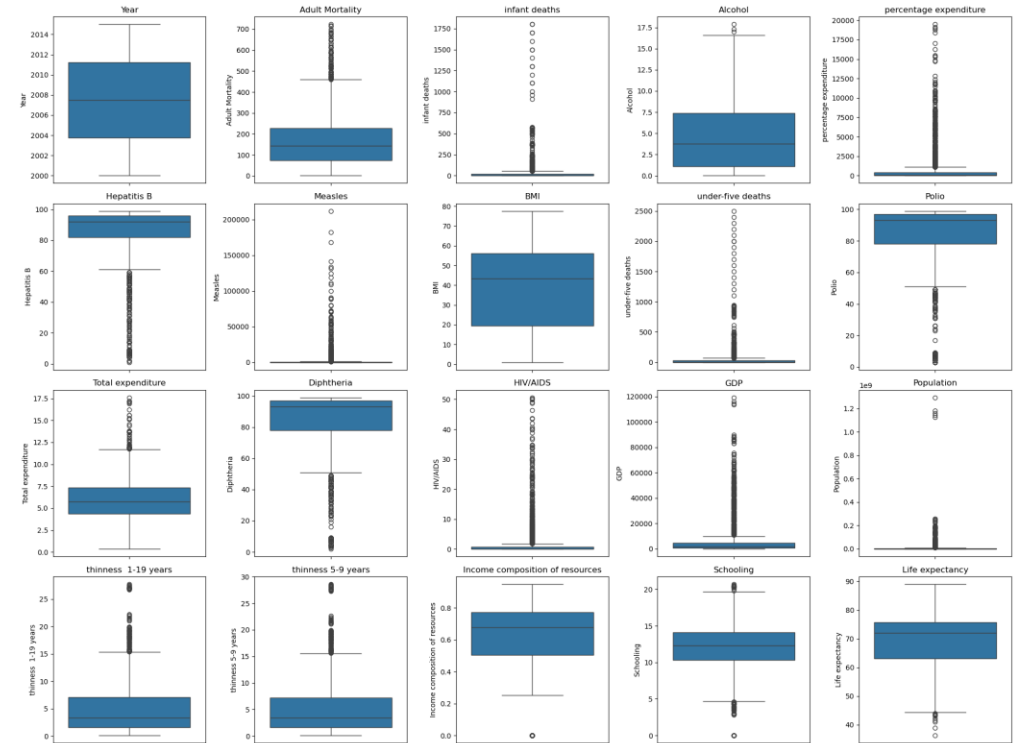
- 1. Clean and preprocess data**
- 2. Perform EDA and statistical analysis.**
- 3. Identify key influencing factors.**

DATA CLEANING

- Handling missing values using median imputation
- Treated outliers using IQR, Winsorization, and Log Transformation
- Normilized and standaridized relevant features.

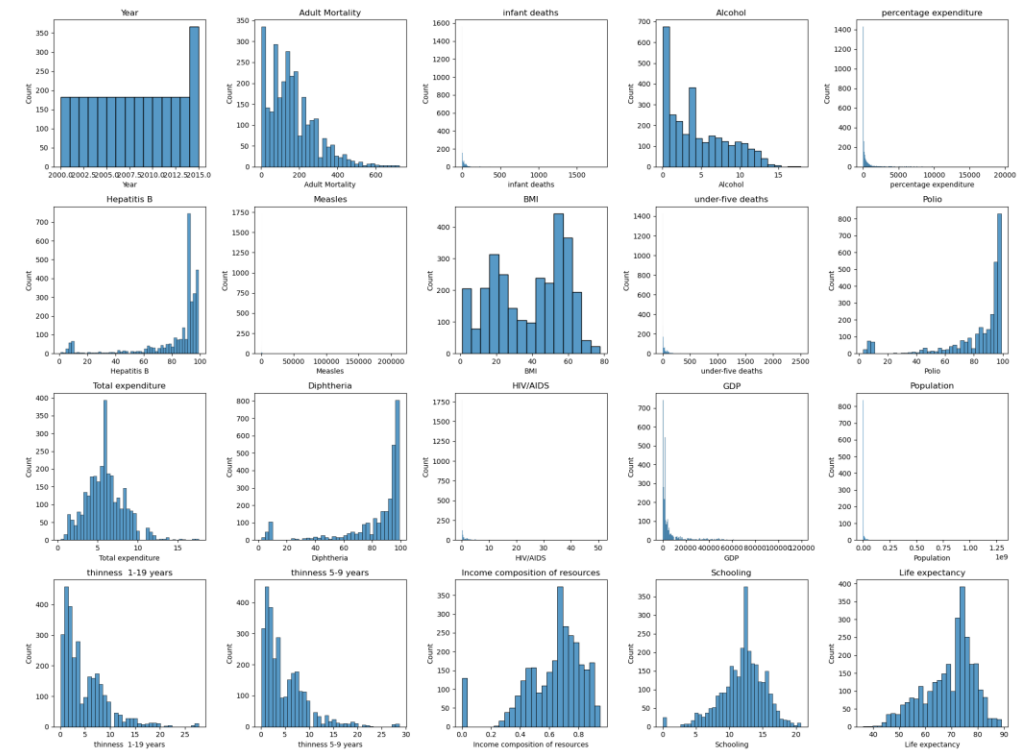
WHY WINSORIZATION ?

