Exploratory Data Analysis on

ENERGY CONSUMPTION

by

Group 16

Dev Vyas
ID: 202201453
Course: BTech(ICT-CS)

Dharmi Patel
ID: 202201467
Course: BTech(ICT-CS)

Preksha Shah
ID: 202203004
Course: BTech(MnC)

Course Code: IT 462
Semester: Autumn 2024

Under the guidance of

Dr. Gopinath Panda

Dhirubhai Ambani Institute of Information and Communication Technology

December 2, 2024

Link to Github Repository.

Acknowledgment

I am writing this letter to express my heartfelt gratitude for your guidance and support throughout the duration of my project titled “ENERGY CONSUMPITON.” Your invaluable assistance has played a pivotal role in shaping the successful completion of this endeavor.

I am extremely fortunate to have had the opportunity to work under your mentorship. Your expertise, encouragement, and willingness to share your knowledge have been instrumental in elevating the quality and scope of my project. Your constructive feedback and insightful suggestions have helped me overcome challenges and develop a deeper understanding of the subject matter.

Furthermore, I would like to extend my appreciation to the entire team at Dhirubhai Ambani Institute of Information and Communication Technology for fostering an environment of collaboration and innovation. The resources and facilities provided have been crucial in conducting comprehensive research and analysis.

I would also like to express my gratitude to my peers and colleagues who have been supportive throughout this journey. Their valuable input and camaraderie have been a constant source of motivation.

Completing this project has been a tremendous learning experience, and I am confident that the knowledge and skills acquired during this endeavor will serve as a solid foundation for my future endeavors.

Once again, thank you for your unwavering guidance and belief in my abilities. Your mentorship has been invaluable, and I am truly grateful for the opportunity to work with you.

Sincerely,
Dev Vyas,202201453
Dharmi Patel,202201467
Preksha Shah,202203004

DECLARATION

We, 202201453,202201467,202203004 hereby declare that the EDA project work presented in this report is our original work and has not been submitted for any other academic degree. All the sources cited in this report have been appropriately referenced.

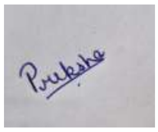
We acknowledge that the data used in this project is obtained from the data.gov site. We also declare that we have adhered to the terms and conditions mentioned in the website for using the dataset. We confirm that the dataset used in this project is true and accurate to the best of our knowledge.

We acknowledge that we have received no external help or assistance in conducting this project, except for the guidance provided by our mentor Prof. Gopinath Panda. We declare that there is no conflict of interest in conducting this EDA project.

We hereby sign the declaration statement and confirm the submission of this report on 2nd December, 2024.

Dev Vyas
ID: 202201453
Course:
BTech(ICT-CS)

Dharmi Patel
ID: 202201467
Course:
BTech(ICT-CS)



Preksha Shah
ID: 202203004
Course:
BTech(MnC)

CERTIFICATE

This is to certify that Group 16 comprising Dev Vyas, Dharmi Patel, and Preksha Shah has successfully completed an exploratory data analysis (EDA) project on the Energy Consumption, which was obtained from data.gov.in.

The EDA project presented by Group 16 is their original work and has been completed under the guidance of the course instructor, Prof. Gopinath Panda, who has provided support and guidance throughout the project. The project is based on a thorough analysis of the Energy Consumption dataset, and the results presented in the report are based on the data obtained from the dataset.

This certificate is issued to recognize the successful completion of the EDA project on the Energy Consumption, which demonstrates the analytical skills and knowledge of the students of Group 16 in the field of data analysis.

Signed,
Dr. Gopinath Panda,
IT 462 Course Instructor
Dhirubhai Ambani Institute of Information and Communication Technology
Gandhinagar, Gujarat, INDIA.

December 2, 2024

Contents

List of Figures

List of Tables

Abstract

This report presents a comprehensive analysis of global energy consumption trends, leveraging monthly and annual datasets sourced from data.gov. The analysis encompasses data cleaning, transformation, exploratory data analysis (EDA), feature engineering, and model fitting, with a focus on uncovering insights into energy production, consumption, and distribution patterns over time.

The EDA section highlights seasonal trends, long-term patterns, and variability across key energy metrics, such as fossil fuels, nuclear power, and renewable energy. Advanced visualization techniques—including histograms, density plots, seasonal heatmaps, and 3D scatter plots—illustrate shifts in energy production and consumption by sector, as well as global adoption of renewables versus continued reliance on fossil fuels. Year-on-year growth insights, rolling averages, and correlation analyses further emphasize emerging patterns and interdependencies among energy metrics.

Feature engineering methods, including feature selection, feature extraction, and outlier handling, play a critical role in preparing the datasets for predictive modeling. Seasonal decomposition and derived features are employed to capture temporal and structural nuances in the data.

The modeling chapter evaluates the performance of machine learning models such as Linear Regression, XGBRegressor, and XGBoost Classifier in predicting key energy metrics. Techniques like SelectKBest, Principal Component Analysis (PCA), and Recursive Feature Elimination (RFE) are applied to improve model accuracy. Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R^2, and classification metrics (AUC, precision, recall, and F1-score) are used to compare and interpret model results.

The findings provide actionable insights into global energy consumption, highlighting critical challenges like seasonal demand variability and the slow transition toward renewable energy. By combining robust data analytics with predictive modeling, this report aims to inform policymakers and stakeholders, offering a data-driven foundation for strategic energy planning and forecasting.

Chapter 1. Introduction

1.1 Project idea

One of the most important measures of a country's energy landscape is **Energy Consumption**. Energy Consumption is a comprehensive measure of all energy consumed by end users, including transformations and distribution losses for electricity generation. Energy Consumption helps us understand how energy resources are being utilized and what patterns emerge in energy demand and supply. The energy consumption is made up of various components, with the major ones being:

- **Total Fossil Fuels Consumption**
- **Nuclear Electric Power Consumption**
- **Total Renewable Energy Consumption**

Energy production and consumption metrics provide insights into a country's energy self-sufficiency and sustainability. The relationship between production and consumption is given as:

Total Primary Energy Consumption = Total Primary Energy Production + Primary Energy Net Imports + Primary Energy Stock Changes

Where:

- Total Primary Energy Production includes fossil fuels, nuclear, and renewable energy production
- Primary Energy Net Imports represents the difference between imports and exports
- Primary Energy Stock Changes account for variations in energy reserves

The objectives of this project are to:

- Analyze monthly and annual energy consumption and production patterns
- Forecast energy metrics for upcoming years using time series analysis
- Study the relationship between different energy sources in both production and consumption
- Evaluate seasonal patterns and long-term trends

This comprehensive analysis would help stakeholders understand current energy patterns and make informed decisions about future energy infrastructure and sustainability initiatives.

1.2 Data Collection

The dataset used in this analysis is sourced from `data.gov`, a U.S. government platform that provides a wide range of open data across various sectors, including energy, environment, and economics. The platform ensures transparency and facilitates access to high-quality datasets for public use, promoting research, innovation, and informed decision-making.

For this analysis, the dataset titled *Monthly and Annual Energy Consumption by Sector* has been utilized. It provides comprehensive energy consumption data by sector, including energy use by different sectors of the economy, such as residential, commercial, and industrial, across monthly and annual periods.

	Month	Total Fossil Fuels Production	Nuclear Electric Power Production	Total Renewable Energy Production	Total Primary Energy Production	Primary Energy Imports	Primary Energy Exports	Primary Energy Net Imports	Primary Energy Stock Change and Other	Total Fossil Fuels Consumption	Nuclear Electric Power Consumption	Total Renewable Energy Consumption	Total Primary Energy Consumption
0	1973-01-01	4.932632	0.068103	0.219839	5.220574	1.173080	0.125781	1.047299	0.771858	6.747651	0.068103	0.219839	7.039731
1	1973-02-01	4.729582	0.064634	0.197330	4.991545	1.168005	0.120883	1.047122	0.390129	6.163095	0.064634	0.197330	6.428796
2	1973-03-01	4.946902	0.072494	0.218686	5.238082	1.309473	0.139950	1.169523	-0.067640	6.044647	0.072494	0.218686	6.339964
3	1973-04-01	4.716271	0.064070	0.209330	4.989672	1.085169	0.194185	0.890984	-0.110067	5.493184	0.064070	0.209330	5.770589
4	1973-05-01	4.956995	0.062111	0.215982	5.235087	1.162804	0.196775	0.966029	-0.305335	5.813551	0.062111	0.215982	5.895781
5	1973-06-01	4.716096	0.073968	0.208249	4.998313	1.129614	0.178871	0.950743	-0.315356	5.347479	0.073968	0.208249	5.633700
6	1973-07-01	4.748276	0.075890	0.207800	5.031966	1.216246	0.164134	1.052112	-0.291695	5.504556	0.075890	0.207800	5.792383
7	1973-08-01	5.041997	0.084883	0.203432	5.330313	1.297848	0.191049	1.106799	-0.395780	5.748879	0.084883	0.203432	6.041332
8	1973-09-01	4.731504	0.085724	0.185300	5.002528	1.226145	0.146492	1.079652	-0.430371	5.376782	0.085724	0.185300	5.651809
9	1973-10-01	4.998321	0.081971	0.193514	5.273806	1.316774	0.207655	1.108119	-0.337003	5.765300	0.081971	0.193514	6.044923

Figure 1.1: Montly Energy Dataset

```
energy_annual.head(10)
```

	Annual Total	Total Fossil Fuels Production	Nuclear Electric Power Production	Total Renewable Energy Production	Total Primary Energy Production	Primary Energy Imports	Primary Energy Exports	Primary Energy Net Imports	Primary Energy Stock Change and Other	Total Fossil Fuels Consumption	Nuclear Electric Power Consumption	Total Renewable Energy Consumption	Total Primary Energy Consumption
0	1949.0	28.740479	0	1.872627	30.613106	1.448158	1.59176	-0.143602	0.396915	28.988371	0	1.872627	30.866419
1	1950.0	32.553205	0	1.906525	34.45973	1.912887	1.465322	0.447565	-1.379921	31.614755	0	1.906525	33.527374
2	1951.0	35.782118	0	1.8908	37.672919	1.892425	2.621545	-0.729119	-1.057801	33.987736	0	1.8908	35.885997
3	1952.0	34.964662	0	1.848694	36.813356	2.145984	2.365131	-0.219147	-0.959659	33.778116	0	1.848694	35.63455
4	1953.0	35.338361	0	1.792616	37.130976	2.313042	1.866013	0.447029	-0.976252	34.802286	0	1.792616	36.601753
5	1954.0	33.752228	0	1.775242	35.52747	2.347876	1.696301	0.651575	-0.544362	33.851458	0	1.775242	35.634683
6	1955.0	37.346797	0	1.82074	39.167537	2.789908	2.285508	0.504399	-0.4569	37.380419	0	1.82074	39.215038
7	1956.0	39.75301	0	1.843178	41.596188	3.206741	2.945377	0.261364	-1.143199	38.855656	0	1.843178	40.714353
8	1957.0	40.11844	0.000112	1.788598	41.90715	3.528983	3.439441	0.089541	-1.303379	38.892315	0.000112	1.788598	40.693313
9	1958.0	37.205122	0.001915	1.813136	39.020172	3.884371	2.049839	1.834533	-0.347378	38.680955	0.001915	1.813136	40.507326

Figure 1.2: Annual Energy Dataset

1.3 Dataset Description

This analysis utilizes two datasets, the *Monthly Energy Consumption by Sector* and the *Annual Energy Consumption by Sector*, both sourced from the U.S. government's open data platform, data.gov. These datasets, provided by the U.S. Energy Information Administration (EIA), offer comprehensive energy consumption statistics across various sectors of the U.S. economy.

- **Monthly Dataset:** This dataset provides detailed energy consumption data on a monthly basis, beginning from 1973 onwards. It captures energy usage trends and patterns in various sectors, offering valuable insights into the seasonal and monthly variations in energy demand.
- **Annual Dataset:** This dataset aggregates the monthly data into yearly totals, offering a broader view of energy consumption and production patterns. It provides a summary of the total energy consumed and produced, as well as import/export statistics on a yearly basis.

1.3.1 Monthly Energy Dataset

DATASET NAME: US-Monthly-Energy-Data 1973-Present

This dataset contains detailed monthly energy consumption and production values from 1973 onwards. It tracks 12 key energy metrics on a monthly basis, providing granular insights into seasonal patterns and short-term trends. The values are in Quadrillion BTU to maintain consistency across different energy types and time periods.

- **Month (Time period: YYYY-MM format)**: The time period for each data entry.
- **Total Fossil Fuels Production**: The total amount of energy produced from fossil fuels.
- **Nuclear Electric Power Production**: The total amount of energy produced from nuclear power sources.
- **Total Renewable Energy Production**: The total amount of energy produced from renewable sources.
- **Total Primary Energy Production**: The total energy produced from all sources.
- **Primary Energy Imports**: The total energy imported into the country.
- **Primary Energy Exports**: The total energy exported from the country.
- **Primary Energy Net Imports**: The net balance of energy imports and exports.
- **Primary Energy Stock Change and Other**: Variations in energy reserves and other related metrics.
- **Total Fossil Fuels Consumption**: The total energy consumed from fossil fuels.
- **Nuclear Electric Power Consumption**: The energy consumed from nuclear electricity generation.
- **Total Renewable Energy Consumption**: The total energy consumed from renewable sources.
- **Total Primary Energy Consumption**: The total energy consumed across all sectors.

1.3.2 Annual Energy Dataset

DATASET NAME: US-Annual-Energy-Data 1973-Present

This dataset represents the yearly aggregated energy metrics from 1973 to the present. It provides a comprehensive view of annual energy patterns, making it suitable for long-term trend analysis and strategic planning. The data is aggregated from monthly values to create annual totals.

- **Year (Time period: YYYY format)**: The time period for each data entry.
- **Annual Total Fossil Fuels Production**: The total amount of energy produced from fossil fuels on an annual basis.
- **Annual Nuclear Electric Power Production**: The total energy produced from nuclear power sources annually.
- **Annual Total Renewable Energy Production**: The total energy produced from renewable sources on an annual basis.
- **Annual Total Primary Energy Production**: The total energy produced from all sources on an annual basis.
- **Annual Primary Energy Imports**: The total energy imported into the country annually.
- **Annual Primary Energy Exports**: The total energy exported from the country on an annual basis.
- **Annual Primary Energy Net Imports**: The net balance of energy imports and exports annually.
- **Annual Primary Energy Stock Change**: Changes in energy reserves and other related metrics for the year.
- **Annual Total Fossil Fuels Consumption**: The total energy consumed from fossil fuels on an annual basis.
- **Annual Nuclear Electric Power Consumption**: The energy consumed from nuclear electricity generation annually.
- **Annual Total Renewable Energy Consumption**: The total energy consumed from renewable sources on an annual basis.
- **Annual Total Primary Energy Consumption**: The total energy consumed across all sectors on an annual basis.

1.4 Packages required

We have used various packages during our project to help us analyse and visualise the data efficiently. The packages are:

1. **NumPy (np)**: NumPy is a fundamental package for scientific computing with Python. We have used it for numerical computations, array manipulation, and linear algebra operations.

2. **Pandas (pd)**: Pandas is a powerful data manipulation and analysis library. We have used it for data cleaning, transformation, exploration, and analysis tasks.

3. **Seaborn (sns)**: Seaborn is a statistical data visualization library based on Matplotlib. We have used it for creating visually appealing plots for exploring relationships in data.

4. **Matplotlib.pyplot (plt)**: Matplotlib is a comprehensive plotting library for Python. We have used it for creating various types of plots and visualizations, including line plots, scatter plots, histograms, etc.

5. **Missingno (msno)**: Missingno is a Python library for visualizing missing data in datasets. It provides informative visualizations such as bar charts, heatmaps, and dendrograms to identify missing values in datasets. It helped us to know the completeness of data.

6. **Scikit-learn (sklearn)**: Scikit-learn is a machine-learning library for Python. We have used it for tasks such as regression, and model selection.

7. **Plotly**: Plotly is a graphing library that enables the creation of interactive, web-based visualizations. We have used it for creating a wide range of plots, including 3D plots, geographical maps, and statistical visualizations. Plotly provides both high-level and low-level interfaces for creating and customizing interactive visualizations with ease.

Chapter 2. Data Cleaning

Data Cleaning is the process of analyzing missing values and outliers in a dataset and handling them according to the needs of the objective. [Missing value Description]
From the datasets mentioned above, we encountered missing values in some of them, while some of them didn't have any missing values. Given below is the missing data analysis for datasets in which we have encountered missing values.

For outliers, we can't replace them with some other values as they will not accurately represent reality.

2.1 Missing data analysis

2.1.1 Monthly Energy Dataset

Here, we have encountered no missing values

```
energy.isnull().sum()
```

	0
Month	0
Total Fossil Fuels Production	0
Nuclear Electric Power Production	0
Total Renewable Energy Production	0
Total Primary Energy Production	0
Primary Energy Imports	0
Primary Energy Exports	0
Primary Energy Net Imports	0
Primary Energy Stock Change and Other	0
Total Fossil Fuels Consumption	0
Nuclear Electric Power Consumption	0
Total Renewable Energy Consumption	0
Total Primary Energy Consumption	0

dtype: int64

Figure 2.1: Monthly Energy Dataset - Missing values Table

2.1.2 Annualy Energy Dataset

Here,also we haven't encountered any missing values

```
energy_annual.isnull().sum()
```

	0
Annual Total	0
Total Fossil Fuels Production	0
Nuclear Electric Power Production	0
Total Renewable Energy Production	0
Total Primary Energy Production	0
Primary Energy Imports	0
Primary Energy Exports	0
Primary Energy Net Imports	0
Primary Energy Stock Change and Other	0
Total Fossil Fuels Consumption	0
Nuclear Electric Power Consumption	0
Total Renewable Energy Consumption	0
Total Primary Energy Consumption	0

dtype: int64

Figure 2.2: Annualy Energy Dataset - Missing values Table

2.2 Imputation

As no missing values have been encountered, no imputation is required.

2.3 Changing Data Types

2.3.1 Monthly Energy Dataset

All the columns are present in dame data types so no need to change that.

```
energy.dtypes
```

	0
Month	datetime64[ns]
Total Fossil Fuels Production	float64
Nuclear Electric Power Production	float64
Total Renewable Energy Production	float64
Total Primary Energy Production	float64
Primary Energy Imports	float64
Primary Energy Exports	float64
Primary Energy Net Imports	float64
Primary Energy Stock Change and Other	float64
Total Fossil Fuels Consumption	float64
Nuclear Electric Power Consumption	float64
Total Renewable Energy Consumption	float64
Total Primary Energy Consumption	float64

dtype: object

Figure 2.3: Monthly Energy Dataset - Data types

2.3.2 Annualy Energy Dataset

All the columns are present in dame data types so no need to change that.

```
# Convert all object columns to float64
energy_annual = energy_annual.astype({col: 'float64' for col in energy_annual.select_dtypes(include='object').columns})
energy_annual.dtypes
```

	0
Annual Total	float64
Total Fossil Fuels Production	float64
Nuclear Electric Power Production	float64
Total Renewable Energy Production	float64
Total Primary Energy Production	float64
Primary Energy Imports	float64
Primary Energy Exports	float64
Primary Energy Net Imports	float64
Primary Energy Stock Change and Other	float64
Total Fossil Fuels Consumption	float64
Nuclear Electric Power Consumption	float64
Total Renewable Energy Consumption	float64
Total Primary Energy Consumption	float64

dtype: object

Figure 2.4: Annually Energy Dataset - Data types

Chapter 3. Visualization

Data visualization is a critical step in the data analysis process, as it enables us to understand the data's underlying patterns, trends, and relationships effectively. Through graphical representations, we can simplify complex datasets, making them easier to interpret and communicate. In this section, we explore various visualization techniques to examine and summarize the energy consumption and production data. These visualizations not only provide insights into seasonal variations, correlations, and long-term trends but also serve as a foundation for identifying key factors that influence energy metrics.

The visualizations are organized into sections to provide a structured overview:

3.1 Univariate analysis

Univariate analysis is a statistical technique used to analyze and summarize the characteristics of a single variable at a time. Its primary goal is to describe the variable's distribution, central tendency (mean, median, mode), dispersion (range, variance, standard deviation), and shape (skewness, kurtosis). This method is often visualized through plots like histograms, box plots, and bar charts, or summarized using numerical measures. By isolating and studying each variable independently, univariate analysis provides valuable insights into its behavior and patterns, serving as a foundational step in data exploration and hypothesis generation.

3.1.1 Production

Histogram

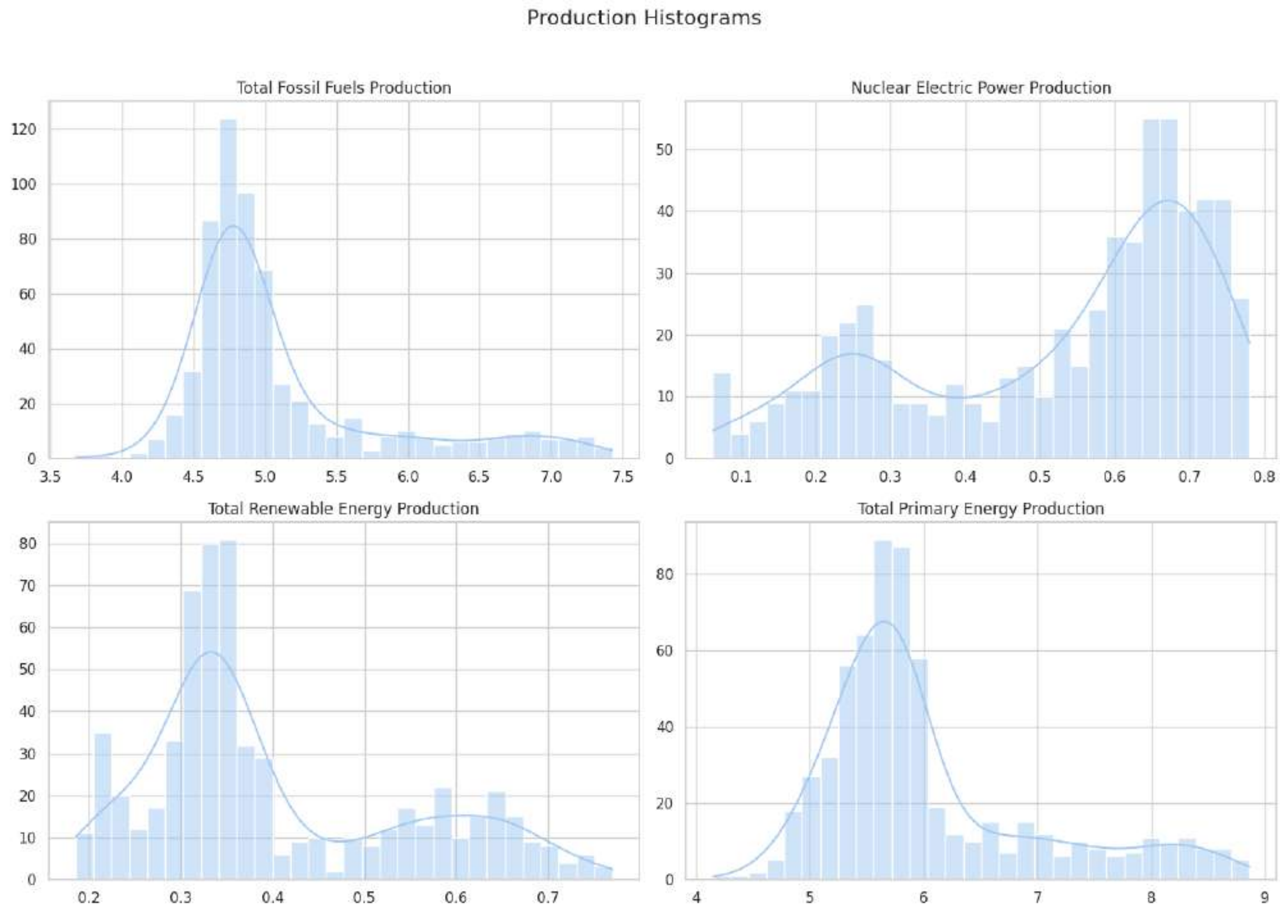


Figure 3.1: Production Histograms

- **Total Fossil Fuels Production**: Most production values lie between 4.5 and 5.5, showing consistency, with a few outliers producing significantly more, highlighting global production disparities.
- **Nuclear Electric Power Production**: The bimodal distribution indicates two groups: one with limited nuclear capability and another with advanced nuclear infrastructure, reflecting uneven reliance on nuclear energy.
- **Total Renewable Energy Production**: Concentrated around 0.3 to 0.4, renewable energy production shows a few leaders contributing large-scale outputs, underscoring adoption gaps.
- **Total Primary Energy Production**: Most production lies between 5 and 6, with a few entities producing much more, reflecting energy production disparities driven by resource availability and economic size.

Boxplot

A boxplot is a statistical visualization tool used to summarize the distribution of a dataset. It displays the dataset's minimum, first quartile (Q1), median (Q2), third quartile (Q3), and maximum values. The box represents the interquartile range (IQR), which contains the middle 50 percent of the data. The whiskers extend to the smallest and largest values within 1.5 times the IQR, while data points beyond the whiskers are considered outliers. This makes boxplots useful for identifying variability, central tendencies, and potential anomalies in the data.

