School of Computer Science and Engineering

Lovely Professional University
Phagwara, Punjab (India)

Month: December Year: 2025

EDA PROJECT

Global Energy Consumption Analysis

Final Report

Submitted By

Bablu Prasad

Reg no: 12301443

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the B.Tech Dissertation/dissertation proposal entitled "**Global Energy Consumption Analysis**", submitted by **BABLU PRASAD** at **Lovely Professional University, Phagwara, India** is a bonafide record of his / her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

Hritwiz Yash

Date:

Acknowledgement

I hereby declare that the research work reported in the dissertation/dissertation proposal entitled "Global Energy Consumption Analysis" in partial fulfilment of the requirement for the award of Degree for Bachelor of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University's Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

Signature of Candidate

Name of the Candidate:

Bablu Prasad

Registration No: 12301443

1. PROBLEM STATEMENT

Global energy consumption has become a critical concern worldwide as industries, transportation, households, and digital infrastructure continue to expand. The growing demand for energy has led to increasing reliance on fossil fuels, rising carbon emissions, and uneven distribution of renewable energy adoption across different countries. Understanding global energy patterns is essential for governments, researchers, and policymakers to plan sustainable development strategies, forecast future energy needs, and address the challenges of climate change and energy security.

However, the global energy system is highly complex and heterogeneous. Countries differ significantly in terms of energy resource availability, economic development, technological advancement, environmental regulations, population size, and levels of industrialization. These differences result in diverse energy consumption behaviors and emission profiles across regions.

The challenge lies in analyzing these variations using real-world data and extracting meaningful insights that can reveal hidden patterns, long-term trends, correlations, and anomalies in energy usage. Such analysis can help in understanding the impact of energy choices on environmental outcomes and economic growth.

This project focuses on conducting an Exploratory Data Analysis (EDA) on a global energy consumption dataset that includes variables such as electricity consumption, fossil fuel share, renewable energy share, natural gas usage, CO_2 emissions, and population. Through systematic EDA techniques, including data cleaning, visualization, and statistical analysis, this study aims to understand global energy consumption patterns, identify outliers, analyze temporal and regional trends, and investigate the relationships between energy usage, economic indicators, and environmental impact.

INTRODUCTION

Energy plays a fundamental role in economic development, industrial growth, and overall improvement in living standards. Almost every sector—manufacturing, transportation, healthcare, agriculture, and digital services—depends heavily on reliable energy availability. With the rapid pace of globalization and technological advancement, global energy demand continues to rise steadily.

At the same time, growing global concern surrounding climate change, environmental degradation, and resource depletion has shifted attention toward sustainability and clean energy transition. Governments and international organizations are increasingly focusing on reducing carbon emissions, improving energy efficiency, and promoting renewable energy sources such as solar, wind, and hydropower.

Different countries adopt different energy strategies based on their economic conditions, natural resource availability, population growth, and policy frameworks. While some nations are actively investing in renewable energy and low-carbon technologies, others still rely heavily on fossil fuels to meet their energy demands. This variation results in significant differences in energy consumption patterns and environmental impact across regions.

Therefore, examining global energy usage is essential to understand the progress toward sustainability, the influence of energy consumption on economic growth, and its impact on environmental indicators such as CO_2 emissions. Analyzing these patterns also helps identify high-emission regions, energy inefficiencies, and opportunities for cleaner energy adoption.

This project applies Exploratory Data Analysis (EDA) techniques to a global energy consumption dataset to uncover meaningful relationships between key energy variables. By using data visualization and statistical analysis, the study aims to derive actionable insights into global energy patterns, trends, and correlations, providing a data-driven foundation for future research and policy formulation.

OBJECTIVE

The objective of this project is to analyze global energy consumption data using Exploratory Data Analysis (EDA) techniques to identify key trends and patterns related to electricity consumption, fossil fuel dependency, renewable energy adoption, and environmental impact. The study aims to gain a comprehensive understanding of how different energy sources contribute to overall energy usage across countries and over time.

This analysis involves examining statistical distributions, visualizing energy consumption trends, and detecting outliers that may indicate unusual energy behavior or data anomalies. Additionally, relationships between multiple variables—such as energy consumption, population, and CO_2 emissions—are explored to understand their interdependencies and influence on sustainability outcomes.

Beyond exploratory analysis, basic Machine Learning techniques are applied to evaluate the predictive potential of the dataset. These models aim to assess how well energy-related variables can predict outcomes such as emissions or energy consumption levels, providing preliminary insights into forecasting capabilities. Overall, the project seeks to combine data exploration and simple predictive modeling to support data-driven understanding of global energy consumption patterns and sustainability challenges.

SCOPE

The scope of this project includes acquiring and understanding a global energy consumption dataset, followed by systematic data preprocessing to ensure high-quality analysis. This involves handling missing values through appropriate imputation techniques, treating outliers to reduce distortion, and performing data cleaning to improve reliability and consistency of the dataset.

Exploratory Data Analysis (EDA) is conducted in multiple stages, including univariate, bivariate, and multivariate analysis, to examine the distribution, variation, and relationships among key energy-related variables across different countries and time periods. Visualization techniques are extensively used to highlight trends, correlations, and regional differences in energy consumption patterns.

The project further investigates critical factors such as fossil fuel dependency, renewable energy growth, GDP influence, population impact, and carbon emission patterns. These analyses help in understanding how economic and environmental indicators interact with energy usage and sustainability efforts.

To enable meaningful comparison across variables with different scales, standardization and normalization techniques are applied to numerical features. In addition, a basic predictive modeling approach using machine learning algorithms is implemented to assess the feasibility of forecasting energy consumption or emissions based on historical data.

The study is limited to the scope of the available dataset and does not incorporate external influences such as geopolitical events, policy interventions, climatic variations, or real-time energy consumption changes. Despite these limitations, the project provides a strong analytical foundation for understanding long-term global energy trends and serves as a baseline for future research in energy sustainability, demand forecasting, and data-driven policy analysis.

LITERTURE REVIEW

Several research studies have highlighted the critical importance of analyzing energy consumption trends to achieve global sustainability goals and climate change mitigation objectives. Reports from international agencies and research organizations indicate that energy consumption has direct and far-reaching implications for environmental performance, economic growth, and overall global development.

A significant portion of the existing literature emphasizes that excessive dependence on fossil fuels is a major contributor to rising carbon emissions, global warming, and long-term environmental degradation. Studies consistently show a strong relationship between fossil fuel usage and increased CO_2 emissions, reinforcing the need for cleaner and more sustainable energy alternatives.

Researchers have also extensively examined renewable energy adoption policies and evaluated their effectiveness in reducing greenhouse gas emissions. Findings suggest that countries investing in renewable energy infrastructure tend to experience lower emission growth rates and improved energy security. Additionally, literature highlights that economic growth plays a crucial role in influencing energy demand, particularly in rapidly industrializing and developing nations where energy consumption rises alongside industrial output and urbanization.

