**GLOBAL RENEWABLE ENERGY TRANSITION ANALYSIS AND FORECAST**

**Introduction**

Energy is like the lifeblood that courses through the veins of modern infrastructure. It serves as the powerhouse for industries, fuels the engines of transportation, and illuminates communities. The very quality and accessibility of energy resources are intricately woven into the tapestry of infrastructure development. As we confront a global infrastructure deficit, which is further complicated by pressing environmental concerns, a profound transformation becomes not just desirable but imperative. This project delves deep into the intricate dance between energy and infrastructure, all in harmony with the overarching theme of **"Infrastructure Deficit: AI's Transformative role in Planning."**

**Problem Statement**

The problem at hand is the urgent need to transition from non-renewable to renewable energy sources to address infrastructure deficits sustainably. Non-renewable energy exacerbates environmental challenges and hampers infrastructure growth. It's critical to understand renewable energy trends, barriers, and the path to eliminating non-renewables.

**Aim of the Project**

The project aims to analyze global energy data and offer data-driven insights for accelerating the transition to renewable energy. It seeks to facilitate informed infrastructure planning that aligns with sustainability, reducing our dependence on non-renewable sources.

**Data Source**

The dataset was obtained from Kaggle, to view the dataset click [here](https://www.kaggle.com/datasets/jamesvandenberg/renewable-power-generation?select=Continent_Consumption_TWH.csv)

**Dataset description**

1. **Renewable Total Power Generation**: This dataset ranks energy sources in annual energy consumption, with tidal waves leading, followed by hydro, wind, and biofuel. Renewable waste and geothermal sources contribute less, providing an overview of their importance.
2. **Non Renewables Total Power Generation**:Focusing on non-renewable sources from 1990 to 2020, it covers Coal, Natural Gas, Nuclear, Oil, and more. It reveals how different countries rely on these sources and their potential environmental impacts.
3. **Top 20 Countries Power Generation**: Highlights the top 20 countries in renewable energy adoption, focusing on Hydro, Biofuel, Solar PV, and Geothermal. This dataset shows their commitment to sustainable energy and global trends in adopting cleaner alternatives.
4. **Renewable Power Generation (1997-2017)**: Tracks renewable energy's growth between 1990 and 2017 in categories like Hydro, Biofuel, Solar PV, and Geothermal. It demonstrates the increasing role of renewables and their environmental benefits.
5. **Country\_Consumption\_TWH**: Spans electricity consumption from 1990 to 2020 for numerous countries. It offers insights into global energy consumption trends, revealing shifts from non-renewable to renewable sources.6. Continent\_Consumption\_TWH: Covers continental electricity consumption from 1990 to 2020, providing a comprehensive view of global energy consumption trends on a continental scale. It's essential for assessing the transition from non-renewable to renewable energy sources worldwide.

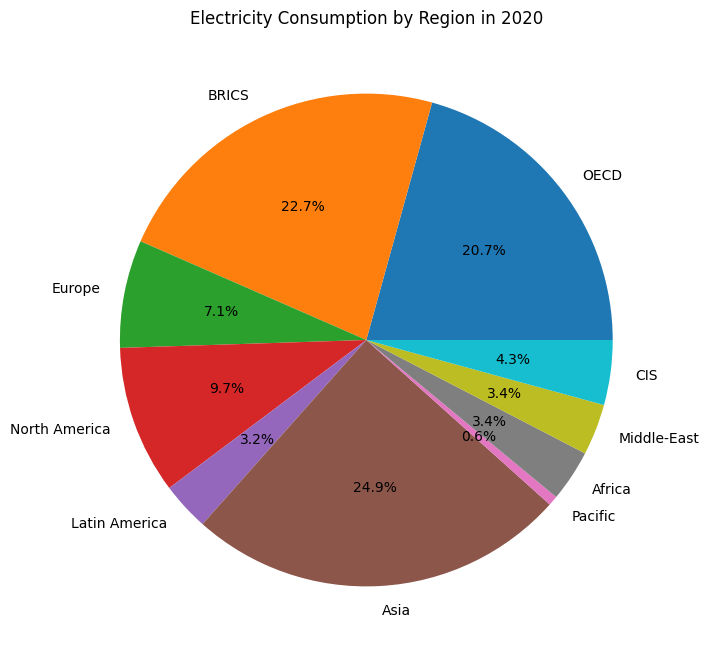
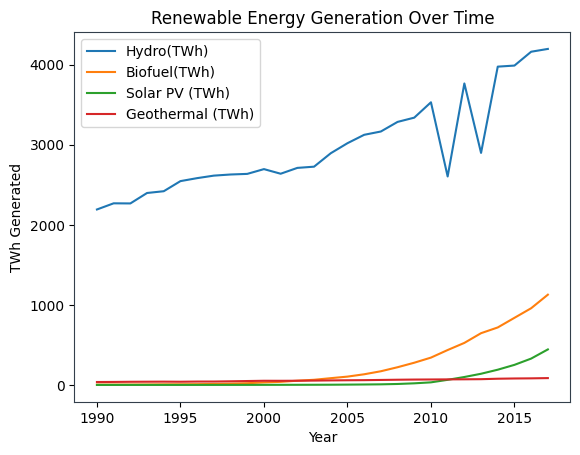
**Our Approach**

