Introduction

This project delves into the economic analysis of countries key sectors and their contributions to its Gross Domestic Product (GDP). By using a dataset that includes critical indicators such as Population, Exports, Agriculture, Manufacturing, and Government Expenditure, the study seeks to provide insights into how these factors interact and contribute to the overall economic health of the country.

Defining the columns

* Country: This column lists the names of countries for which the economic data is recorded. Each row corresponds to a specific country.
* Year: This column indicates the specific year the data is collected. It allows for time-series analysis, showing how the indicators change over time.
* AMA Exchange Rate: This refers to the average market exchange rate of the country's currency against a major currency (usually the USD) for a given year. This rate reflects the value of the local currency in international markets.
* Population: The total population of the country in the specified year. This figure is essential for understanding other per capita indicators and the size of the market.
* Per Capita GNI (Gross National Income): This indicator represents the average income of a country's residents, calculated as the Gross National Income divided by the population. It reflects the overall economic well-being of the population.
* Agriculture, Hunting, Forestry, Fishing: This column shows the contribution of these sectors to the country's GDP. It reflects the importance of primary industries like agriculture, hunting, forestry, and fishing in the economy.
* Construction: This column represents the contribution of the construction sector to the GDP. It includes activities related to building infrastructure, residential, and non-residential structures.
* Exports: This measures the total value of goods and services exported by the country. Exports are a key component of a country’s economy and GDP.
* General Government Financial Consumption Expenditure: This is the expenditure by the government on goods and services consumed by the public sector. It includes spending on public services, defense, and public administration.
* Household Consumption Expenditure: This is the total expenditure by households on goods and services, including durable goods (like cars), non-durable goods (like food), and services (like healthcare).
* Import of Goods and Services: This measures the total value of goods and services imported by the country. Imports are subtracted from GDP as they represent spending on foreign goods.
* Manufacturing: This represents the contribution of the manufacturing sector to the GDP. It includes the production of goods in factories and other industrial facilities.
* Transportation: This measures the contribution of these sectors to the GDP, including the transportation of goods and people, warehousing, and communication services.
* GDP (Gross Domestic Product): This represents the total value of all goods and services produced within a country during a specific period. It is a broad measure of a nation's overall economic activity and health.

DATA CLEANING AND TRANSFORMATION

The dataset had 10,512 rows and 17 columns, the data was thoroughly cleaned, addressing missing values by replacing some with the mean and removing rows with excessive missing entries thereby reducing it to 10,418 rows; additionally, I created a column (GDP IN DOLLARS) which is the multiplication of a country’s GDP and the AMA exchange rate . This ensured that the analysis would yield more accurate and reliable results. Using visualization techniques like pie charts, the contributions of different sectors to the GDP were clearly represented for better understanding, especially for non-technical audiences.

Moreover, a correlation matrix was employed to identify the relationships between the key economic indicators. This was followed by a linear regression analysis that predicted GDP based on Population and Exports as primary drivers. This regression model highlighted the influence of these factors on economic performance, offering insights for future economic planning.

Overall, this study provides a holistic view of how various economic sectors and factors contribute to GDP, offering a foundation for deeper exploration and decision-making in economic development.

1. Descriptive Statistics

- Calculate basic statistics like mean, median, mode, standard deviation, and variance for indicators like GDP, GNI, Per Capita GNI, etc., to understand the central tendency and spread of the data.

- Identify outliers in economic indicators using box plots.

2. Time Series Analysis

- Trend Analysis: Examine how GDP, GNI, and other indicators have evolved over time for individual countries or groups of countries.

- Growth Rate Calculation: Calculate the annual growth rates of GDP, population, exports, etc., to assess economic growth.

- Forecasting: Use models like ARIMA to predict future values of GDP or population based on historical data.

3. Correlation and Regression Analysis

- Correlation Matrix: Compute the correlation matrix to identify relationships between different economic indicators.

- Linear Regression: Perform regression analysis to predict GDP or GNI based on other variables such as population, exports, and manufacturing output.

- Multiple Regression: Develop a model to predict GDP using multiple predictors like population, exports, and household consumption expenditure.

4. Comparative Analysis

- Country Comparisons: Compare economic indicators across countries, such as comparing GDP, GNI, and exports between two or more countries.

8. Visualization

- Heatmaps: Visualize the correlation matrix to quickly identify relationships between indicators.

- Line Plots: Plot time series data of GDP, exports, etc., for different countries.

- Bar Charts: Compare economic indicators across countries or regions.

- Scatter Plots: Explore relationships between two indicators, like GDP vs. exports.

- Bubble Charts: Use bubble charts to represent three dimensions, such as GDP, population, and exports.

9. Economic Impact Analysis

- Impact of Population on GDP: Analyze how changes in population size impact GDP using regression models.

- Effect of Exchange Rates on Trade: Investigate how fluctuations in exchange rates (AMA and IMF-based) affect imports and exports.

10. Hypothesis Testing

- T-tests or ANOVA: Test if there are statistically significant differences in economic indicators (e.g., GDP, exports) between different groups of countries (e.g., developed vs. developing).

11. Cross-Sectional Analysis

- Year-specific Comparisons: Analyze and compare the economic indicators for all countries in a specific year to identify global trends or anomalies.

Python Libraries for Analysis:

- Pandas: For data manipulation and analysis.

- NumPy: For numerical operations.

- Matplotlib/Seaborn: For data visualization.

- Scikit-learn: For machine learning tasks like clustering, regression, and PCA.

- Statsmodels: For statistical modeling and hypothesis testing.

- SciPy: For advanced statistical analysis.

These analyses can provide insights into economic performance, trends, and relationships between key indicators across countries and time.