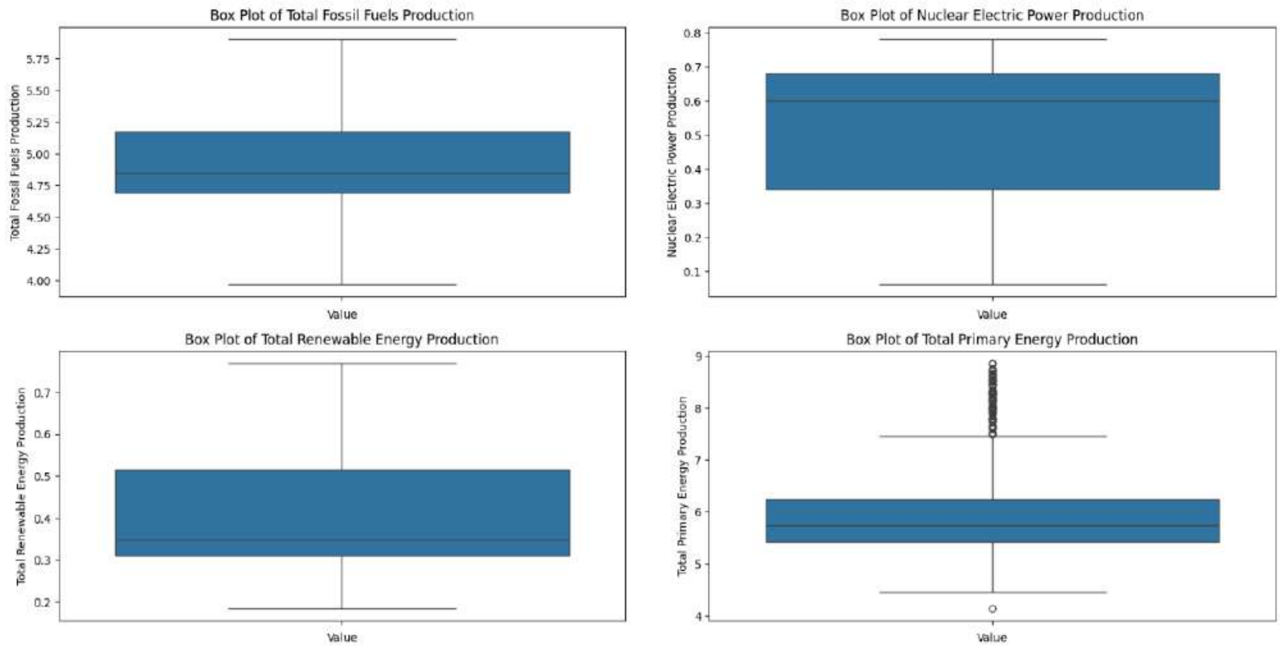


Figure 3.2: Production Boxplot

The production boxplots illustrate the variability in energy production metrics. Fossil fuels and nuclear electric power production show narrow boxes, indicating consistent production levels with minimal variability. Renewable energy production has a wider box, suggesting more variability in its distribution. For primary energy production, several outliers are present, highlighting extreme values that deviate from the majority of the data. These insights reveal stability in traditional energy sources and variability in renewable energy, with potential anomalies in primary energy production.

Density Plot

A density plot is a data visualization tool that shows the distribution of a dataset by estimating its probability density function. It is a smoothed version of a histogram, where the area under the curve represents the entire dataset and integrates to 1. Unlike histograms, density plots are continuous and provide a clearer view of the data's shape, including peaks, valleys, and spread. They are particularly useful for comparing multiple distributions or identifying skewness, multimodality, and central tendencies in the data.

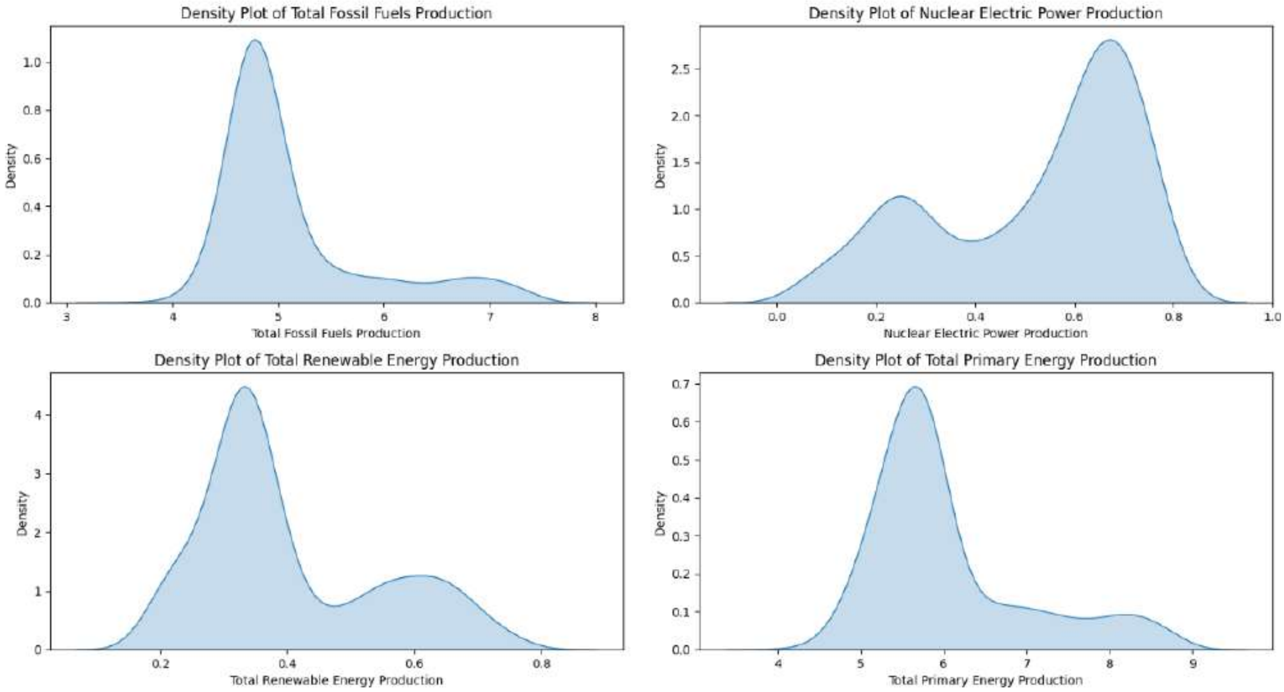


Figure 3.3: Production Density Plot

Violin Plot

A violin plot is a statistical visualization that combines features of a boxplot and a density plot to show the distribution, central tendencies, and variability of data. It consists of a mirrored density plot on either side of a central axis, providing a full view of the data's distribution shape (peaks, valleys, and spread). Inside the violin, a boxplot or summary statistics (median, quartiles, etc.) is often displayed to highlight key metrics. Violin plots are particularly useful for comparing distributions across multiple categories while retaining details about data density and variability.

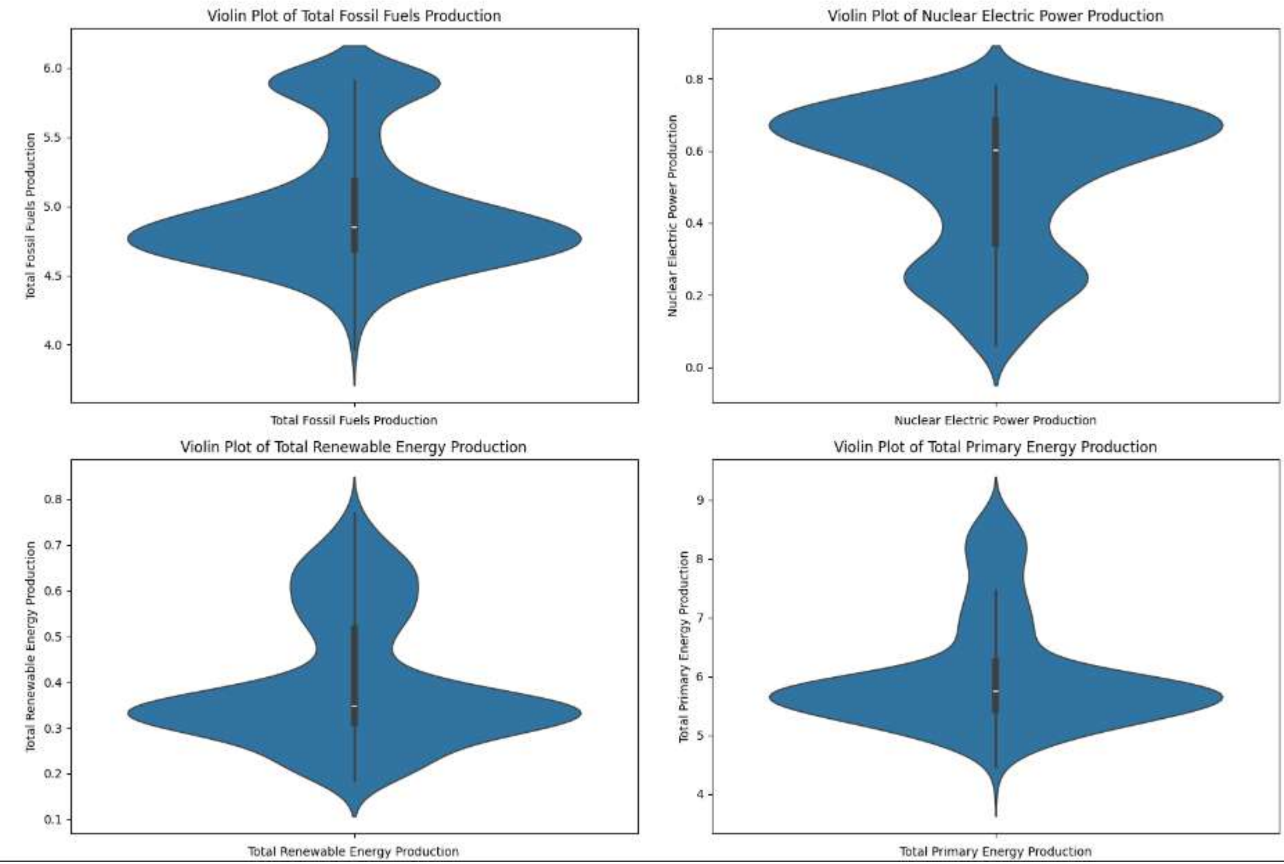


Figure 3.4: Production Violin Plot

Dot Plot for Total Production

A dot plot is a type of visualization used to display individual data points along an axis. Each point represents a single observation, allowing patterns, trends, or clusters to be easily identified. It is particularly useful for highlighting the distribution and changes in data over time or across categories.

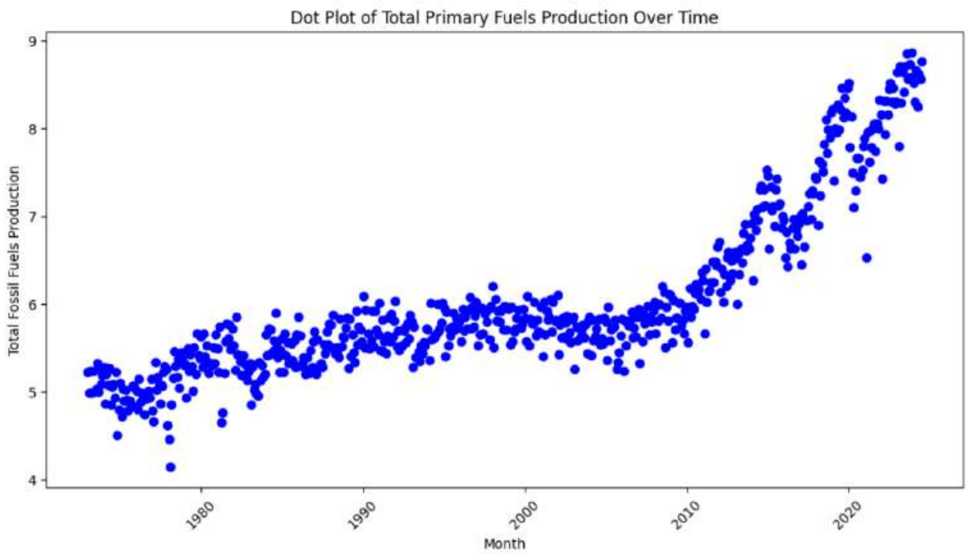


Figure 3.5: Dot Plot for Total Production

The dot plot of total fossil fuels production over time reveals a clear upward trend, indicating a steady increase in production. The plot also shows periods of fluctuations and accelerations, particularly after 2000, suggesting changes in energy policies, demand, or technology advancements contributing to the growth.

QQ Plot

A Quantile-Quantile (QQ) plot is a graphical tool to assess whether a dataset follows a specified theoretical distribution, often the normal distribution. Data quantiles are plotted against theoretical quantiles, and if the data aligns along the diagonal red line, it suggests the data closely follows the chosen distribution.

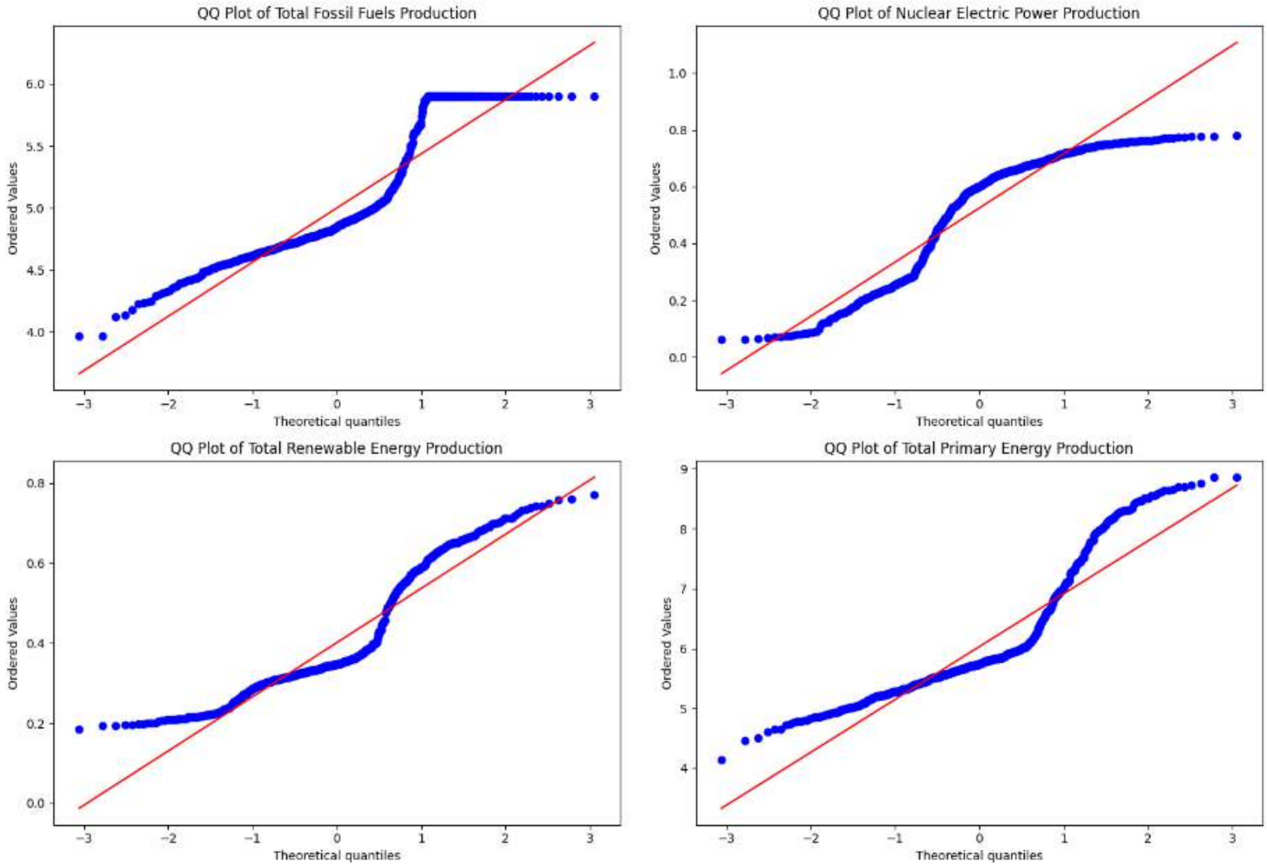


Figure 3.6: Production QQ Plot

The QQ plots for different energy production types reveal deviations from normality. While some datasets, like nuclear electric power production, are relatively close to the line, others, such as fossil fuels and renewable energy production, exhibit significant departures, particularly in their tails. This suggests the presence of skewness or heavy-tailed distributions in certain datasets.

3.1.2 Consumption

Histogram

The histograms for energy consumption types reveal distinct distribution patterns. Total fossil fuels and primary energy consumption appear slightly right-skewed, indicating higher concentrations of lower values. Renewable energy and nuclear power consumption show multimodal behavior, hinting at varying trends or shifts in consumption over time. The KDE curves highlight the underlying density trends, providing a clearer view of the data's structure.

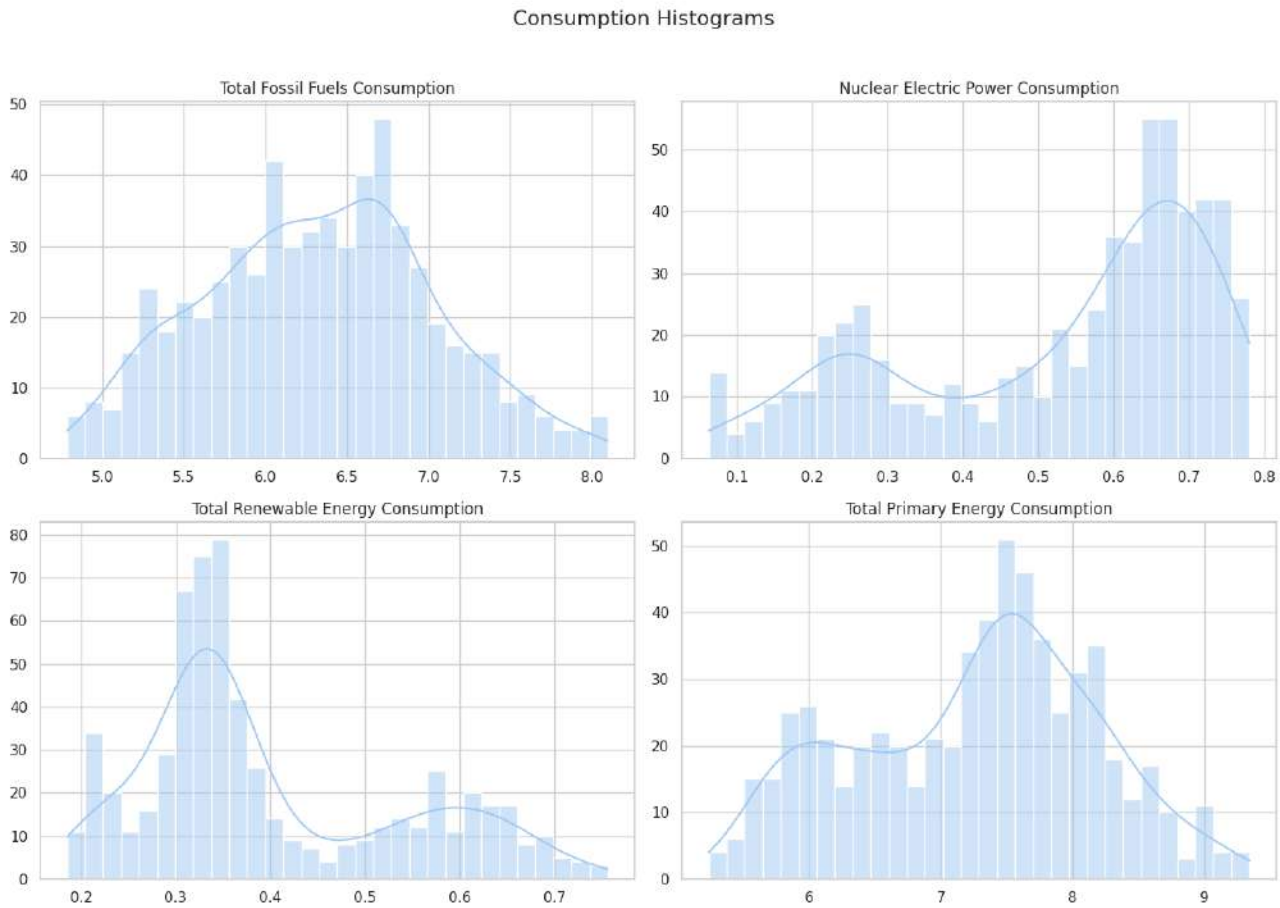


Figure 3.7: Consumption Histogram

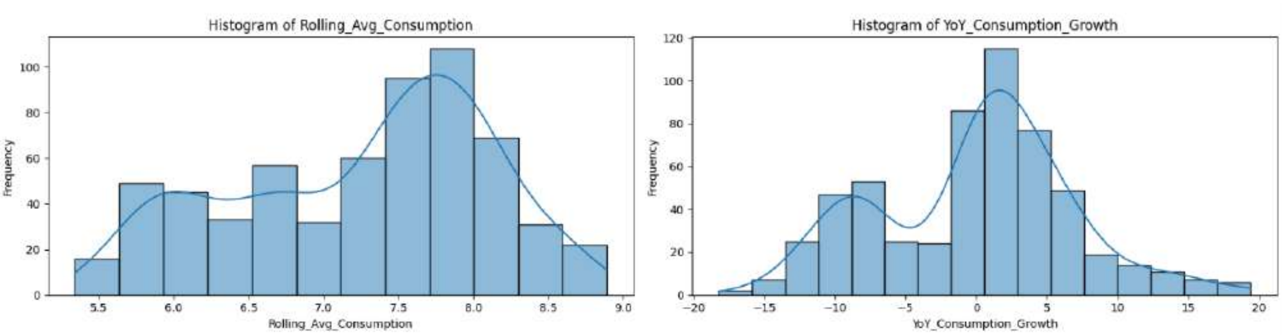


Figure 3.8: Total Consumption Histogram

This analysis examines the distribution of energy consumption types across countries, focusing on fossil fuels, nuclear power, renewable energy, and primary energy. Each histogram provides insights into global energy trends and national dependencies on different energy sources.

- **Total Fossil Fuels Consumption:** The distribution is slightly left-skewed, with a peak around 6.5–7, indicating that most countries have moderate fossil fuel usage, with fewer relying heavily on it. This suggests a common but gradually declining dependency on fossil fuels as some countries transition towards alternative energy.

- **Nuclear Electric Power Consumption:** The bimodal shape, with peaks at 0.2 and 0.6, shows two distinct groups of countries: one group with significant reliance on nuclear power and another using it sparingly or not at all. This reflects diverse national policies and acceptance levels of nuclear energy across the globe.
- **Total Renewable Energy Consumption:** The right-skewed distribution, with a peak at 0.3, shows that renewable energy adoption is still limited. A few countries lead in renewable consumption, while most have low levels, indicating that the global transition to renewable sources is underway but in early stages for many nations.
- **Total Primary Energy Consumption:** The left-skewed distribution, with a peak around 7, implies that most countries have moderate overall energy needs, with a few exhibiting very high consumption. This is likely due to differences in industrialization and population size.

Fossil fuels remain the dominant source of energy globally, though some countries are gradually reducing their reliance on them. Nuclear power has variable adoption, with certain countries heavily relying on it while others avoid it altogether. Renewable energy adoption is on the rise but still limited in many places, indicating that the transition to renewables is ongoing. Primary energy consumption varies significantly based on economic and demographic factors, with some countries having notably higher energy needs.

Boxplot

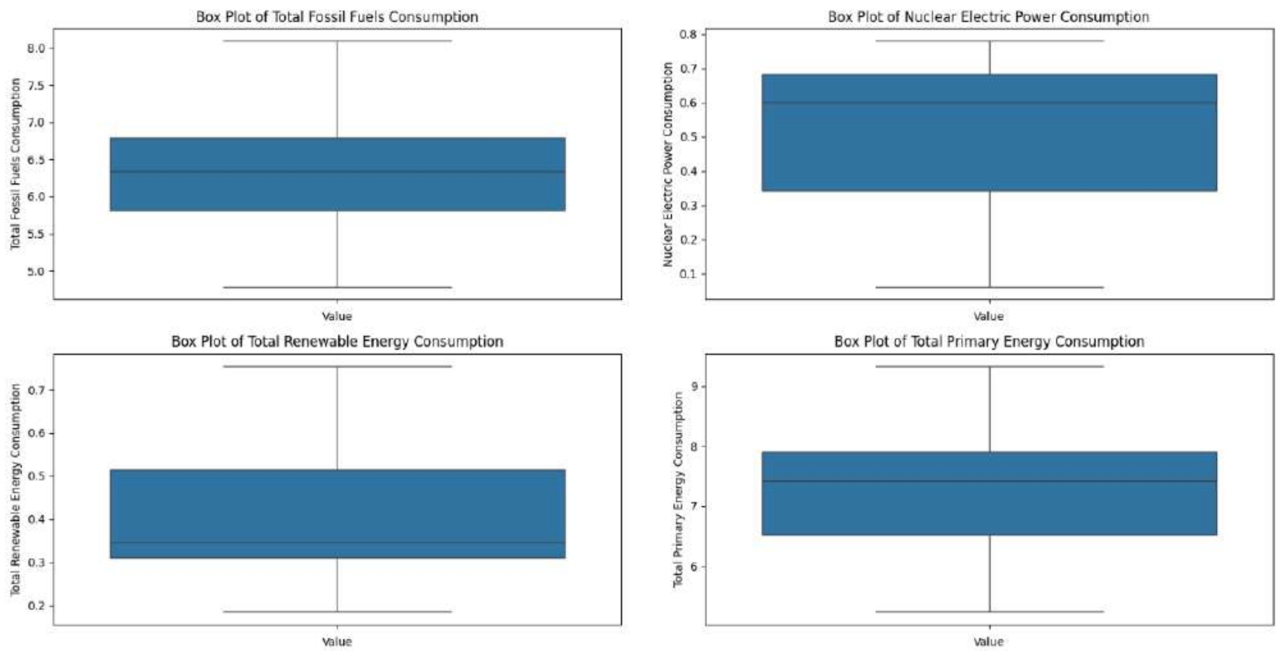


Figure 3.9: Consumption Boxplot

The box plots illustrate the distribution of various energy consumption types (fossil fuels, nuclear, renewable, and primary energy). Key observations include the relatively tight interquartile ranges across all plots, indicating consistency in energy usage trends. Notably, the nuclear electric power consumption has a narrow range compared to fossil fuels and renewables, suggesting higher stability or limited variation in its usage. There are no significant outliers visible, pointing to overall uniform consumption patterns across the analyzed data.

