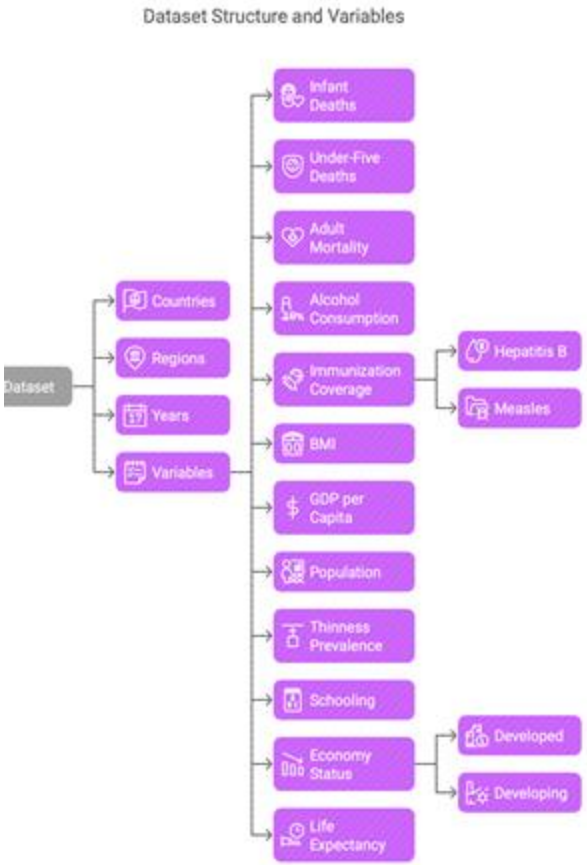
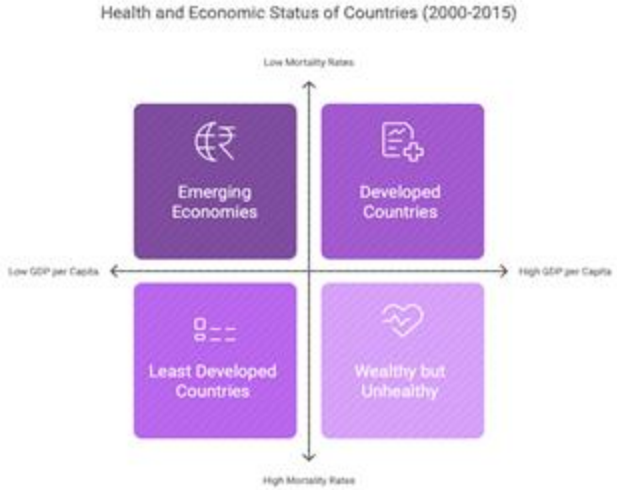
Global Health in Numbers

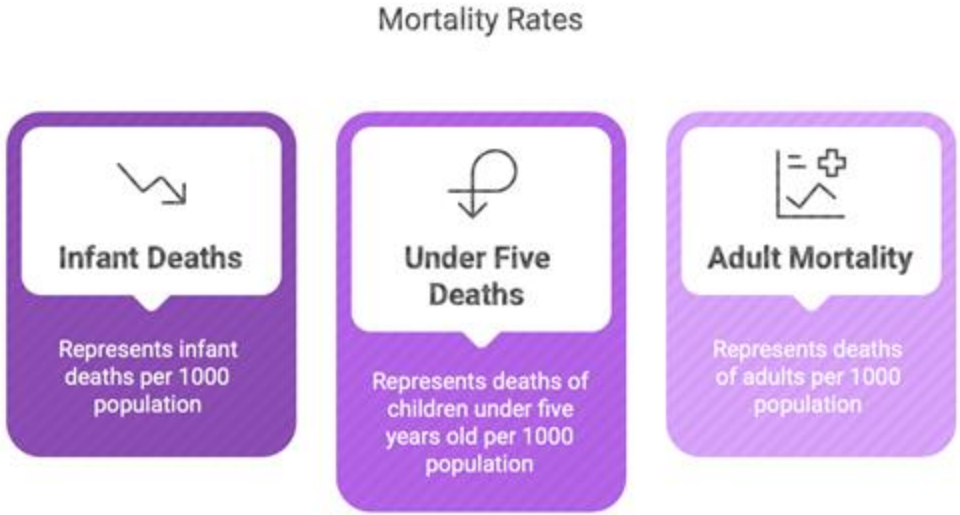
A Statistical Dive into Life Expectancy (2000–2015)