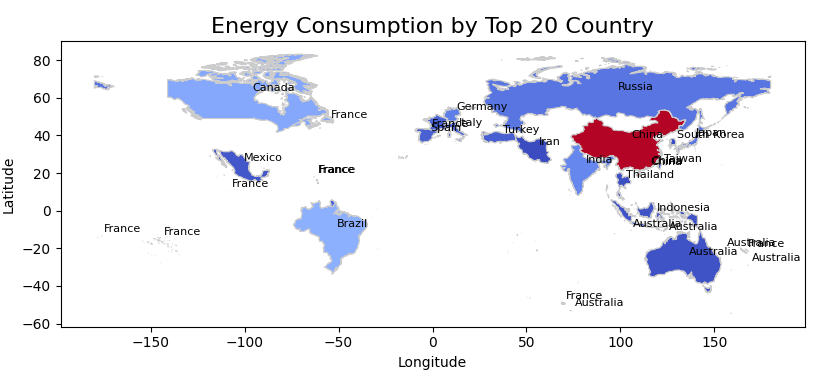
**Exploratory Data Analysis**

During the initial phase of the exploratory data analysis (EDA), univariate analyses were conducted to gain a comprehensive understanding of each column as a feature contributing to the overall narrative. Subsequently, in the second phase, bivariate and multivariate analyses were performed, examining relationships between two or more distinct datasets, allowing for a deeper and more insightful exploration of the narrative encapsulated within the table.

Some of the key data visualizations done includes:

* Contribution of renewable energy sources to the total Energy Consumption value for each country.
* Top countries generating renewable energy (TWH).
* Top countries in various energy adoption
* Renewable energy generation over Time (Renewable Energy trends analysis for some selected countries).
* Energy Consumption in various countries, continents and regions of the world.
* Yearly Energy consumption growth.
* Geospatial analysis of the top 20 countries in Energy growth





**DATA CLEANING AND PREPROCESSING**

The transition from yearly to daily granularity in energy consumption data represents a pivotal step in our data analysis process, ushering in a wealth of opportunities for in-depth exploration. This granulation process involves breaking down the traditionally aggregated annual values into daily increments. This approach empowers us to discern and record even the most subtle day-to-day variations in energy consumption patterns. By doing so, we not only enrich our understanding of these consumption patterns but also equip ourselves with the tools necessary for producing more accurate and robust time series forecasts.

The shift to daily granularity results in a dataset with a significantly increased volume of data points. This expanded dataset, teeming with daily observations, offers a profound advantage. It allows our predictive models to finely attune themselves to the intricacies of daily fluctuations in energy usage. As a result, our forecasting models become more adept at capturing the nuances of daily changes. This precision enhances the accuracy of our predictions, ultimately leading to superior decision-making within the energy sector.