Energy transitions—such as shifting from coal- and oil-based systems toward solar, wind, hydro, and nuclear energy—are being studied intensively as viable pathways for sustainable global development. These transitions are viewed as essential for balancing economic progress with environmental responsibility.

Many analytical studies and government reports rely on large-scale energy datasets to observe consumption patterns, identify sector-wise energy usage, and estimate future energy demand. Exploratory Data Analysis (EDA) has emerged as a widely adopted approach in energy research due to its ability to reveal consumption behavior, visualize temporal trends, examine correlations among variables, and identify potential drivers of increasing energy demand. Overall, prior literature strongly supports continuous, data-driven analysis of global energy trends as a foundation for developing effective sustainability strategies and guiding energy transition policies.

DATASET SELECTION & PREPROCESSING

DATASET SELECTION:

The dataset contains global historical records related to multiple energy indicators such as electricity consumption, energy sources, emissions, economic indicators, and demographic information recorded across several countries over multiple years.

Dataset Description:

- Number of rows and columns: ~12,000 rows, 17 columns.
- Feature types:
  - Categorical: Country
  - Numerical: All remaining columns

The dataset covers energy consumption patterns, energy source distribution, economic variables, and environmental indicators. These diverse attributes enable detailed exploratory data analysis across countries, time periods, and energy sources, supporting global sustainability and energy policy evaluation.

Column Descriptions:

- Country: Name of the country to which the statistical data belongs.
- Total_Energy_Consumption_TWh: Total energy consumption measured in terawatt-hours.
- Electricity_Consumption_TWh: Total electricity usage measured in terawatt-hours.
- Oil_Consumption_Barrels: Total oil consumption measured in barrels.
- Cal_Consumption_Tonnes: Total coal usage measured in tons.
- NaturalGas_Consumption_CubicMeters: Total natural gas consumption measured in cubic meters.
- Renewable_Share_Percent: Percentage contribution of renewable energy sources in total energy mix.
- Nuclear_Share_Percent: Percentage share of nuclear energy in total energy production.
- CO2_Emissions_Mt: Carbon dioxide emissions measured in millions of tons.
- Population_Millions: Total population of the country measured in millions.
- GDP_BillionUSD: Gross domestic product measured in billions of US dollars.
- Energy_Per_Capita_MWh: Energy consumed per individual measured in megawatt-hours.
- Energy_Imports_Percent: Percentage of energy imported by the country.
- FossilFuel_Share_Percent: Percentage share of fossil fuels (oil, coal, gas) in energy mix.

HANDLING MISSING VALUES

Rows with missing values:

```
df.isnull().sum()
```

```
Country                              600
Year                                 600
Total_Energy_Consumption_TWh         600
Electricity_Consumption_TWh          600
Oil_Consumption_Barrels              600
Coal_Consumption_Tonnes              600
NaturalGas_Consumption_CubicMeters   600
Renewable_Share_Percent              600
Nuclear_Share_Percent                600
CO2_Emissions_Mt                     600
Population_Millions                  600
GDP_BillionUSD                       600
Energy_Per_Capita_MWh               1166
Energy_Imports_Percent               600
Energy_Exports_Percent               600
FossilFuel_Share_Percent             600
Electricity_Price_USD_kWh            600
dtype: int64
```

Strategy for handling missing values:

Missing values were identified across several numerical columns. Since the missing values were relatively small compared to the entire dataset and were uniformly distributed, the rows containing NaN values were removed using the **dropna()** function to maintain a clean dataset for analysis and machine learning.

OUTLIER DETECTION AND TREATMENT

Outliers were detected using the Interquartile Range (IQR) technique.

outlier detection

```
# === Detect outliers in each numeric column using IQR ===

num_cols = df.select_dtypes(include=['float64', 'int64']).columns

outlier_counts = {}

for col in num_cols:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1

    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR

    count = ((df[col] < lower) | (df[col] > upper)).sum()
    outlier_counts[col] = count

# print results
for col, cnt in outlier_counts.items():
    print(f"{col}: {cnt}")
```

```
Year: 0
Total_Energy_Consumption_TWh: 384
Electricity_Consumption_TWh: 369
Oil_Consumption_Barrels: 339
Coal_Consumption_Tonnes: 415
NaturalGas_Consumption_CubicMeters: 364
Renewable_Share_Percent: 0
Nuclear_Share_Percent: 0
CO2_Emissions_Mt: 419
Population_Millions: 45
GDP_BillionUSD: 409
Energy_Per_Capita_MWh: 790
Energy_Imports_Percent: 0
Energy_Exports_Percent: 0
FossilFuel_Share_Percent: 0
Electricity_Price_USD_kWh: 57
```

Strategy for handling missing values:

- Outliers in numerical variables were detected using the Interquartile Range (IQR) method.
- Instead of deleting records, outlier values were capped to the lower and upper IQR limits using the clip() function.
- The following line of code was used for capping:
  df[num_cols] = df[num_cols].clip(lower_bound, upper_bound, axis=1)
- Capping helped preserve the dataset size while reducing the influence of extreme values.
- This approach avoided distortion caused by outliers and improved the reliability of statistical analysis.

EXPLORATORY DATA ANALYSIS

Univariate Analysis

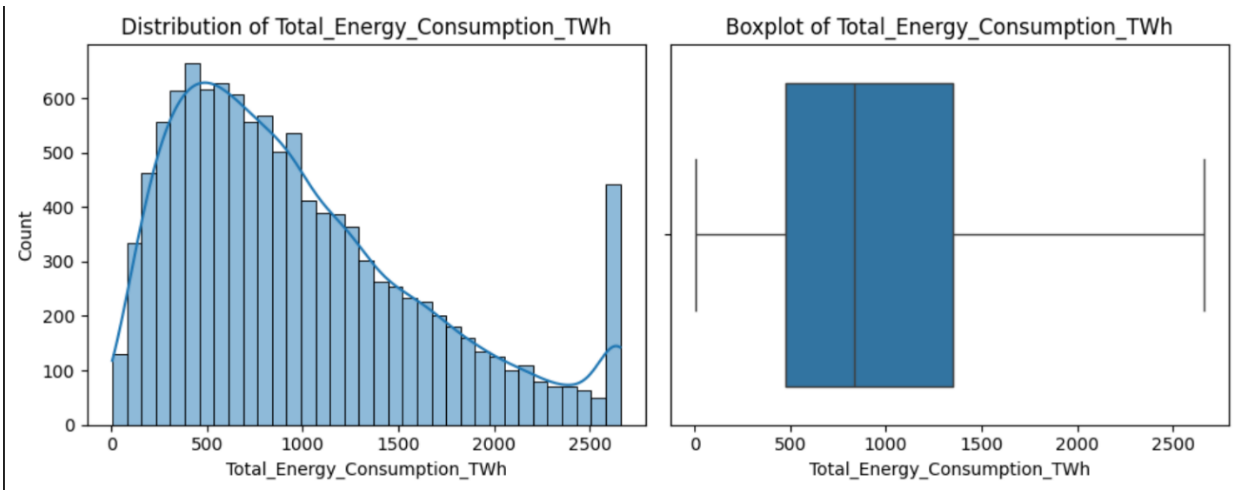
Univariate analysis was performed to understand the individual behavior of each variable in the dataset without considering its relationship with other variables. This step helps in identifying data distribution, central tendency, variability, skewness, and the presence of extreme values. It also provides a foundational understanding of the dataset before proceeding to bivariate and multivariate analysis.

