

World Bank Development Indicators Data Analysis

Name: Abdul Rehman

Student Number: 23113350

GitHub Repository: [<https://github.com/AR-KASHMIRI47/REFERRAL-Statistics-and-Trends.git>]

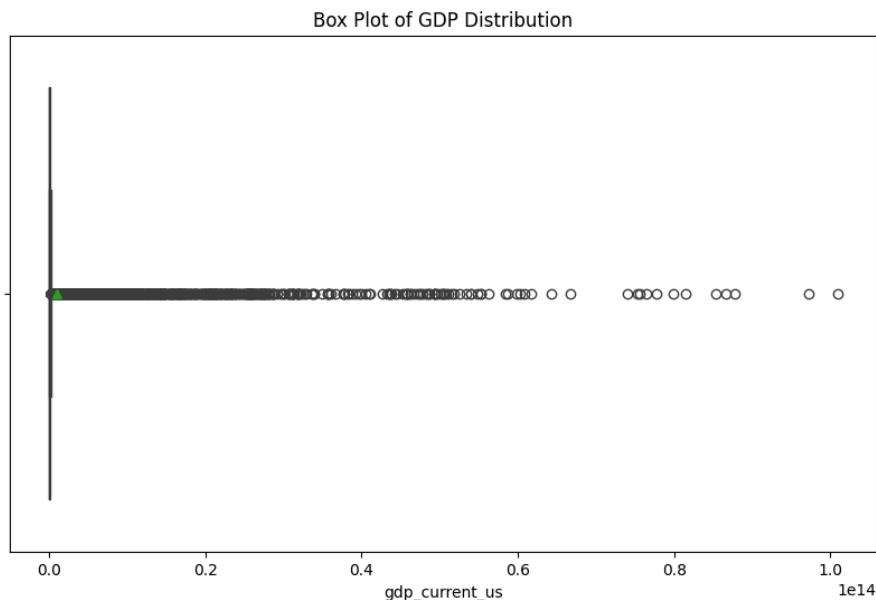
Introduction

This report explores the World Bank Development Indicators dataset, a comprehensive collection of economic, social, and environmental metrics across countries. The objective was to analyze global development trends using data cleaning, descriptive statistics, and visualization techniques. The analysis focuses on key indicators such as GDP, life expectancy, and various correlations among economic factors.

Analysis and Visualization

1. GDP Distribution (Box Plot)

The GDP distribution is visualized using a box plot.

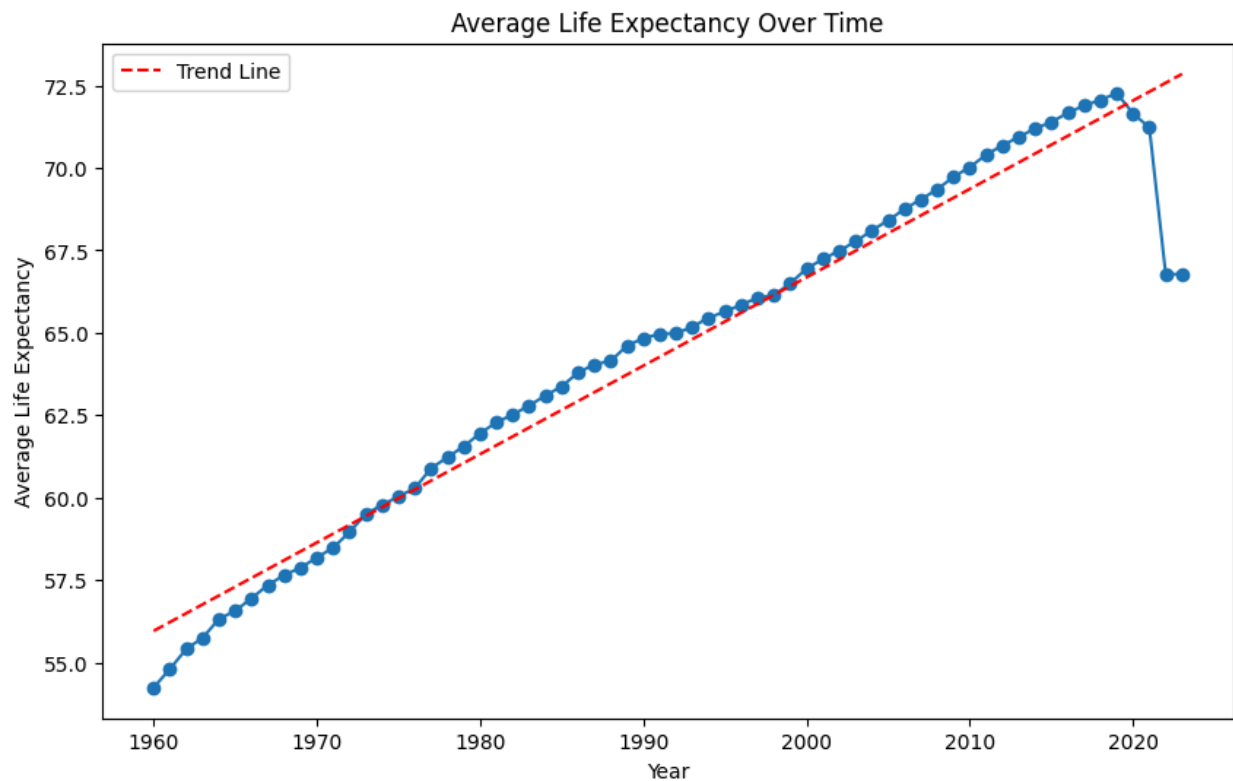


This plot provides a clear view of the data spread, including the median, interquartile range, and outliers. Most countries exhibit lower GDP values, while a few outliers display exceptionally high GDP. The distribution is right-skewed (skewness = 9.30, kurtosis = 111.63), indicating significant

economic disparity. The median GDP (17.86 billion USD) is lower than the mean (939.35 billion USD), emphasizing the impact of outliers.