Following the granulation process, we turn our attention to the datetime column. We parse this column to convert it into the appropriate date type format, ensuring uniformity and consistency in our dataset. Any gaps in the data, arising from missing values, are meticulously addressed using sound statistical techniques. This meticulous data imputation process guarantees data completeness and integrity.

Moreover, we delve into a comprehensive evaluation of the datetime columns' quality. This assessment encompasses an exploration of the chronology, seasonality, and orderliness inherent in the dataset. By thoroughly examining these temporal aspects, we gain deeper insights into the time-dependent dynamics of energy consumption.

In addition to these steps, we employ a suite of statistical tests and techniques to unlock hidden insights within the data. One such test is the *Augmented Dickey-Fuller (ADF*) test, which aids in assessing the stationarity of time series data. Furthermore, we apply log transformations and other time series analysis methods to extract valuable patterns and trends, further enhancing the richness of our dataset and our ability to make informed decisions within the energy sector.

**FEATURE ENGINEERING**

Our journey through feature engineering commences with the extraction of new seasonality attributes. These include essential temporal elements such as year, month, day, day of the year, week of the year, quarter, and the prevailing season. These newly minted features, after a meticulous study, reveal their cyclical nature. To imbue our models with the knowledge of their cyclical essence, we employ sine and cosine transformations, enabling our algorithms to grasp the periodicity inherent in these features.

One of the cornerstones of our time series analysis is the decomposition of time series data. This intricate process involves partitioning the time series into three fundamental components: the level, the trend, and the seasonality, along with the residual noise. Such a decomposition offers a powerful abstract model that aids in our comprehension of time series data and enhances our capabilities in the realm of time series analysis and forecasting.

To undertake this decomposition, we turn to the *seasonal\_decompose()* function within the esteemed statsmodels library. With this tool at our disposal, we meticulously dissect our time series data, unveiling its inherent structures and patterns.

Furthermore, in the quest for deeper insights and understanding, we apply a *shift()*, or *lag*, to each variable. This operation permits us to explore the correlations and relationships between these variables. By calculating these shifts and assessing their effects, we gain a more profound comprehension of the dynamic interplay within our dataset, ultimately strengthening our ability to make informed decisions in the realm of time series analysis and forecasting.

**MODELING**

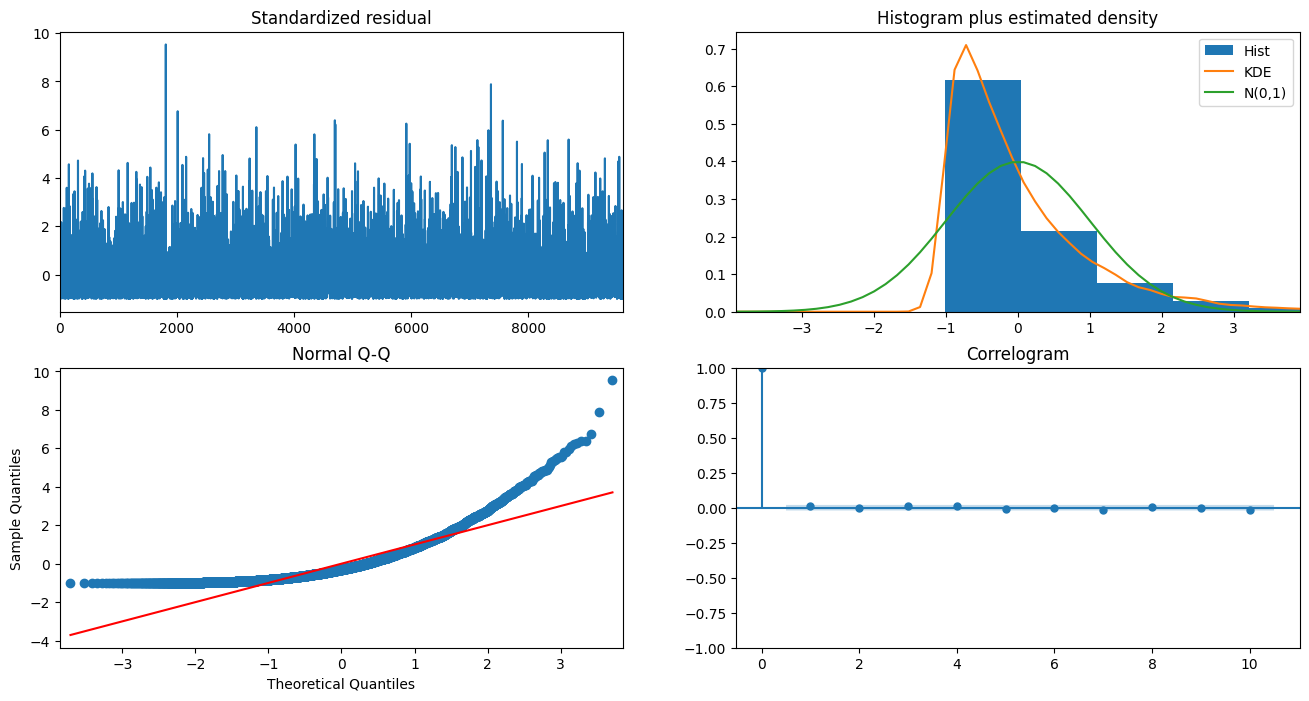
We applied various methods to fit and train the model, starting with Model\_1. For this model, we employed `auto\_arima`, a popular time series forecasting model known for its speed, low memory usage, and auto-fitting capabilities.

