OPIM 5671 – Data Mining and Business Intelligence

Electricity Generation

and

Demand Forecasting

<u>Group 7</u>

EXECUTIVE SUMMARY

The energy market is an essential component of modern society, and reliable electricity generation and consumption are critical for maintaining economic growth and development. However, with the increasing demand for energy and concerns over environmental impact, it is crucial to understand and predict future trends in energy generation and consumption. This project aimed to provide insights into these trends through the analysis of electricity generation and consumption datasets. The datasets included information on electricity generated from different sources, including fossil fuels and renewable sources, as well as end-use from various sectors such as industrial, commercial, and residential.

Through the use of statistical analysis, we found that there is both seasonality and trend in all the components we analyzed. To achieve our objectives, we developed four best models for the components, including winters additive, seasonal multiplicative model + sarima (3,0,3), winters multiplicative for non-renewable, renewable, and total energy generated, respectively. The total end-use had the best model of additive seasonal with an accuracy of 97%, derived by dividing the MAPE value from 100.

Overall, our analysis indicates that energy demand is increasing over time, and there is a potential for electricity blackouts in the future. We also compared electricity generation from renewable and non-renewable sources and found that using renewable sources can help decrease carbon footprints.

In conclusion, this project has helped us better understand the difference between electricity generation and consumption, and provided insights into the energy market. We hope that our findings will help stakeholders make informed decisions that can lead to a more sustainable future.

Table of Contents

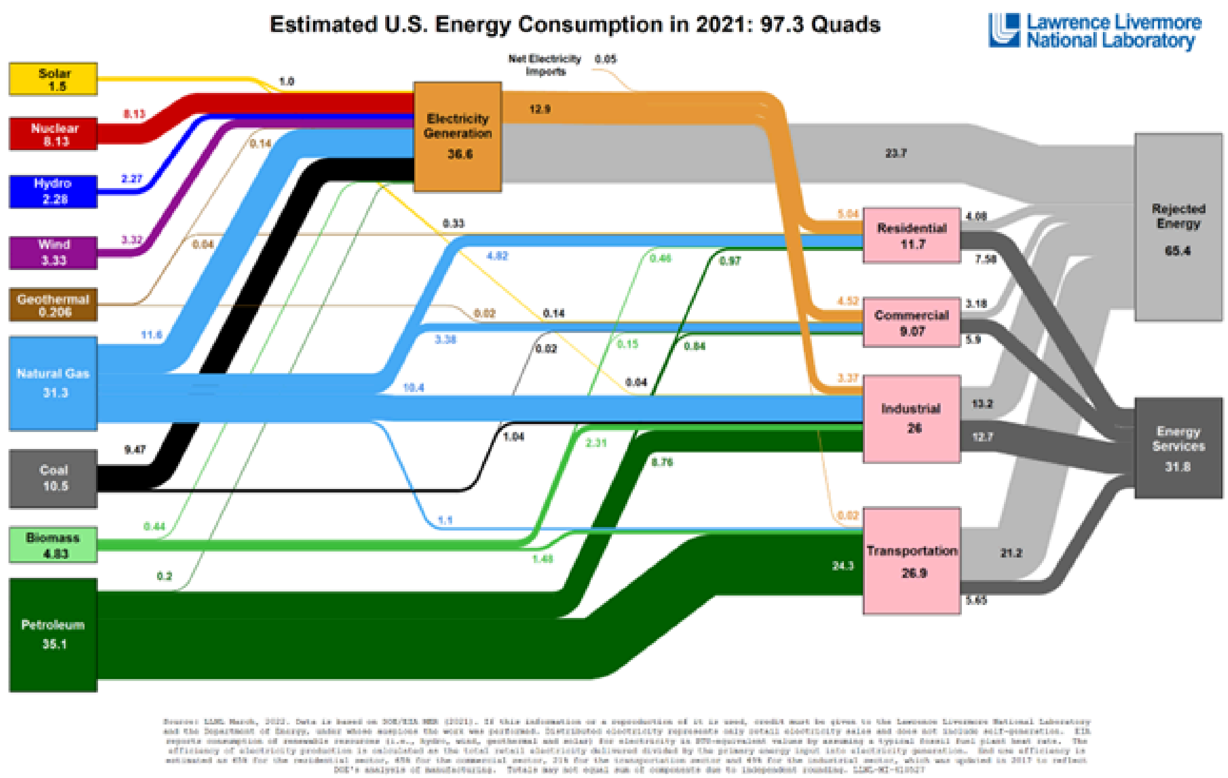
1. INTRODUCTION

Electricity generation and supply is a critical aspect of modern life, and understanding its complexities is essential for ensuring a sustainable and reliable power supply. Energy is a fundamental component of modern society, and it is necessary for powering homes, businesses, and industries. Accurate knowledge of electricity generation and supply enables energy providers to meet the demands of the consumers, maintain the stability and reliability of the energy system, and make informed decisions about investments in new energy sources and infrastructure development. Not knowing the electricity generation and supply can lead to various real-world problems, including blackouts, brownouts, and electricity shortages. These problems can occur due to a mismatch between the supply and demand for electricity, leading to disruptions in industrial processes, loss of productivity, and increased costs for consumers. Moreover, the lack of accurate information about electricity generation and supply can lead to overuse of non-renewable energy sources and over-dependence on a single source of energy, leading to environmental degradation and climate change. Renewable resources such as solar, wind, hydropower, and geothermal energy have the potential to provide a sustainable and environmentally friendly source of electricity.