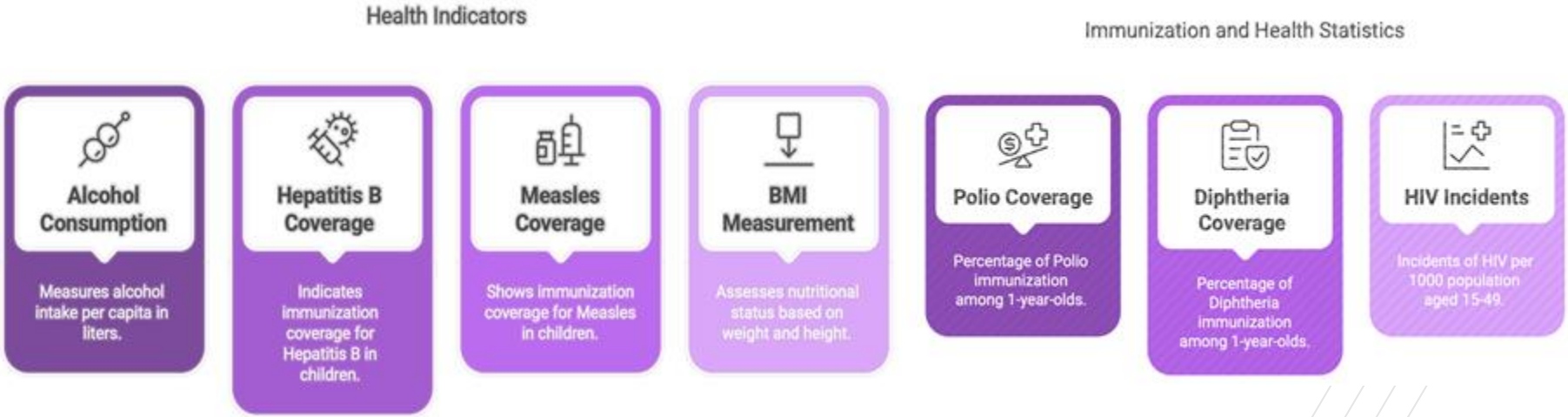


A Brief Overview about the Dataset

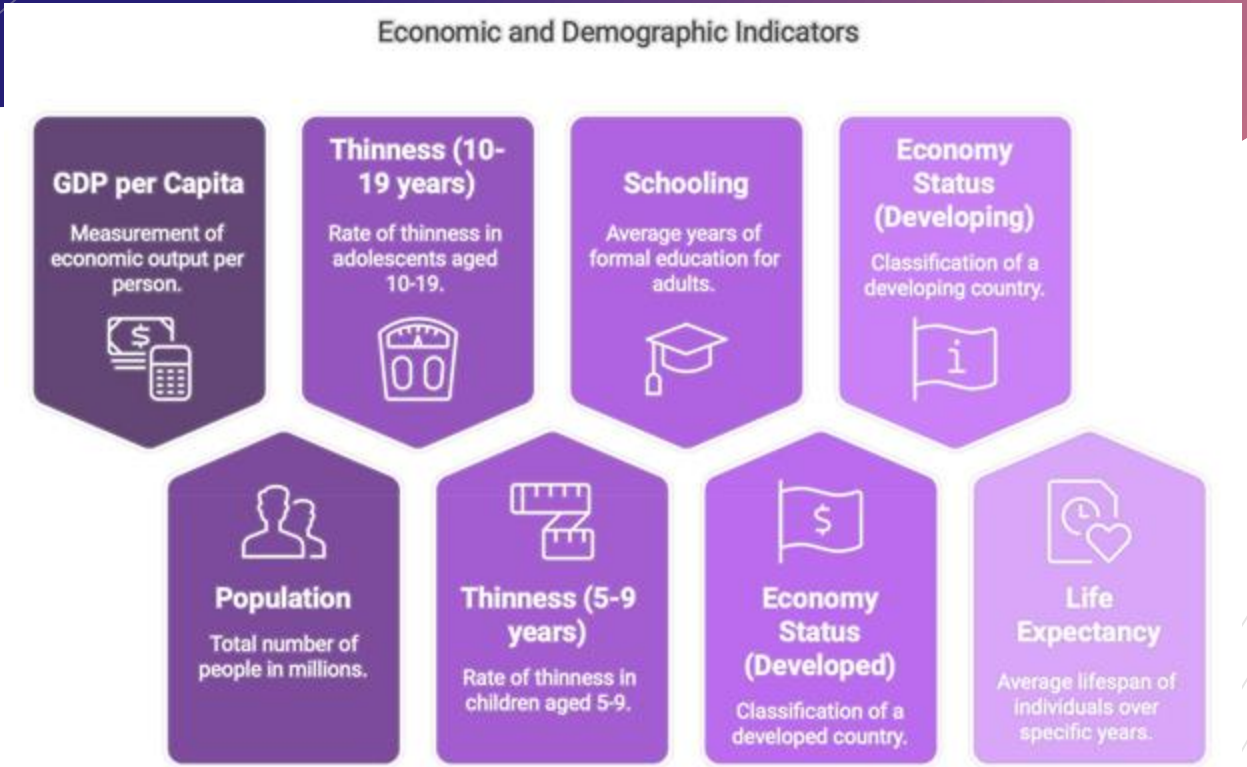
Understanding the data



Understanding the data



Understanding the data



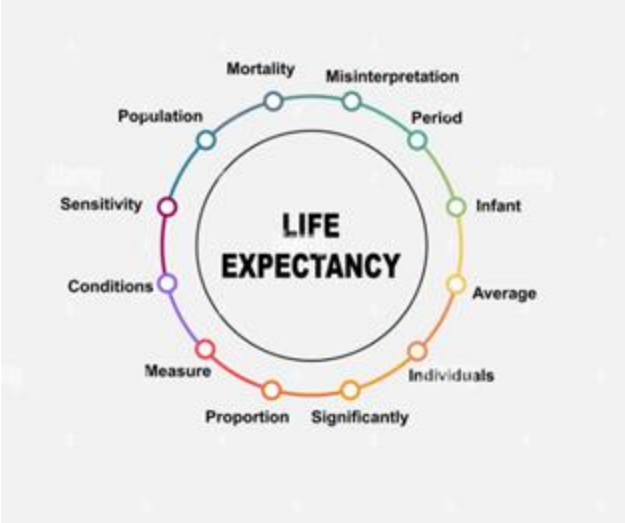
Why Life Expectancy?