- Because completely removing them might reduce important variations in data.
- Winsorization is used when outliers are extreme but we don't want to remove data points because they might contain useful information.
- Instead of removing them, we cap extreme values to a certain percentile, preventing them from having too much influence.



WHY LOG TRANSFORMATION ?

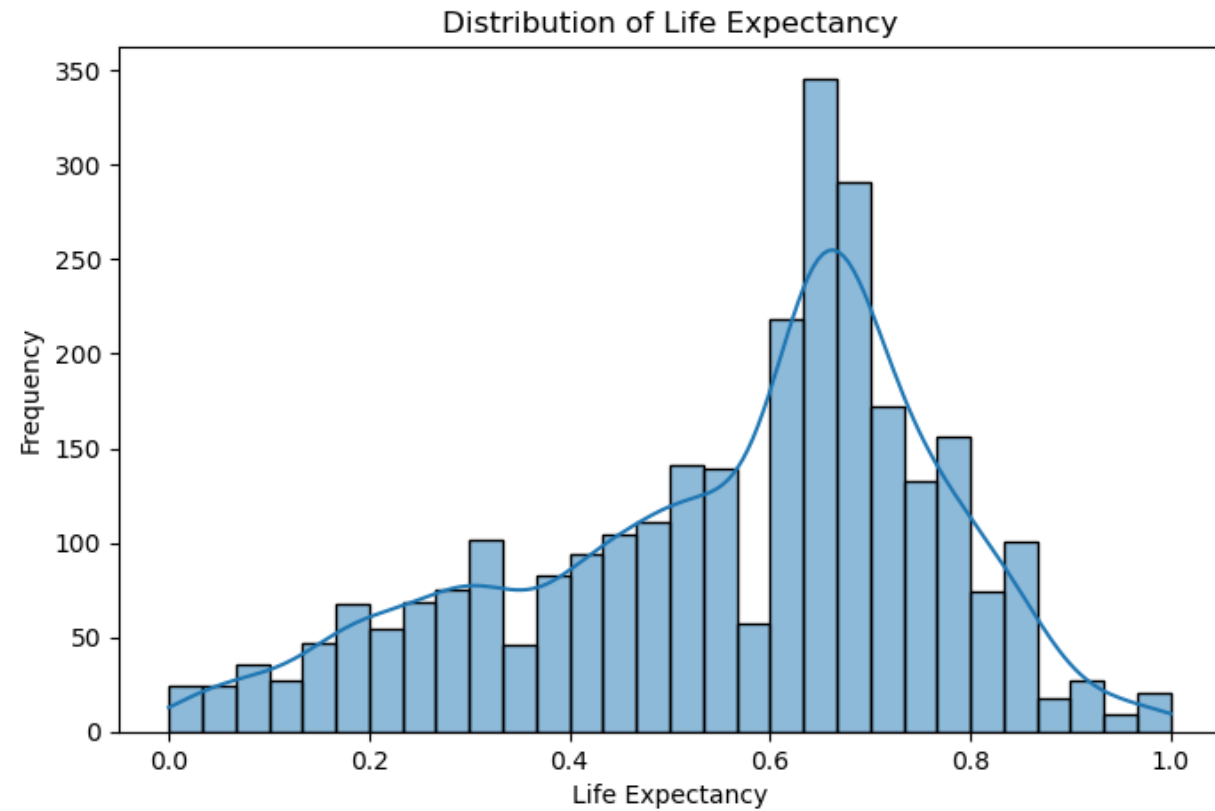
- Log transformation is useful when data is highly skewed
- It reduces the effect of large values and makes the data more normally distributed
- It was applied to features with a very high number of outliers where simple trimming wouldn't be effective.