The dataset was divided into training and validation sets in an 85:15 ratio. We fine-tuned the hyperparameters, initializing them with the following values:

*start\_p=1 start\_q=1 test='adf' max\_p=3 max\_q=3 m=1 d=None seasonal=False start\_P=0 D=0 trace=True error\_action='ignore' suppress\_warnings=True stepwise=True*

After fitting the model, it selected the best model as **ARIMA(0,1,0)(0,0,0)[0]** with an intercept. The total fit time for this model was **11.891** seconds.

In addition, we visualized the model's diagnostic plots, which can provide insights into its performance and any areas for improvement.



Examining the diagnostic plots some of this were found to be true:

**Top left**: The residual errors appear to fluctuate around a mean of zero and maintain a fairly consistent variance between 0 and 2, indicating a reasonably constant level of error.

**Top right**: The density plot suggests that the distribution of the residual errors is approximately normal with a mean close to zero, indicating a good fit.

**Bottom left**: Most of the blue dots do not align closely with the red line, implying that the distribution is slightly skewed, albeit not significantly.

**Bottom right**: The Correlogram, also known as the ACF (AutoCorrelation Function) plot, reveals that the residual errors exhibit some degree of autocorrelation, suggesting that there might be patterns or dependencies in the data that the model has not fully captured. This information is important for further model improvement.

**PERFORMANCE EVALUATION**

To assess the performance of the individual country models, we calculated three key evaluation metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics provided insights into the model's accuracy and allowed for the identification of areas for improvement.

In the pursuit of enhancing model performance, several key improvements were implemented:

1. **Maximizing Optimal Intercepts**: Efforts were made to optimize the model's intercepts, fine-tuning this parameter to improve its ability to capture the data's underlying patterns and trends.

2. **Dropping Redundant and Low-Importance Columns**: Columns that contributed minimally to the predictive power of the model were identified and removed. This not only streamlined the model but also reduced the risk of overfitting, potentially leading to improved accuracy.

3. **Reducing Noise techniques**: Preprocessing techniques that introduced noise or extraneous information were minimized, focusing the model on the most relevant and significant features. This refinement process aimed to enhance the model's ability to make accurate predictions by reducing unnecessary complexity.

These enhancements collectively aimed to boost the model's performance and increase the accuracy of energy consumption forecasts for each country.

**SAVING THE MODEL**

In our univariate time series modeling approach, each country's energy consumption data was treated as a separate model. This means that we trained a distinct model for each country using its corresponding energy consumption column and datetime intervals. Once trained, each model was saved as a pickle file with a .pkl extension, ensuring that the models are preserved and can be easily accessed and utilized for future analyses and predictions.