A key indicator of overall health and quality of life

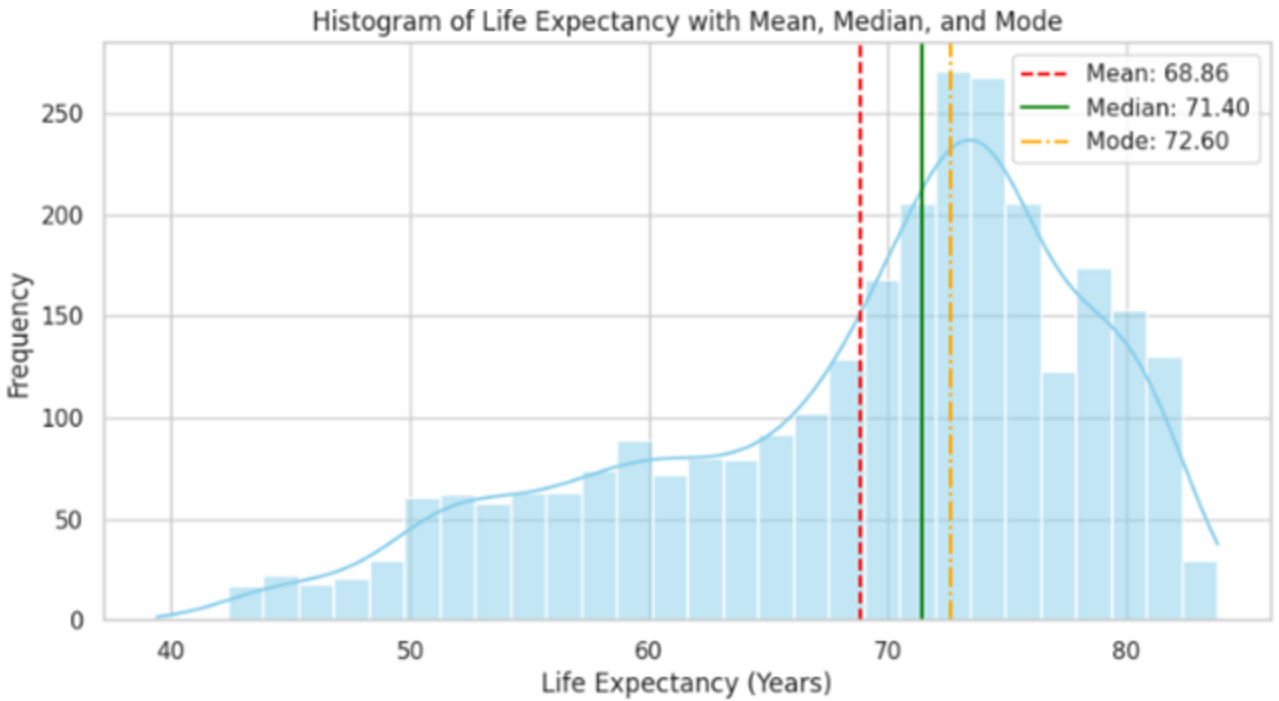
Reflects nutrition, healthcare, sanitation, income, and education

Shows how well nations support health and longevity

Links social, economic, and health-related factors globally



Histogram of Life Expectancy Vs Frequency

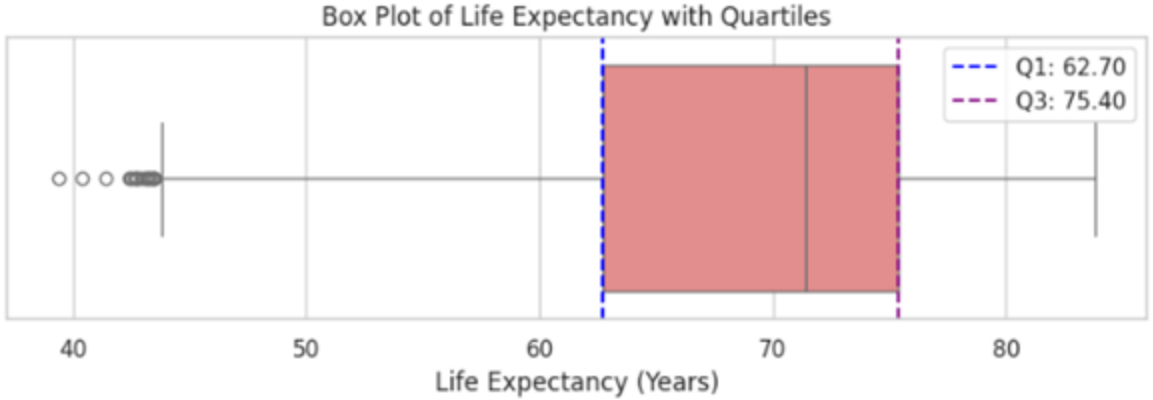


Majority of countries fall between **65 – 80 years**
Mean year is **68.86** years

Distribution is slightly left-skewed (more low outliers)