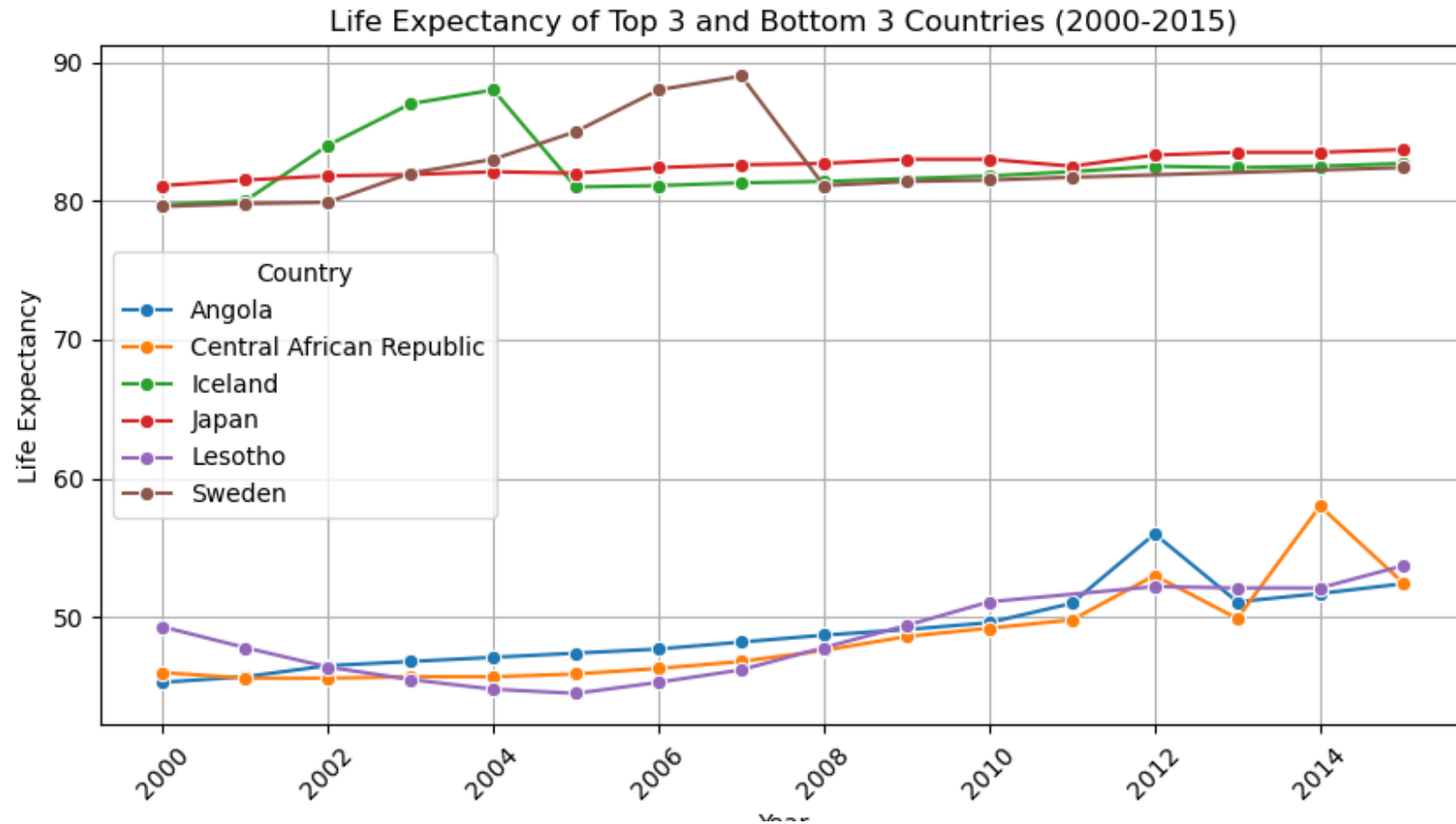
EXPLORATORY DATA ANALYSIS

- Distribution analysis of features
- Correlation heatmap to identify relationships
- Visualization for trends and patterns

