World Population Analysis Report

# Introduction

* The objective of this project is to analyze historical world population data and predict future population trends.
* Understanding population dynamics is crucial for planning and policy-making in various sectors such as healthcare, education, and infrastructure.

# Data Collection

* The data is gathered from websites which are:
  + World Population Review
  + World Bank
  + MIF
* The data contains information about the birth rate, death rate, fertility rate, sex ratio, population growth, area occupied by each country, Human Development Index (HDI), etc.

# 3. Exploratory Data Analysis (EDA)

This section explores population data to understand trends and distributions across countries and continents.

## Summary Statistics

The analysis examined factors such as Human Development Index (HDI), fertility rate, GDP growth, and birth/death rates to assess their impact on population growth. Most countries showed high HDI, indicating a generally high standard of living. The fertility rate distribution was right-skewed, and GDP growth followed a normal distribution centered around 2.5% to 5.0%.

## Visualization of Population Trends

The analysis employed various visualizations, including bar charts, pie charts, and line graphs. Key trends observed:  
- Asia holds more than 50% of the global population.  
- Asia's population has increased exponentially, while Africa's growth has been steady.  
- Europe experienced slight population decreases post-2000.

## Analysis of Key Factors Affecting Population Growth

Population growth is influenced by factors like fertility rate, birth rate, and GDP. Countries with high birth and fertility rates (e.g., Niger) showed rapid population growth. Conversely, countries with low fertility and birth rates faced stagnant or declining growth.

# 4. Model Building

I used Linear Regression to make Population Prediction. The Data had lots of features that were correlated and according to the assumption of Linear Regression the features should be independent of each other .So with the help of RFE(Recursize Feature Elimination) and VIF(Variance Inflation Factor ), we took the most stable and important features to make prediction which were **HI\_Avg\_4decades, fertility\_rate, BirthToDeathRatio, SexRatioAllAges, gdpgrowth\_10yr\_avg, World Population Percentage**.

# 5. Model Evaluation

I used R2\_score for the model evaluation and achieved a R2\_score of 99 %.I also plotted the distribution of the residuals which is called the difference in prediction and the true value. Ideally it should show a normal distribution but their was some distortion at the mean.

# 6. Results and Discussion

The analysis indicates a strong correlation between population growth and factors such as birth and fertility rates. Key insights include:  
- High birth rates and fertility are predictors of rapid population growth.  
- Economic factors like GDP growth vary across regions, influencing regional population stability.  
  
The results highlight regional demographic diversity, with population concentration in Asia and rapid growth in Africa.

# 7. Conclusion

The study provided a comprehensive view of global population trends, revealing significant regional variation. High fertility and birth rates were central to growth in certain regions, while economic factors influenced stability in others.Most of the countries which are not recognized as countries such as American Samoa their information is not available so due to lack of data the model is not robust . Future research could incorporate timeseries forecasting models to predict population dynamics more accurately, accounting for socioeconomic and environmental factors.