Peak around 70–75 years shows global concentration

Box Plot of Life Expectancy



Q1 - 25th Percentile
Q2 - 50th Percentile = Median
Q3 - 75th Percentile

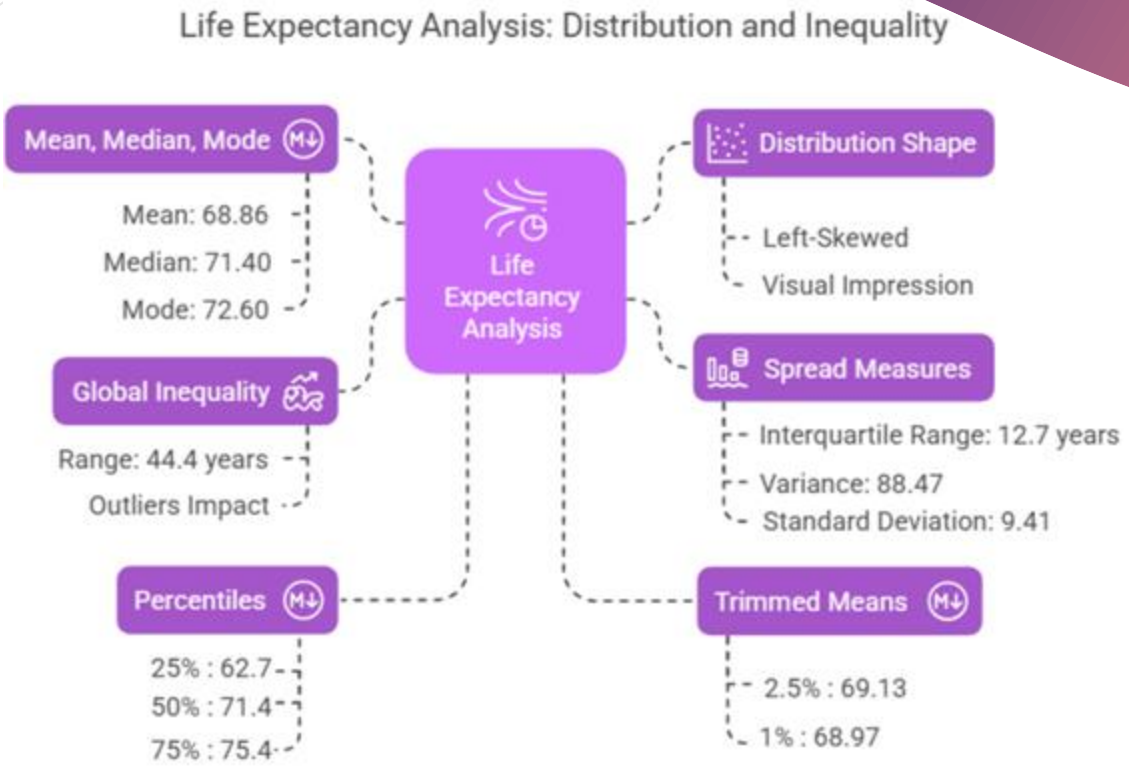
IQR = Q3 - Q1
Low Outliers = Q1 - (1.5 X IQR)
High Outliers = Q3 + (1.5 X IQR)

Majority of countries fall between 65–80 years

Median life expectancy ≈ 68.86 years

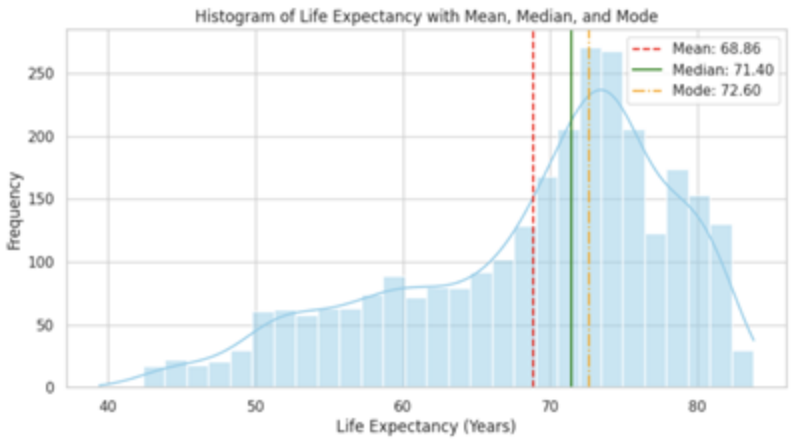
Boxplot shows a few low-end outliers
Box shifted right → confirms left-skew

Numerical Calculations and Comparison

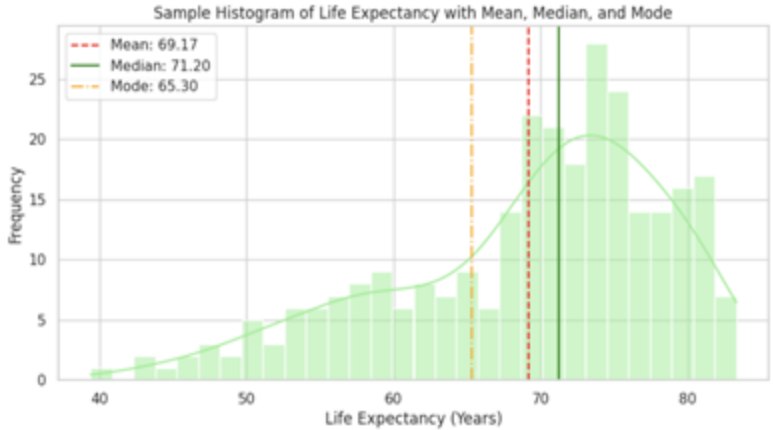


Comparison between Original and Sample (10%) Data

Original



10% Data

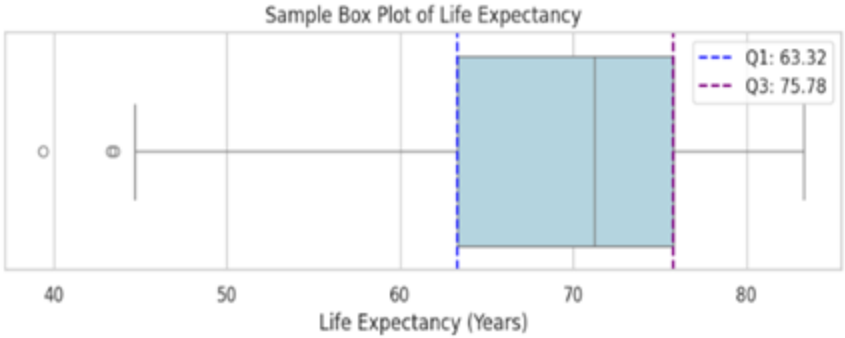


- The histogram of life expectancy is left-skewed in both datasets.
- The full dataset has a mean of **68.86 years** and a median of **71.40 years**, while the 10% sample has a mean of **69.17 years** and a median of **71.20 years**, showing minimal deviation. The mode shifts from **72.60 years** in the full dataset to **65.30 years** in the sample due to random selection. The Kernel Density Estimate (KDE) confirms that the sample retains the overall distribution shape.