1.1. Problem Statement

The problem that we are addressing in this project is the need to ensure a balance between the generation and consumption of electricity in the energy market. From the figure 1 that shows Estimated U.S. Energy Consumption in 2021 we can see that 70% of generated electricity is rejected. Rejected energy is the portion of energy that is generally wasted heat to the environment. Energy Services is the portion of energy that is put to useful applications. By accurately forecasting the demand and electricity generation, rejected energy can be reduced.

The Project also focuses on forecasting the usage of fossil fuels and renewable energy sources in the future from which we can analyze different aspects of energy usage.



Data Exploration

The Project entails an exploration of two datasets, namely energy consumption and electricity generation.

1. Energy Generation Dataset:

The electricity generation dataset covers the period from January 1973 to October 2022 and contains information on Net Energy Generation in the United States. The dataset is structured around energy generation sources such as Fossil fuels (including Coal, Petroleum, Natural Gas, and Other Gas), Renewable Energy Sources (including conventional hydroelectric power, biomass, geothermal, solar, and wind), and Nuclear Electric Power.

We have columns Electricity Generation in all sectors from Coal, Petroleum, Electricity, Natural Gas, Other Gases, Nuclear electric power, hydroelectric power, conventional hydroelectric power, wood, waste, geothermal, solar, wind, Total Renewable, Total Non-renewable, Total from all sources.

The data cleaning process ensured that there were no missing values in the dataset. As we planned to forecast renewable and nonrenewable resources usage for electricity generation, we have created two distinct columns in the dataset to be able to analyze and forecast them separately for more insights. Now the columns we opted to work on from this dataset are Non-renewable energy, Renewable energy and Total electricity generation.

Electricity Generation Dataset

Electricity Net Generation Data Link

(https://www.eia.gov/totalenergy/data/browser/index.php?tbl=T07.02A#/?f=M&start=197301&end=202210&charted=15)

This dataset consists of Net Energy Generation in the United States again spanning from January 1973 to October 2022 via energy generation sources like Fossil Fuels which includes Coal, Petroleum, Natural Gas, Other Gas; Renewable Energy Source which includes conventional hydroelectric power, biomass, Geothermal, Solar, Wind; and Nuclear Electric power.

Column	Datatype	Description
Month	Datetime	Contains Month and Year
Coal	Decimal	Unit-Million Kilo-Watt hour
Petroleum	Decimal	Unit-Million Kilo-Watt hour
Natural Gas	Decimal	Unit-Million Kilo-Watt hour
Other Gases	Decimal	Unit-Million Kilo-Watt hour
Nuclear Electric Power	Decimal	Unit-Million Kilo-Watt hour
Hydroelectric pump storage	Decimal	Unit-Million Kilo-Watt hour
Renewable Energy	Decimal	Unit-Million Kilo-Watt hour
Conventional Hydroelectric Power	Decimal	Unit-Million Kilo-Watt hour
Wood	Decimal	Unit-Million Kilo-Watt hour
Waste	Decimal	Unit-Million Kilo-Watt hour
Geothermal	Decimal	Unit-Million Kilo-Watt hour
Solar	Decimal	Unit-Million Kilo-Watt hour
Wind	Decimal	Unit-Million Kilo-Watt hour
Total	Decimal	Unit-Million Kilo-Watt hour

DATA MANIPULATION:

In the Electricity Generation dataset, we have combined the columns related to renewable energy. The columns for renewable energy sources for electricity generation are:

- Hydroelectric Pump Storage
- Renewable Energy
- Conventional Hydroelectric Power
- Wood
- Waste
- Solar
- Geothermal
- Solar
- Wind

Similarly for the Electricity Generation from non renewable sources, we have added the following:

- Coal
- Petroleum
- Natural Gas
- Other Gases
- Nuclear Electric Power

Final Dataset Description:

Column Name	Data Type
DATE	DATETIME
TotalElectricityGeneration(Million KWh)	Float
RenewableSources(Million KWh)	Float

NonRenewableSources(Million KWh)	Float

Link for transformed dataset:

Electricity Net Generation.xlsx

In the Electricity End Use data, we have used the total end use column for electricity demand.

Final Dataset Description:

Column Name	Data Type
Month	DATETIME
TotalEndUse(Million KWh)	Float

Link for transformed dataset:

ELECTRICITY_END_USE.xlsx

3. Methodology

SAS Studio is a powerful data analysis and statistical modeling tool that is widely used in the industry for various applications including forecasting.