Columns Used for Univariate Analysis
Univariate analysis was conducted on both **numerical** and **categorical** variables:

Numerical columns analyzed:

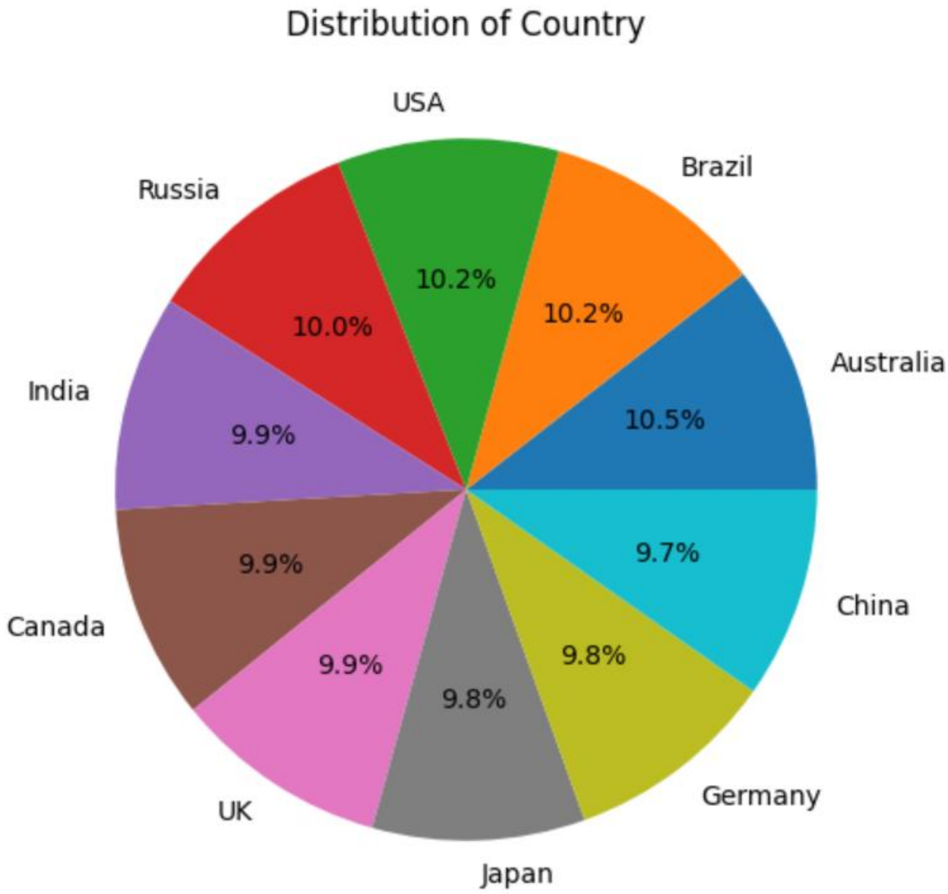
- Year
- Total_Energy_Consumption_TWh
- Electricity_Consumption_TWh
- Oil_Consumption_Barrels
- Coal_Consumption_Tonnes
- NaturalGas_Consumption_CubicMeters
- Renewable_Share_Percent
- Nuclear_Share_Percent
- CO_2_Emissions_Mt
- Population_Millions
- GDP_BillionUSD
- Energy_Per_Capita_MWh
- Energy_Imports_Percent
- Energy_Exports_Percent
- FossilFuel_Share_Percent
- Electricity_Price_USD_kWh

Example:



Categorical column analyzed:

- Country



Findings from Univariate Analysis:

- The distribution of electricity consumption and total energy consumption shows significant variation across countries, indicating unequal energy usage globally.
- Fossil fuel share and renewable energy share display widespread, highlighting differences in energy dependency and adoption of sustainable sources.
- CO_2 emissions show a right-skewed distribution, suggesting that a small number of countries contribute disproportionately to global emissions.
- GDP and population also exhibit skewness, reflecting economic and demographic Inequality among countries.
- The pie chart for the Country variable revealed that the dataset contains balanced representation from multiple countries, ensuring fair comparative analysis.
- The presence of extreme values in several numerical variables justified the need for outlier treatment in later stages.