Comparison between Original and Sample (10%) Data

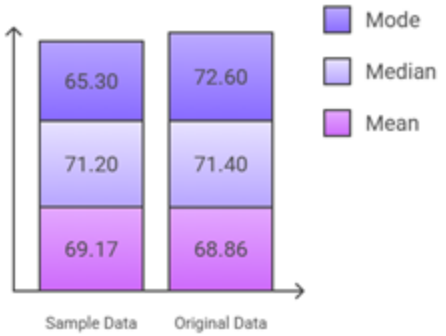
Original

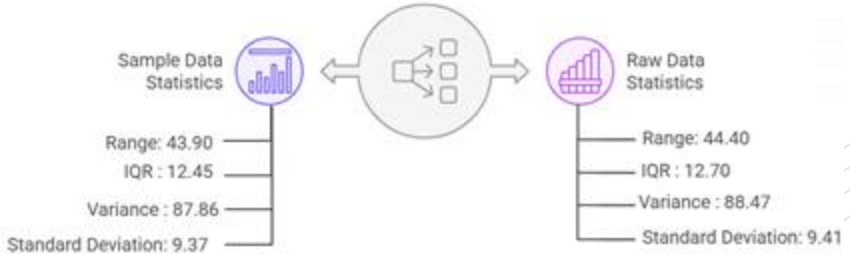
10% Data



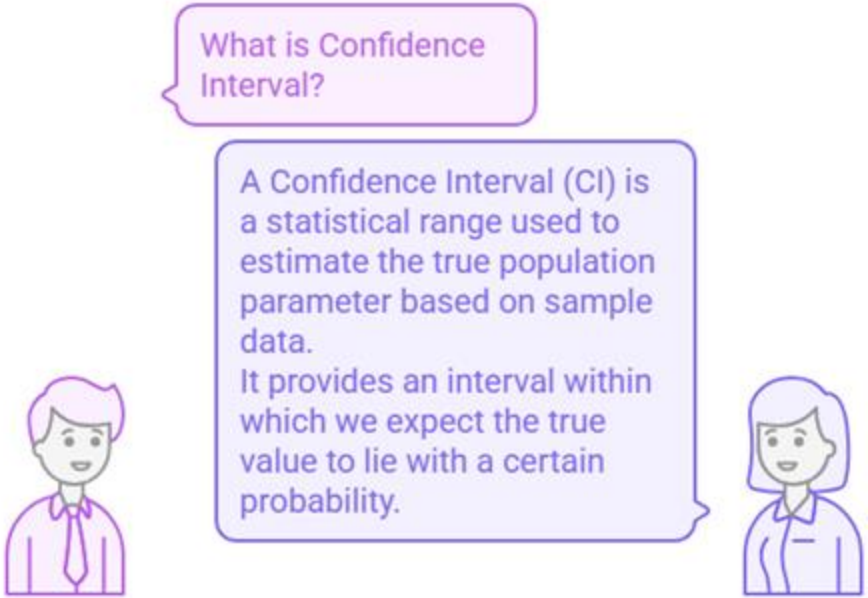
- The box plot shows similar spread and variability between datasets.
- The first quartile (Q1) is **62.70 years** in the full dataset and **63.32 years** in the sample, while the third quartile (Q3) is **75.40 years** and **75.78 years**, respectively. The interquartile range (IQR) remains stable at **12.70 years** (full) and **12.46 years** (sample). The full dataset contains more low-end outliers, while the sample captures fewer due to its size. Despite minor variations, the sample preserves the original dataset's key statistical properties.

Numerical Finding Analysis





WHAT IS CONFIDENCE INTERVAL ?



$$CI = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

CI = confidence interval

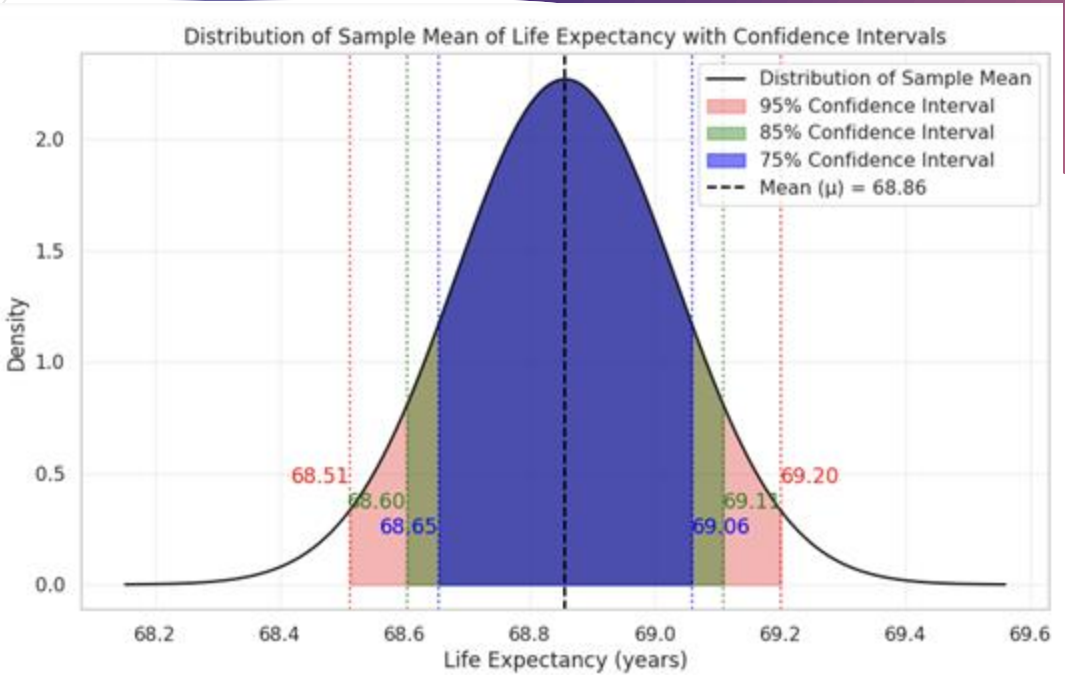
$\bar{x}$ = sample mean

z = confidence level value

s = sample standard deviation

n = sample size

Constructing Confidence Interval



This plot demonstrates how confidence intervals work in estimating the **true mean life expectancy** based on sample data. The shaded regions show different levels of certainty (from 75% to 95%) about where the true mean lies. The more confident you want to be, the wider the interval needs to be.