Density Plot

The density plots display the distribution of energy consumption across different sources. Fossil fuels and primary energy consumption exhibit unimodal distributions, suggesting a dominant range of usage levels. In contrast, renewable energy and nuclear electric power consumption display bimodal patterns, indicating the presence of two distinct consumption trends or groups. These variations highlight differences in energy reliance and suggest areas for deeper exploration regarding energy diversification and consumption patterns.

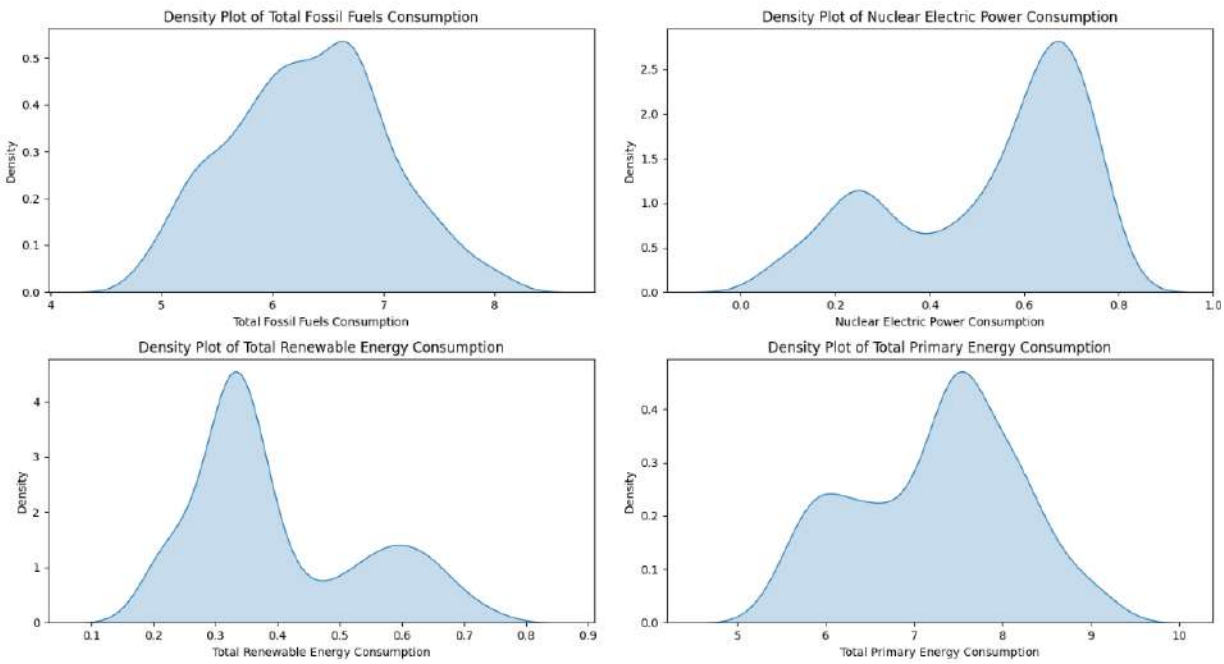


Figure 3.10: Consumption Density Plot

Violin Plot

The violin plots provide a detailed view of the distribution and density of energy consumption. Fossil fuels and primary energy exhibit symmetrical distributions with concentrated densities around the median, reflecting consistent consumption patterns. Renewable energy shows a more varied distribution with distinct density peaks, indicating heterogeneous adoption levels. Nuclear energy displays a relatively narrow density spread, suggesting uniformity in its consumption. These insights highlight variations in dependency and adoption trends across energy sources.

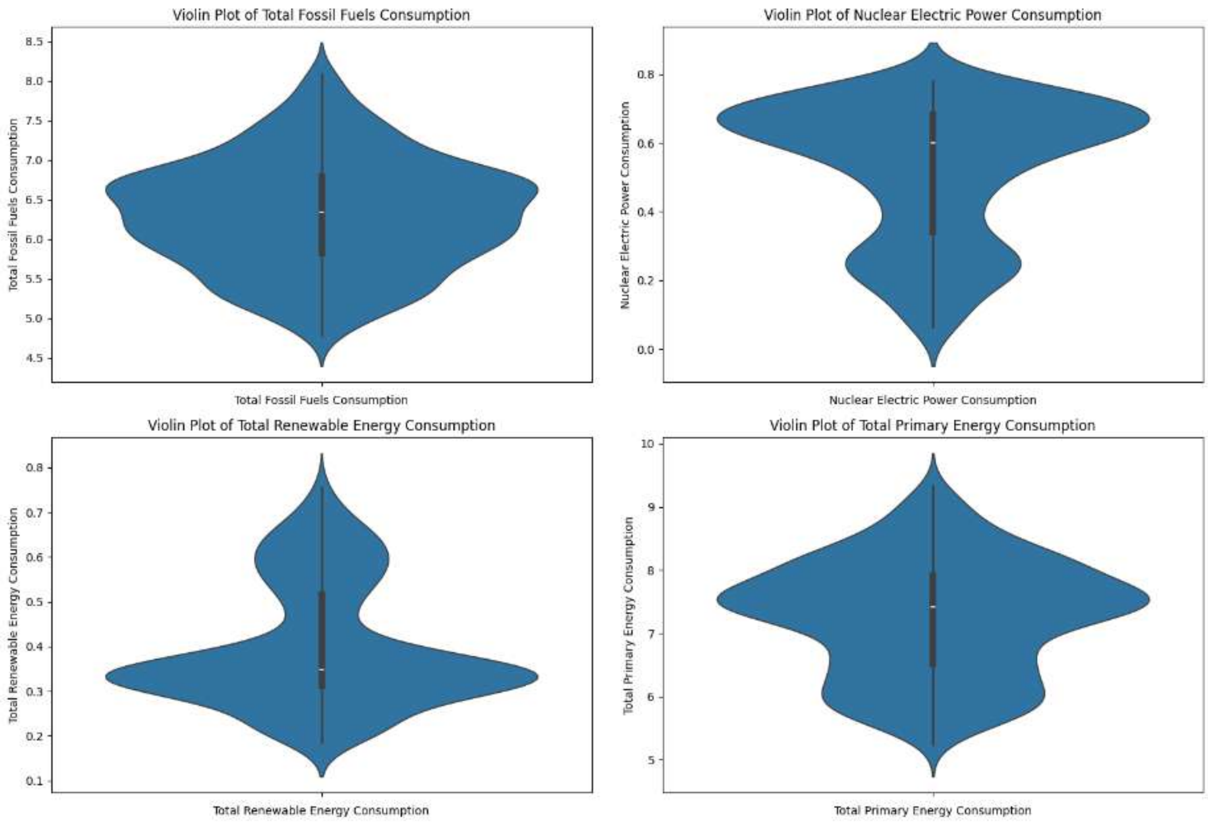


Figure 3.11: Consumption Violin Plot

Dot Plot for Total Consumption

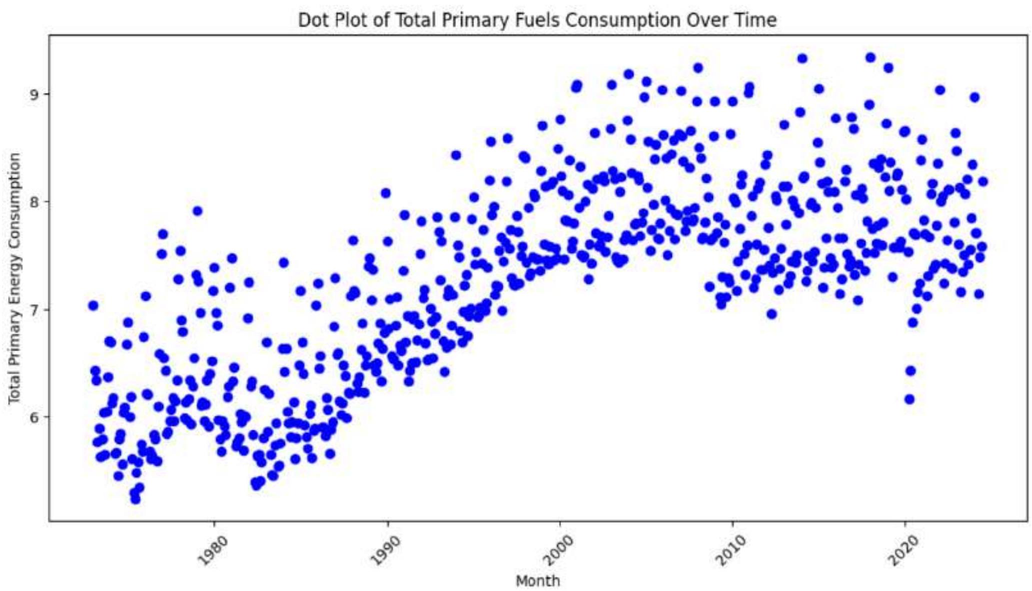


Figure 3.12: Dot Plot for Total Consumption

The dot plot shows total primary energy consumption increasing consistently from the 1980s to around 2010, reflecting long-term growth in energy demand. After 2010, the trend appears more variable, with occasional declines, suggesting shifts in consumption patterns or potential impacts of efficiency measures, economic factors, or energy transitions. The overall trend highlights a historical rise in energy usage with recent fluctuations.

QQ Plot

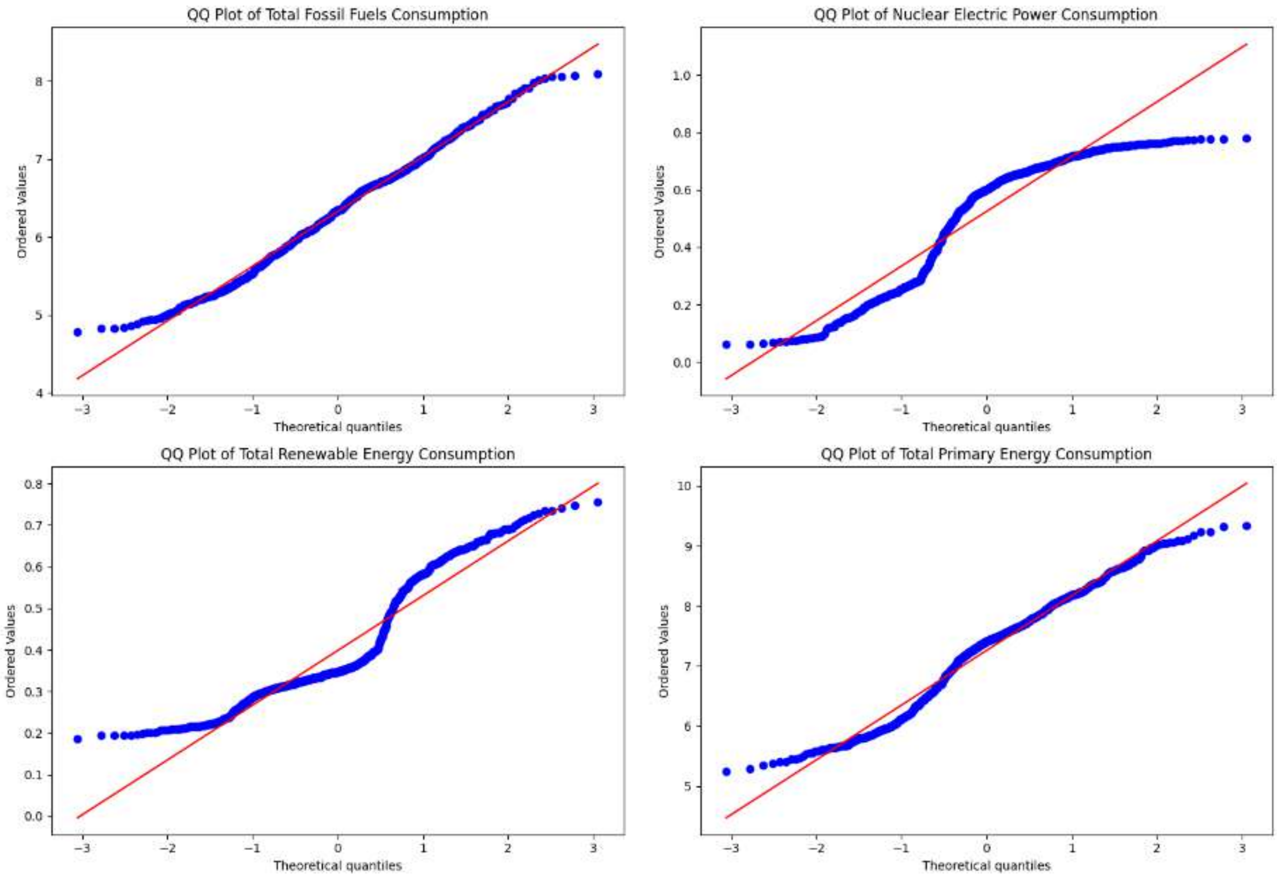


Figure 3.13: QQ Plot for Consumption

The QQ plots evaluate the normality of different energy consumption types. Total fossil fuel and primary energy consumption align closely with the theoretical quantiles, indicating near-normal distributions. In contrast, nuclear and renewable energy consumption deviate, showing nonlinear patterns that suggest non-normality. These variations highlight differences in the statistical behavior of energy types.

3.1.3 Imports, Exports, Net Imports, and Stock Change

Histogram

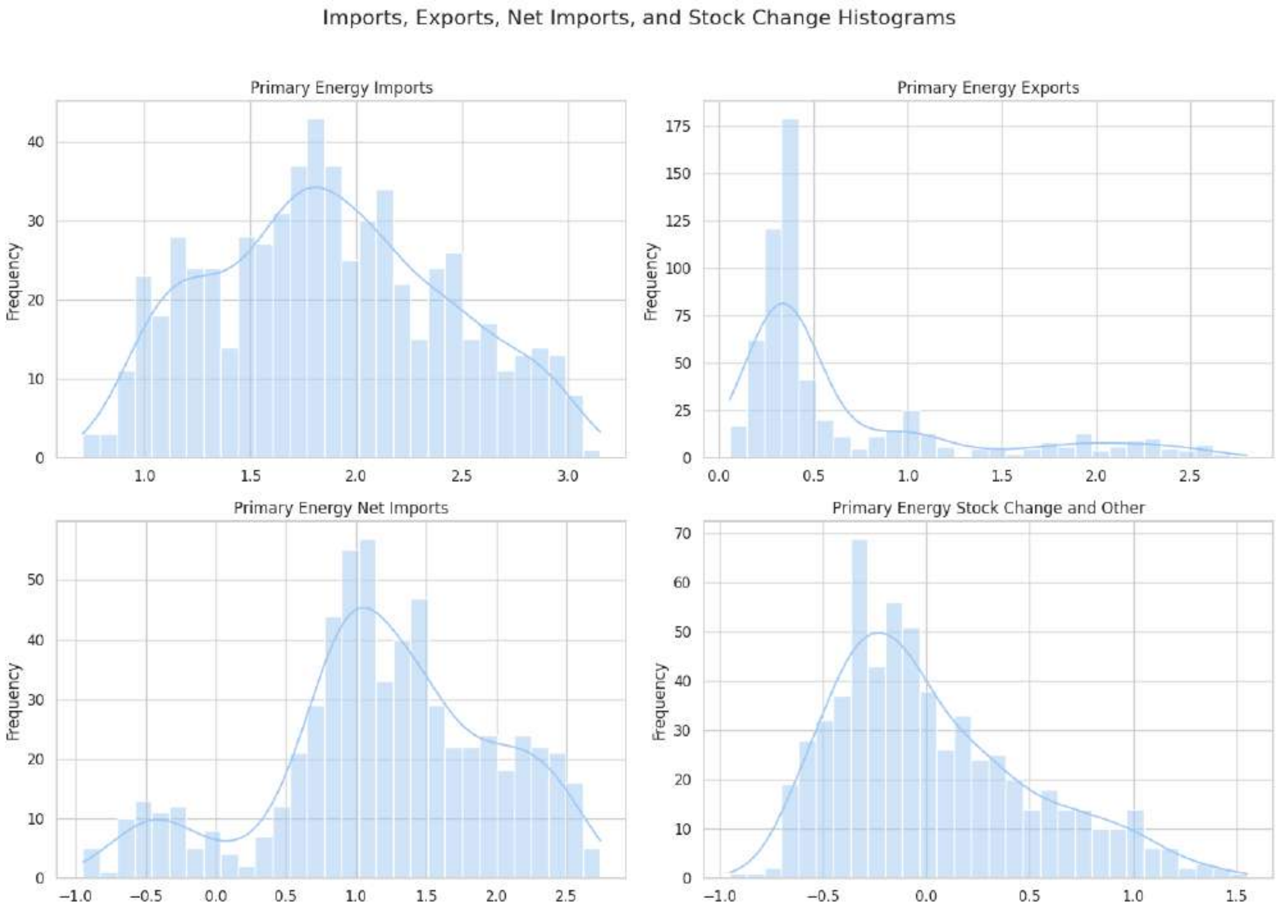


Figure 3.14: Imports, Exports, Net Imports, and Stock Change Histograms

- **Primary Energy Imports:** The distribution is slightly right-skewed, with a peak around 1.5–2.0, indicating that many countries have moderate energy import levels, while fewer rely heavily on imports. This may reflect efforts by some nations to maintain energy independence or rely on domestic sources.

- **Primary Energy Exports:** The distribution is highly right-skewed with a peak near 0, indicating that most countries have low or negligible energy exports. Only a few countries are significant energy exporters, which highlights the global concentration of energy production in certain regions.

- **Primary Energy Net Imports:** The distribution is right-skewed with a peak around 1.0–1.5, suggesting that most countries are net energy importers to some extent. However, some countries have higher levels of net imports, potentially due to limited domestic resources or high energy demand.

- **Primary Energy Stock Change and Other:** This histogram has a left-skewed distribution with a peak near -0.5, implying that many countries have a net decrease in stock change or limited storage, likely managing energy through consistent imports or other balancing mechanisms.

Boxplot

The box plots reveal the distribution of energy imports, exports, net imports, and stock changes. Imports and exports exhibit relatively symmetric distributions with a few outliers, particularly in exports. Net imports show notable variability with several outliers below the lower range, indicating occasional negative values. Stock changes display a tighter distribution, suggesting more stable trends. These insights highlight the dynamics and variability in energy trade and inventory.

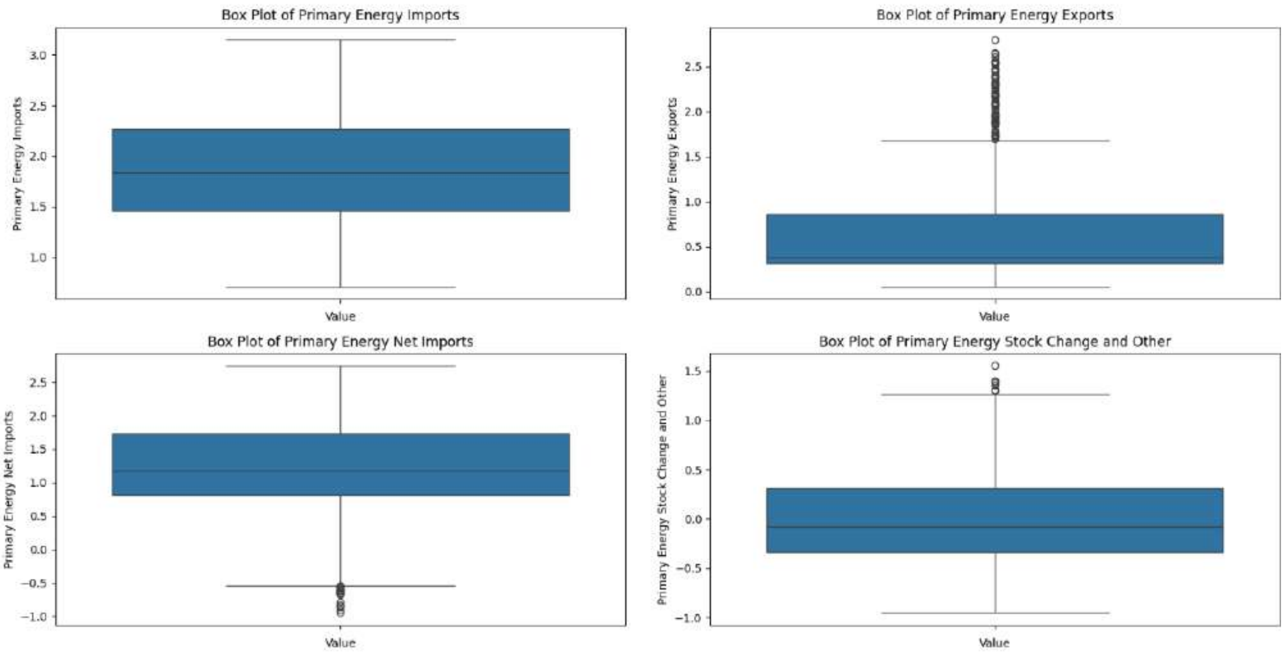


Figure 3.15: Energy Imports,Exports,Net Imports, Stock Change

Density Plot

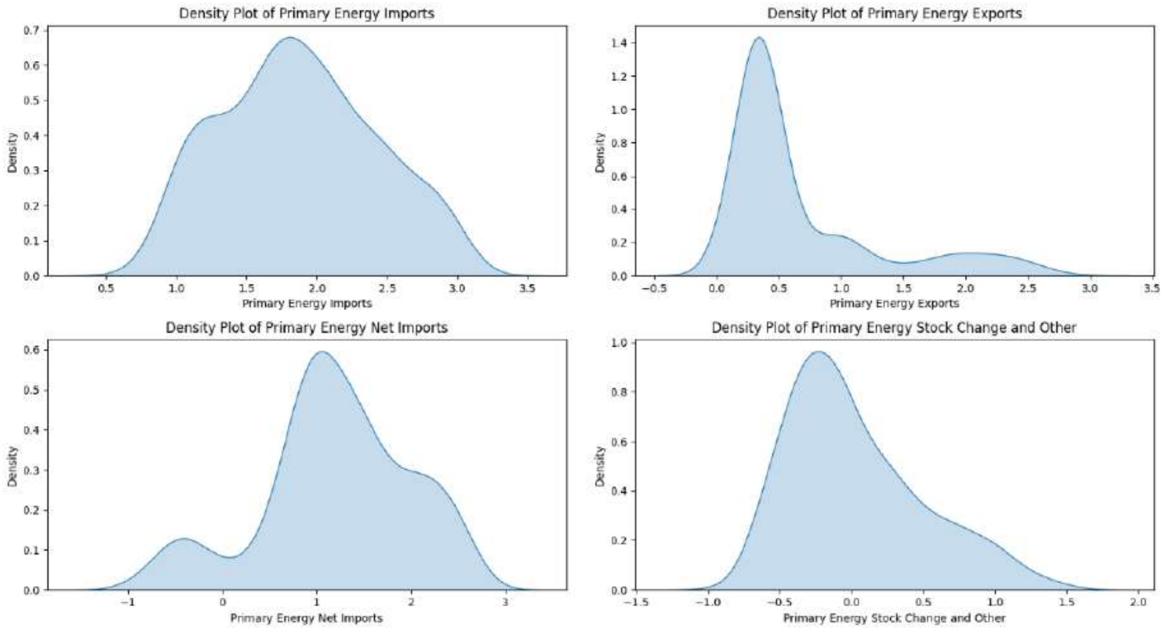


Figure 3.16: Denisty Plot for Imports,Exports,Net Imports and Stock Change

The density plots illustrate the distribution of primary energy imports, exports, net imports, and stock changes. Imports and exports show relatively unimodal distributions with imports peaking around 2 and exports around 0. Net imports have a wider distribution with a peak near 1, while stock changes display a less symmetrical shape, suggesting variability. These visualizations reveal distinct patterns in energy flow components.

Violin Plot

The violin plots display the distribution of primary energy imports, exports, net imports, and stock changes, along with their median and interquartile ranges. Imports and exports have a symmetric distribution, with imports showing a wider spread. Net imports exhibit a bimodal pattern, while stock changes have a broader spread and variability, indicating diverse energy storage and adjustments.