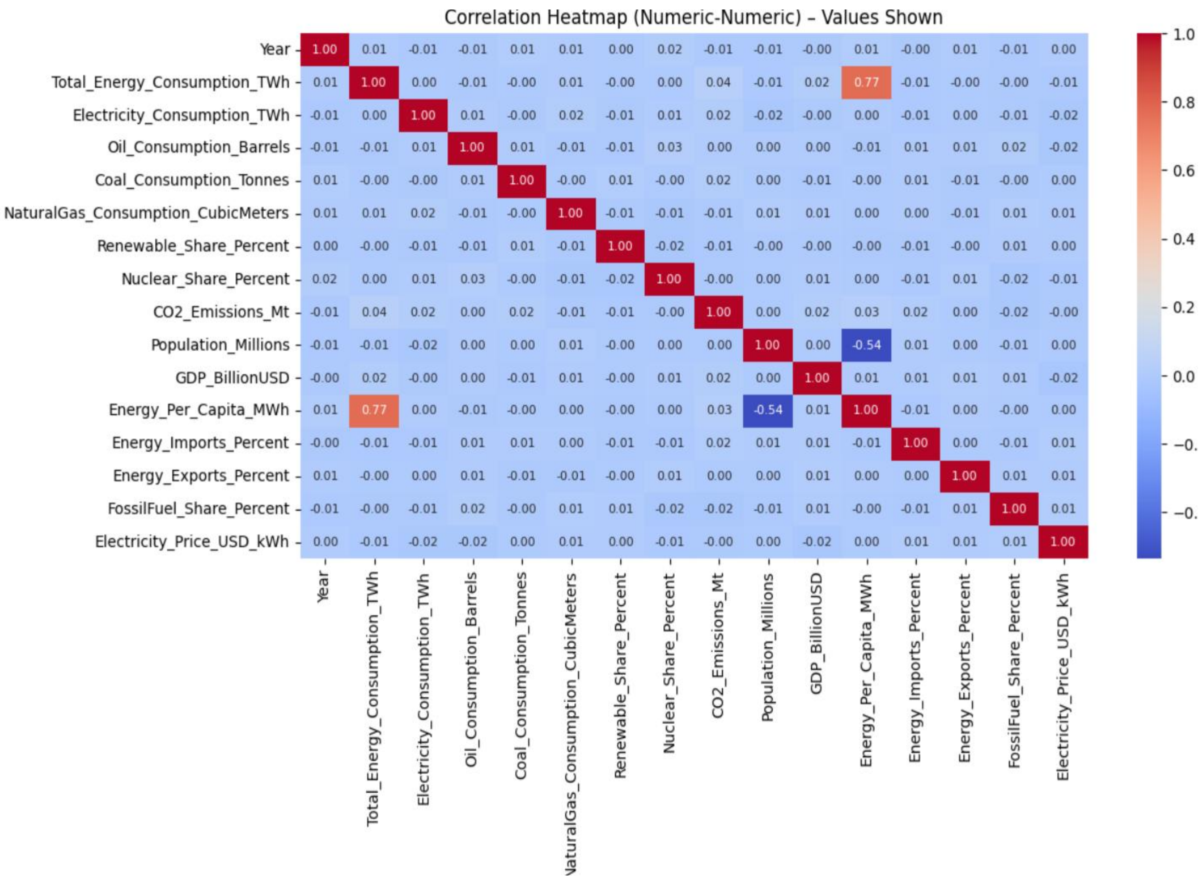
Bivariate Analysis

Bivariate analysis was performed to examine the relationship between two variables at a time and to understand how changes in one variable influence another. This analysis helps in identifying correlations, patterns, and dependencies between variables, which cannot be observed through univariate analysis alone.

Types of Bivariate Analysis Performed

1. Numerical – Numerical Analysis

Numerical–numerical analysis was conducted using scatter plots and correlation heatmaps to examine relationships between continuous variables.



Findings:

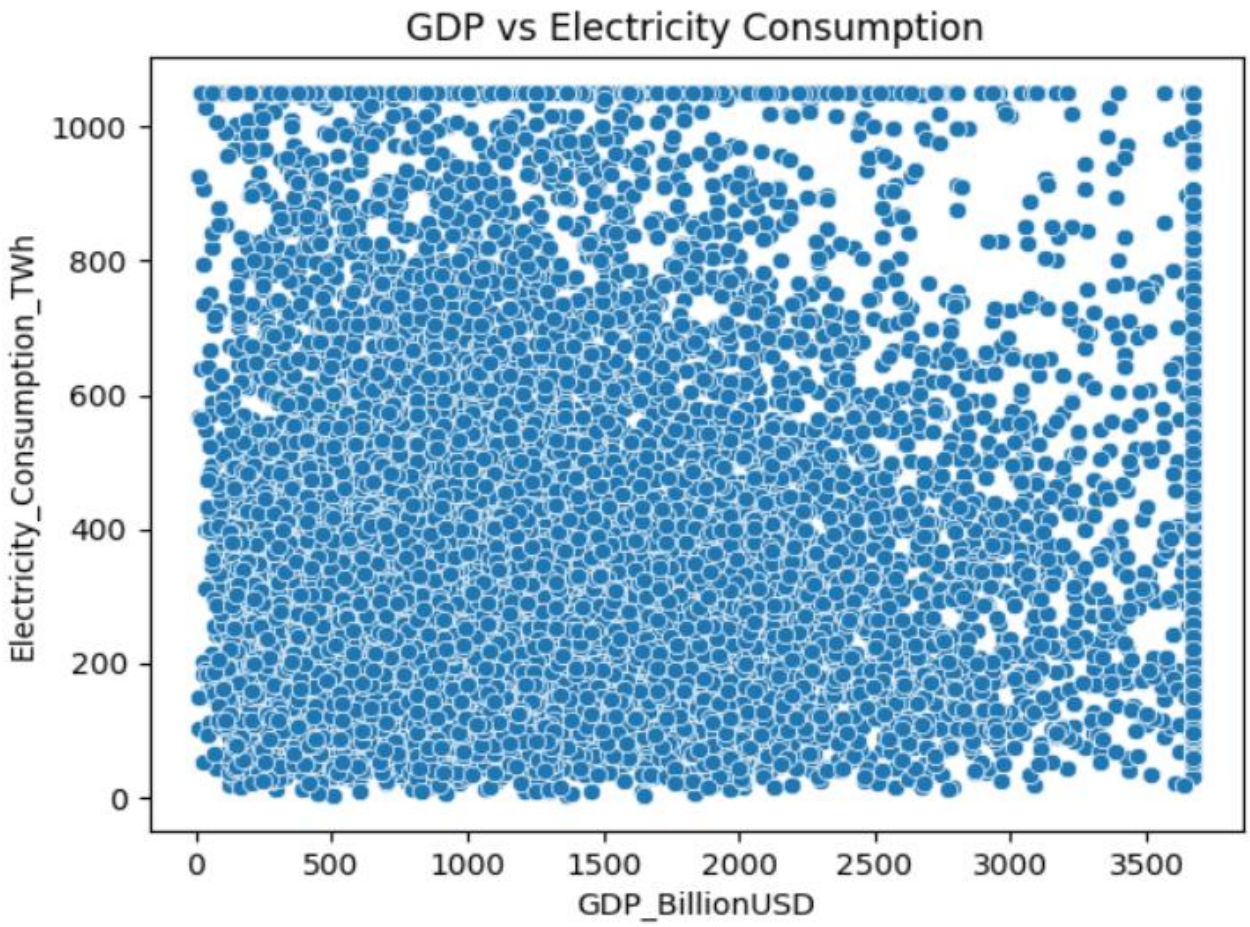
- A positive relationship was observed between **GDP and electricity consumption**, indicating that economically developed countries tend to consume more electricity.
- **Fossil fuel share shows a strong association with CO_2 emissions**, confirming that higher fossil fuel usage leads to increased carbon emissions.
- Energy consumption increases with population, though the relationship is not strictly linear due to differences in industrialization and energy efficiency.

2. Numerical – Categorical Analysis

Numerical–categorical analysis was performed using boxplots to compare numerical variables across different categories.

Columns analyzed:

- Electricity_Consumption_TWh across Country
- GDP_BillionUSD across Country



Findings:

- Significant variation in electricity consumption was observed across countries.
- Some countries consistently show higher median electricity usage, reflecting industrial and economic differences.
- Boxplots highlighted differences in spread and central tendency among countries, reinforcing global energy inequality.