Hypothesis Testing

Should we reject the null hypothesis that life expectancy is 70 years?

Fail to Reject H_0

The p-value is greater than the significance level, indicating no strong evidence against H_0.

Reject H_0

The p-value is less than the significance level, suggesting strong evidence against H_0.

Hypothesis Testing

Two-Tailed Z-Test Process

01 Calculate Z-Statistic

02 Determine P-Value

03 Compare P-Value to α

04 Decision to Reject or Not

05 Interpret Results

Observed value

Mean value

$$z = \frac{x - \mu}{\sigma}$$

z-Score

Standard deviation

Hypothesis Testing

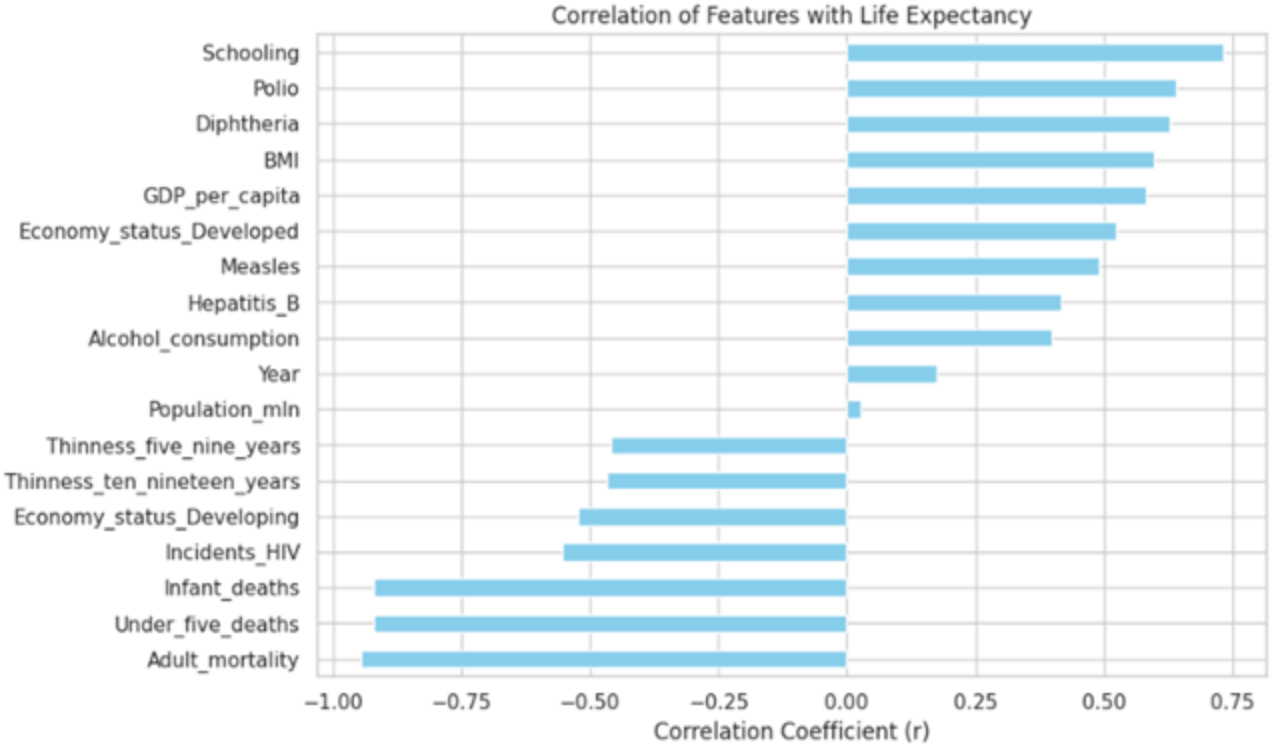
P-value
0.1365

Z-statistic
-1.4890

Decision Rule

As P-Value is not less than α (0.05)
So we fail to reject the null hypothesis.

Correlation of Life Expectancy with All Other Columns



$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}$$

Positive Correlation -When one variable increases, the other also increases.