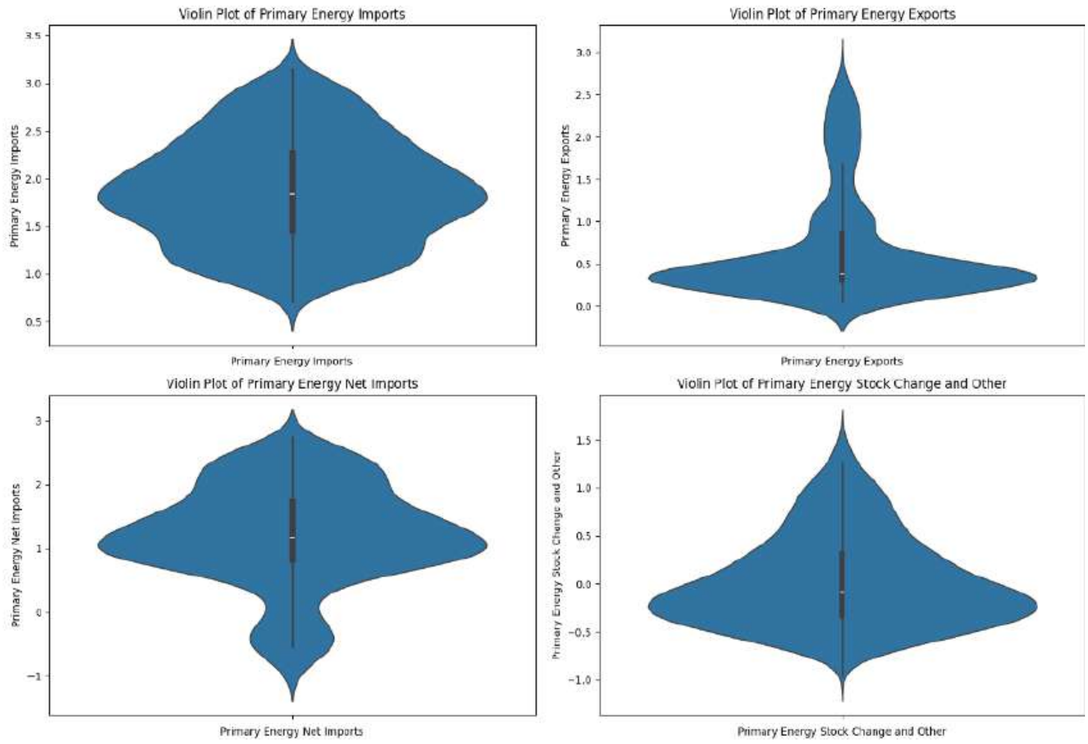


Figure 3.17: Violin Plot for Imports,Exports,Net Imports and Stock Change

Dot Plot for Imports, Exports, Net Imports, and Stock Change

The dot plot shows the trend of primary energy net imports over time. Net imports steadily increased from the 1980s to around 2005, peaking before undergoing a sharp decline post-2010. This pattern suggests significant shifts in energy trade dynamics, possibly due to changing domestic production, consumption patterns, or policy interventions.

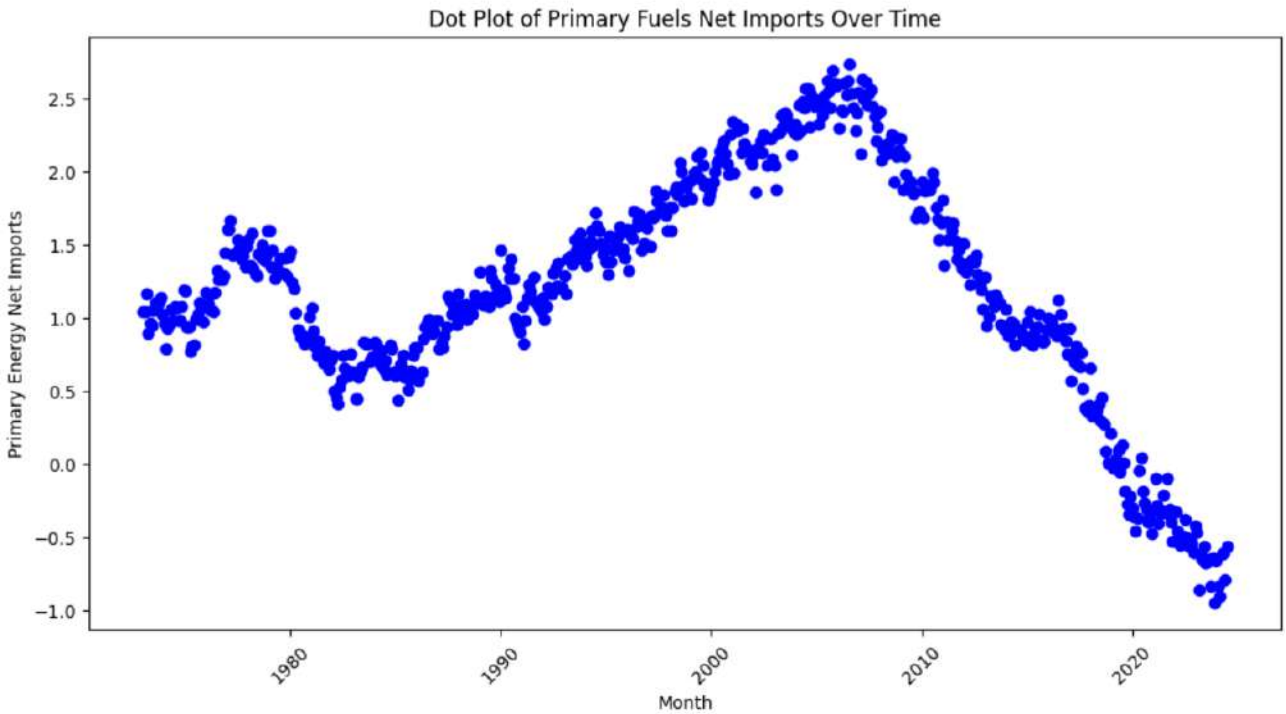


Figure 3.18: Dot Plot for Imports, Exports, Net Imports, and Stock Change

QQ Plot

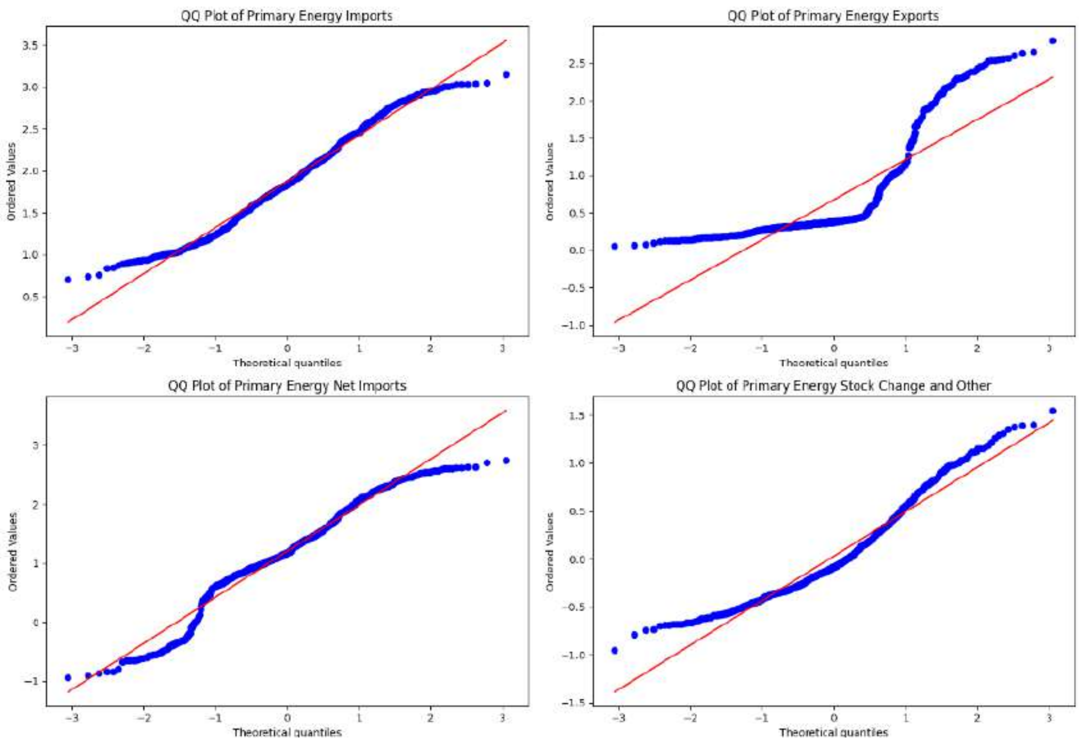


Figure 3.19: QQ Plot of Imports, Exports, Net Imports, and Stock Change

The QQ plots depict the distribution of imports, exports, net imports, and stock changes for primary energy sources compared to a theoretical normal distribution. Deviations from the red line indicate non-normality, with exports and stock changes showing noticeable departures, suggesting potential skewness or outliers in the data. In contrast, imports and net imports align more closely with the normal distribution.

3.1.4 Energy Deficit

Histogram

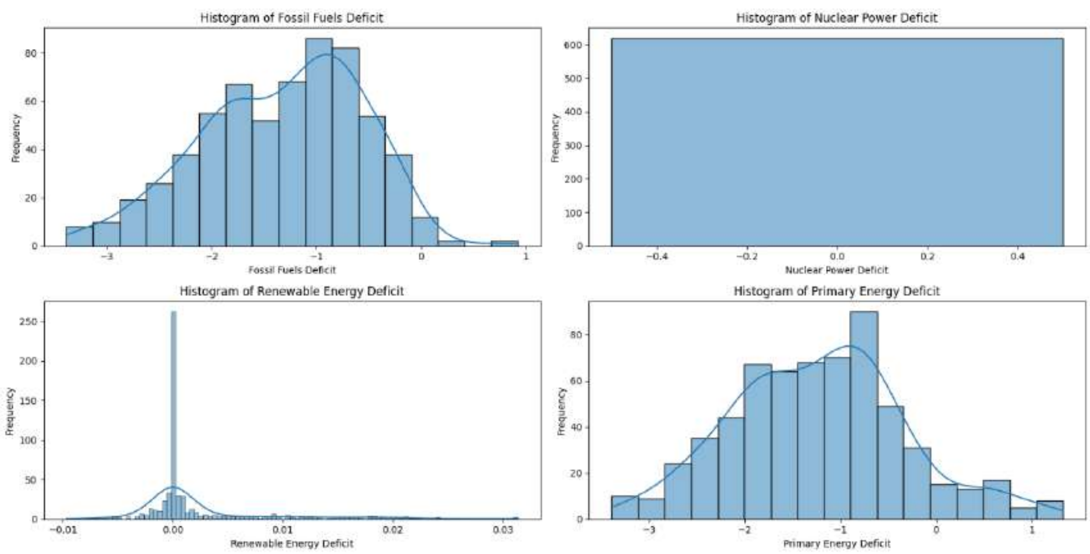


Figure 3.20: Energy Deficit Histograms

- **Fossil Fuels Deficit:**
  - Approximately normal distribution, centered around zero.
  - Slight right skew, indicating a few larger deficits.
  - Most values fall between -1 and 1, indicating small deficits.
- **Nuclear Power Deficit:**
  - Flat histogram, suggesting many zero or near-zero values.
  - Data concentrated at a single point, showing little variation.
  - Indicates rare deficits or possible data recording issues.
- **Renewable Energy Deficit:**
  - Sharp peak at zero, indicating minimal or no deficit for many data points.
  - Highly right-skewed, with occasional large positive deficits.
  - Suggests typically low deficits but occasional substantial shortages.
- **Primary Energy Deficit:**

– Approximately normal distribution, centered slightly below zero.
– Slight right skew, with a few significant deficits.
– Wide spread of values, indicating higher variability in deficits.

Box Plot

The box plots illustrate the distribution of deficits across different energy types: fossil fuels, nuclear power, renewable energy, and primary energy. Fossil fuels and primary energy deficits appear to have a compact distribution without significant outliers, whereas renewable energy shows several outliers, indicating irregularities or extreme values in its deficit data. The nuclear power deficit shows very little variation, suggesting minimal fluctuation.

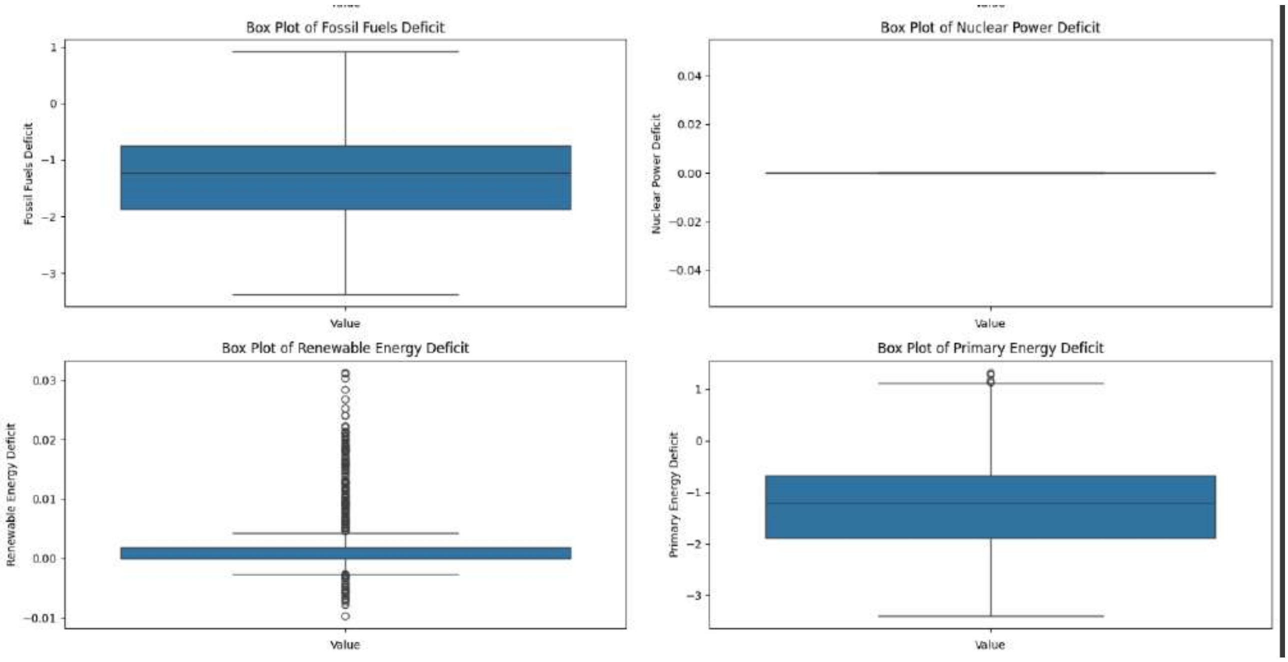


Figure 3.21: Box Plot of Energy Deficit

Density Plot

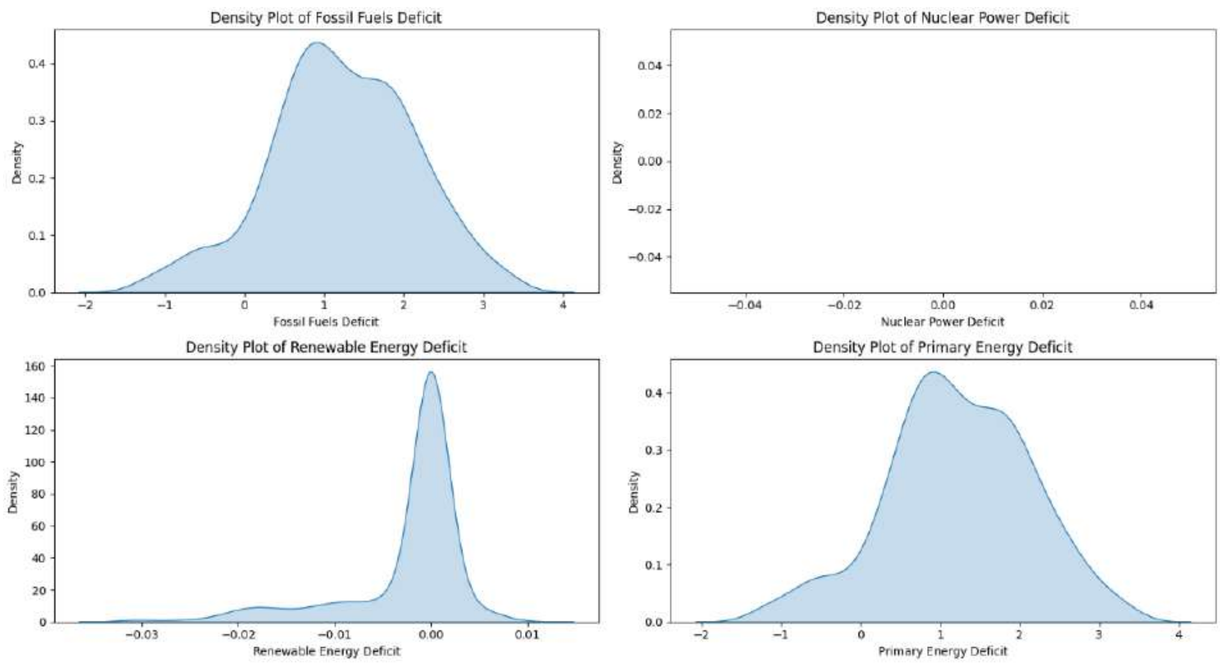


Figure 3.22: Deficit Density Plot

The density plots provide a smoothed view of the distribution of energy deficits. Fossil fuels and primary energy deficits show unimodal distributions, with fossil fuels being slightly skewed. The renewable energy deficit density exhibits a longer tail, reflecting variability and potential extreme values, while the nuclear power deficit shows almost no significant spread, confirming its low variability.

Violin Plot

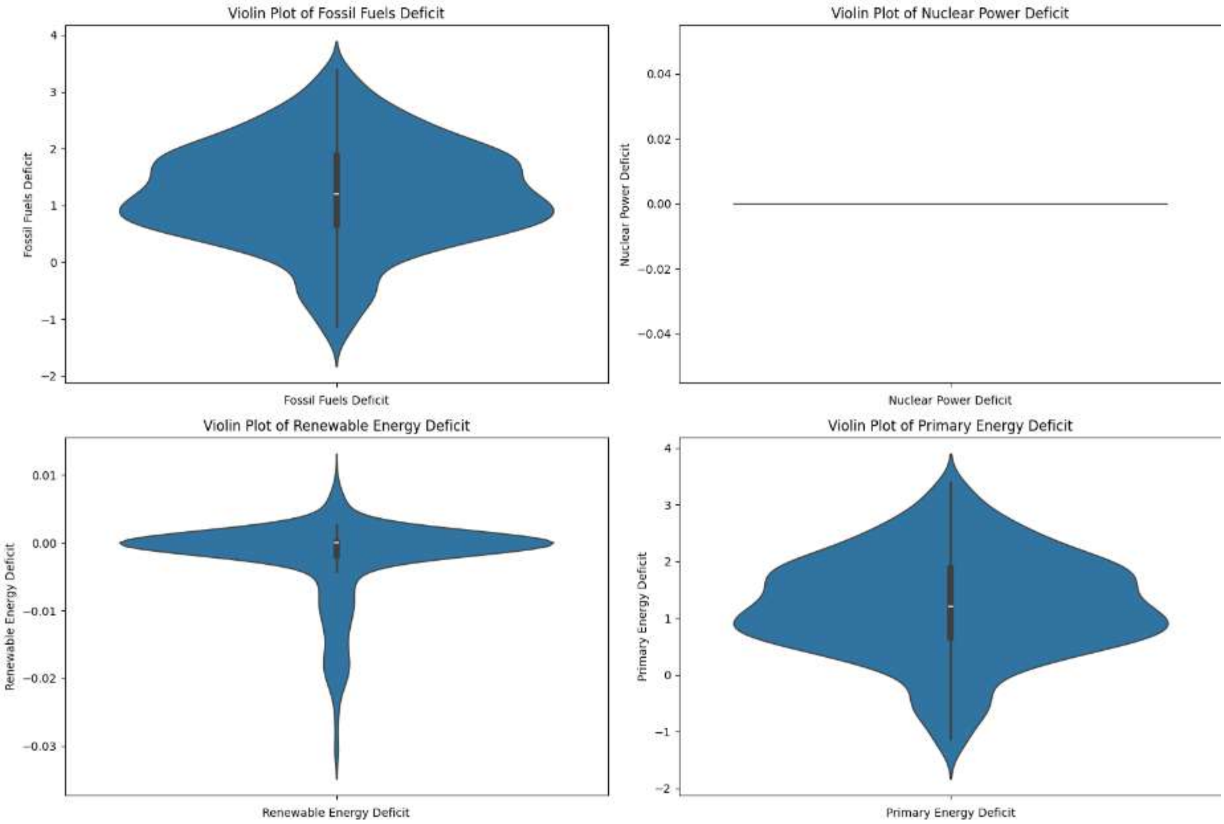


Figure 3.23: Deficit Violin Plot

The violin plots provide insights into the distribution and density of deficits for various energy types. Fossil fuels and primary energy show symmetrical distributions with a concentration around the median, whereas renewable energy exhibits a wider spread and a potential presence of extreme values. The nuclear power deficit shows minimal variation, aligning with prior analyses.

Dot Plot

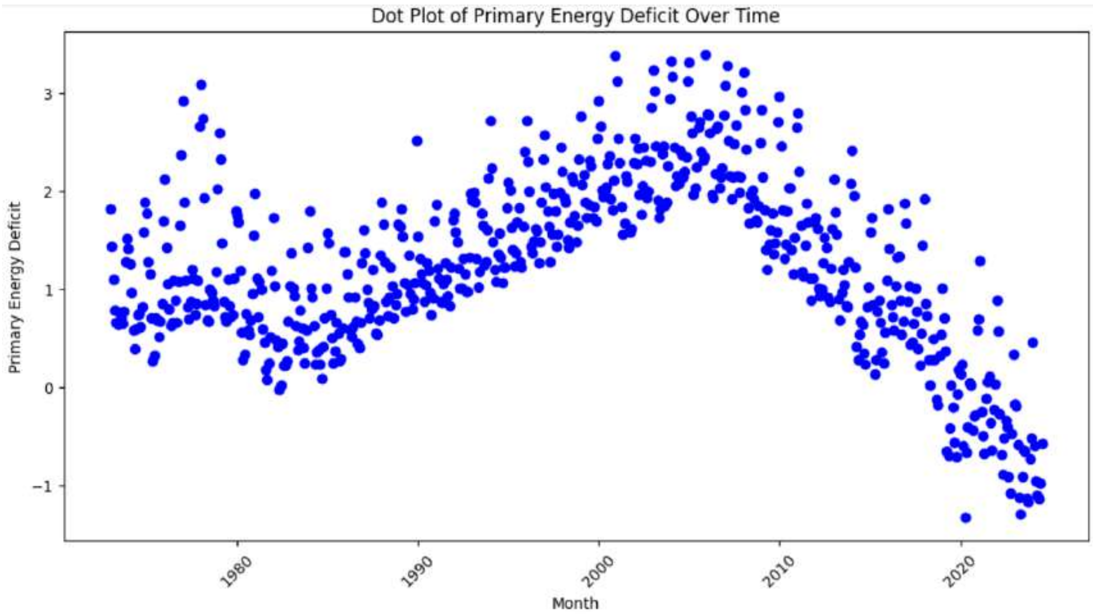


Figure 3.24: Dot Plot for total enegry deficit

The dot plot for total energy deficit highlights the fluctuation of deficit values over time. The scattered points indicate a clear trend or pattern, with variability increasing in certain regions. This could imply external factors affecting the energy deficit during specific periods.

QQ Plot

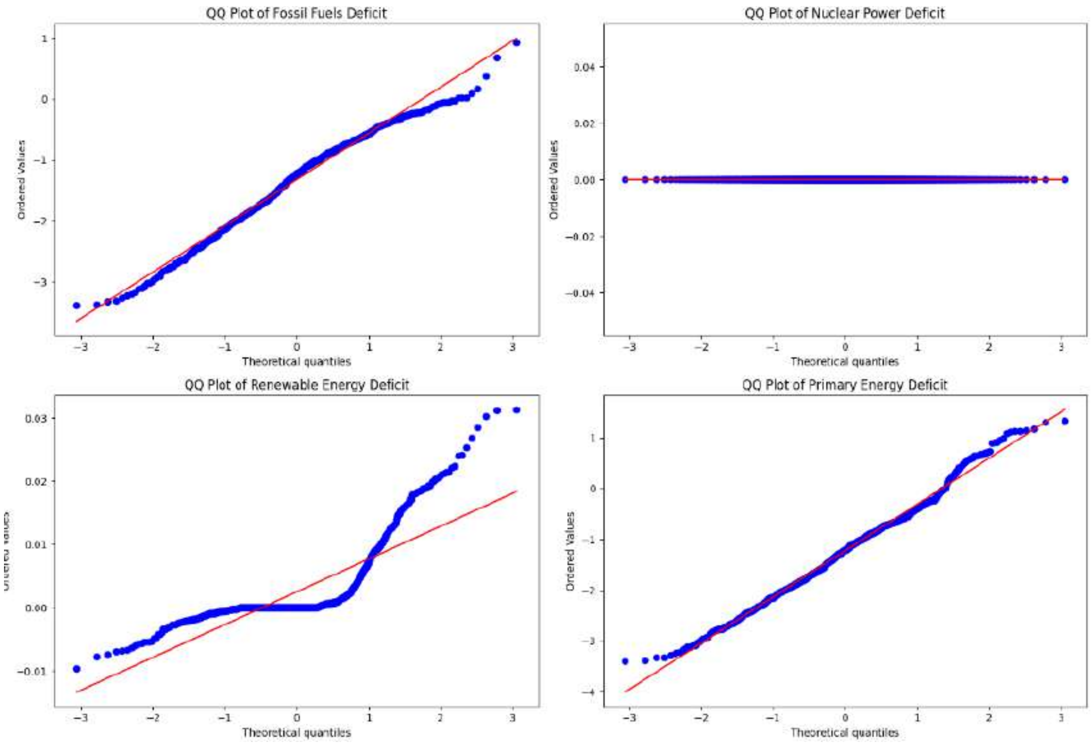


Figure 3.25: QQ Plot for Deficit Energy

The QQ plots for deficit energy types assess their adherence to a normal distribution. Fossil fuels and primary energy deficits align closely with the theoretical normal line, except for slight deviations at the extremes. Renewable energy deficits show significant deviations, indicating potential skewness or heavy tails, while nuclear power deficits maintain near-normal behavior.