**FORECASTING (2020-2030)**

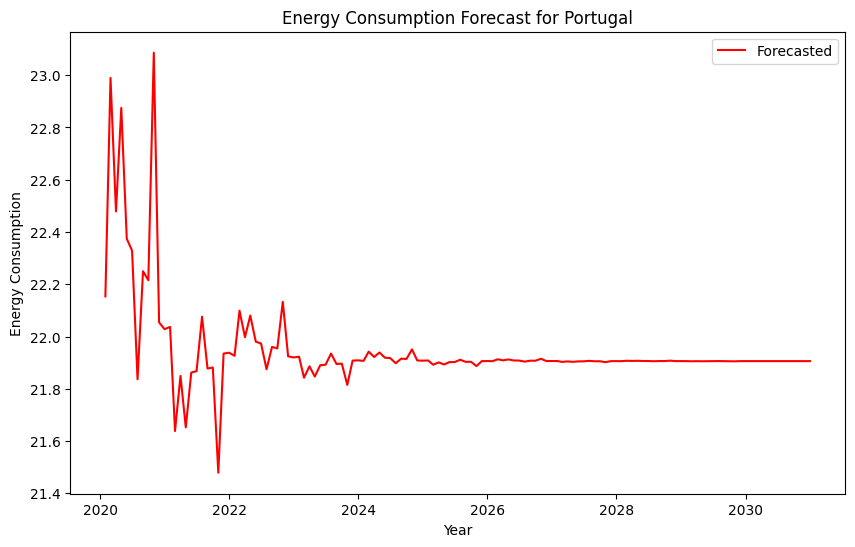
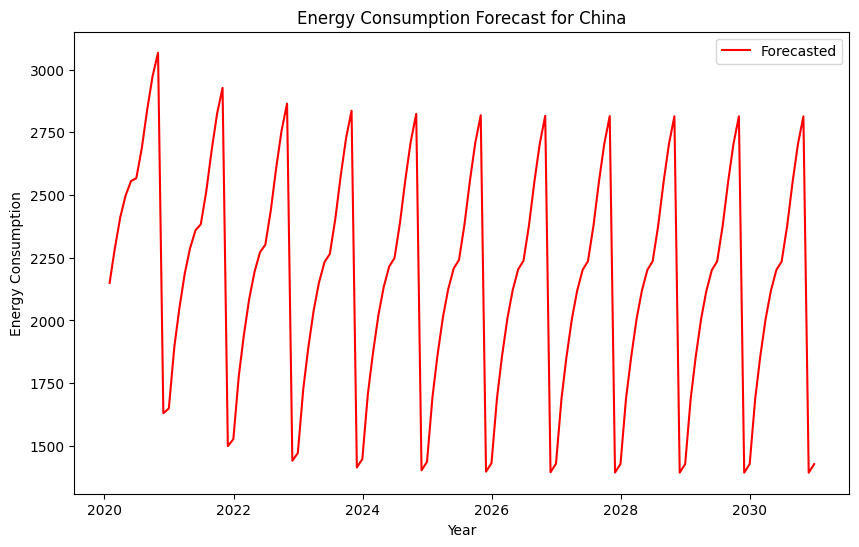
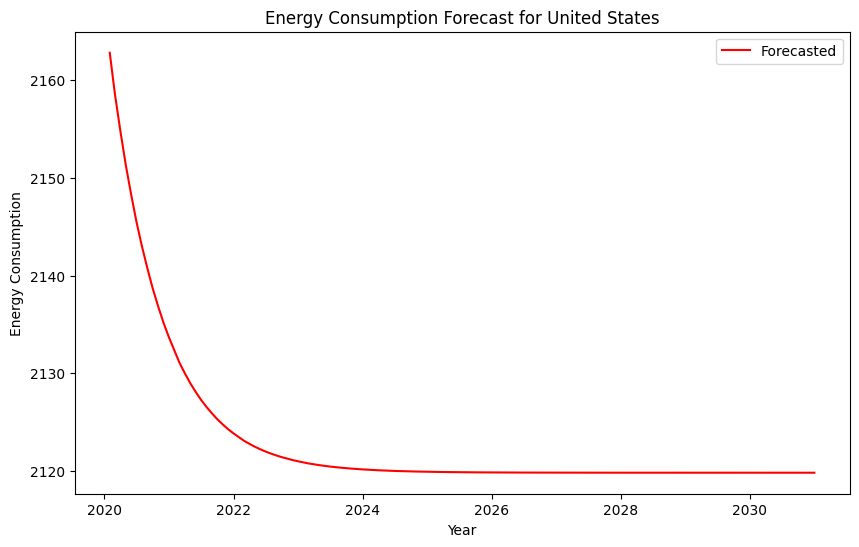
One of the primary objectives of this project is to visualize and determine the energy consumption trends for each country. To achieve this, models were developed for each country, and these models were used to forecast energy consumption for a period of 10 years, spanning from 2020 to 2030. This extensive forecasting exercise serves several important purposes:

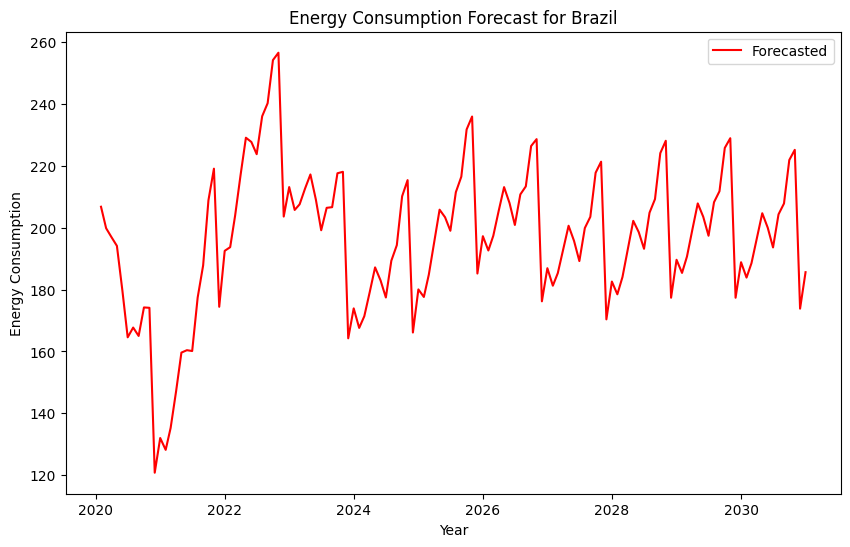
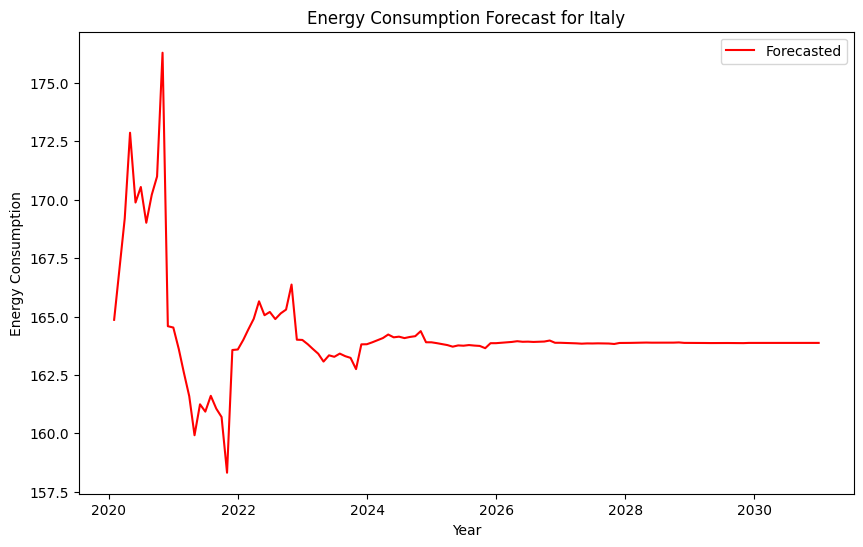
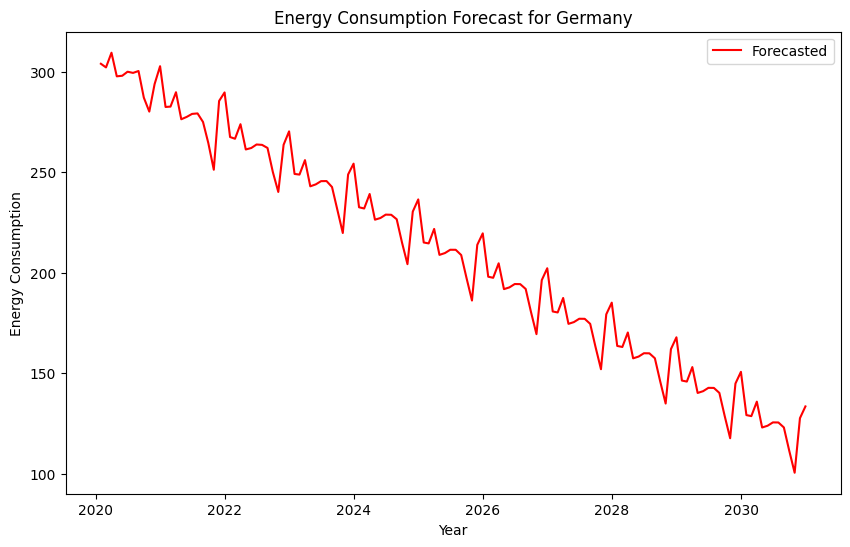
1. **Setting Energy Goals**: By projecting energy consumption over a 10-year period, it becomes possible to establish clear energy goals for each country. These goals can be aligned with sustainability targets and energy efficiency objectives.