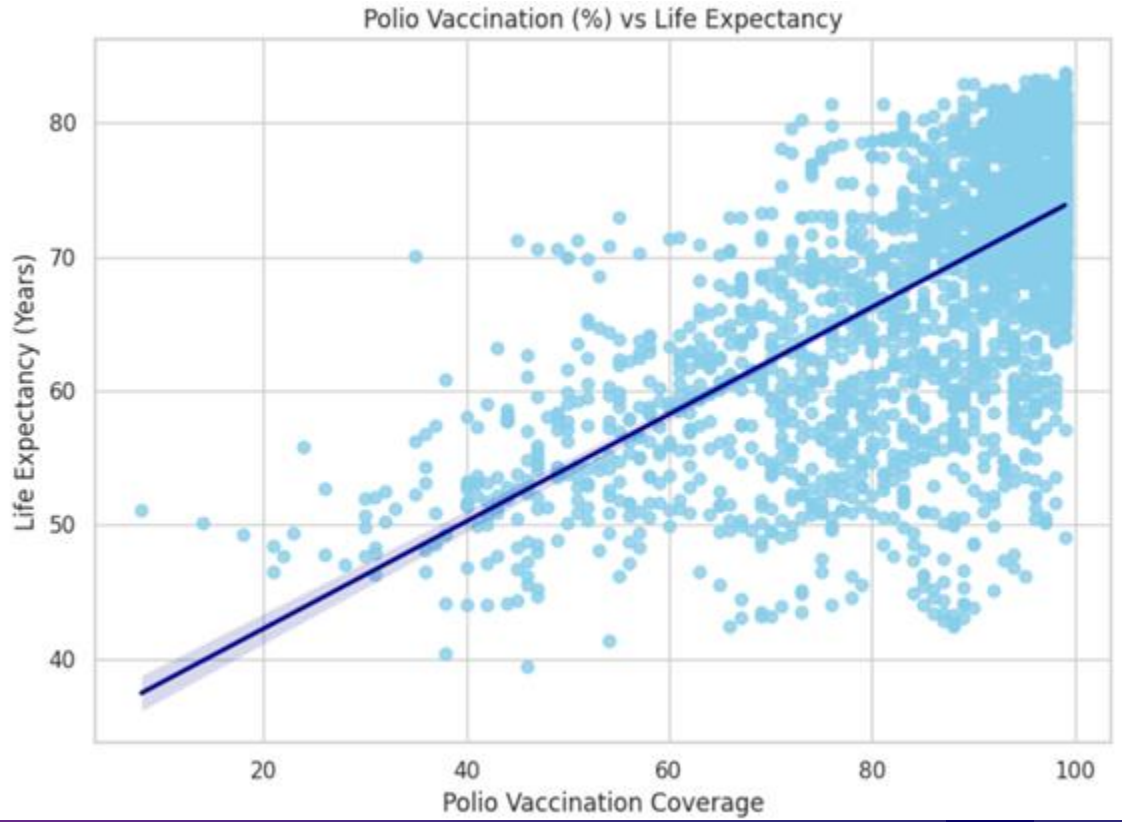
Negative Correlation -When one variable increases, the other decreases.

Polio Vaccination vs Life Expectancy

Strong positive correlation: **r = 0.6412**

R^2 = 0.4112: Polio coverage explains ~41% of variance in life expectancy

Vaccines reduce preventable diseases, boosting public health.
Emphasizes importance of **immunization programs** globally.

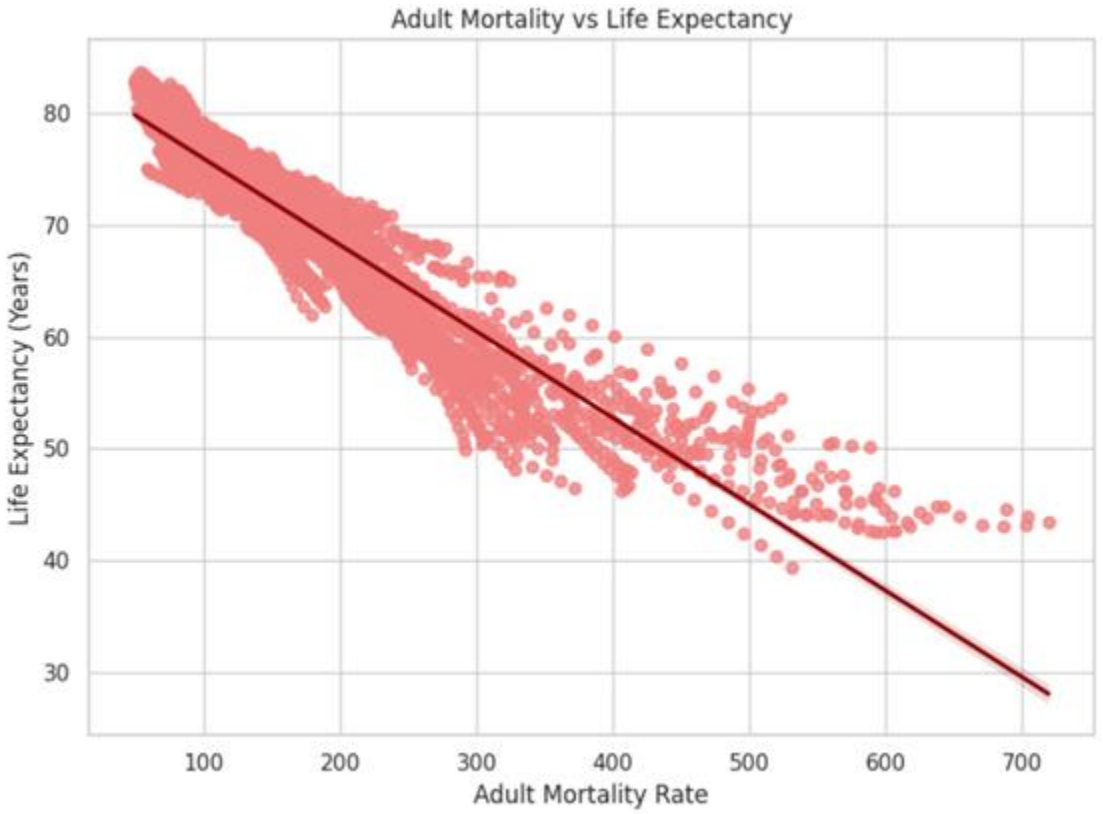


Adult Mortality vs Life Expectancy

Very strong **negative correlation**:
r = -0.9454

R^2 = 0.89: Explains 89% of life expectancy variance

Reflects **healthcare quality**, disease burden, and socio-economic conditions
Key indicator of **adult health and survival impact**