3.1.5 Trend Analysis of Monthly Total Renewable Energy Production

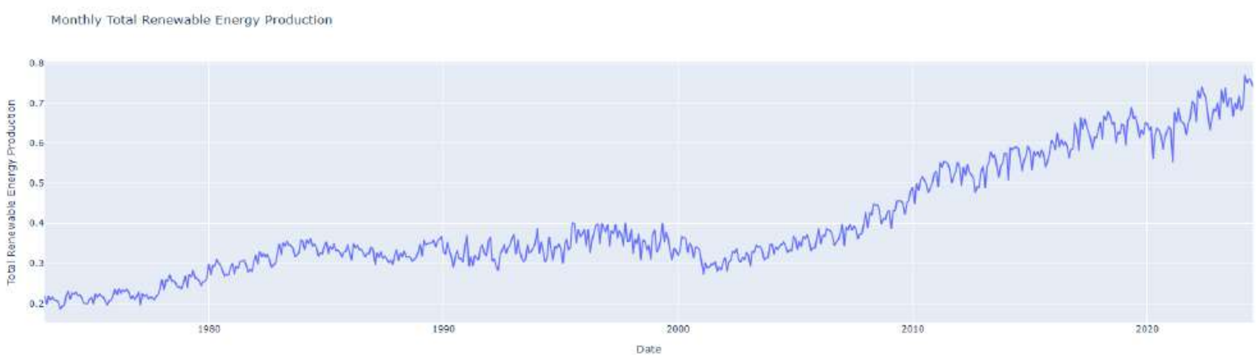


Figure 3.26: Trend Analysis of Monthly Total Renewable Energy Production

1. **Low Initial Production Levels**:

- In the early years, renewable energy production started at relatively low levels.
- There were gradual increases with some fluctuations, possibly due to seasonal effects or varying energy demands.

2. **Shift in the Early 2000s**:
   - Around the early 2000s, there is a visible shift in the trend.
   - Production began to rise more consistently, indicating a change in growth dynamics.
3. **Steep Increase Post-2010**:
   - After 2010, the trend steepens, showing a substantial increase in renewable energy adoption.
   - This rise is likely driven by technological advancements and policy incentives for sustainable energy.
4. **Growing Reliance on Renewable Energy**:
   - The increasing trend highlights the expanding reliance on renewable sources as a significant component of the energy mix.
5. **Seasonal Fluctuations**:
   - The periodic fluctuations observed each year may reflect seasonal influences on energy production.
6. **Long-Term Trend**:
   - The long-term trend underscores a steady shift towards higher renewable energy production over the years.
7. **Implications for Further Analysis**:
   - This analysis provides an initial understanding of the temporal trends in renewable energy.
   - It serves as a foundation for more detailed exploration of factors influencing growth in renewable energy production.

3.1.6 Annual Analysis of Total Primary Energy Production Using Candlestick Chart

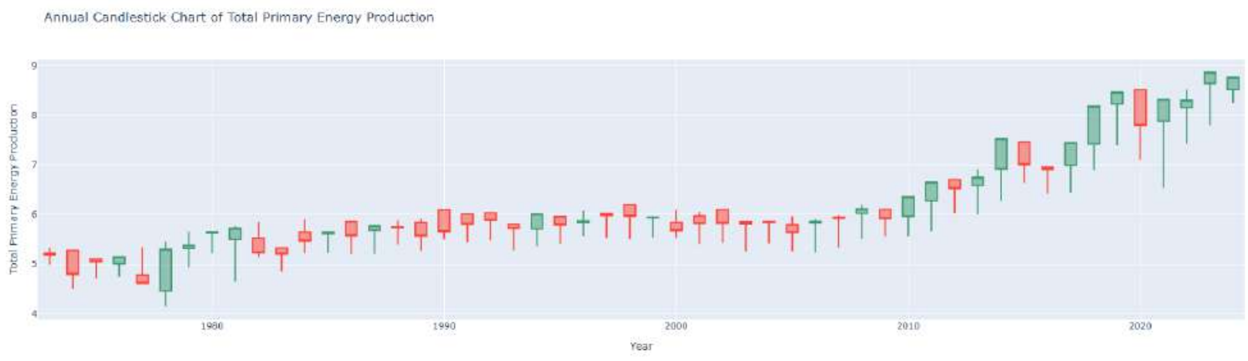


Figure 3.27: Annual Analysis of Total Primary Energy Production Using Candlestick Chart

1. **Candlestick Representation**:
   - Each candlestick represents the yearly open, close, high, and low values of energy production.
   - Shows fluctuations in production over time.
2. **1970s to 1990s**:
   - Production remains stable with smaller candlesticks.
   - Limited growth and low variability.
3. **2000s**:
   - Upward trend begins with taller candlesticks.
   - Increased production and higher variability.
4. **Post-2010**:
   - Accelerated growth with predominantly green candlesticks.
   - Represents years of production increases.
5. **Declines**:
   - Occasional red candlesticks indicate minor declines.
   - These do not disrupt the upward trend.
6. **Key Insights**:
   - Shows both seasonal variability and a long-term upward trajectory.

3.1.7 Rolling Average of Total Primary Energy Consumption

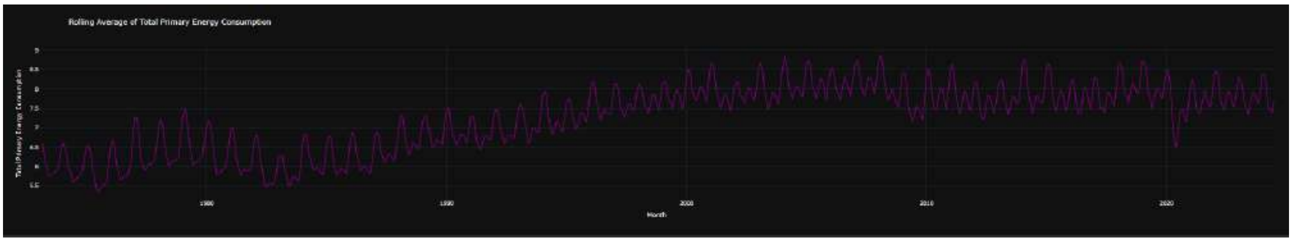


Figure 3.28: Rolling Average of Total Primary Energy Consumption

The rolling average plot effectively highlights the increasing trend, seasonal patterns, and potential disruptions in the total primary energy consumption data over time. This kind of analysis helps in understanding long-term trends and making informed predictions.

3.1.8 Year on year Growth

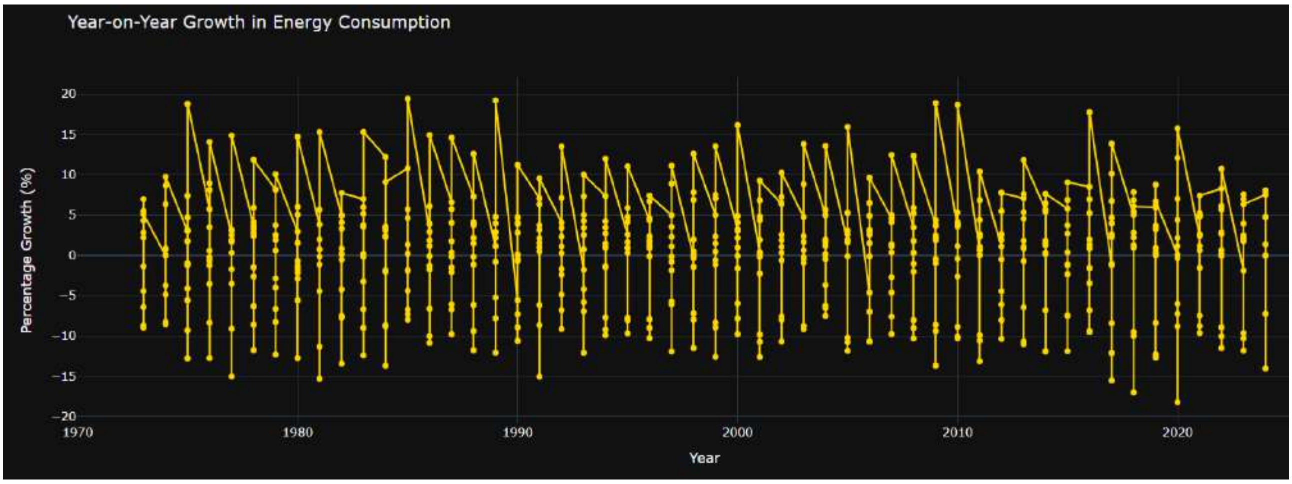


Figure 3.29: Year on Year Growth in Energy Consumption

The graph illustrates the year-on-year percentage growth in energy consumption from 1970 to 2023. It displays significant fluctuations, with periods of sharp increases followed by steep declines, reflecting varying economic activities, energy policies, and global events over time.

3.2 Bivariate Analysis

3.2.1 Scatter Plots

A scatter plot may indicate the key insights into how Total Primary Energy Consumption is related to Total Renewable Energy Production. A positive trend would suggest that high energy-consuming regions or periods are also the ones producing the most renewable energy. No trend means a poor correlation. Seasonal color coding may emphasize variations in renewable energy production by seasons, with more renewable energy during sunnier or windier seasons. The size of the points represents energy consumption, which can easily identify energy-intensive regions. In addition, outliers, such as low renewable production in high consumption areas, would point to policy or infrastructure gaps. Clusters in the data will indicate similar energy profiles across regions or sectors. This means the overall plot can help explain visually how primary and renewable energy metrics correlate and provide insight into trends, anomalies, and potential areas of improvement regarding renewable energy adoption.

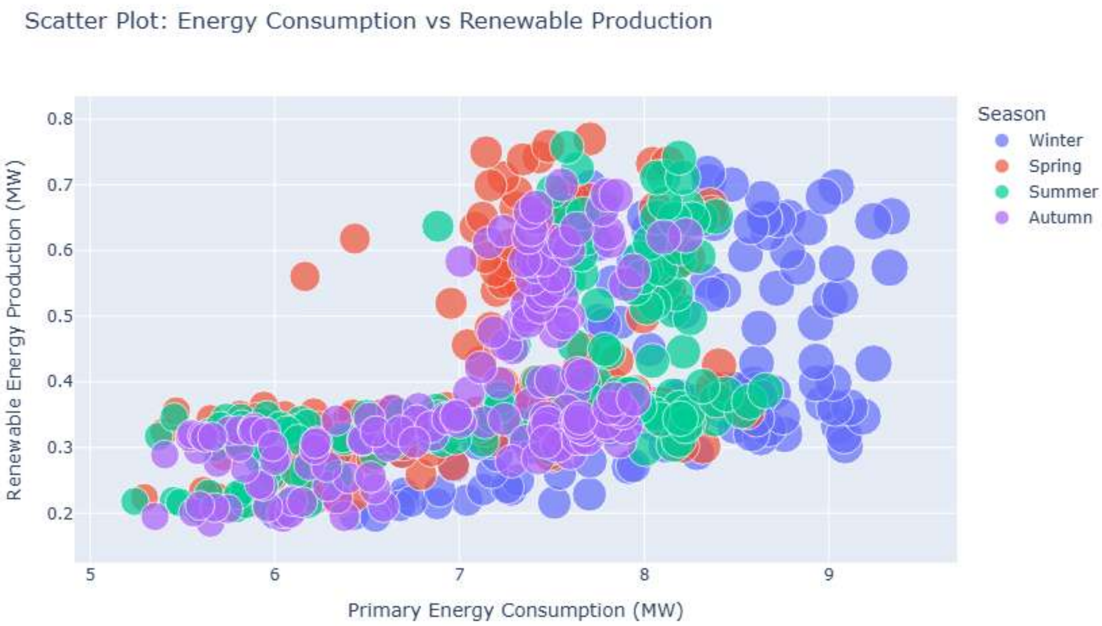


Figure 3.30: Scatter: Bubble Plot

The bubble chart highlights the relationship between total primary energy consumption and renewable energy production across seasons. Larger bubbles indicate higher energy deficits, showing a clear variation in deficits across different consumption levels. Seasonal trends reveal distinct clusters, with renewable energy production relatively low compared to consumption, particularly for larger deficits. The visualization helps identify seasonal and consumption-related trends in energy balance. The scatter plot illustrates the relationship between primary energy consumption and renewable energy production across seasons. Data points are clustered, showing that renewable energy production remains consistently lower than consumption. Seasonal variations are evident, with slight overlaps, suggesting trends in renewable energy contribution during different periods. This aids in identifying seasonal energy production patterns.

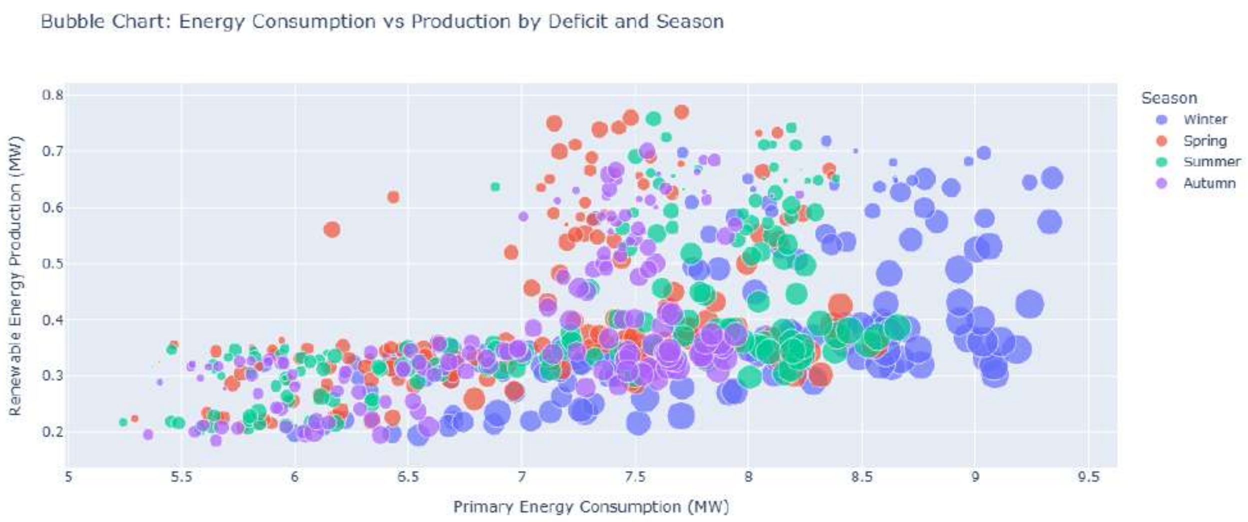


Figure 3.31: Scatter: Bubble Plot

3.2.2 Hexbin Plots

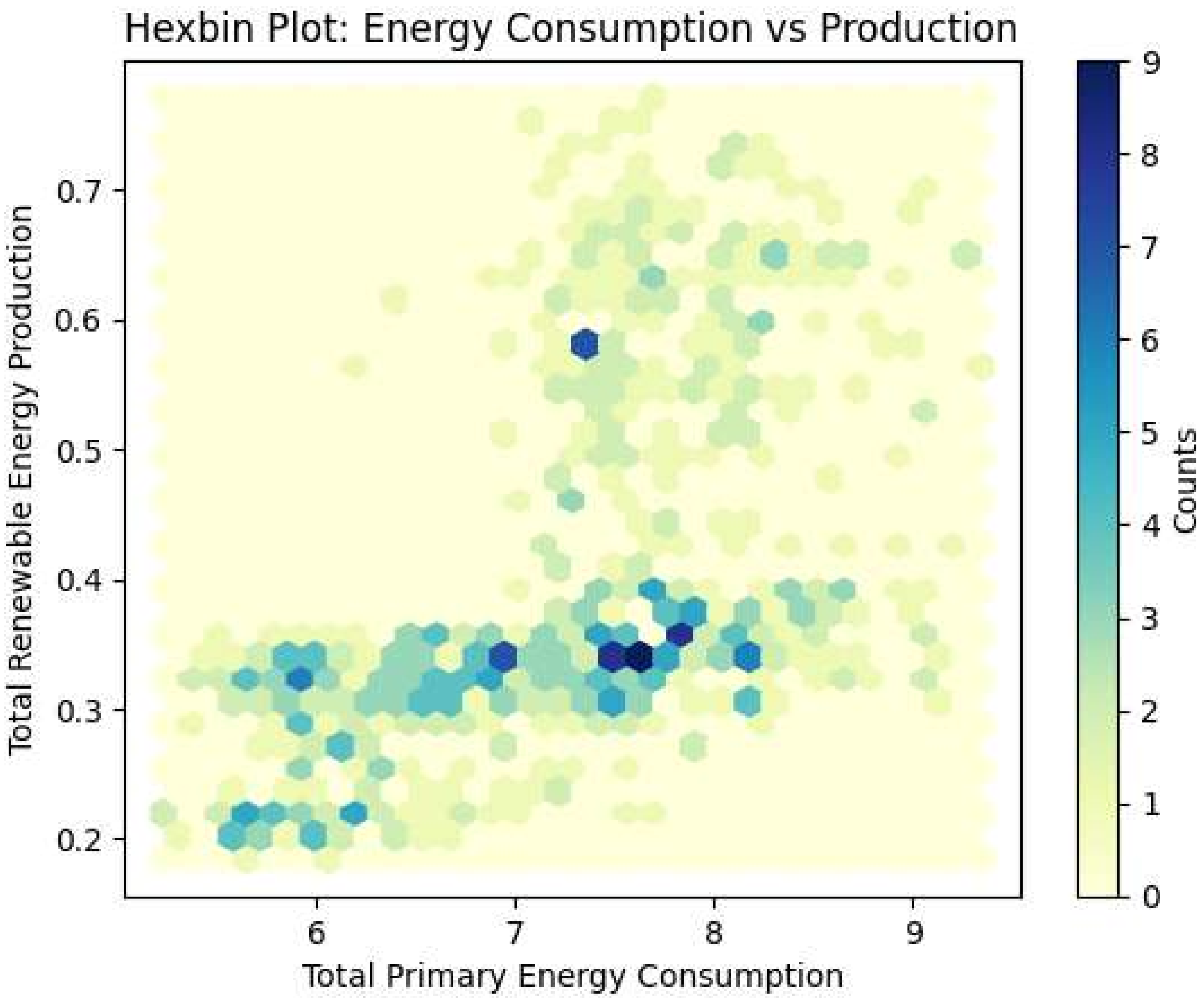


Figure 3.32: Hexbin Plot

The hexbin plot visualizes the density of data points for total primary energy consumption and renewable energy production. Darker hexagons indicate higher densities, with most values concentrated in the range of 6–8 MW consumption and 0.3–0.5 MW production. This provides insights into common consumption-production levels and highlights areas with fewer observations.

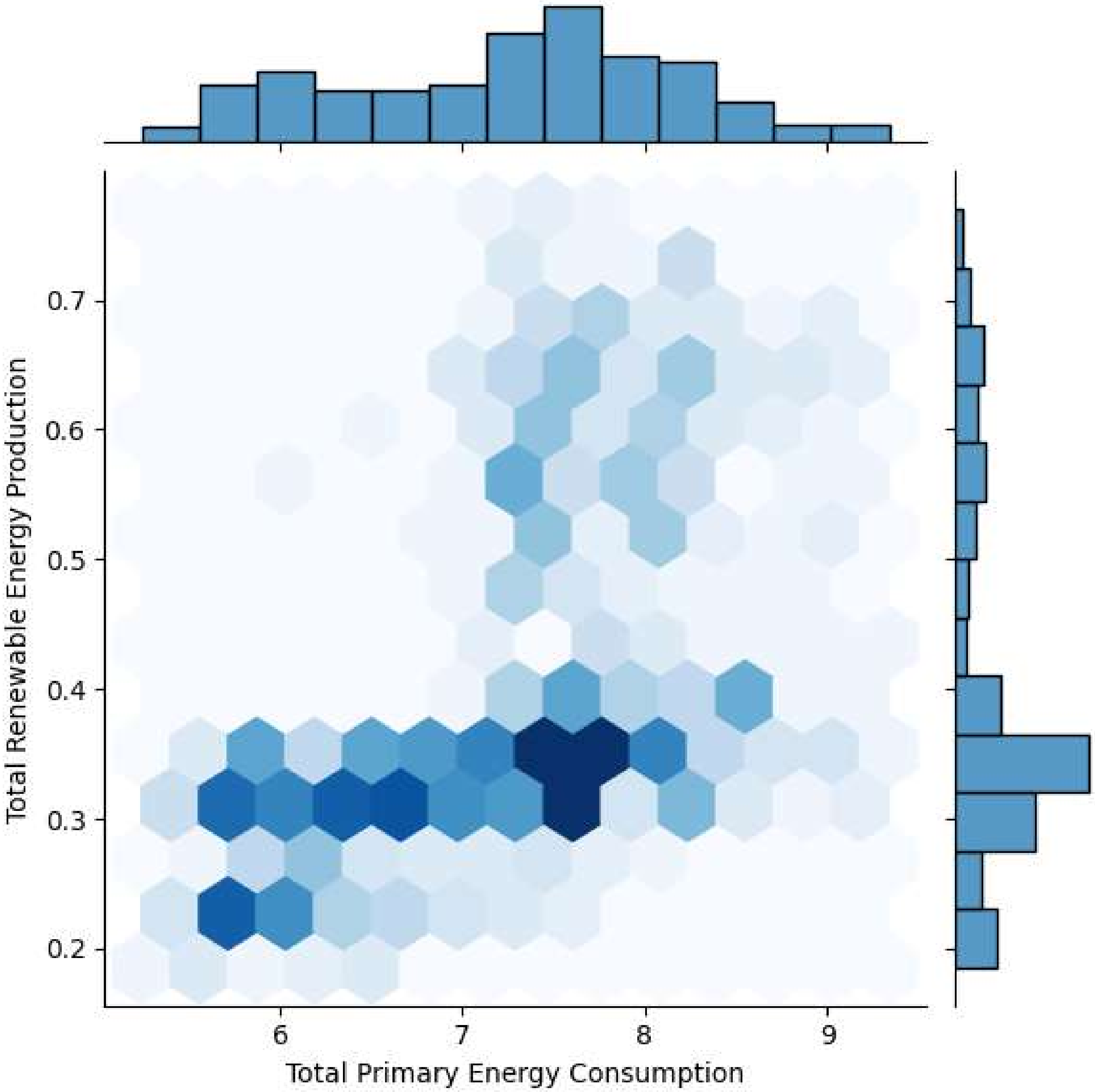


Figure 3.33: Hexbin Plot

The hexbin plot visualizes the relationship between Total Primary Energy Consumption and Total Renewable Energy Production, showing clusters of denser data points in darker blue. There is a concentration of data in regions where primary energy consumption is moderate to high, and renewable energy production ranges between 0.3 and 0.4, suggesting a potential correlation or trend. Outliers or sparsely populated hexagons indicate less frequent combinations of these variables. This analysis highlights the variation in energy usage patterns across different contexts or datasets.

3.2.3 Principal Component Analysis

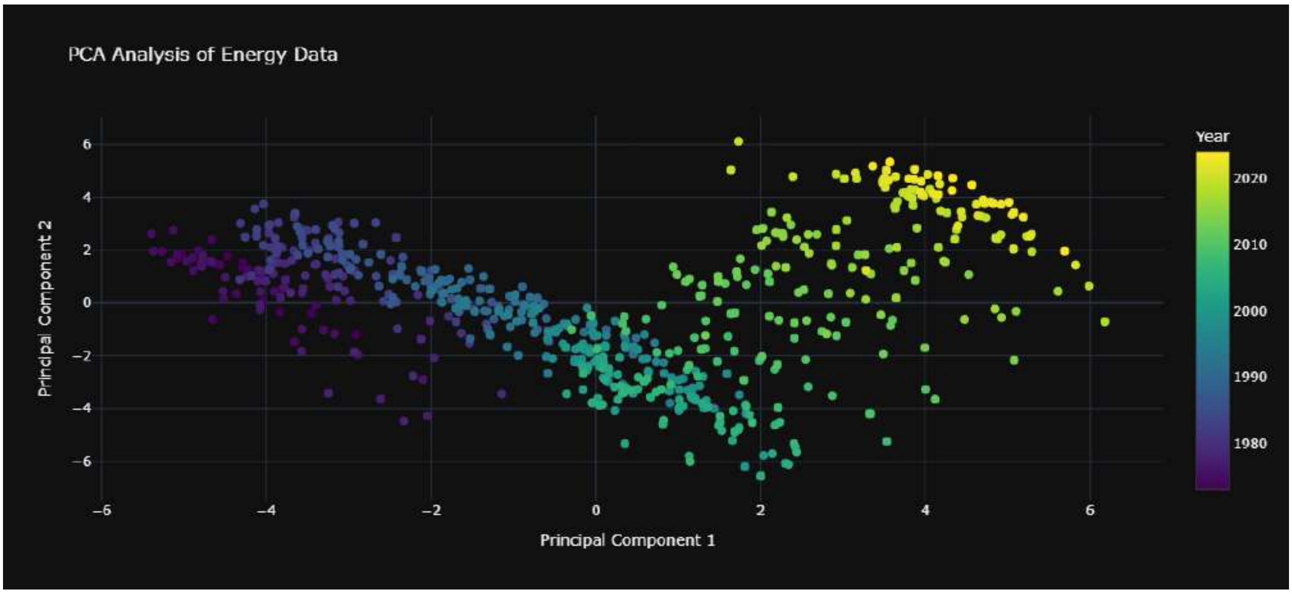


Figure 3.34: PCA analysis: PCA1- Rolling Average and PCA2- Year on Year Growth

The PCA plot reveals insights into energy data trends, with PCA1 (rolling average) showing smoother long-term variations, and PCA2 (year-on-year growth) capturing shorter-term fluctuations. Recent years display higher growth variability, while earlier years cluster more consistently, reflecting steady energy consumption patterns. The temporal gradient emphasizes evolving dynamics over time.