3. Numerical vs Categorical

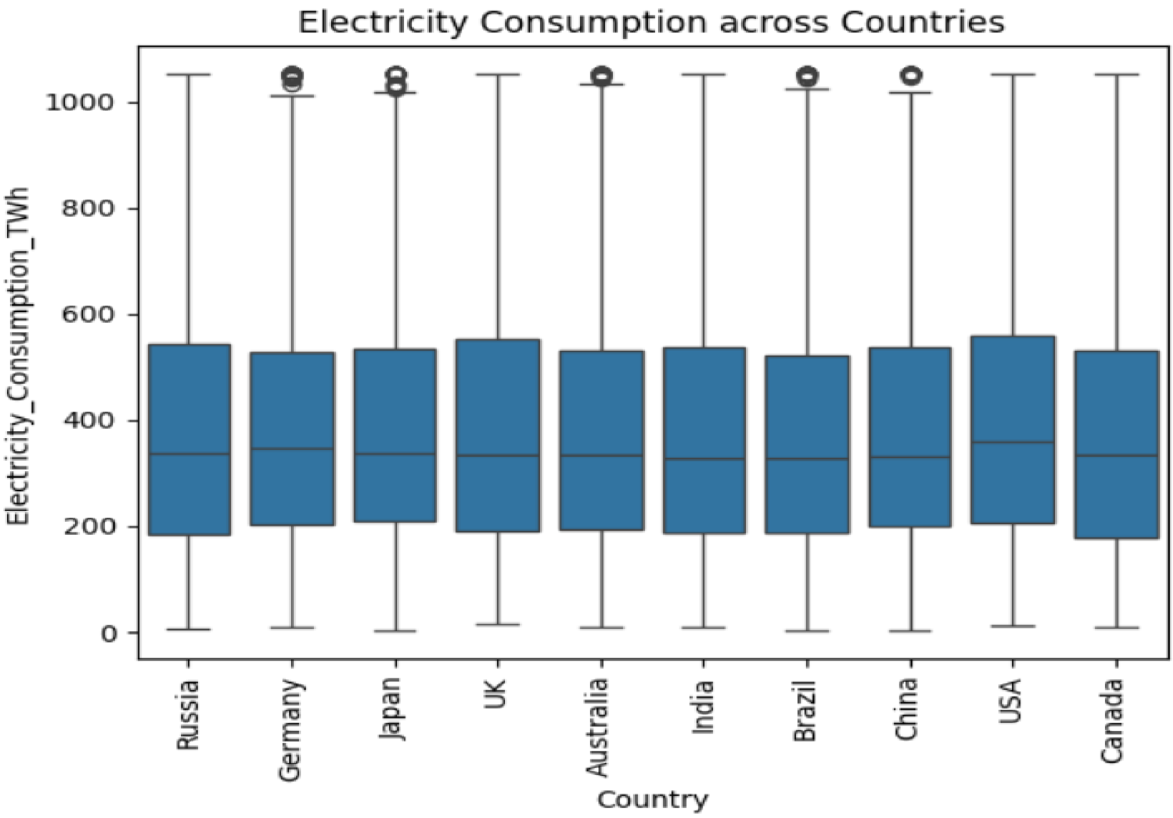
Numerical–categorical bivariate analysis was performed to understand how numerical energy indicators vary across different categories. In this project, the categorical variable **Country** was analyzed against multiple numerical energy and economic variables to observe country-wise variations.

Columns Used
Categorical Column:

- Country

Numerical Columns:

- Electricity_Consumption_TWh
- GDP_BillionUSD
- FossilFuel_Share_Percent
- CO_2_Emissions_Mt



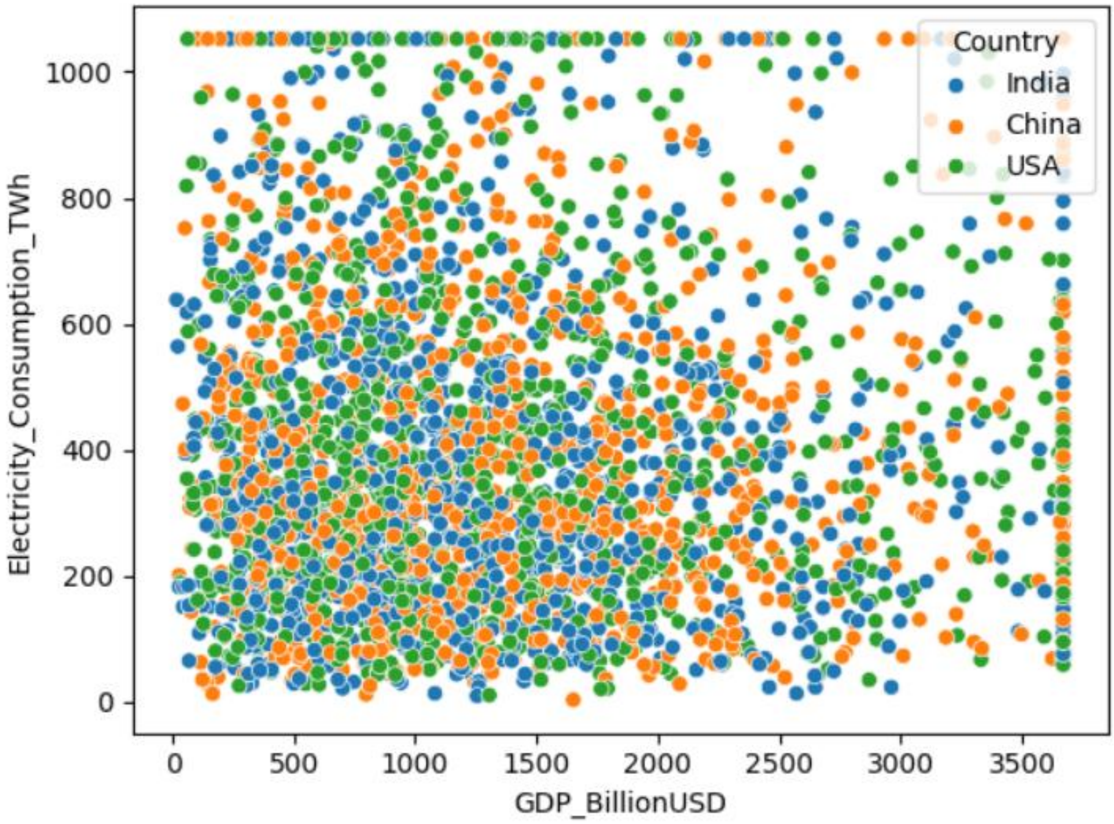
Findings from Numerical–Categorical Analysis

- Electricity consumption shows significant variation across different countries, indicating unequal energy demand and usage patterns globally.
- Economically stronger countries generally exhibit higher median electricity consumption compared to developing nations.
- Fossil fuel dependency differs considerably among countries, reflecting varying energy policies and resource availability.
- CO_2 emissions are higher in countries with greater fossil fuel usage, further highlighting environmental impact differences across regions.

Multivariate Analysis

The main objectives of performing multivariate analysis were:

- to identify correlations among multiple energy-related variables
- to understand combined effects of economic and environmental factors
- to detect multicollinearity between numerical features
- to support feature selection for machine learning modeling



Findings from Multivariate Analysis

- Strong positive correlations were observed between total energy consumption, electricity consumption, and GDP, indicating that economic growth is closely linked with higher energy demand.
- Fossil fuel share shows a strong positive relationship with CO_2 emissions, confirming the environmental impact of fossil-based energy usage.
- Renewable energy share exhibits an inverse relationship with fossil fuel share, suggesting a shift toward cleaner energy sources in some countries.
- Population shows moderate correlation with energy consumption, indicating that population size alone does not fully determine energy usage; industrialization and efficiency also play key roles.
- Electricity price shows relatively weak correlation with consumption, implying that pricing alone may not strongly influence overall electricity demand at a global level.