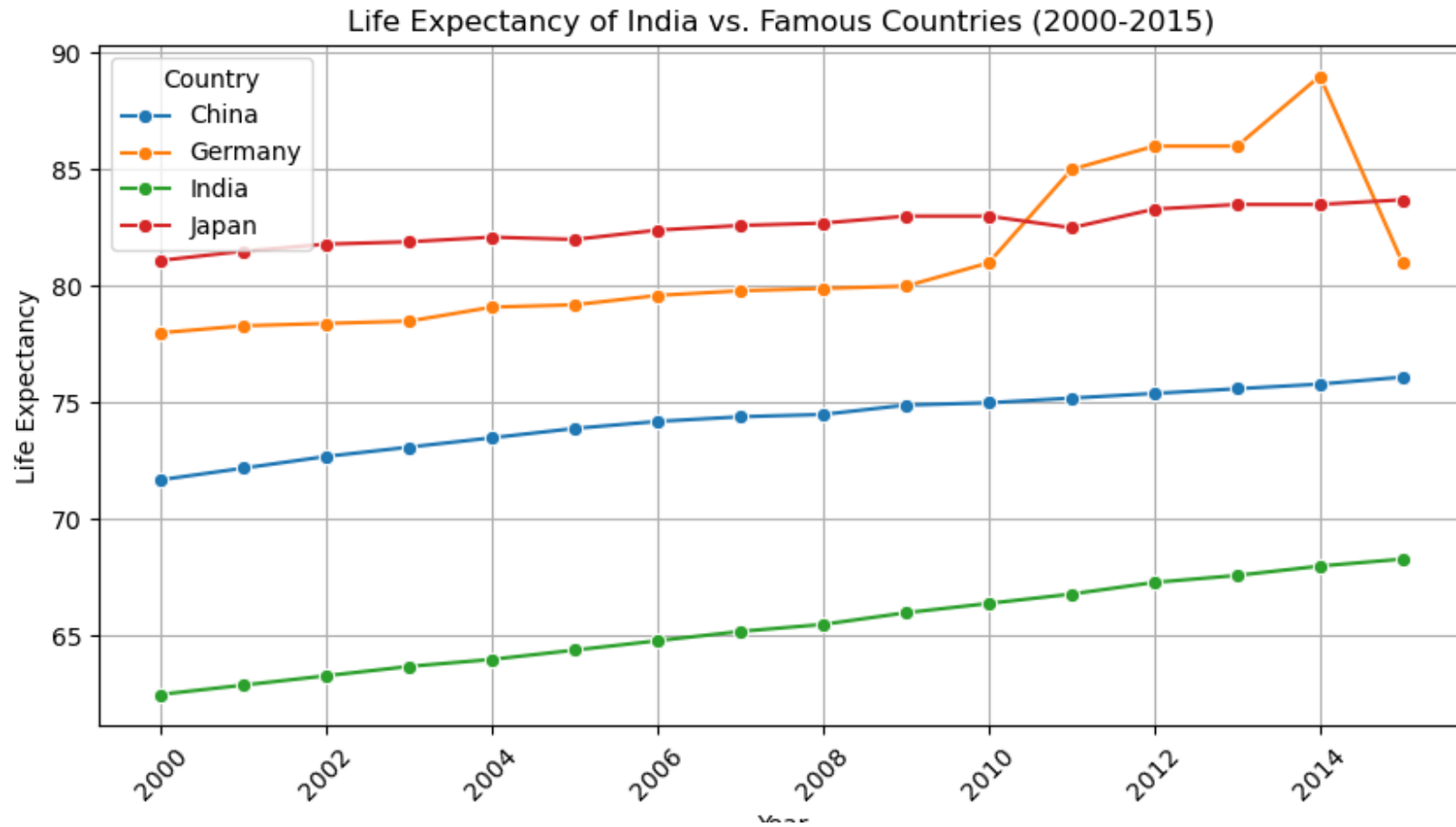
DISTRIBUTION OF LIFE EXPECTANCY



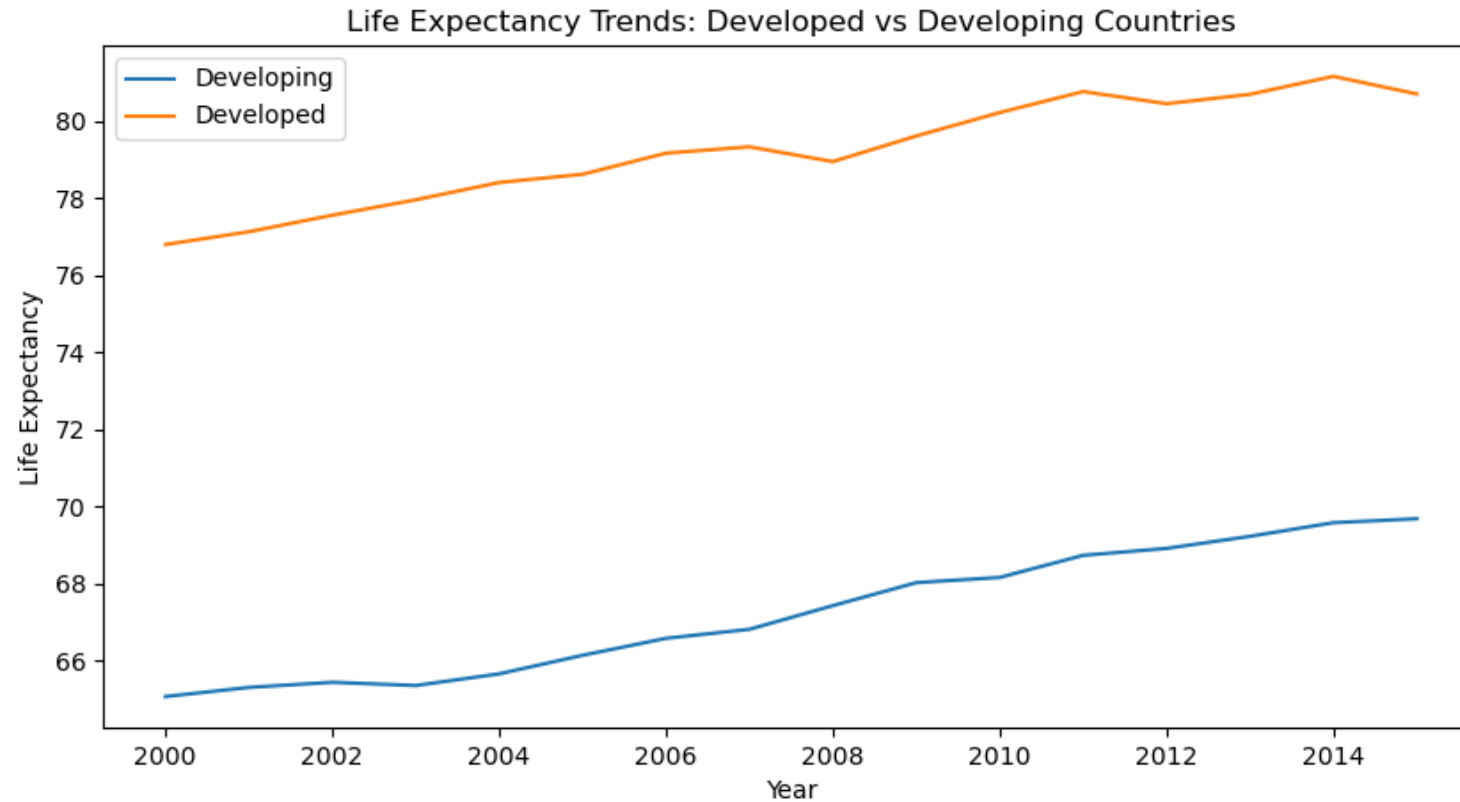
The distribution appears slightly right-skewed, meaning more countries have higher life expectancies