SAS is often used for time forecasting due to its time series analysis capabilities, a wide range of statistical modes, integration with other SAS products, and ease of use. We will be using SAS studio throughout our project to determine which model suits the best for power generation time series forecasting. For this part, we will be only taking only non-renewable resources as the dependent variable.

Time Series Analysis:

Time series analysis is a statistical technique used to analyze and interpret time-dependent data. In this project, we used time series analysis to study the historical patterns in energy generation, energy consumption, energy generation through non-renewable sources, and electricity generation through renewable sources. Time series analysis helps to identify trends, seasonality, and irregular components in the data.

Forecasting:

Forecasting is the process of predicting future values based on historical data. In this project, we used forecasting to estimate future energy generation, energy consumption, energy generation through non-renewable sources, and electricity generation through renewable sources. Forecasting techniques help in decision making, planning, and budgeting.

Exponential Smoothing Models:

Exponential smoothing models are a family of time series models that use a weighted average of past observations to forecast future values. In this project, we used the Holt-Winters exponential smoothing model to forecast the energy consumption and energy generation through renewable sources. Exponential smoothing models are easy to understand, computationally efficient, and provide reasonably accurate forecasts. However, these models assume that the underlying data has a constant trend and seasonality, which may not always be the case.

ARIMA Models:

ARIMA (Autoregressive Integrated Moving Average) models are a class of time series models that can capture both trend and seasonality in the data. In this project, we used ARIMA models to forecast energy generation through non-renewable sources. ARIMA models are versatile and can handle a wide range of time series data. However, these models require a stationary time series, which may require differencing the data to remove trends and seasonality.

Seasonal ARIMA Models:

Seasonal ARIMA models are a type of ARIMA model that can handle time series data with seasonal patterns. Seasonal ARIMA models incorporate both trend and seasonality in the data, making them suitable for forecasting seasonal time series data. However, these models can be computationally intensive and may require a large amount of data to estimate the model parameters.

Hybrid Models:

Hybrid models are a combination of two or more time series models. In this project, we used a hybrid model that combined the Holt-Winters exponential smoothing model and the Seasonal ARIMA model to forecast energy generation through renewable sources. Hybrid models can improve the accuracy of forecasts by combining the strengths of different models. However, building hybrid models can be complex and time-consuming.

Machine Learning Models for Time Series:

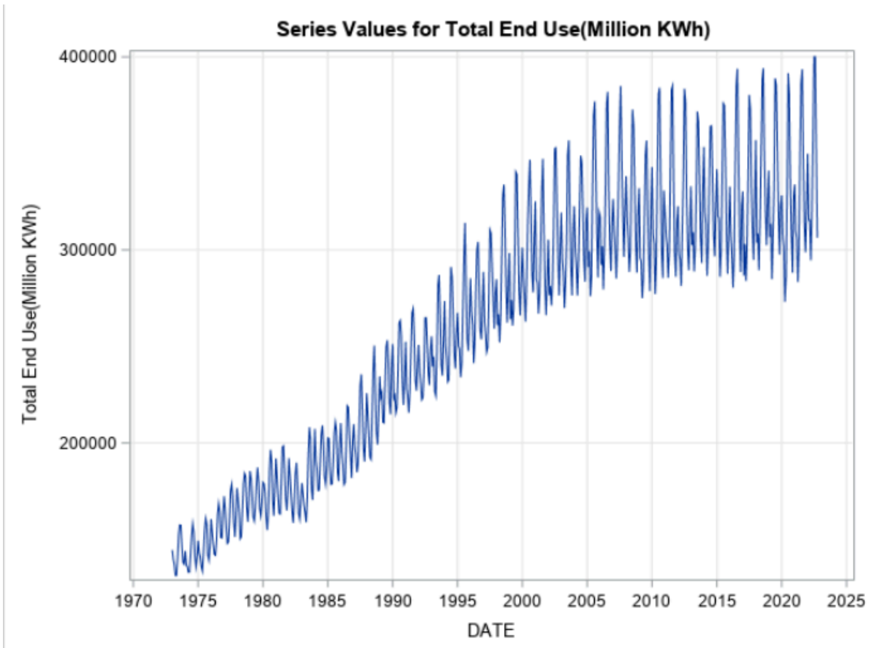
Machine learning models, such as neural networks, support vector machines, and random forests, can be used for time series forecasting. In this project, we did not use machine learning models for time series forecasting, but they can be useful for forecasting complex, non-linear time series data. Machine learning models require large amounts of data and can be computationally intensive, but they can provide highly accurate forecasts.

In time series exploration we first performed basic decomposition analysis with the required component column as dependent variable and Month as time ID. Below are the results of decomposition analysis.