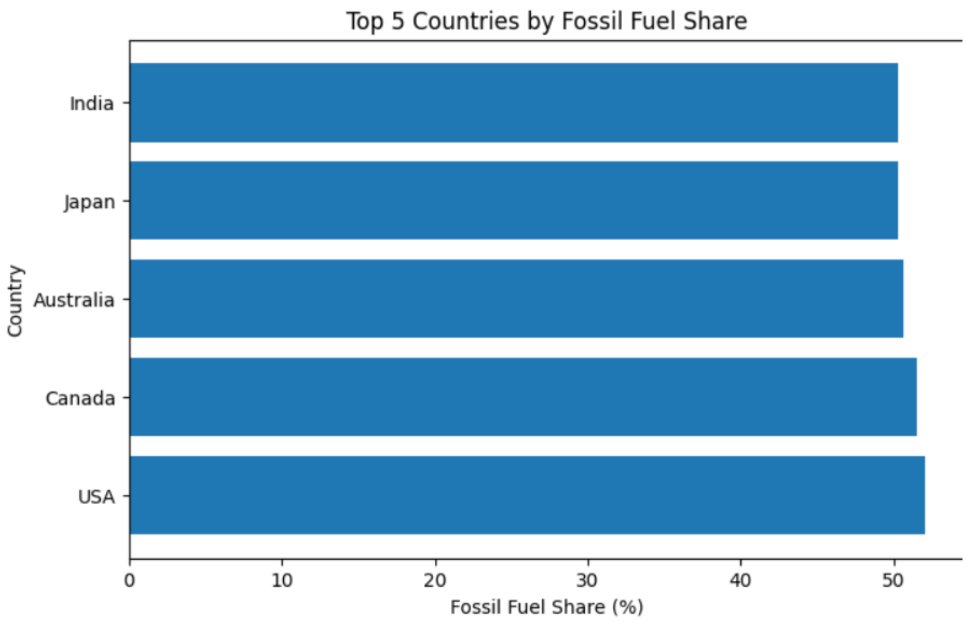
Additional Analysis Questions

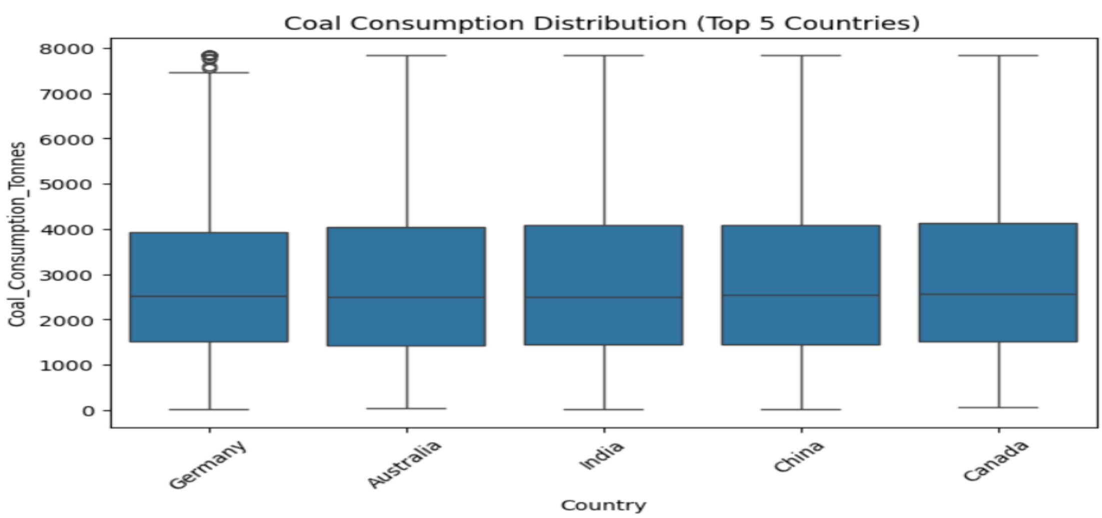
After completing univariate, bivariate, and multivariate analysis, additional exploratory questions were framed to gain deeper insights into country-wise energy consumption patterns. These questions focus on identifying major energy-consuming countries and understanding their dependence on different energy sources.

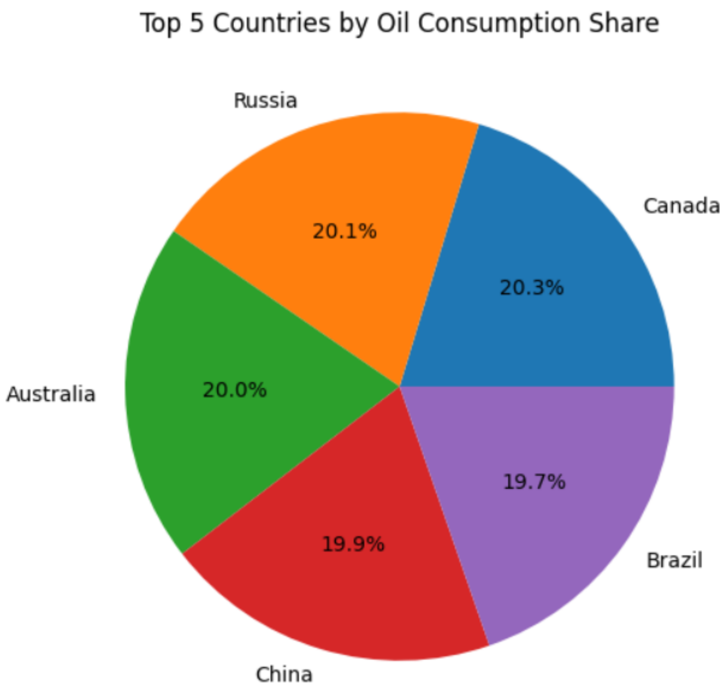
The following additional analysis questions were explored:

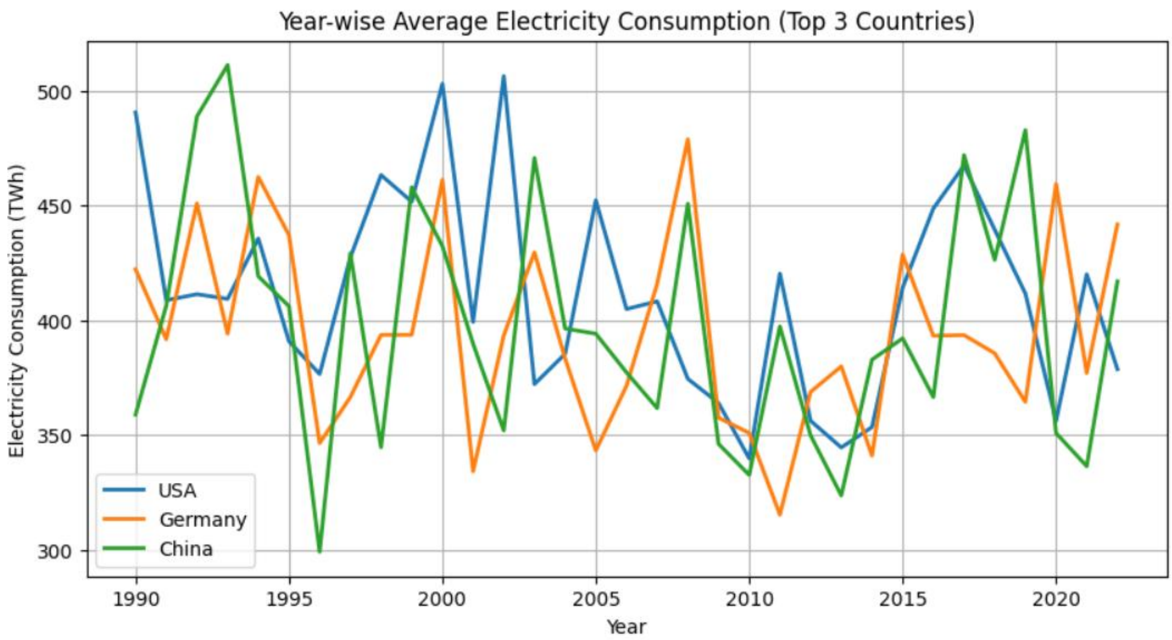
1. Which countries have the highest average fossil fuel consumption?
2. Which countries consume the highest amount of coal on average?
3. Which countries have the highest oil consumption?
4. Which countries consume the most electricity on average?
5. Which countries have the highest share of renewable energy in their total energy mix?

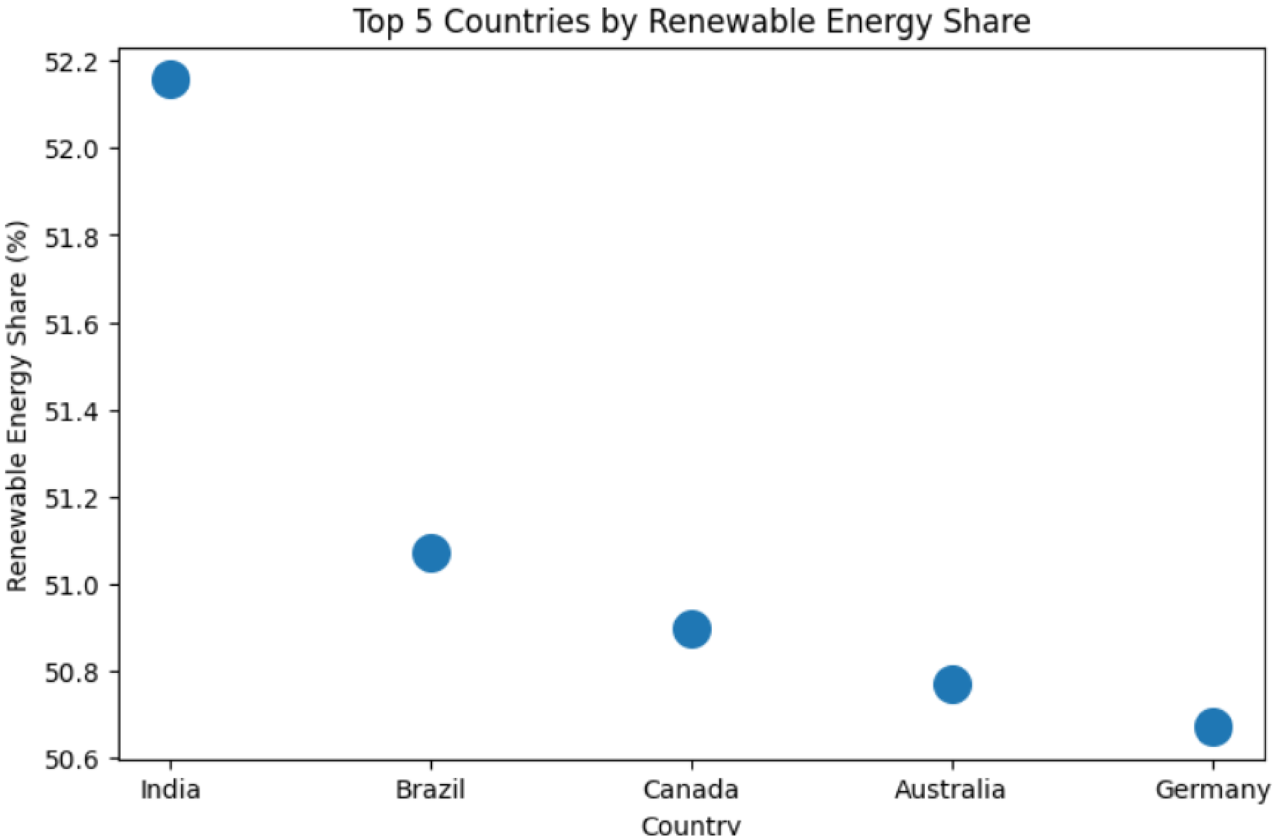
These questions were analyzed using a combination of visualizations such as bar charts, boxplots, pie charts, line charts, and scatter plots to provide a comprehensive comparison across countries.






Top 5 Countries by Oil Consumption Share
Russia
Canada
Australia
Brazil
China
20.1%
20.3%
20.0%
19.7%
19.9%


Year-wise Average Electricity Consumption (Top 3 Countries)
Electricity Consumption (TWh)
Year
USA
Germany
China
500
450
400
350
300
1990
1995
2000
2005
2010
2015
2020



Findings from Additional Analysis:

• Countries such as China, USA, and India show the highest fossil fuel share, indicating strong dependence on traditional energy sources to meet large industrial and population-driven energy demands.
• Coal consumption is highest in countries like China and India, reflecting continued reliance on coal-based power generation due to availability of resources and existing infrastructure.
• Oil consumption is dominant in countries such as USA and Russia, primarily driven by transportation, industrial activities, and economic development.
• Electricity consumption is highest in economically strong and industrialized countries such as USA, China, and Germany, highlighting the close relationship between economic growth and energy demand.
• Countries such as Germany and Canada show a relatively higher renewable energy share, indicating stronger adoption of sustainable energy practices and supportive energy policies.

Data Standardization (Z-Score Scaling)

Why Z-Score Standardization was applied
• The dataset contains numerical variables with different units and scales (e.g., GDP in billions, electricity consumption in TWh, population in millions).
• Without scaling, variables with larger values could dominate analysis and machines. Learning models.
• Z-score standardization ensures fair comparison among all numerical features.
• It improves numerical stability and performance of machine learning algorithms.
• Standardization prepares the dataset for predictive modeling.