3.3 Multivariate analysis

3.3.1 Annual Trends in Energy Consumption by Sector

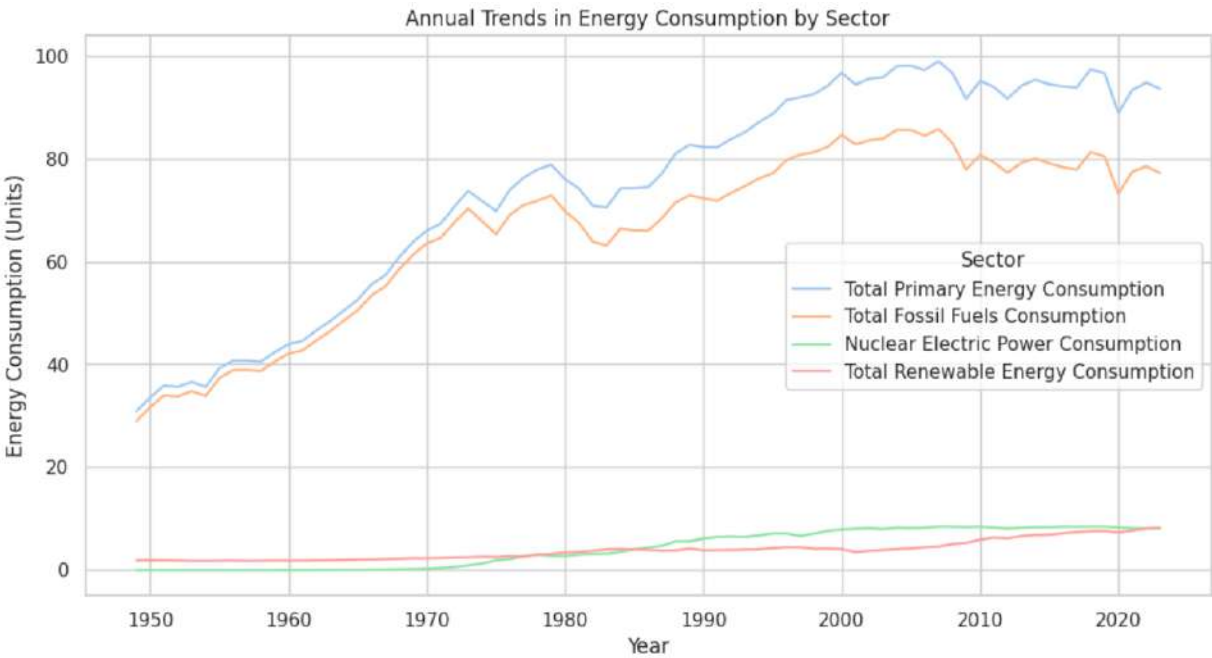


Figure 3.35: Annual Trends in Energy Consumption by Sector

The graph shows a significant increase in total primary energy consumption from 1950 to around 2010, with fossil fuels remaining the dominant source. However, consumption of nuclear and renewable energy exhibits modest growth, reflecting gradual diversification in energy sources.

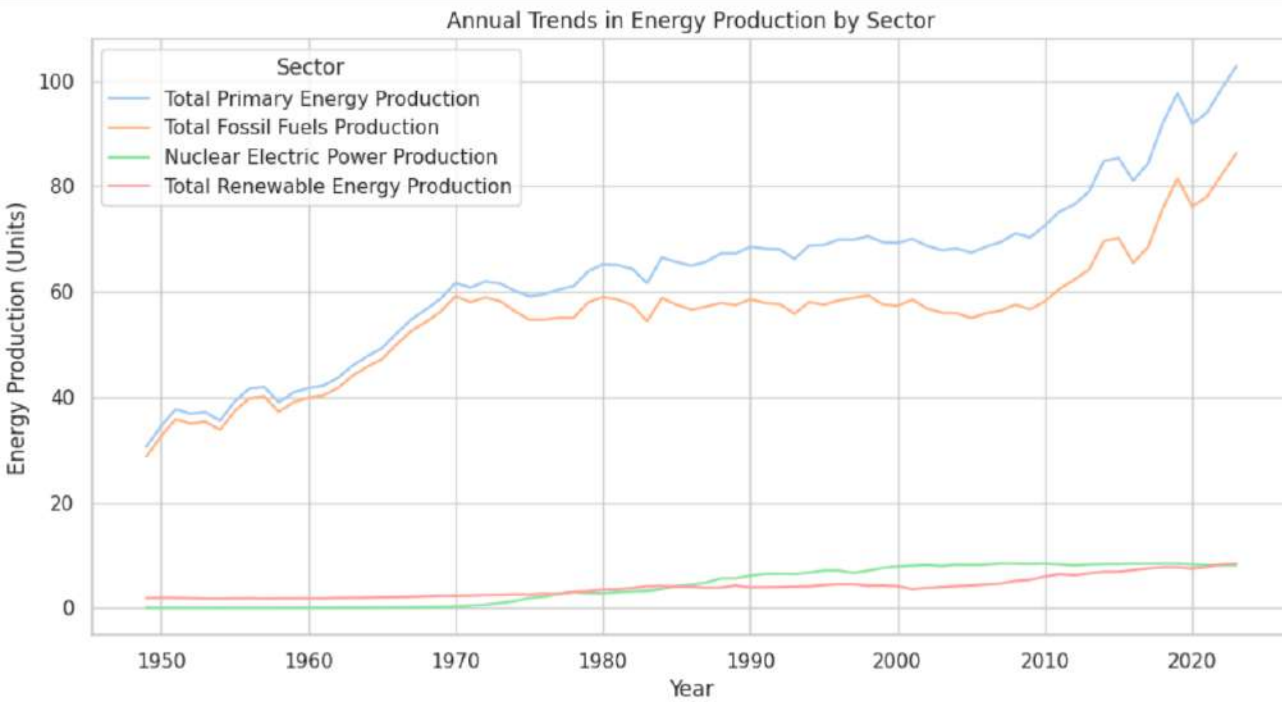


Figure 3.36: Annual Trends in Energy Production by Sector

The production of total primary energy follows a similar trend to consumption, with fossil fuels leading throughout. Noticeable growth in renewable energy production post-2000 suggests increased adoption of sustainable energy technologies, while nuclear energy shows steady but slower growth.

3.3.2 Annual Correlation Matrix

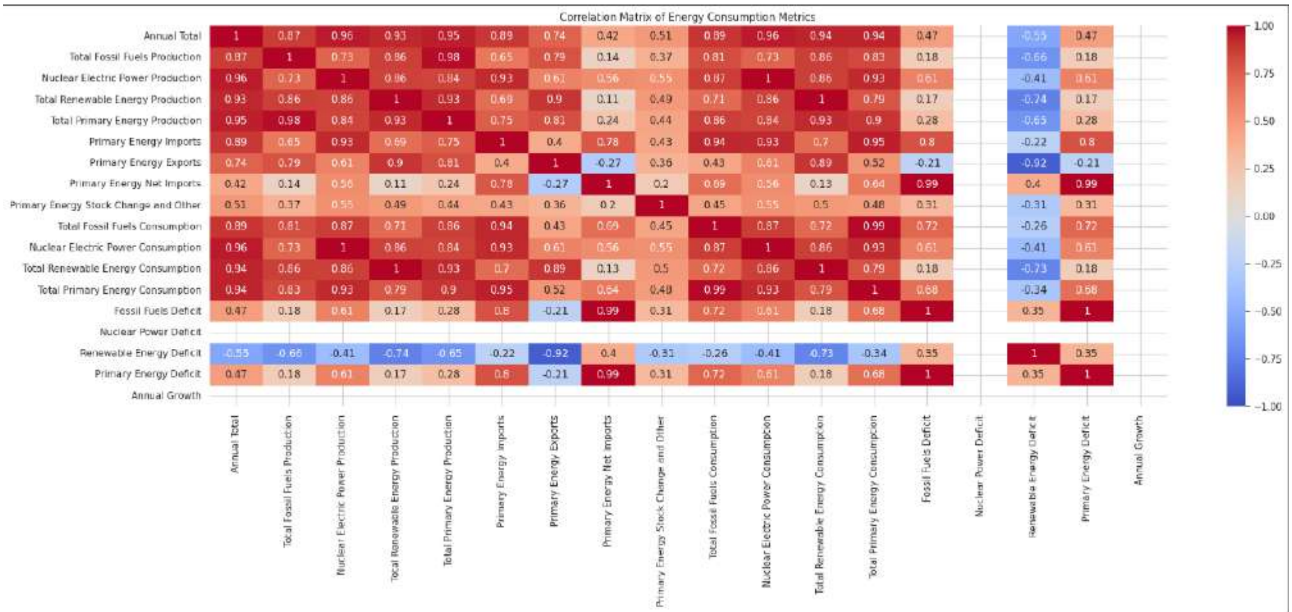


Figure 3.37: Annual Correlation Matrix

The annual correlation matrix highlights strong positive correlations between total fossil fuel production and overall energy consumption, signifying their interdependence. Renewable energy sources show weaker correlations with traditional energy metrics, indicating a less integrated role in the energy mix at the annual scale.

3.3.3 Montly Correlation Matrix

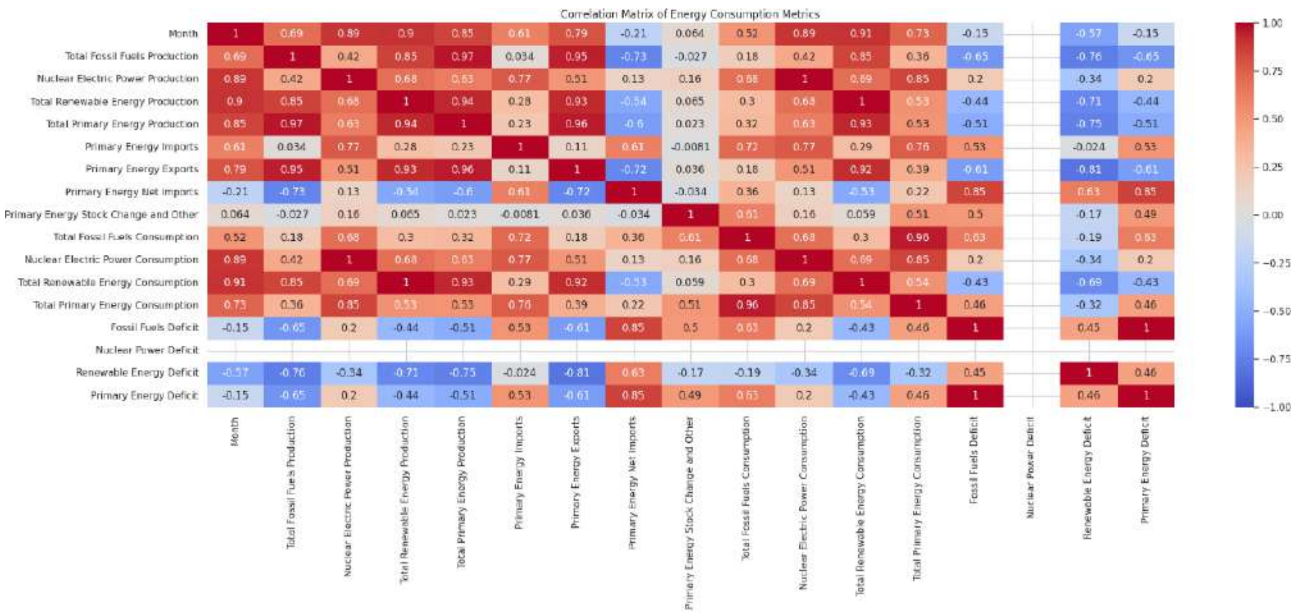


Figure 3.38: Montly Correlation Matrix

The monthly correlation matrix reveals more granular trends, with strong seasonal or cyclical relationships among energy metrics. For example, renewable energy production may correlate positively with specific consumption patterns, reflecting seasonal variations in demand and production capacity.

3.3.4 Montly Energy HeatMap

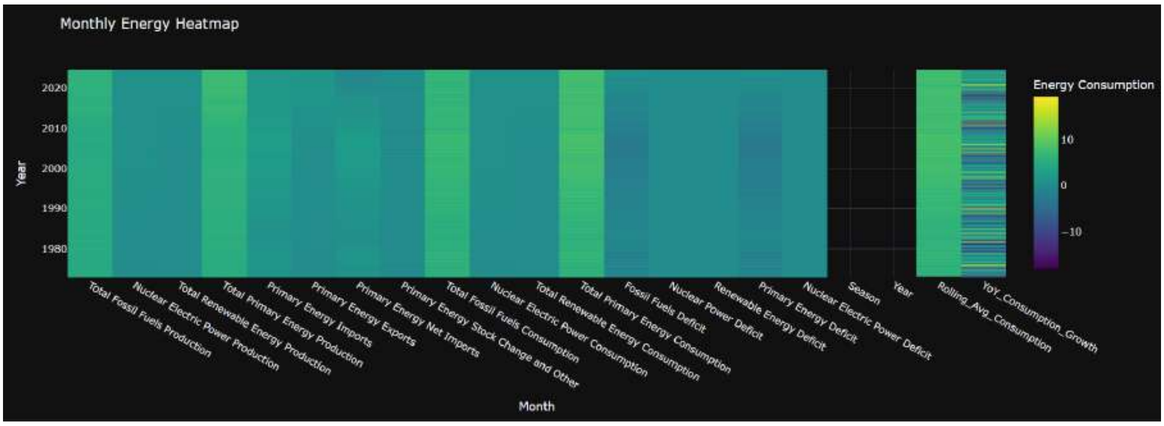


Figure 3.39: Montly Energy Heatmap

The heatmap visualizes variations in energy consumption across months and categories, with noticeable periodic patterns. The intensity of color suggests peak energy usage aligns with specific months, likely reflecting seasonal demand changes such as heating in winter or cooling in summer.

Additionally, renewable energy sources display a steadier pattern compared to fossil fuels, indicating consistent production trends, possibly influenced by natural factors like sunlight or wind availability.

3.3.5 3D Scattar Plot

The 3D scatter plot shows the relationship between Total Primary Energy Consumption (x-axis), Total Renewable Energy Production (y-axis), and Primary Energy Exports (z-axis) over time. Earlier years (darker colors) are concentrated at lower values, while recent years (brighter colors) exhibit higher variability, particularly in energy exports. This suggests evolving trends in energy utilization and export strategies over time.

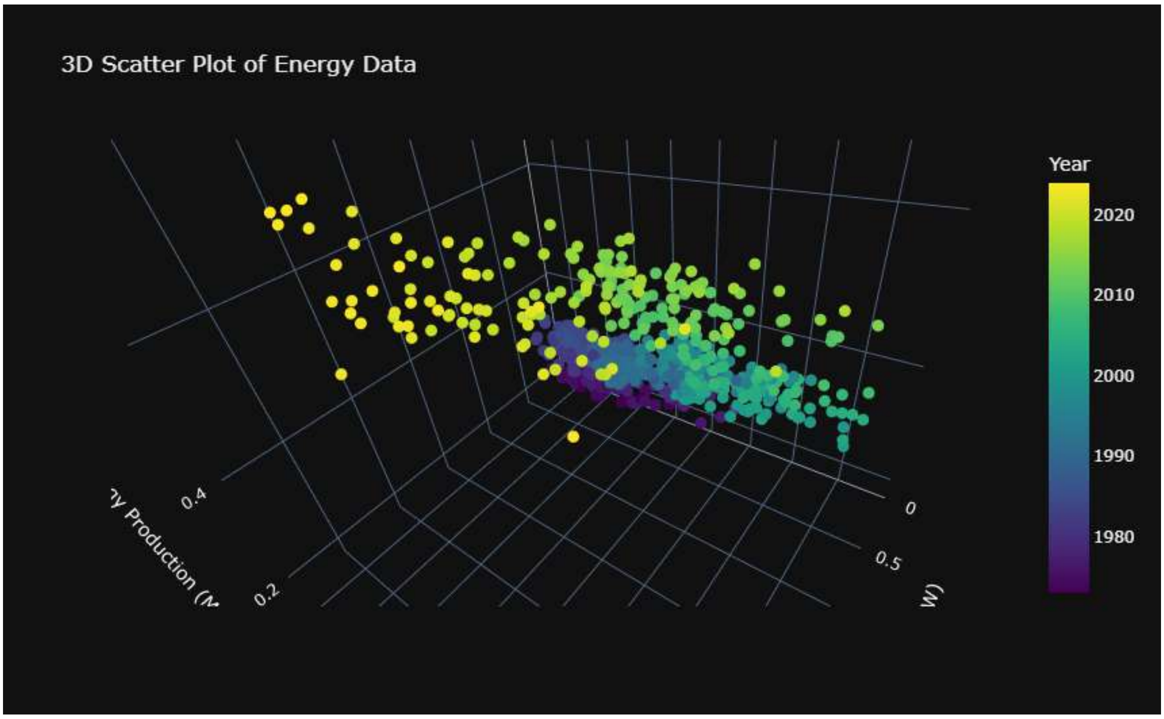


Figure 3.40: 3D Scattar Plot: Primary Energy Consumption - Total Renewable Energy Production - Primary Energy Exports

3.3.6 Radar Plot

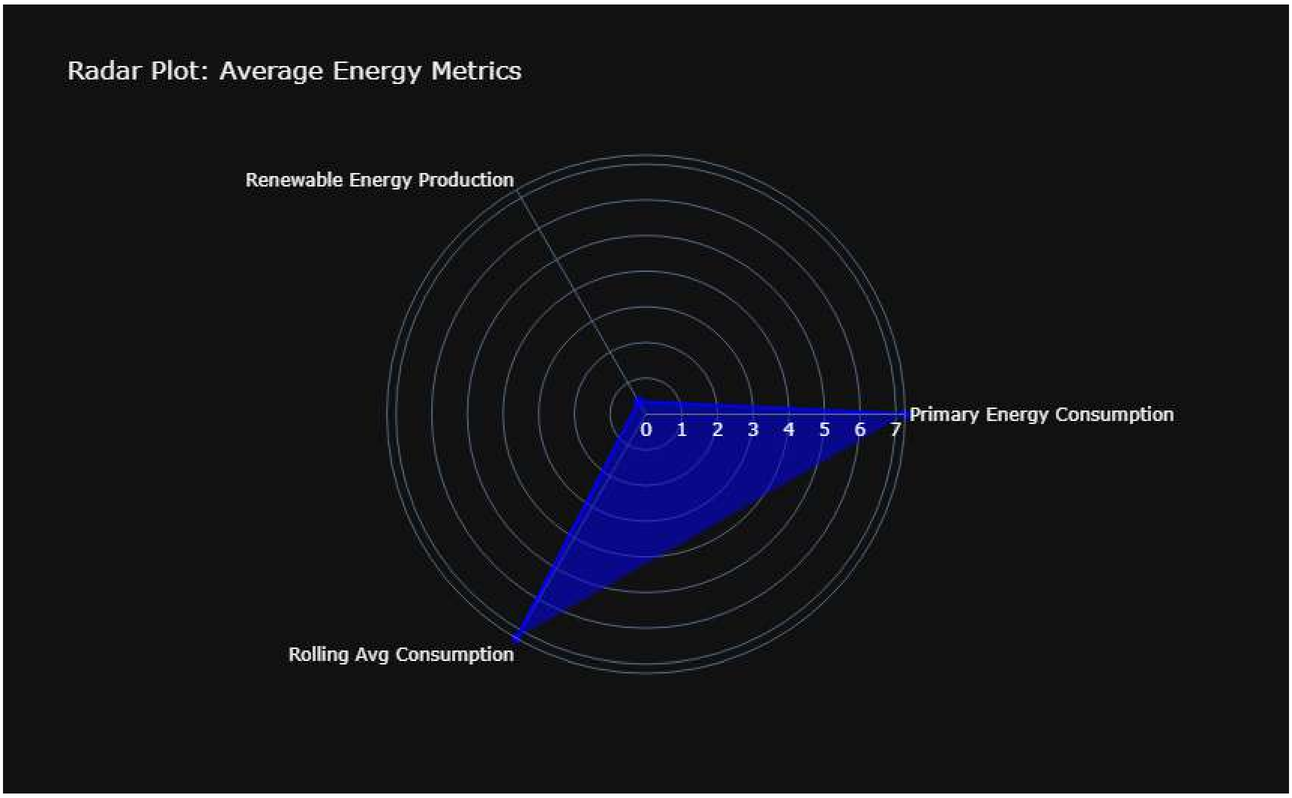


Figure 3.41: Radar Plot

This radar plot visualizes three key energy metrics: Primary Energy Consumption, Renewable Energy Production, and Rolling Average Consumption. The largest segment corresponds to Primary Energy Consumption, indicating its significantly higher average value compared to the other metrics. Renewable Energy Production and Rolling Average Consumption are comparatively lower, suggesting opportunities for growth in renewable energy. The balanced shape of the plot highlights disparities in energy usage patterns and potential areas for sustainable energy initiatives.

3.3.7 Ternary Plot

This ternary plot illustrates the distribution of energy metrics: Primary Energy Consumption, Renewable Energy Production, and Fossil Fuels Consumption. The clustering of points near the Fossil Fuels Consumption axis highlights the dominance of fossil fuels in the energy mix. Renewable energy production occupies a smaller share, indicating the potential for expansion. The gradual shift in point distribution over time, shown by the color gradient, suggests changes in energy trends, likely reflecting increasing emphasis on renewable sources.