2. Life Expectancy Trend (Regression)

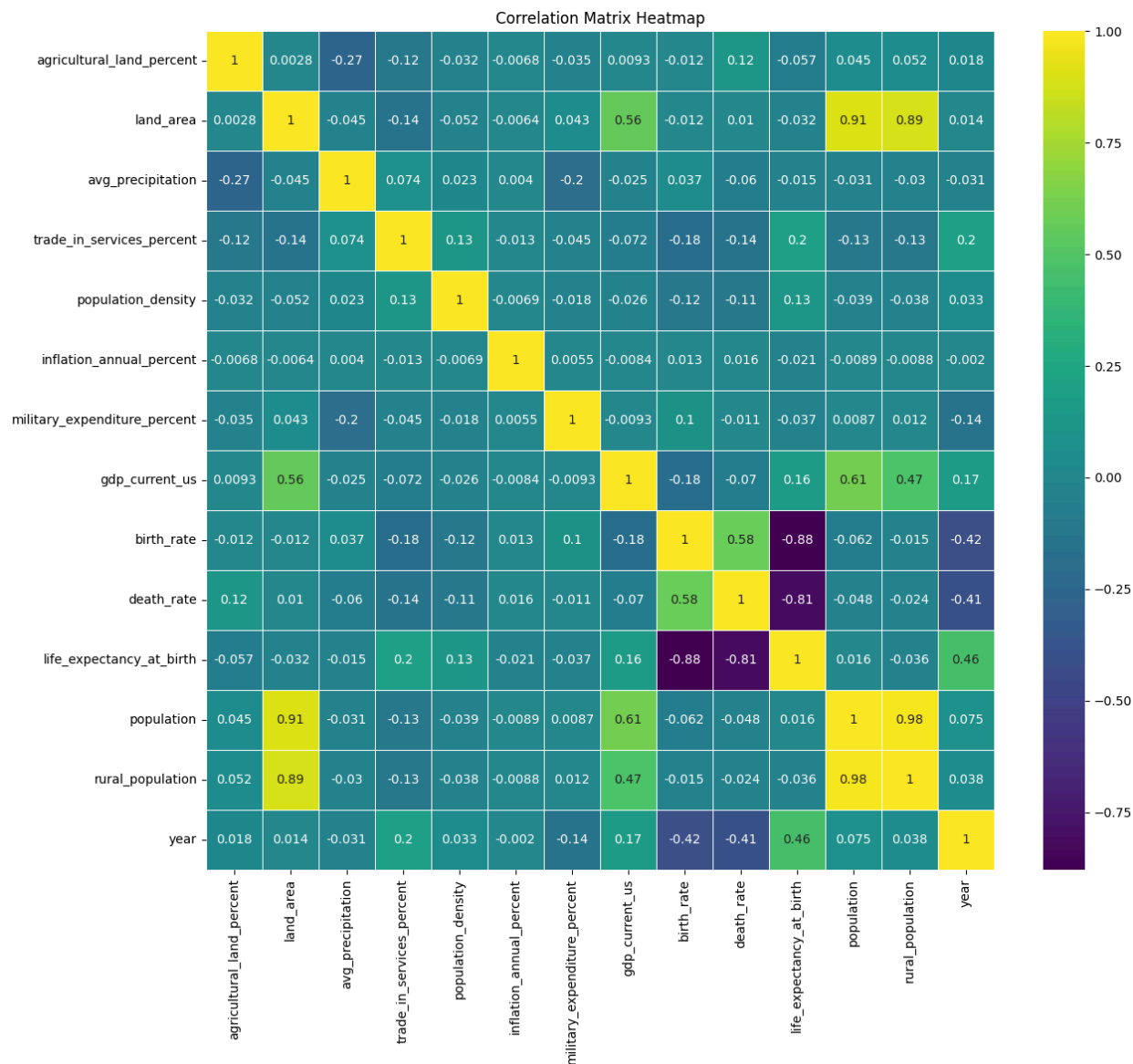
The second plot is a line graph displaying the trend of average life expectancy over time, enhanced with a regression trend line.



This reveals a consistent upward trajectory, reflecting global improvements in healthcare, nutrition, and living conditions. The regression line highlights the positive correlation ($r = 0.46$) between time and life expectancy.

3. Correlation Matrix Heatmap

The heatmap visualizes the correlation between various numerical indicators.



It highlights strong positive correlations, such as between GDP and land area ($r = 0.56$), indicating direct relationships. Conversely, negative correlations, such as between trade in services and agricultural land ($r = -0.12$), reveal inverse associations. This matrix provides a clear overview of inter-variable dependencies.

Descriptive Statistics

This section provides a comprehensive overview of the dataset's statistical properties:

- **Mean, Median, Standard Deviation, Skewness, and Kurtosis** are calculated for all numeric columns.

- **Median Absolute Deviation (MAD)** is calculated to measure variability.

- The data is right-skewed with high kurtosis, indicating significant outliers.

Conclusion

This analysis provided insights into global development trends, showing a significant disparity in GDP distribution, a consistent improvement in life expectancy, and complex relationships between various economic indicators. Future studies could focus on identifying factors influencing GDP growth or investigating country-specific trends.

GitHub Repository

You can access the complete analysis and code at:

[<https://github.com/AR-KASHMIRI47/REFERRAL-Statistics-and-Trends.git>]]