Parameters used in Z-Score Scaling
• Mean (μ): Average value of each numerical feature
• Standard Deviation (σ): Measure of spread of each feature
• Formula used:

$$\boldsymbol{Z = \frac{X - \mu}{\sigma}}$$

Findings after Standardization
• All numerical variables were transformed to have a mean approximately equal to 0.
• The standard deviation of each scaled feature became approximately 1.
• No single feature dominated due to scale differences after transformation.
• The standardized dataset was suitable for machine learning modeling.

	mean	std
Year	5.038309e-15	1.000044
Total_Energy_Consumption_TWh	-1.978924e-16	1.000044
Electricity_Consumption_TWh	-6.419816e-17	1.000044
Oil_Consumption_Barrels	-1.246566e-16	1.000044
Coal_Consumption_Tonnes	-1.647026e-16	1.000044
NaturalGas_Consumption_CubicMeters	6.232831e-19	1.000044
Renewable_Share_Percent	3.103950e-16	1.000044
Nuclear_Share_Percent	3.552714e-17	1.000044
CO2_Emissions_Mt	-2.792308e-16	1.000044
Population_Millions	-6.631732e-16	1.000044
GDP_BillionUSD	5.484891e-17	1.000044
Energy_Per_Capita_MWh	4.820463e-17	1.000046
Energy_Imports_Percent	-2.486900e-16	1.000044
Energy_Exports_Percent	1.483414e-16	1.000044
FossilFuel_Share_Percent	1.246566e-18	1.000044
Electricity_Price_USD_kWh	-2.739329e-16	1.000044

MACHINE LEARNING ANALYSIS

After completing data preprocessing, exploratory data analysis, and standardization, a basic Machine Learning model was applied to examine whether global energy indicators could be used to predict electricity consumption. This step was included to demonstrate predictive capability of the cleaned dataset and to move beyond descriptive analysis.

Purpose of Machine Learning
• To test whether energy, economic, and environmental variables can predict electricity consumption
• To validate insights obtained from exploratory data analysis

Target Variable
• **Electricity_Consumption_TWh**
This variable was selected as the target because electricity consumption is a key indicator of energy demand and economic activity.

Input Features
All standardized numerical variables except the target were used as input features, including:

- GDP_BillionUSD
- FossilFuel_Share_Percent
- Renewable_Share_Percent
- CO_2_Emissions_Mt
- Population_Millions
- Energy_Per_Capita_MWh
- Other energy consumption indicators

The categorical column *Country* was excluded since the model requires numerical input.

Model Used
• **Linear Regression**
Linear Regression was chosen because:

- It is simple and easy to interpret
- It helps identify linear relationships between variables
- It is suitable for demonstrating basic machine learning concepts

Model Training
• The dataset was split into **80% training data** and **20% testing data**
• The model was trained using standardized numerical features
• Rows containing missing values were removed before training

Evaluation Metric
• **R^2 Score (Coefficient of Determination)**
R^2 measures how much variation in the target variable can be explained by the input features.

Model Output and Findings
• The Linear Regression model produced a **very low (slightly negative) R^2 score**
• This indicates that the linear model was unable to effectively capture the complex relationships in global energy data
• The result suggests that electricity consumption depends on multiple non-linear and country-specific factors

CONCLUSION

This project successfully conducted a comprehensive Exploratory Data Analysis (EDA) on a global energy consumption dataset to understand energy usage patterns across different countries and time periods. The study focused on key energy indicators such as electricity consumption, fossil fuel dependency, renewable energy share, GDP, population, and carbon emissions. Through systematic data preprocessing, missing values were handled, outliers were treated using the IQR capping method, and numerical features were standardized using Z-score scaling to ensure consistency and reliability in analysis.

Univariate analysis helped in understanding the individual distribution and characteristics of each variable, highlighting variations in energy usage and economic indicators across countries. Bivariate analysis revealed meaningful relationships, particularly between electricity consumption and GDP, as well as between fossil fuel usage and CO_2 emissions. Multivariate analysis using correlation heatmaps further demonstrated the interconnected nature of energy, economic, and environmental factors, emphasizing the complexity of global energy systems.

A basic machine learning model using Linear Regression was implemented to explore the predictive capability of the dataset. The low R^2 score obtained from the model indicates that global electricity consumption is influenced by multiple complex and non-linear factors that cannot be fully captured by a simple linear model. This outcome highlights the importance of advanced modeling techniques for real-world energy prediction tasks.

Overall, this project demonstrates how effective data preprocessing, visualization, and analytical techniques can be applied to extract meaningful insights from large-scale energy datasets. The findings emphasize the need for sustainable energy planning and further data-driven analysis to support global energy policy and decision-making.

FUTURE SCOPE

The current study provides a strong foundation for understanding global energy consumption patterns; however, there are several opportunities to extend and enhance the scope of this work. One important direction is the application of advanced machine learning and deep learning models such as Random Forest, Gradient Boosting, Support Vector Machines, and Neural Networks. These models can better capture complex and non-linear relationships between energy, economic, and environmental variables, potentially improving prediction accuracy.

Time-series analysis can also be incorporated to study long-term energy trends and seasonal patterns. Techniques such as ARIMA, LSTM, and Prophet can be used to forecast future electricity demand at country or regional levels, which would be valuable for energy planning and infrastructure development. Incorporating temporal dynamics would provide deeper insights into how energy consumption evolves over time.

The dataset can be enriched by integrating external variables such as climate conditions, energy policies, technological adoption, carbon pricing mechanisms, and geopolitical factors. Including such features would allow a more holistic analysis of energy consumption behavior and help evaluate the impact of policy interventions on energy usage and emissions.

From a visualization perspective, the development of more advanced interactive dashboards using tools such as Tableau, Power BI, or web-based frameworks can further enhance user engagement. Dashboards can be extended to support real-time data updates, scenario analysis, and country-wise comparisons, enabling policymakers and researchers to explore insights dynamically.

Future research may also focus on sustainability analysis by evaluating renewable energy transition rates, emission reduction pathways, and progress toward global climate targets. Comparative studies between developed and developing nations could reveal structural differences in energy consumption and highlight best practices for sustainable development. Overall, extending this project in these directions would contribute to more informed decision-making and support global efforts toward sustainable energy management.

REFERENCES

Kaggle. *Global Energy Consumption Dataset*.
Available at: https://www.kaggle.com
(Accessed for global energy statistics and indicators used in this project)

International Energy Agency (IEA). *World Energy Statistics and Balances*.
Available at: https://www.iea.org
Used as a reference for understanding global energy trends and consumption patterns.

International Renewable Energy Agency (IRENA). *Global Renewable Energy Statistics*.
Available at: https://www.irena.org
Referenced for renewable energy adoption and sustainability insights.

United Nations. *Energy and Climate Change Reports*.
Available at: https://www.un.org
Used for understanding the global environmental impact of energy consumption.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013).
An Introduction to Statistical Learning. Springer.
Referenced for concepts related to regression analysis and machine learning fundamentals.

Pandas Development Team. *pandas Documentation*.
Available at: https://pandas.pydata.org
Used for data manipulation and preprocessing techniques.
Scikit-learn Developers. *scikit-learn User Guide*.
Available at: https://scikit-learn.org
Used for machine learning model implementation and evaluation.