2. **Expectations Management**: The forecasts provide a basis for managing expectations regarding energy consumption. Understanding the likely trajectory of energy usage helps policymakers, businesses, and consumers prepare for future energy needs.

3. **Continuous Improvement**: Monitoring energy consumption trends over the coming decade allows for ongoing assessment and improvement of energy policies and practices. By identifying potential areas of growth or reduction in energy consumption, countries can adapt and refine their strategies accordingly.

The results of these 10-year energy consumption forecasts were visualized using line plots, making it easier to interpret and communicate the expected trends and patterns for each country. This visual representation is a valuable tool for decision-makers and stakeholders in the energy sector.





**CONCLUSION**

The energy consumption forecasts for several countries, including China, Brazil, Italy, Sweden, Kazakhstan, Colombia, Japan, and Taiwan, highlight the critical challenges and opportunities they face in the coming decade. These forecasts demonstrate that economic development, population growth, and rising living standards are driving increased energy demand. However, the transition to more sustainable energy sources and enhanced energy efficiency measures is crucial to mitigate the environmental and economic impacts of rising consumption. It is evident that countries pursuing net-zero scenarios show a path towards a more secure, environmentally responsible, and economically resilient future. Monitoring progress towards these targets will be essential to ensure a sustainable energy future for these nations.

**SUMMARY**  
This project is dedicated to addressing the critical nexus of energy and infrastructure, recognizing the urgent need for a transition from non-renewable to renewable energy sources. By analyzing global energy data, identifying leading countries in renewable energy adoption, forecasting the path to global renewable energy goals, and pinpointing barriers to rapid adoption, it aims to provide actionable recommendations for sustainable infrastructure development. This multifaceted approach aligns with the overarching goal of creating a more resilient, environmentally responsible, and sustainable global infrastructure landscape.

**External resources**  
GitHub: [Insight\_IQ HDSC ‘23 Premiere Project](https://github.com/Gbekoilias/Insights_IQ-Premiere-Project-HDSC-23)

Presentation: [Presentation\_Premiere\_Project\_Insight\_IQ\_HDSC ‘23](https://docs.google.com/presentation/d/1B54dB3QmDO3VRtx4LHX718F--SizLYlL5ybZvuqR-_8/edit?usp=sharing)