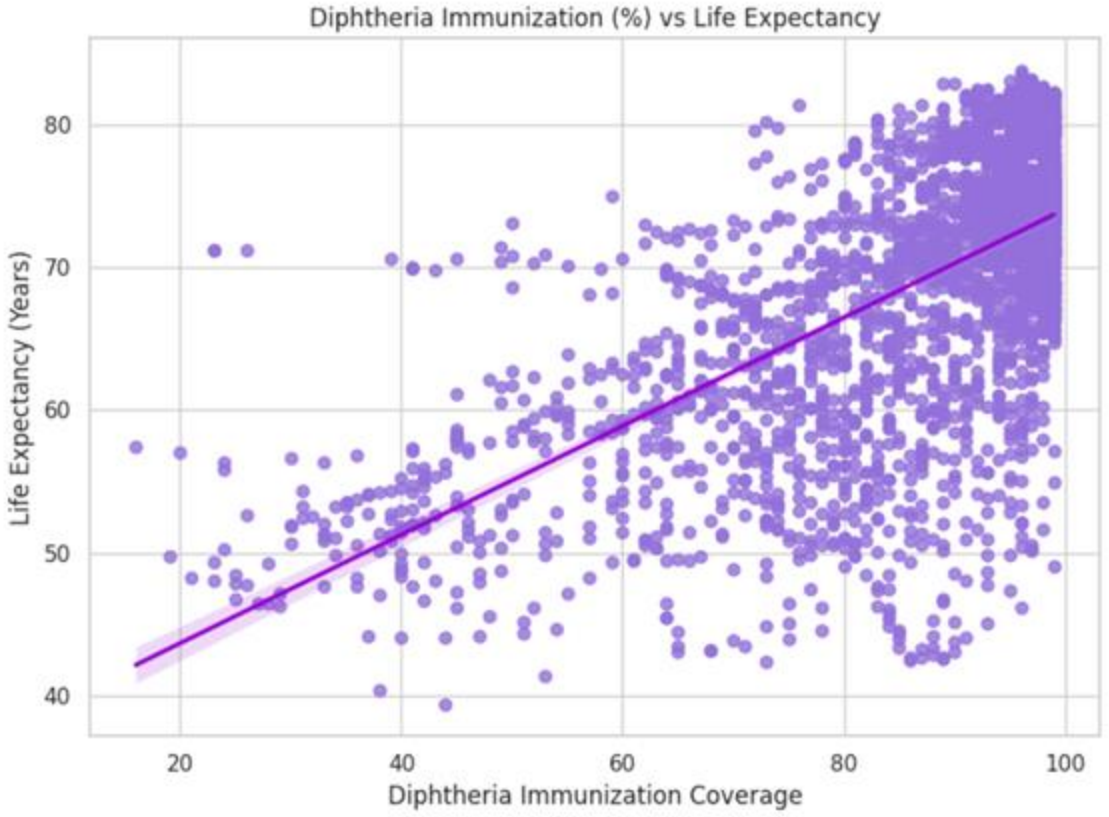


Diphtheria Immunization vs Life Expectancy

Strong positive correlation: **r = 0.6275**

R^2 = 0.39: Explains 39% of life expectancy variance.

Consistent pattern across regions supports global relevance
Highlights the **impact of childhood vaccines** on reducing mortality



Schooling vs Life Expectancy

Strong positive correlation: **r = 0.7325** (highest in dataset)

R^2 = 0.54: Schooling explains 54% of life expectancy variance

- Education improves **health awareness**, income, and lifestyle
- Highlights **education as a long-term health investment**

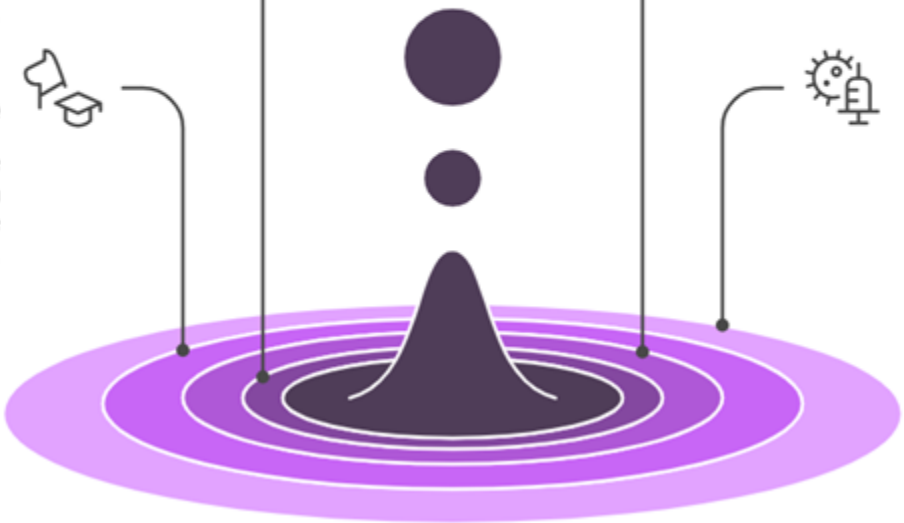
Factors Influencing Life Expectancy

Polio Vaccination Impact

Highlights the role of polio vaccination in increasing life expectancy.

Education's Role

Shows the positive correlation between education and life expectancy.



Adult Mortality Influence

Demonstrates how adult mortality rates affect life expectancy negatively.

Diphtheria Immunization Effect

Illustrates the impact of diphtheria vaccination on life expectancy.

References

World Health Organization. (2024). Global Health Observatory Data. Retrieved from https://www.who.int/data/gho

United Nations. (n.d.). UN Data. Retrieved from https://data.un.org

The Pandas Development Team. (n.d.). Pandas documentation. Retrieved from https://pandas.pydata.org

The Seaborn Development Team(n.d.). Seaborn: Statistical data visualization. Retrieved from https://seaborn.pydata.org

Matplotlib documentation. Retrieved from https://matplotlib.org

University of Illinois Urbana-Champaign. (n.d.). Exploratory data analysis: Quartiles and box plots. Retrieved from https://discovery.cs.illinois.edu/learn/Exploratory-Data-Analysis/Quartiles-and-Box-Plots/

Life expectancy WHO (updated) (2023). [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/lashagoch/life-expectancy-who-updated

THANK YOU