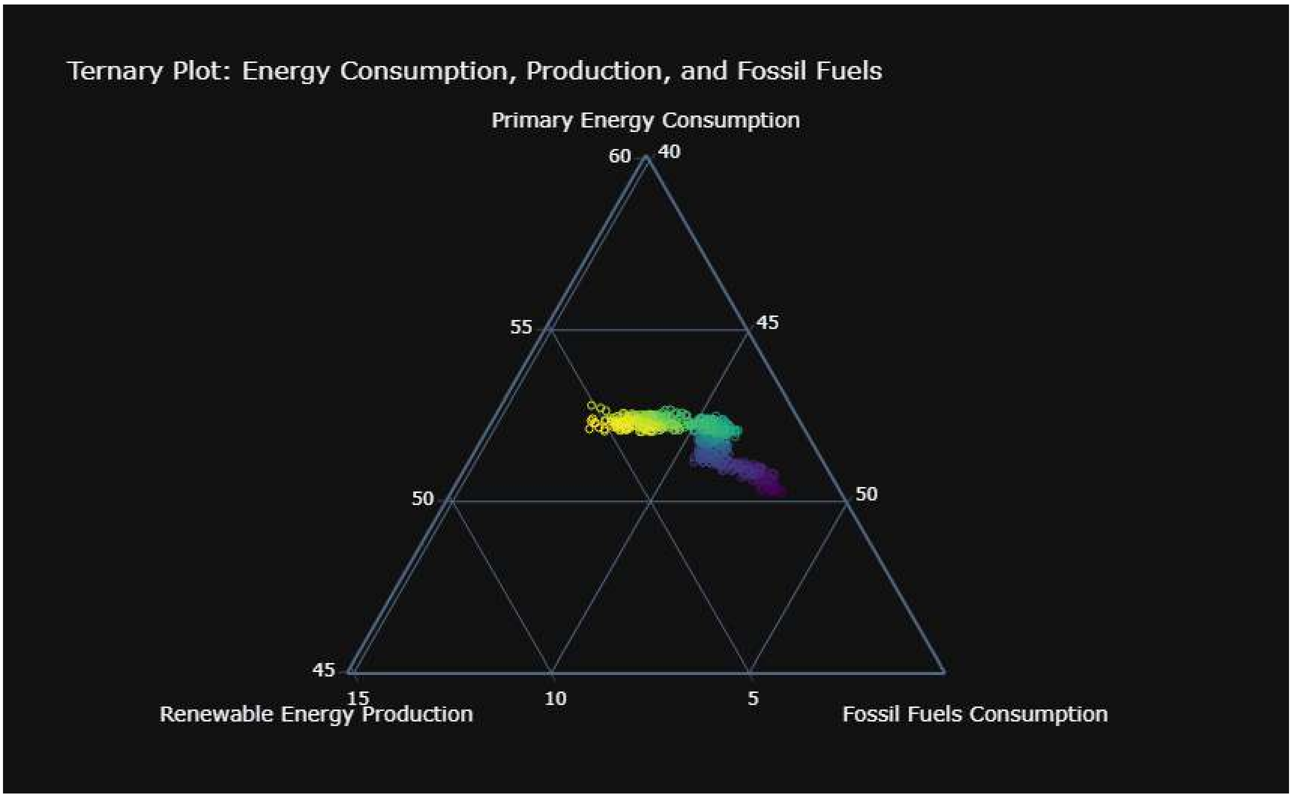


Figure 3.42: Ternary Plot: Primary Energy Consumption - Renewable Energy Production - Fossil Fuels Consumption

Chapter 4. Feature Engineering

4.1 Feature selection

4.1.1 For Monthly Energies

	Month	Total Fossil Fuels Production	Nuclear Electric Power Production	Total Renewable Energy Production	Total Primary Energy Production	Primary Energy Imports	Primary Energy Exports	Primary Energy Net Imports	Primary Energy Stock Change and Other	Total Fossil Fuels Consumption	Nuclear Electric Power Consumption	Total Renewable Energy Consumption	Total Primary Energy Consumption	Fossil Fuels Deficit	Nuclear Power Deficit	Renewable Energy Deficit	Primary Energy Deficit
0	1973-01-01	4.932632	0.068103	0.219839	5.220574	1.173080	0.125781	1.047299	0.771858	6.747651	0.068103	0.219839	7.039731	1.815019	0.0	0.0	1.819157
1	1973-02-01	4.729582	0.064634	0.197330	4.991545	1.168005	0.120883	1.047122	0.390129	6.163095	0.064634	0.197330	6.428796	1.433513	0.0	0.0	1.437251
2	1973-03-01	4.946902	0.072494	0.218686	5.238082	1.309473	0.139950	1.169523	-0.067640	6.044647	0.072494	0.218686	6.339964	1.097745	0.0	0.0	1.101882
3	1973-04-01	4.716271	0.064070	0.209330	4.989672	1.085169	0.194185	0.890984	-0.110067	5.493184	0.064070	0.209330	5.770589	0.776913	0.0	0.0	0.780917
4	1973-05-01	4.956995	0.062111	0.215982	5.235087	1.162804	0.196775	0.966029	-0.305335	5.613551	0.062111	0.215982	5.895781	0.656556	0.0	0.0	0.660694

Figure 4.1: Feature Creation: Adding Columns for Energy Deficits (Fossil Fuels, Nuclear Power, Renewable Energy, Primary Energy)

Total Primary Energy Consumption	Fossil Fuels Deficit	Nuclear Power Deficit	Renewable Energy Deficit	Primary Energy Deficit	Nuclear Electric Power Deficit	Season	Year	Rolling_Avg_Consumption	YoY_Consumption_Growth
7.039731	-1.815019	0.0	0.0	-1.819157	0.0	Winter	1973-01-01	NaN	NaN
6.428796	-1.433513	0.0	0.0	-1.437251	0.0	Winter	1973-01-01	NaN	-8.678386
6.339964	-1.097745	0.0	0.0	-1.101882	0.0	Spring	1973-01-01	6.602830	-1.381783
5.770589	-0.776913	0.0	0.0	-0.780917	0.0	Spring	1973-01-01	6.179783	-8.980729
5.895781	-0.656556	0.0	0.0	-0.660694	0.0	Spring	1973-01-01	6.002111	2.169484
5.633700	-0.631383	0.0	0.0	-0.635387	0.0	Summer	1973-01-01	5.766690	-4.445230
5.792383	-0.756280	0.0	0.0	-0.760417	0.0	Summer	1973-01-01	5.773955	2.816675
6.041332	-0.706882	0.0	0.0	-0.711019	0.0	Summer	1973-01-01	5.822472	4.297868
5.651809	-0.645278	0.0	0.0	-0.649281	0.0	Autumn	1973-01-01	5.828508	-6.447634
6.044923	-0.766979	0.0	0.0	-0.771117	0.0	Autumn	1973-01-01	5.912688	6.955543
6.376542	-1.277936	0.0	0.0	-1.281940	0.0	Autumn	1973-01-01	6.024425	5.485909
6.700965	-1.508318	0.0	0.0	-1.512457	0.0	Winter	1973-01-01	6.374143	5.087758
6.697586	-1.417001	0.0	0.0	-1.420679	0.0	Winter	1974-01-01	6.591698	-0.050426
6.124546	-1.250615	0.0	0.0	-1.253938	0.0	Winter	1974-01-01	6.507699	-8.555919
6.173677	-0.964386	0.0	0.0	-0.968065	0.0	Spring	1974-01-01	6.331936	0.802198
5.661961	-0.580461	0.0	0.0	-0.584021	0.0	Spring	1974-01-01	5.986728	-8.288675
5.664620	-0.393312	0.0	0.0	-0.396990	0.0	Spring	1974-01-01	5.833419	0.046963
5.451943	-0.593337	0.0	0.0	-0.596897	0.0	Summer	1974-01-01	5.592841	-3.754480

Figure 4.2: Season, Rolling Avg, and YoY Growth

- **Season**: A column added to categorize each row by the corresponding season (Winter, Spring, Summer, Autumn) based on the date, enabling seasonal trend analysis.
- **YoY Consumption Growth**: Year-on-Year growth rate of total primary energy consumption, highlighting annual changes in energy demand.
- **Rolling Average Consumption**: A moving average of energy consumption, smoothing out short-term fluctuations to reveal longer-term trends.
- **Deficit** Added Energy Deficit Columns for Fossil Fuels, Nuclear Power, Renewable Energy, Primary Energy.

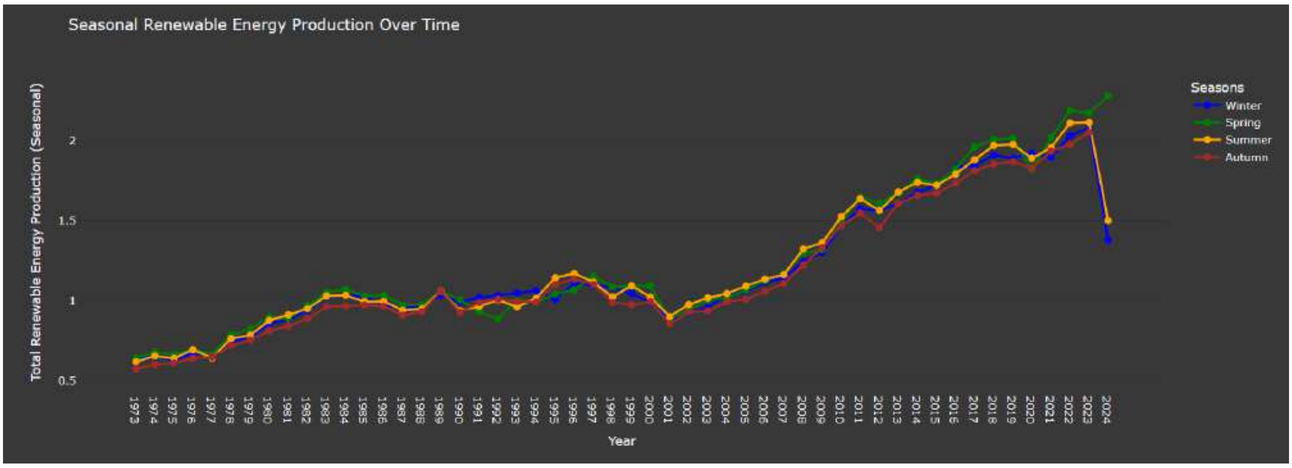


Figure 4.3: Seasonal Renewable Energy Production Over Time

The graph illustrates the trends in renewable energy production across different seasons over time. There is a general upward trend in production, suggesting increasing adoption or efficiency in renewable energy sources. Seasonal fluctuations are evident, with distinct patterns for each season. Toward the end of the timeline, a sharp dip is observed, potentially signaling a disruption or seasonal anomaly.

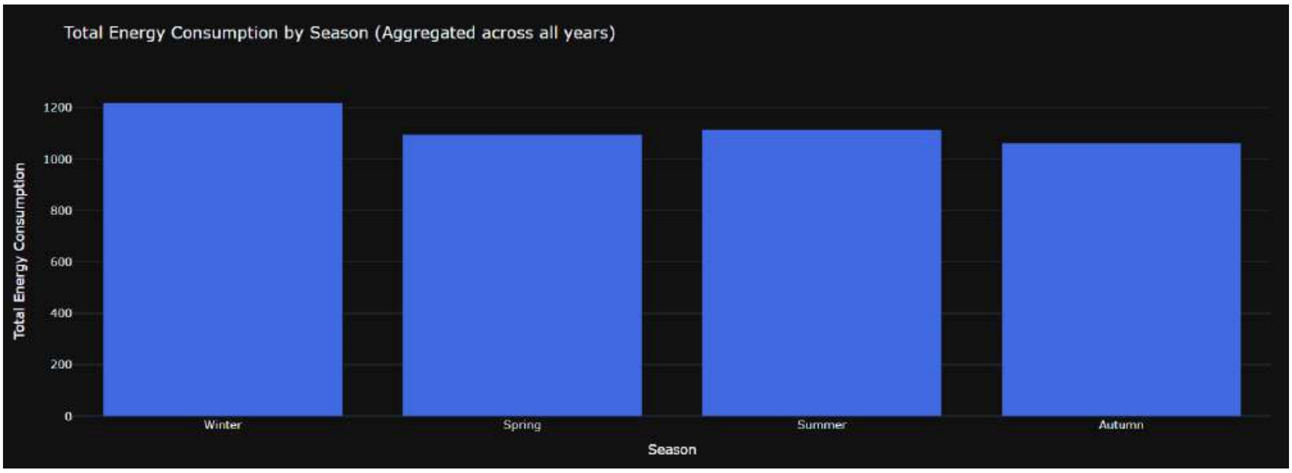


Figure 4.4: Total Energy Consumption by Season

The graph aggregates energy consumption across all years, showing relatively equal distribution among seasons. The similar bar heights indicate no significant seasonal variation in total energy usage. This uniformity suggests that energy demands remain consistent throughout the year, irrespective of season-specific factors.

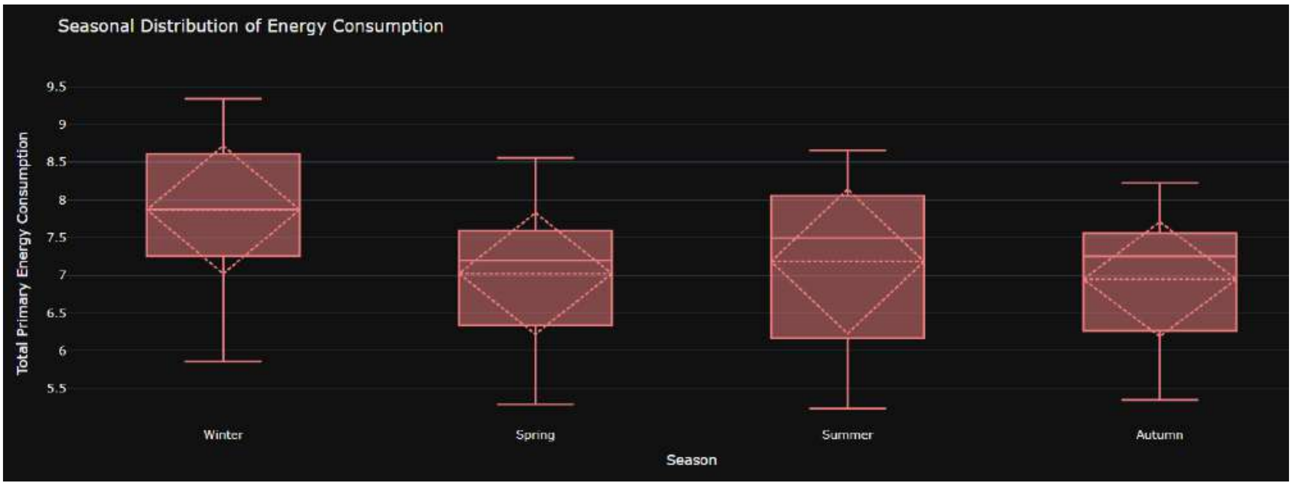


Figure 4.5: Seasonal Distribution of Energy Consumption

The box plot displays the seasonal variations in energy consumption. Winter exhibits the highest median consumption, with a wider range compared to other seasons, indicating greater variability. Summer and autumn have relatively moderate medians, while spring shows the lowest median and the smallest range, suggesting more consistent energy usage during that season.

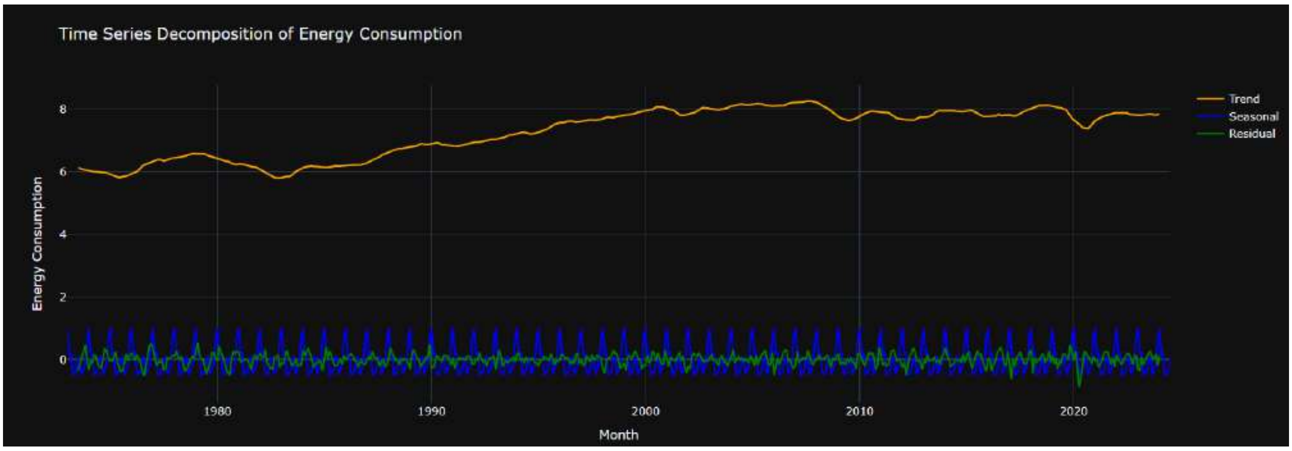


Figure 4.6: Time Series Decomposition of Energy Consumption

This decomposition plot breaks down energy consumption into its trend, seasonal, and residual components. The trend line reveals a gradual increase in energy consumption over time. The seasonal component shows repeating periodic patterns, highlighting consistent fluctuations within a year. Residuals capture irregularities, which appear minimal and do not deviate significantly from expected values.

4.1.2 For Annual Energies

```
energy_annual.head()
```

	Annual Total	Total Fossil Fuels Production	Nuclear Electric Power Production	Total Renewable Energy Production	Total Primary Energy Production	Primary Energy Imports	Primary Energy Exports	Primary Energy Net Imports	Primary Energy Stock Change and Other	Total Fossil Fuels Consumption	Nuclear Electric Power Consumption	Total Renewable Energy Consumption	Total Primary Energy Consumption	Fossil Fuels Deficit	Nuclear Power Deficit	Renewable Energy Deficit	Primary Energy Deficit
0	1949-01-01	28.740479	0.0	1.872627	30.613106	1.448158	1.591760	-0.143602	0.396915	28.988371	0.0	1.872627	30.866419	0.247892	0.0	0.0	0.253313
1	1950-01-01	32.553205	0.0	1.906525	34.459730	1.912887	1.465322	0.447565	-1.379921	31.614755	0.0	1.906525	33.527374	-0.938450	0.0	0.0	-0.932356
2	1951-01-01	35.782118	0.0	1.890800	37.672919	1.892425	2.621545	-0.729119	-1.057801	33.987736	0.0	1.890800	35.885997	-1.794382	0.0	0.0	-1.786922
3	1952-01-01	34.964662	0.0	1.848694	36.813356	2.145984	2.365131	-0.219147	-0.959659	33.778116	0.0	1.848694	35.634550	-1.186546	0.0	0.0	-1.178806
4	1953-01-01	35.338361	0.0	1.792616	37.130976	2.313042	1.866013	0.447029	-0.976252	34.802286	0.0	1.792616	36.601753	-0.536075	0.0	0.0	-0.529223

Figure 4.7: Feature Creation: Adding Columns for Energy Deficits (Fossil Fuels, Nuclear Power, Renewable Energy, Primary Energy)

The calculated deficits (Fossil Fuels, Nuclear Power, Renewable Energy, and Primary Energy) reveal gaps between energy consumption and production, highlighting dependency on imports. These trends help assess self-sufficiency, identify shortfalls, and guide energy policy for enhancing domestic production.