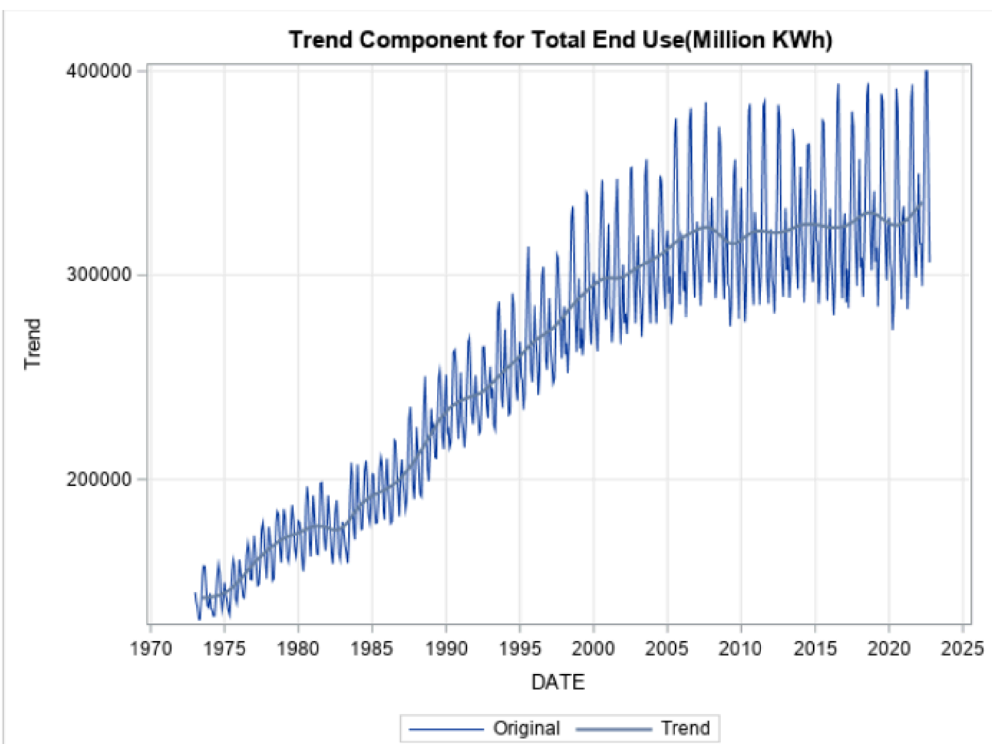
Time Series Exploration:

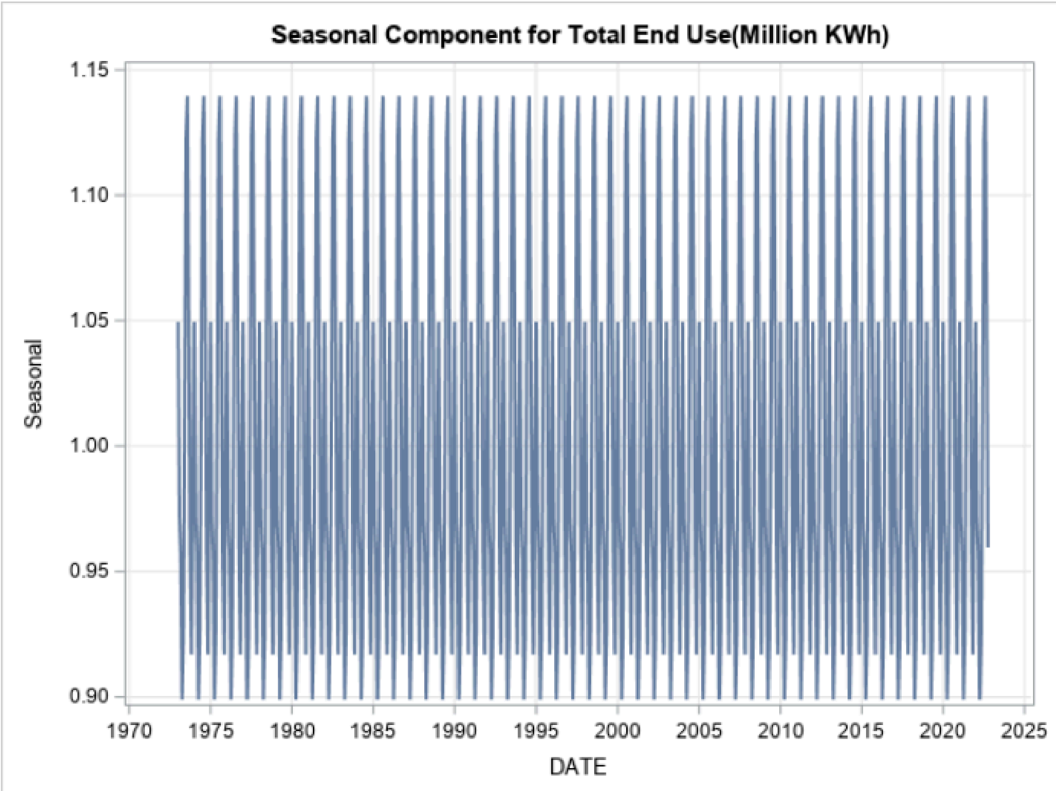
Total Electricity End Use:

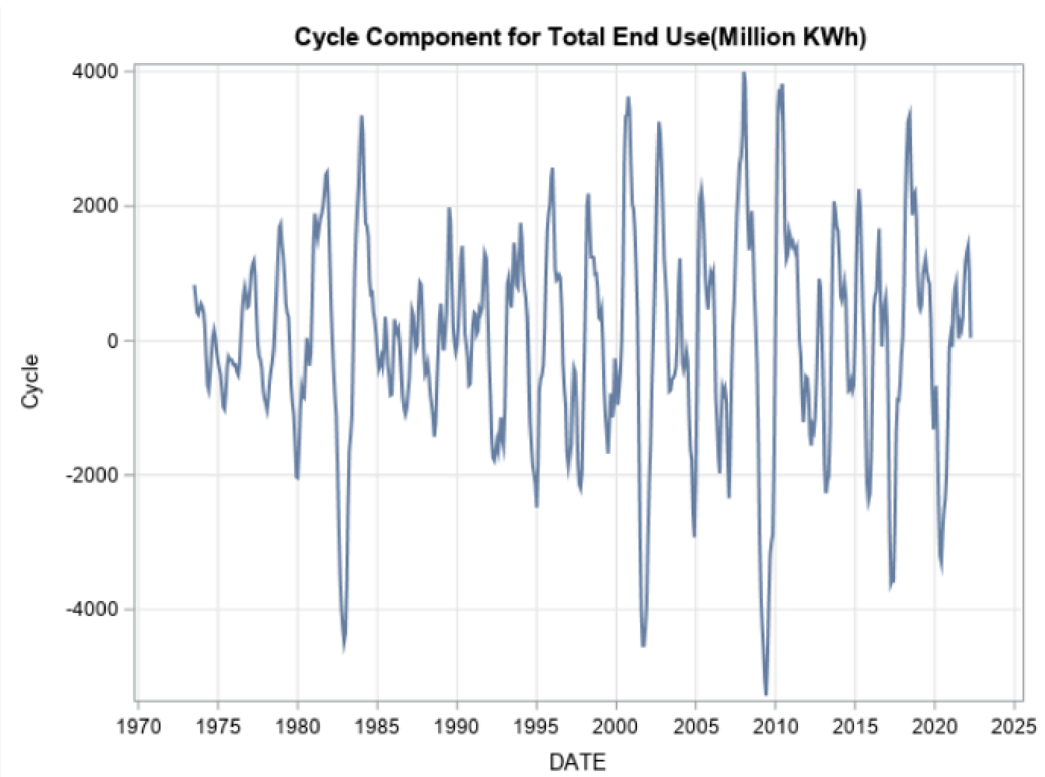
Series Graph for Total End Use:

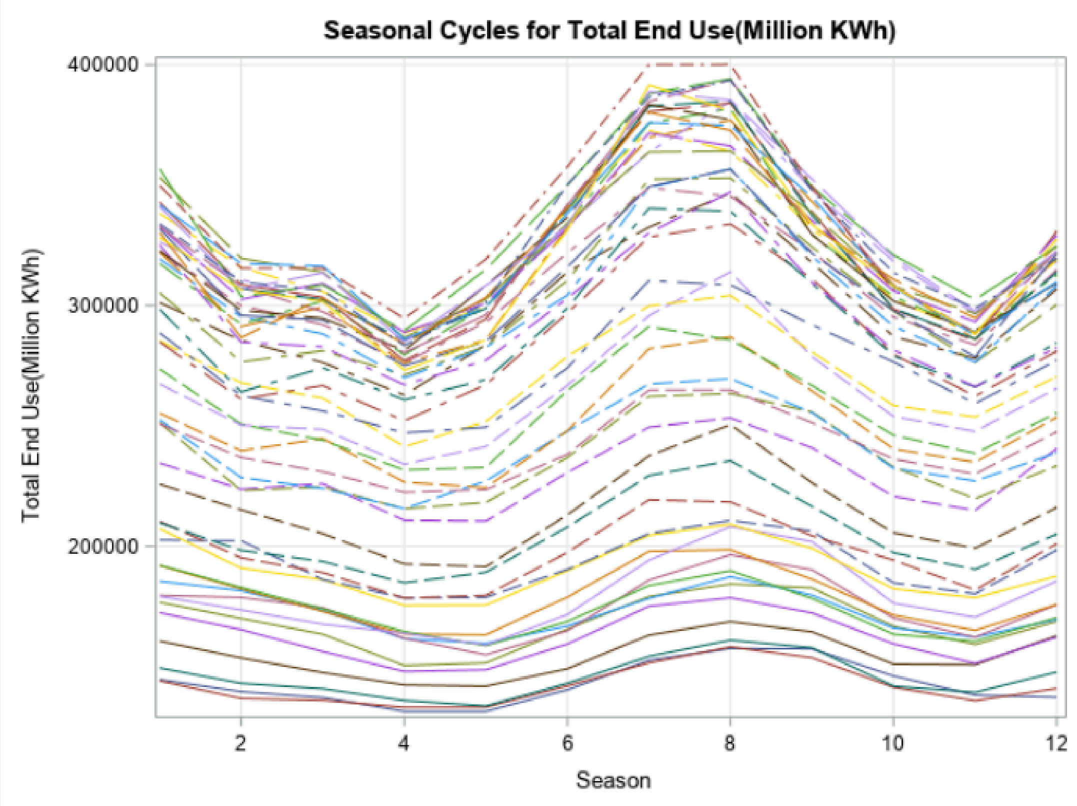


Decompositional Analysis for Total End Use:




Seasonal Component for Total End Use(Million KWh)
Seasonal
1.15
1.10
1.05
1.00
0.95
0.90
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
DATE


Cycle Component for Total End Use(Million KWh)
Cycle
4000
2000
0
-2000
-4000
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
DATE



Based on the decomposition analysis, there was a linear trend up to 2002, however the slope of the trend rapidly dropped after 2002. From the seasonal cycles graph, it is evident that the time series data exhibit seasonality.

Factors affecting the change in trend:

The increase in electricity demand from 1980 to 2002 in a linear trend can be attributed to several factors, such as population growth, urbanization, economic growth, and technological advancements. As the population grew, the demand for electricity increased due to the need for more homes, offices, and public spaces. Urbanization also played a significant role as people moved from rural areas to urban areas, leading to increased demand for electricity. Economic growth also led to increased demand for electricity as industries and businesses expanded. Finally, technological advancements, such as the rise of personal computers and the internet, also contributed to the increase in electricity demand.

However, after 2002, the slope of the linear trend decreased, indicating that the rate of growth of electricity demand has slowed down. This decrease in slope can be attributed to several factors, such as improvements in energy efficiency, changes in consumer behavior, and a shift towards renewable energy sources.

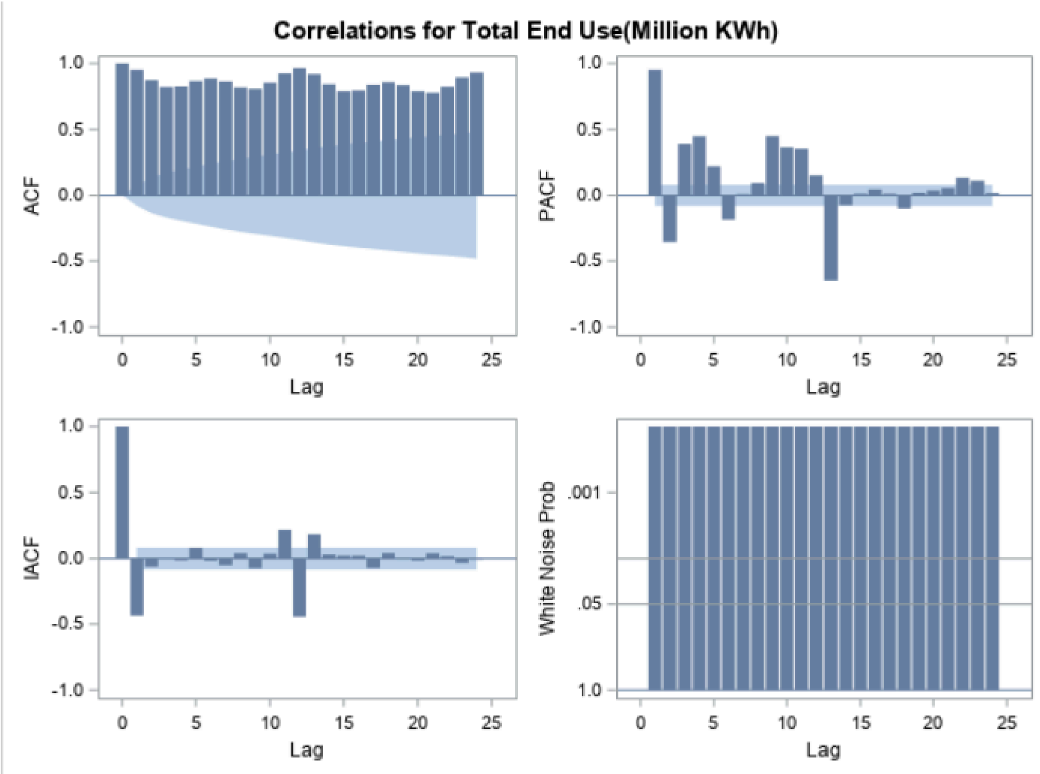
Improvements in energy efficiency have led to a reduction in electricity consumption in many areas. For example, appliances and lighting systems have become more efficient, and buildings have been designed to be more energy-efficient. These improvements have led to a decrease in the rate of growth of electricity demand.