TOP 3 AND BOTTOM 3 COUNTRIES

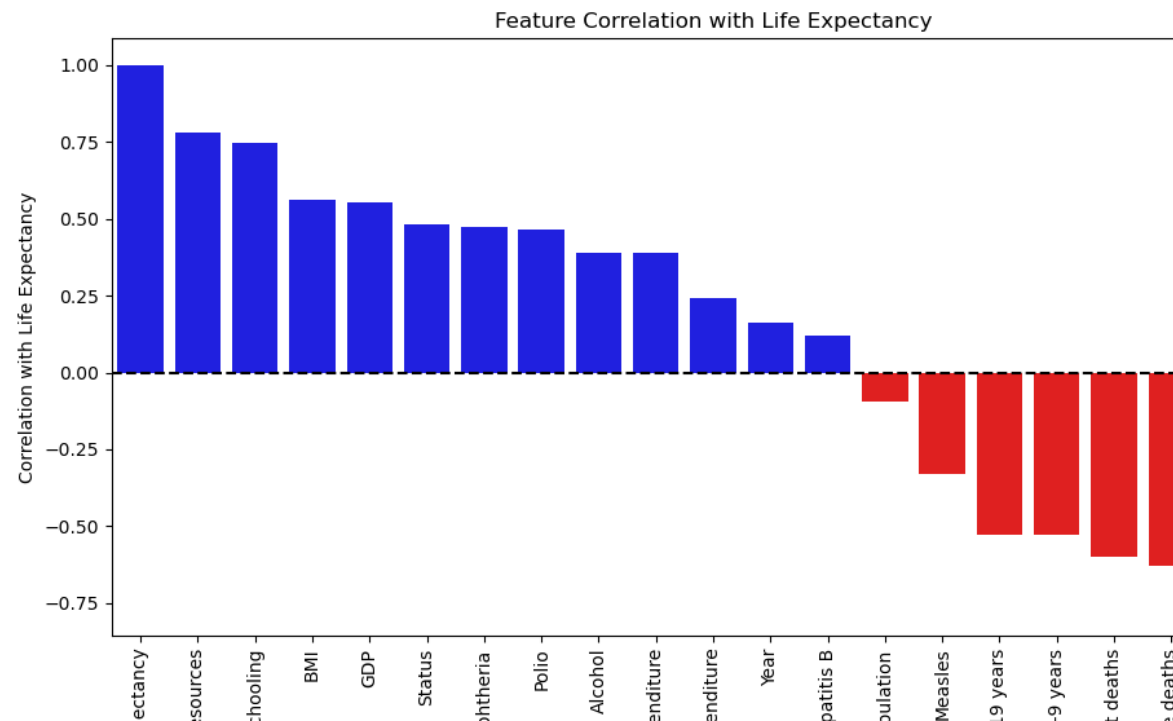


INDIA VS FAMOUS COUNTRIES



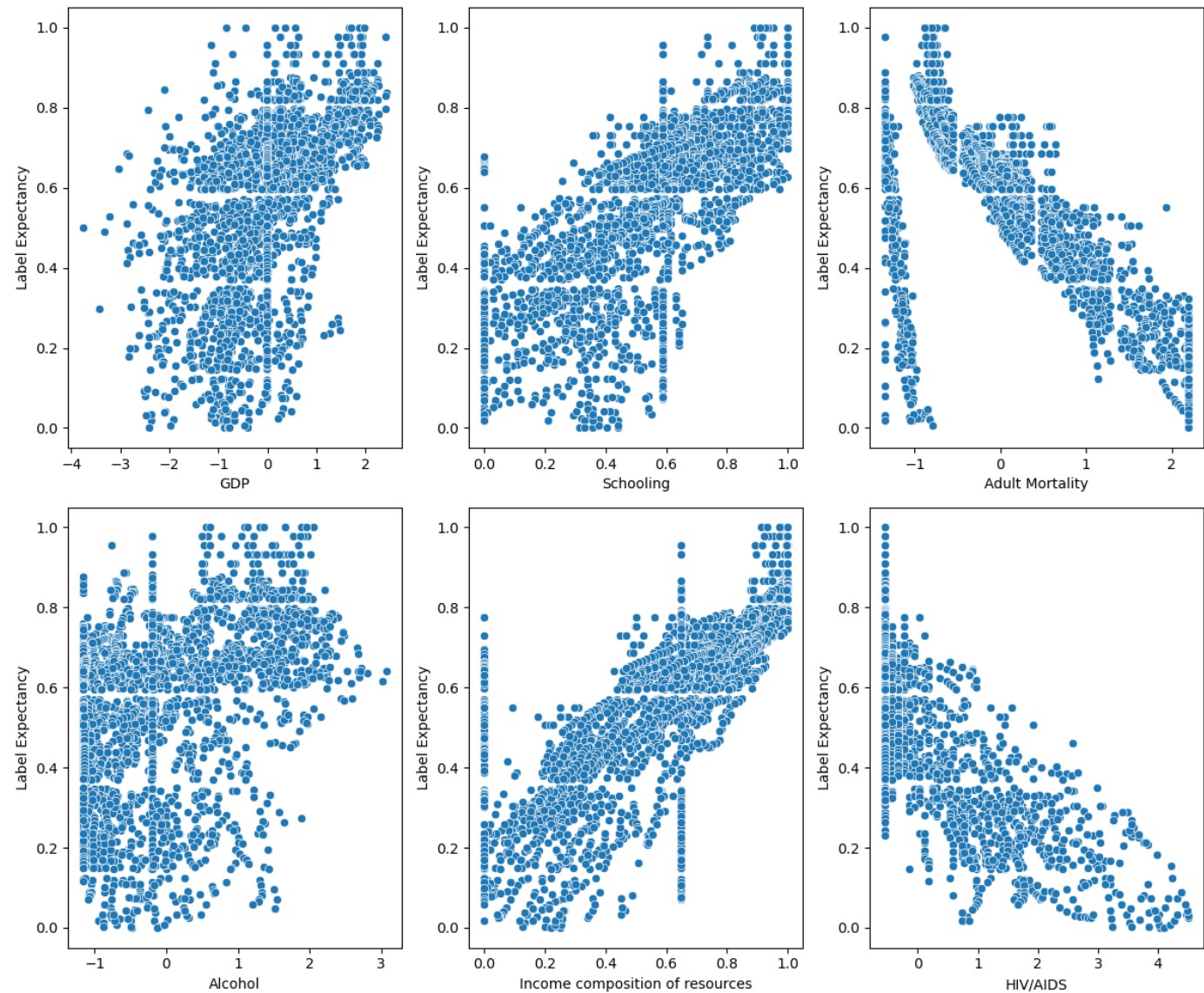
DEVELOPED VS DEVELOPING COUNTRIES

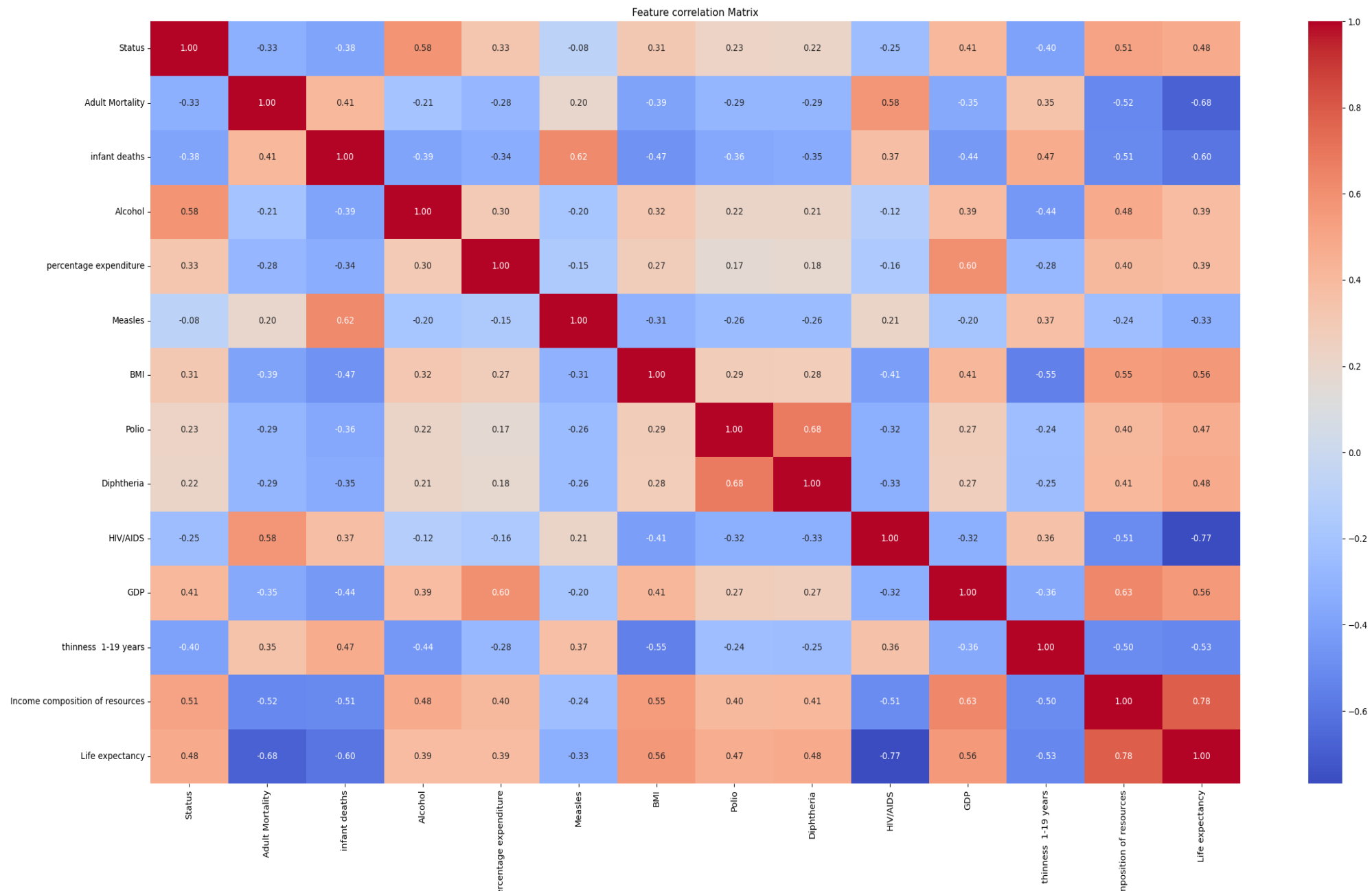
CORRELATION WITH LIFE EXPECTANCY



- Increase Life Expectancy – Income, Composition of Resources, Schooling, GDP, BMI and Polio .
- Decrease Life Expectancy – HIV/AIDS , Infant Deaths & Under-5 deaths, Adult Mortality, Thinness(5-9yrs and 10-19 yrs.

CORRELATION WITH LIFE EXPECTANCY





STATISTICAL ANALYSIS

Hypothesis Testing

- Confidence Interval (95%): The true population mean life expectancy is likely between 68.98 and 69.66 years.
- T-Test: No significant difference between sample mean (67.33) and population mean (69.32).
- Chi-Square Test: Life expectancy is significantly associated with a country's development status.
- ANOVA Test: GDP has a significant impact on life expectancy

KEY INSIGHTS AND FINDINGS

- Higher GDP and schooling improve life expectancy
- High HIV/AIDS rates and infant mortality reduce life expectancy.
- Adult mortality and healthcare expenditure play significant roles.
- Economic development, education, and healthcare improvements play a crucial role in increasing life expectancy.
- Reducing mortality from diseases is key to enhancing public health.

- Economic growth, education, and healthcare access are the strongest drivers of life expectancy
- Preventable diseases, malnutrition, and mortality rates must be addressed to improve life expectancy in developing countries.
- Government policies focusing on healthcare infrastructure, vaccination programs, and nutrition can drastically improve life expectancy worldwide.

CONCLUSION

THANK YOU

Anantha Krishnan K

BCE229