4.2 Feature extraction

4.2.1 Handling Outliers

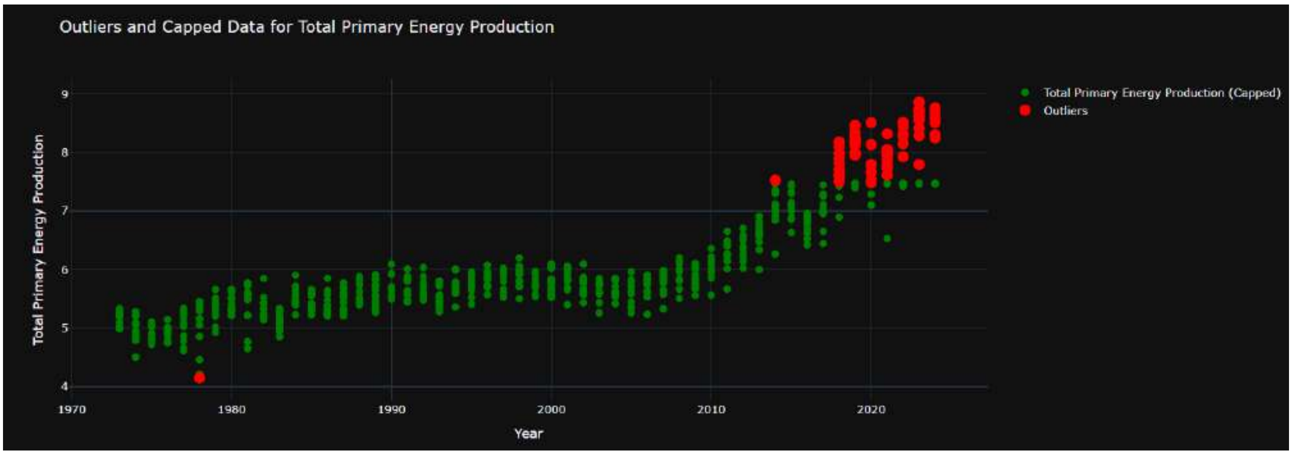


Figure 4.8: Outlier for Total Primary Energy Production

H

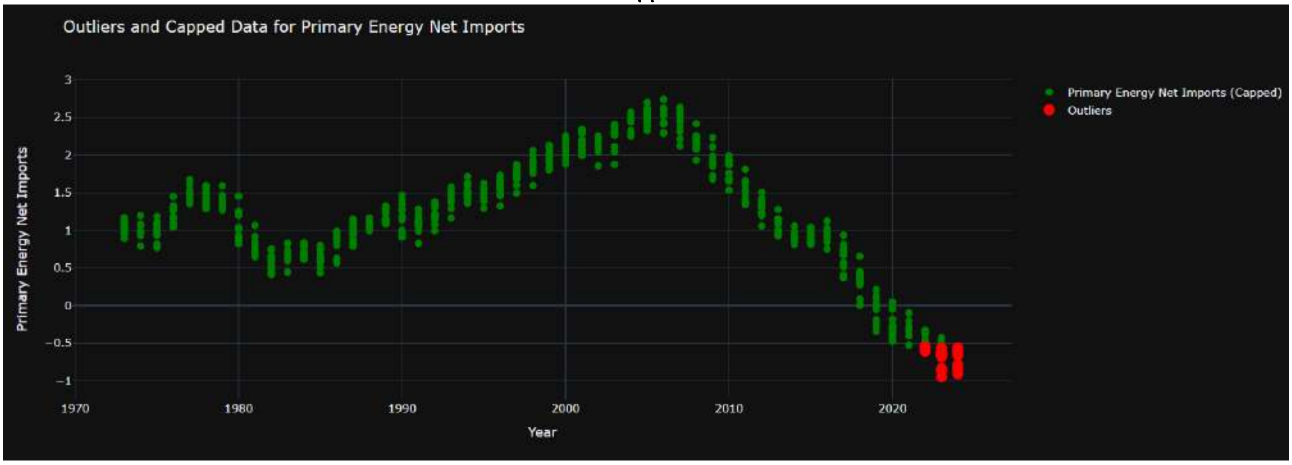


Figure 4.10: Outlier for Total Energy Net Imports

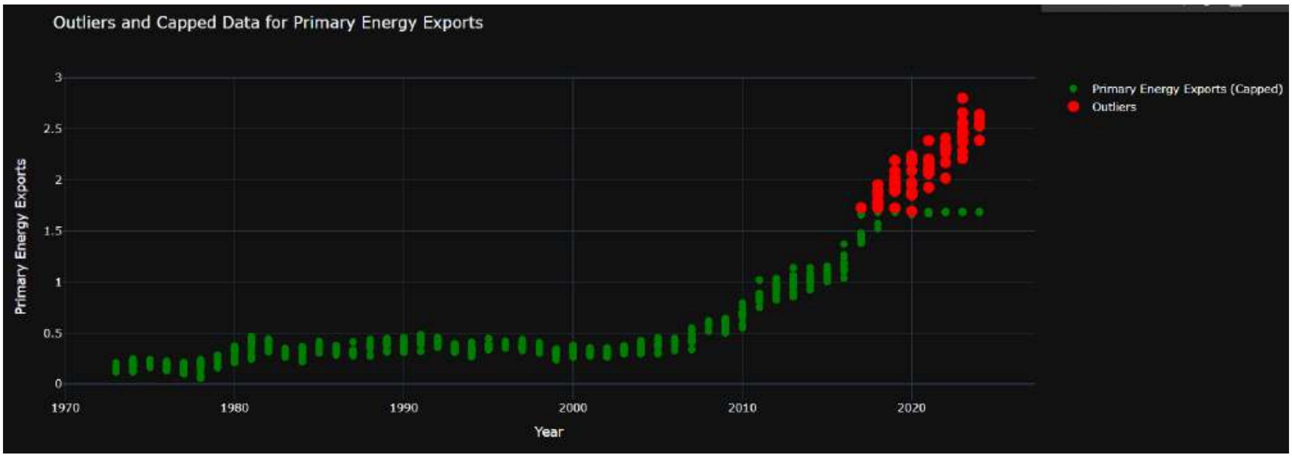


Figure 4.9: Outlier for Total Primary Energy Exports

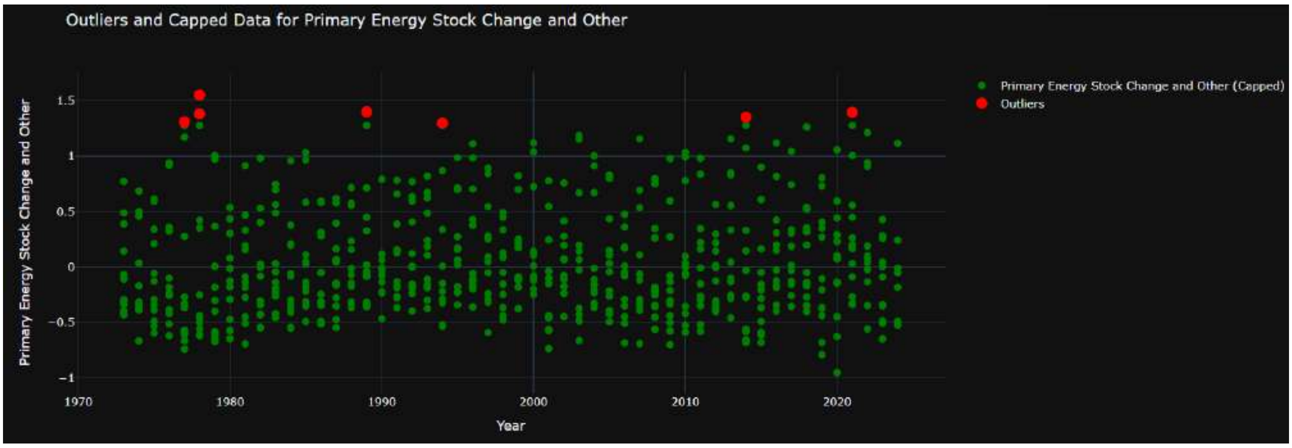


Figure 4.11: Outlier for Primary Energy Stock Change and Other

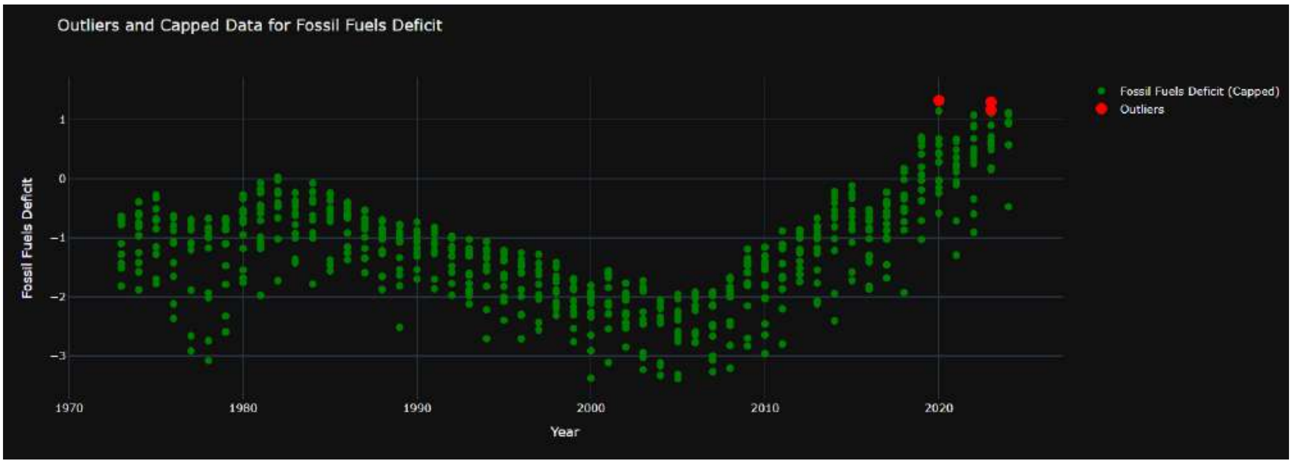


Figure 4.12: Outlier for Fossil Fuels Deficit

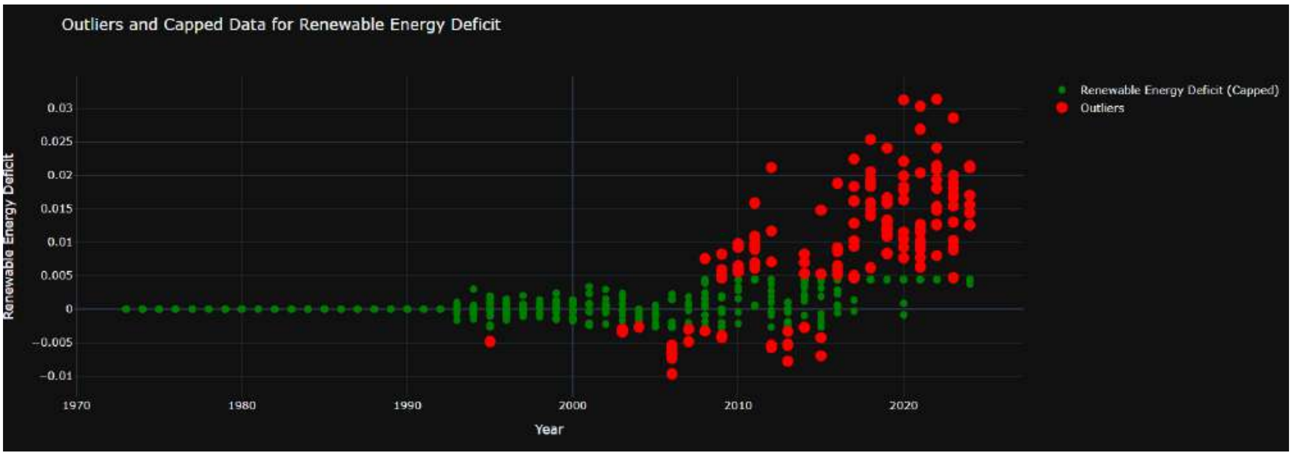


Figure 4.13: Outlier for Renewable Energy Deficit

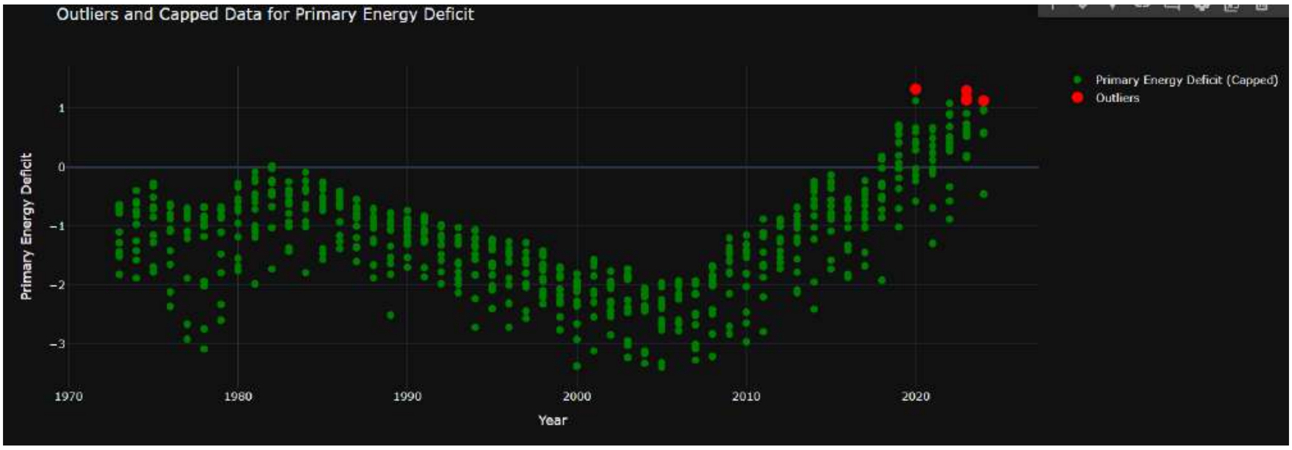


Figure 4.14: Primary Energy Deficit

Chapter 5. Model fitting

5.1 Linear Regression

5.1.1 SelectKBest Performance

SelectKBest demonstrated exceptional performance across all energy consumption metrics:

- Perfect prediction ($R = 1.000$) achieved for **Nuclear Electric Power** and **Total Primary Energy Consumption**.
- Near-perfect prediction for **Total Fossil Fuels** ($R = 0.9999$) and **Renewable Energy** ($R = 0.9991$).
- Consistently low error rates with MAE ≤ 0.0045 across all metrics.
- Strong discriminative ability with **AUC** scores ranging from 0.9839 to 1.0000.

5.1.2 Principal Component Analysis (PCA) Performance

PCA showed good but comparatively lower performance:

- Moderate to high accuracy with R values ranging from 0.8283 to 0.9420.
- Higher error rates compared to other methods (MAE $: 0.0279 - 0.2203$).
- Notable performance in **Nuclear Electric Power** prediction ($R = 0.9191$).
- **Renewable Energy** prediction showed strong accuracy ($R = 0.9420$) despite lower **AUC** (0.7823).

5.1.3 Recursive Feature Elimination (RFE) Performance

RFE demonstrated strong performance comparable to SelectKBest:

- Perfect prediction for **Total Primary Energy Consumption** ($R = 1.000$).
- Near-perfect prediction for **Total Fossil Fuels** ($R = 0.9999$).
- Strong performance for **Nuclear Electric Power** ($R = 0.9800$) and **Renewable Energy** ($R = 0.9603$).
- Consistent **AUC** scores ranging from 0.7823 to 1.0000.

Table 5.1: Model Performance Metrics Using Linear Regression

Method	Metric	Total Fossil Fuels	Nuclear Electric Power	Renewable Energy	Total Primary Energy
SelectKBest	MAE	0.0045	0.0000	0.0028	0.0000
	MSE	0.0000	0.0000	0.0000	0.0000
	R^2	0.9999	1.0000	0.9991	1.0000
	AUC	0.9839	1.0000	0.9839	1.0000
PCA	MAE	0.2203	0.0462	0.0279	0.1769
	MSE	0.0765	0.0033	0.0012	0.0501
	R^2	0.8283	0.9191	0.9420	0.9351
	AUC	0.8790	0.9274	0.7823	0.9194
RFE	MAE	0.0048	0.0236	0.0239	0.0000
	MSE	0.0000	0.0008	0.0008	0.0000
	R^2	0.9999	0.9800	0.9603	1.0000
	AUC	0.9839	0.9435	0.7823	1.0000

5.2 XGBRegressor

SelectKBest

- Performed exceptionally well for Nuclear Electric Power and Renewable Energy, with near-perfect R^2 (0.9997 and 0.9982) and very low errors (MAE = 0.0026 and 0.0039).
- While it showed strong performance for Total Fossil Fuels ($R^2 = 0.9461$) and Total Primary Energy ($R^2 = 0.9571$), it exhibited slightly higher MAE and MSE compared to RFE.

PCA

- Demonstrated good performance but was less effective compared to SelectKBest and RFE.
- Notable strengths were observed in Renewable Energy prediction ($R^2 = 0.9662$), though the AUC (0.8710) and error rates (MAE = 0.0210) were relatively higher.
- Struggled with Total Fossil Fuels ($R^2 = 0.8420$), indicating it may not capture features as effectively for this metric.

RFE

- Showed strong and consistent performance across all metrics.
- Achieved the highest R^2 (0.9687) for Renewable Energy and the lowest MAE (0.0195) compared to PCA.
- Performed comparably to SelectKBest for Total Fossil Fuels and Total Primary Energy, while achieving slightly higher R^2 (0.9633) for the latter.

Table 5.2: Model Performance Metrics Using XGBRegressor

Method	Metric	Total Fossil Fuels	Nuclear Electric Power	Renewable Energy	Total Primary Energy
SelectKBest	MAE	0.1152	0.0026	0.0039	0.1427
	MSE	0.0240	0.0000	0.0000	0.0331
	R^2	0.9461	0.9997	0.9982	0.9571
	AUC	0.9113	0.9919	0.9919	0.9274
PCA	MAE	0.2027	0.0359	0.0210	0.1890
	MSE	0.0704	0.0025	0.0007	0.0630
	R^2	0.8420	0.9384	0.9662	0.9184
	AUC	0.8790	0.9355	0.8710	0.9274
RFE	MAE	0.1122	0.0355	0.0195	0.1213
	MSE	0.0230	0.0026	0.0006	0.0283
	R^2	0.9483	0.9355	0.9687	0.9633
	AUC	0.9274	0.9435	0.8548	0.9435

5.3 XGBoost Classifier

The XGBoost classifier is a powerful tool for weather prediction due to its accuracy, feature reliability, and strong discrimination capabilities. Its ability to handle large-scale datasets, robustly assess feature importance, and differentiate between complex weather patterns makes it a preferred choice for meteorological applications.

Table 5.3: SelectKBest Performance Metrics for Hot Weather Prediction

Metric	Class 0 (Not Hot)	Class 1 (Hot)	Overall / Avg.
Precision	0.84	0.67	0.75 (Macro) / 0.79 (Weighted)
Recall	0.90	0.53	0.71 (Macro) / 0.80 (Weighted)
F1-Score	0.87	0.59	0.73 (Macro) / 0.79 (Weighted)
Accuracy	**79.84%**		
AUC	**0.8556**		

Table 5.4: PCA Performance Metrics for Hot Weather Prediction

Metric	Class 0 (Not Hot)	Class 1 (Hot)	Overall / Avg.
Precision	0.77	0.53	0.65 (Macro) / 0.70 (Weighted)
Recall	0.91	0.26	0.59 (Macro) / 0.73 (Weighted)
F1-Score	0.83	0.35	0.59 (Macro) / 0.70 (Weighted)
Accuracy	**73.39%**		
AUC	**0.8281**		

Table 5.5: RFE Performance Metrics for Hot Weather Prediction

Metric	Class 0 (Not Hot)	Class 1 (Hot)	Overall / Avg.
Precision	0.83	0.65	0.74 (Macro) / 0.78 (Weighted)
Recall	0.90	0.50	0.70 (Macro) / 0.79 (Weighted)
F1-Score	0.86	0.57	0.71 (Macro) / 0.78 (Weighted)
Accuracy	**79.03%**		
AUC	**0.8539**		

Chapter 6. Conclusion Future Scope

6.1 Findings and Observations

- **Energy Consumption Trends:**
  - The global energy consumption has exhibited a steady upward trajectory over the years, driven by economic growth, population increase, and industrialization. This growth, however, is not uniform across all energy sources, with fossil fuels continuing to dominate the energy mix.
  - Renewable energy sources, particularly solar and wind, have seen significant growth, but their adoption is still lagging compared to traditional fossil fuels, especially in developing regions.
- **Seasonal Variations:**
  - Energy consumption patterns reveal strong seasonal trends, with higher demand during colder months (for heating) and warmer months (for cooling). These seasonal fluctuations are more pronounced in regions with extreme weather conditions.
  - Seasonal energy production, particularly from renewables like hydroelectric power, also follows distinct patterns, with certain regions experiencing a drop in energy availability during dry seasons or periods of low wind activity.
- **Fossil Fuels vs. Renewable Energy:**
  - Fossil fuels (coal, oil, and natural gas) still make up the majority of global energy production, but their share is slowly decreasing as renewable energy sources gain ground. This trend is more evident in developed countries that have made strides toward cleaner energy sources.
  - Despite the growing share of renewable energy, fossil fuels continue to account for a significant portion of energy consumption, especially in industrial sectors and countries with limited access to renewable energy infrastructure.
- **Energy Imports and Exports:**
  - Energy trade (imports and exports) plays a crucial role in balancing regional energy deficits and surpluses. Countries with low energy production capacity often rely heavily on imports, while nations with abundant energy resources (e.g., oil and natural gas) serve as major exporters.

- The trend towards energy independence is rising, as countries invest in renewable energy and energy storage systems to reduce their reliance on imported fossil fuels.

- **Net Energy Imports:**
  - Countries with low domestic energy production, particularly in Europe and parts of Asia, show high levels of net imports. However, net energy imports are steadily decreasing in regions with rapid adoption of renewable energy and improved energy efficiency.
  - The trend toward self-sufficiency is expected to continue as technology advances and renewable energy infrastructure expands.

- **Sector-Specific Energy Consumption:**
  - The industrial sector remains the largest consumer of energy, particularly fossil fuels, followed by the transportation and residential sectors.
  - There is a notable shift in some regions toward electrification of transportation and industry, which could have significant long-term implications for energy consumption patterns.

- **Renewable Energy Adoption:**
  - While renewable energy adoption is increasing, it is still insufficient to meet the global energy demands fully. The share of renewables in the energy mix is growing at a faster pace in countries with robust policy frameworks and investment in renewable infrastructure.
  - Countries with abundant renewable resources, like solar and wind, are increasingly focusing on harnessing these sources, yet challenges like intermittency and grid integration continue to hinder their widespread use.

- **Data Quality and Insights:**
  - The data quality is generally high, but some inconsistencies, particularly in historical datasets and in certain regions with less transparency, were observed. These discrepancies highlight the need for better data collection and standardization practices across energy sectors.
  - Missing values in some countries' energy production and consumption records impacted the analysis, requiring thoughtful interpolation techniques to maintain analysis integrity.

- **Energy Efficiency and Future Projections:**
  - Energy efficiency measures, such as advancements in grid infrastructure, energy storage technologies, and smart meters, are likely to have a significant impact on future energy consumption patterns.
  - Future energy consumption will be heavily influenced by the global transition towards cleaner energy, innovations in renewable energy technology, and shifts in energy policy at the national and international levels.

- **Modeling and Forecasting Insights:**
  - Predictive models, including XGBRegressor and XGBoost Classifier, demonstrated strong performance in forecasting energy consumption trends. These models outperformed traditional methods like linear regression, indicating the effectiveness of machine learning in capturing non-linear relationships in energy datasets.

- The feature engineering process, particularly the use of seasonality and sector-specific features, improved model accuracy, making predictions more reliable and actionable.

These findings underscore the dynamic nature of global energy consumption and the ongoing transition towards more sustainable and efficient energy systems. The adoption of renewable energy, coupled with technological advancements, presents both challenges and opportunities for the future energy landscape.

6.2 Challenges

- **Data Inconsistencies:**
  - Despite the availability of extensive global energy consumption data, inconsistencies in data collection and reporting across different countries pose a significant challenge. Inaccurate or missing data, particularly from developing regions, can hinder comprehensive analysis and lead to skewed conclusions.
  - Historical data for energy consumption often lacks standardization, making it difficult to compare data across different time periods and regions.
- **Intermittency of Renewable Energy:**
  - Renewable energy sources, such as solar and wind, are subject to intermittent availability, making it challenging to ensure a reliable and continuous energy supply. This intermittency requires investment in energy storage technologies and grid infrastructure to maintain stability.
  - In regions where renewable energy production is highly dependent on weather conditions, there can be significant fluctuations in energy availability, creating challenges for balancing supply and demand.
- **Energy Storage and Grid Integration:**
  - The integration of renewable energy into existing energy grids requires advanced energy storage systems and grid modernization. Current storage solutions (e.g., batteries) are still limited in terms of capacity and cost-effectiveness, especially for large-scale applications.
  - Grid infrastructure in many regions is outdated and not designed to handle the dynamic and decentralized nature of renewable energy sources, creating challenges for efficient distribution and consumption.
- **Political and Regulatory Barriers:**
  - The adoption of renewable energy often faces political and regulatory challenges, especially in countries with entrenched fossil fuel industries. Subsidies and policy frameworks that favor fossil fuels may hinder the transition to cleaner energy sources.
  - Inconsistent and fragmented energy policies across regions also create challenges for international cooperation on energy production, trade, and climate change mitigation.
- **Technological and Financial Barriers:**

  - The transition to renewable energy requires substantial investments in research and development, particularly for emerging technologies like carbon capture, storage, and advanced nuclear power. Many countries face financial constraints in making these investments.
  - In developing regions, limited access to advanced technology and capital often slows the adoption of cleaner energy technologies, leaving these regions reliant on traditional fossil fuels.

- **Energy Efficiency and Demand Management:**
  - While technological advancements in energy efficiency are progressing, the pace of adoption is slow in many sectors. Industries and households may not always prioritize energy efficiency due to cost concerns or lack of awareness.
  - Effective demand-side management, such as incentivizing consumers to shift their energy use to off-peak hours, requires significant behavioral change and government intervention. Encouraging energy-saving practices across sectors is still a major challenge.

- **Social and Economic Impacts:**
  - The shift towards renewable energy and energy efficiency can have significant social and economic implications. The potential for job losses in the fossil fuel industry, combined with the need for retraining workers, presents challenges for policymakers in managing the transition.
  - The economic costs of shifting to renewable energy, including subsidies, infrastructure investments, and technological development, may strain government budgets, particularly in countries with lower GDPs or those heavily reliant on fossil fuel exports.

- **Global Coordination and Equity:**
  - Achieving global sustainability goals requires international coordination on energy consumption, climate change mitigation, and resource sharing. However, there are significant differences in the priorities and capabilities of developed and developing countries, which may complicate efforts to achieve equity in energy access and emissions reduction.
  - Ensuring that the benefits of renewable energy and energy access are shared equitably across regions and populations remains a key challenge. Developing countries with large populations may struggle to meet energy demands without exacerbating inequality.

- **Uncertainty in Future Energy Demands:**
  - Predicting future energy demands is inherently uncertain due to unpredictable factors such as economic shifts, technological breakthroughs, and changing consumer behaviors. This uncertainty complicates long-term energy planning and infrastructure investments.
  - Additionally, the global focus on decarbonization and the rapid pace of technological change make it difficult to forecast the role of specific energy sources (renewable or fossil) in the future energy mix.

These challenges highlight the complexities associated with transitioning to a more sustainable and efficient global energy system. Addressing these obstacles requires coordinated efforts across political, technological, and social domains to ensure a smooth and equitable transition toward cleaner energy.

6.3 Future Scope

The future scope of energy consumption analysis and the transition towards a more sustainable energy system is vast and encompasses several critical areas of development and innovation. The following points outline some key directions for future research and advancements in energy analysis:

- **Advancements in Renewable Energy Technologies:**
  - The development of more efficient and cost-effective renewable energy technologies such as solar, wind, geothermal, and ocean energy will play a central role in meeting future energy demands. Research into next-generation photovoltaic materials, high-efficiency wind turbines, and advanced geothermal systems will be essential.
  - Innovations in energy storage, such as solid-state batteries and other next-generation energy storage solutions, will help address the intermittency issues of renewable energy and enable a more stable and reliable grid.
- **Smart Grids and Energy Management Systems:**
  - The implementation of smart grids, which use real-time data to monitor and manage energy flows, will be crucial for integrating distributed energy sources, optimizing energy use, and improving grid stability. The future will likely see the proliferation of autonomous energy systems that can dynamically adjust to demand and supply fluctuations.
  - The development of advanced energy management systems that allow for real-time monitoring and optimization of energy use at the household, industrial, and national levels will improve efficiency and reduce waste.
- **Energy Efficiency and Conservation:**
  - Continued progress in energy-efficient technologies across sectors—such as transportation, buildings, and industrial processes—will be essential. The adoption of electric vehicles, energy-efficient appliances, and smart building systems can reduce energy demand significantly.
  - Innovations in energy-efficient manufacturing and industrial processes, such as circular economy practices, will contribute to reducing energy consumption and environmental impact.
- **Decarbonization and Carbon Capture Technologies:**
  - Achieving net-zero emissions will require breakthroughs in carbon capture, utilization, and storage (CCUS) technologies. The future will likely see the scaling up of these technologies to reduce the carbon footprint of existing fossil fuel-based infrastructure.
  - Research into alternative fuels, such as hydrogen and biofuels, will also play a role in decarbonizing hard-to-electrify sectors, such as heavy industry and long-haul transportation.
- **Artificial Intelligence and Machine Learning in Energy Forecasting:**

- The application of artificial intelligence (AI) and machine learning (ML) in energy forecasting, optimization, and demand response will continue to evolve. These technologies can enhance predictive accuracy, optimize energy distribution, and improve the management of renewable energy resources.
- Machine learning models can be used to forecast energy demand more accurately and develop smarter energy management solutions, helping to reduce waste and improve grid efficiency.

- **Integration of Emerging Energy Sources:**
  - The future will likely witness increased exploration of emerging energy sources, such as tidal energy, space-based solar power, and advanced nuclear technologies. These innovations could significantly diversify the global energy mix and reduce reliance on traditional fossil fuels.
  - The integration of decentralized energy sources and community-driven power generation models will become more prominent, empowering local communities to generate and manage their own energy.
- **Global Energy Cooperation and Policy Frameworks:**
  - Global cooperation on energy transition and sustainability will be essential in addressing the challenges posed by climate change. Future international agreements will need to ensure that developing countries receive the necessary support in terms of technology, financing, and expertise to transition to clean energy systems.
  - The establishment of robust and unified policy frameworks for clean energy investments, carbon pricing, and emissions reduction will be crucial to driving global action on climate change.
- **Public Awareness and Education:**
  - Increased public awareness and education on the importance of energy conservation, renewable energy adoption, and sustainable practices will be critical to driving behavioral change at the consumer level. Incentivizing energy-saving behaviors through policy, education, and economic measures will promote wider participation in the energy transition.
- **Sustainability in Energy Supply Chains:**
  - Future energy systems will need to ensure sustainability across the entire supply chain, from resource extraction and manufacturing to energy generation and distribution. Circular economy principles will play a crucial role in reducing waste and improving the sustainability of energy systems.
  - Advancements in the recycling and reuse of materials used in renewable energy technologies, such as solar panels and batteries, will be key to reducing the environmental footprint of energy systems.
- **Localized Energy Solutions:**

- In the coming years, localized energy solutions will become more prominent, especially in rural and off-grid areas. Microgrids and decentralized energy systems can provide communities with greater energy security and reduce reliance on centralized grid infrastructure.
- The development of energy solutions tailored to specific geographical and cultural contexts will help meet the unique needs of diverse populations and ensure equitable access to clean energy.

In conclusion, the future of energy consumption and sustainability will be shaped by technological innovation, improved energy management, and greater global collaboration. The shift towards renewable energy, the adoption of energy-efficient practices, and the integration of new technologies will be central to achieving a sustainable and equitable energy future. As challenges are addressed, opportunities will arise to create a more resilient, efficient, and environmentally friendly global energy system.

Group Contribution

Dev Vyas

- Data Analysis on Monthly and Annual Energy Consumption
- Focused on Multivariate Analysis and Model Fitting
- Latex Report

Dharmi Patel

- Data Analysis on Monthly and Annual Energy Consumption
- Focused on Bivariate Analysis and Model Fitting
- Latex Report

Preksha Shah

- Data Analysis on Monthly and Annual Energy Consumption
- Focused on Univariate Analysis and Model Fitting
- Latex Report

Short Bio

1. **Dev Vyas**is a passionate software engineer with a strong interest in data analysis and web development. He is currently pursuing a BTech in ICT with a minor in Computational Science at DAIICT and has worked on various projects involving optimization and the MERN stack.

Throughout his academic journey, Dev has been actively involved in extracurricular activities, serving as the Deputy Convener of the Cubing Club and a Core Member of the Film Club. He is proficient in programming languages such as Python, JavaScript, and C++, and enjoys solving complex problems related to software engineering and data analysis.

In addition to his technical skills, Dev is an excellent communicator and a collaborative team player. He thrives in team environments, working efficiently to deliver quality solutions within deadlines.

Outside of his academic and technical pursuits, Dev is an avid tennis player who enjoys staying active and engaging in sports.

2. **Dharmi Patel** is a passionate software engineer with expertise in web development, machine learning and research. She is currently pursuing a BTech in ICT with a minor in Computational Science at DAIICT and has worked on various projects involving front-end, back-end development, machine learning, optimization and research.

Throughout her career, Dharmi has actively contributed to open-source projects and has a keen interest in exploring new technologies. She is proficient in programming languages such as C++, Python, JavaScript, and Java, and enjoys tackling challenging problems in software development.

In addition to her technical skills, Dharmi is an effective team player with excellent communication skills. She enjoys collaborating with colleagues and has a track record of delivering high-quality solutions within tight deadlines.

Outside of work, Dharmi enjoys swimmming, dancing, and collaborating at events for software development.

3. **Preksha Shah** is a driven technology enthusiast pursuing a Bachelor's degree in Mathematics and Computing at DA-IICT. With a strong technical foundation, she is passionate about research, web development, and algorithmic problem-solving.

Preksha has worked on diverse projects, ranging from developing interactive web applications to Data Analysis and Human-Computer Interaction. She is proficient in C++, Python, Java, and JavaScript, and adept with tools like MERN STACK and Figma. Her ability to tackle complex challenges and design user-focused solutions underscores her technical expertise.

In addition to her technical pursuits, Preksha channels her creativity as a graphic designer. She has successfully managed technical and cultural events, showcasing her leadership and teamwork abilities. She is the Convener of Muse Club and a core member of Team i.Fest.

Outside of academics, Preksha enjoys sketching, swimming and exploring innovative ideas that merge technology with creativity. She aspires to leverage her skills to build impactful solutions and excel in software engineering.

References

[1] Dataset
URL: https://catalog.data.gov/dataset/monthly-and-annual-energy-consumption-by-sector