Changes in the consumer behavior have also contributed to the decrease in the slope of the linear trend. Many consumers are now more aware of their energy consumption and are taking steps to reduce their energy usage. For example, people are turning off lights when they leave a room, using energy-efficient appliances, and using public transportation instead of driving.

Finally, there has been a shift towards renewable energy sources, such as wind and solar power, which has contributed to a reduction in the rate of growth of electricity demand. This shift towards renewable energy sources has led to a decrease in the use of non-renewable energy sources, such as coal and oil, which has resulted in a reduction in the rate of growth of electricity demand.

Correlation Analysis for Total End Use:

There are significant autocorrelations and partial correlations between the lags, as shown by the correlation plots below. The white noise tests demonstrate that the data are not distributed randomly and contain some signal.



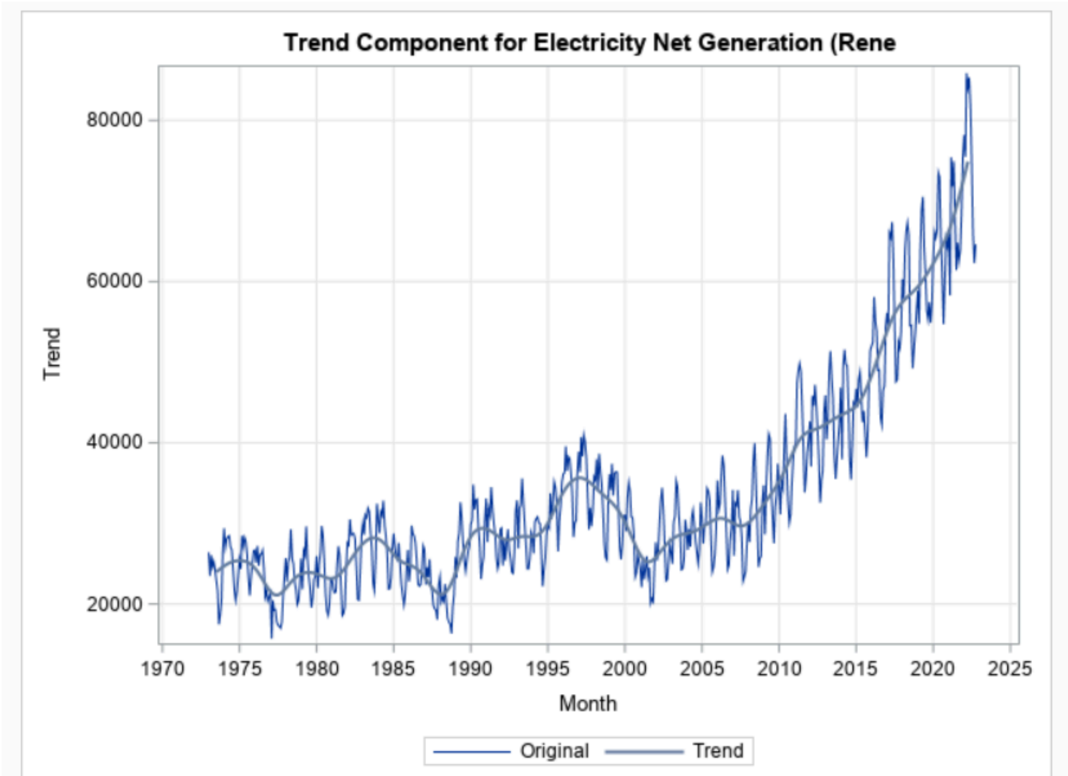
Electricity Generation From Renewable Energy Sources:

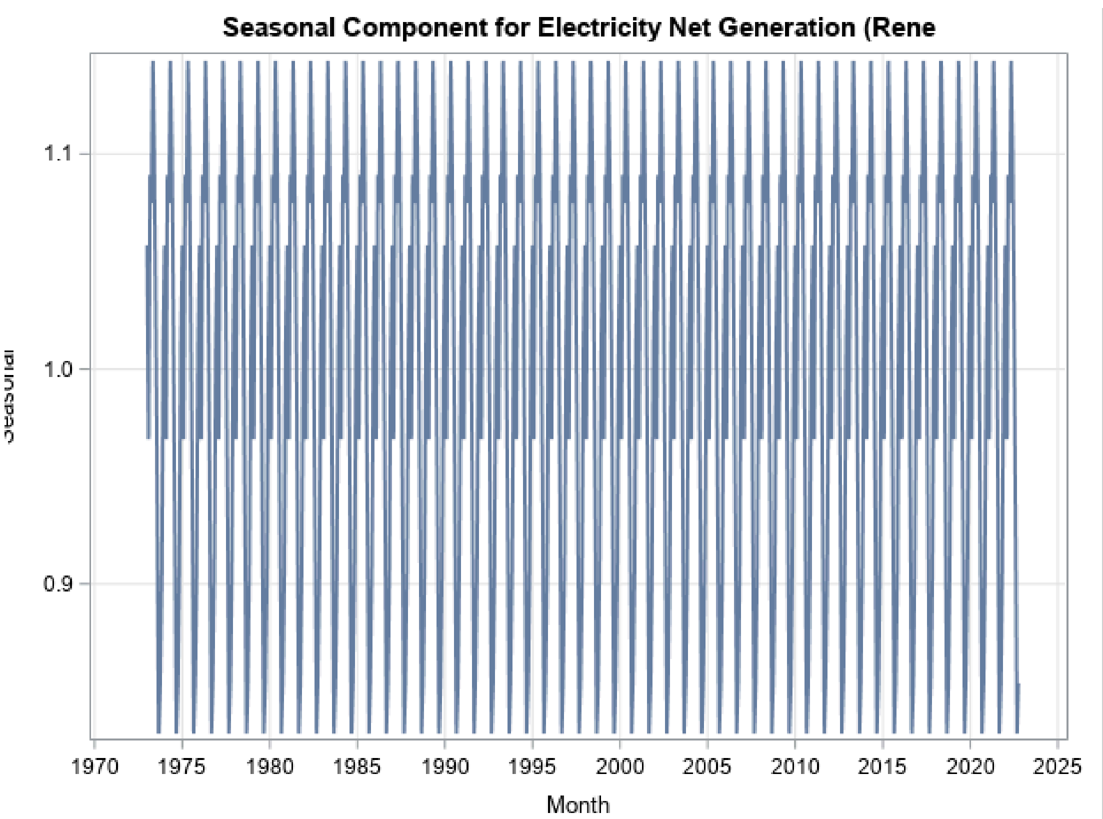
To perform the time series forecasting of renewable energy sources we took the data from the Energy generation dataset that we had and imported it in SAS. Post importing the dataset successfully in SAS Studio 3.8 we performed time series exploration of the dataset to get a basic idea of the nature of the dataset.

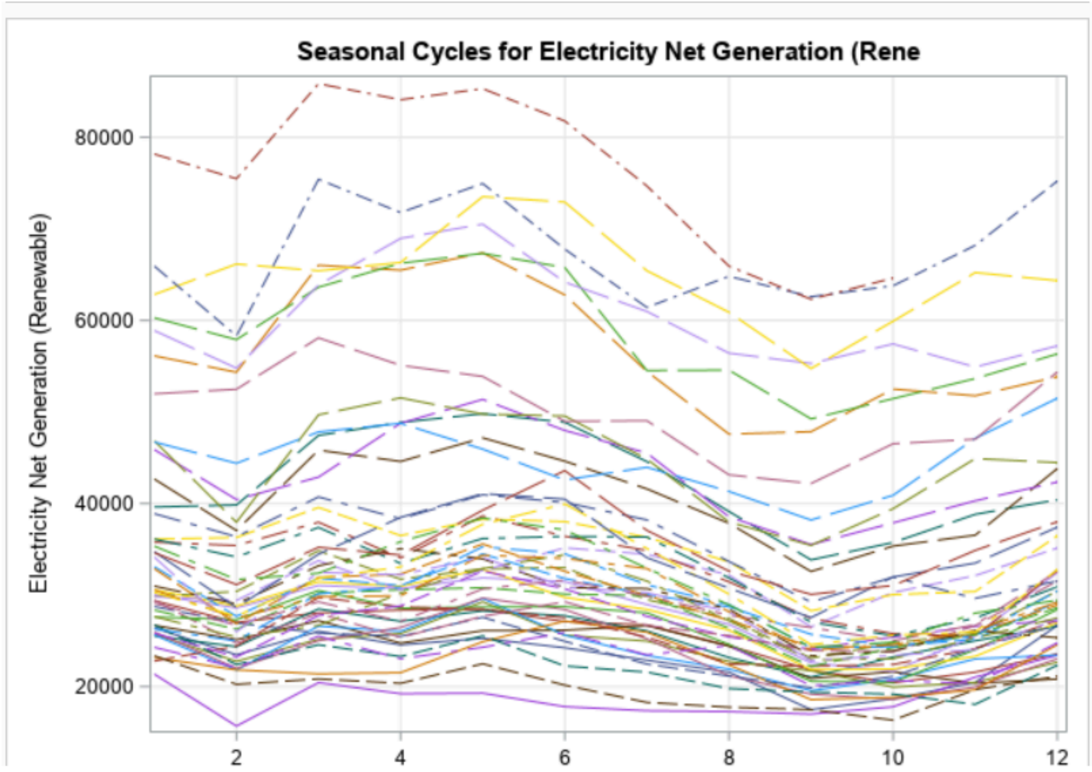
Series Value Graphs for Renewable Energy:

Decomposition Analysis:

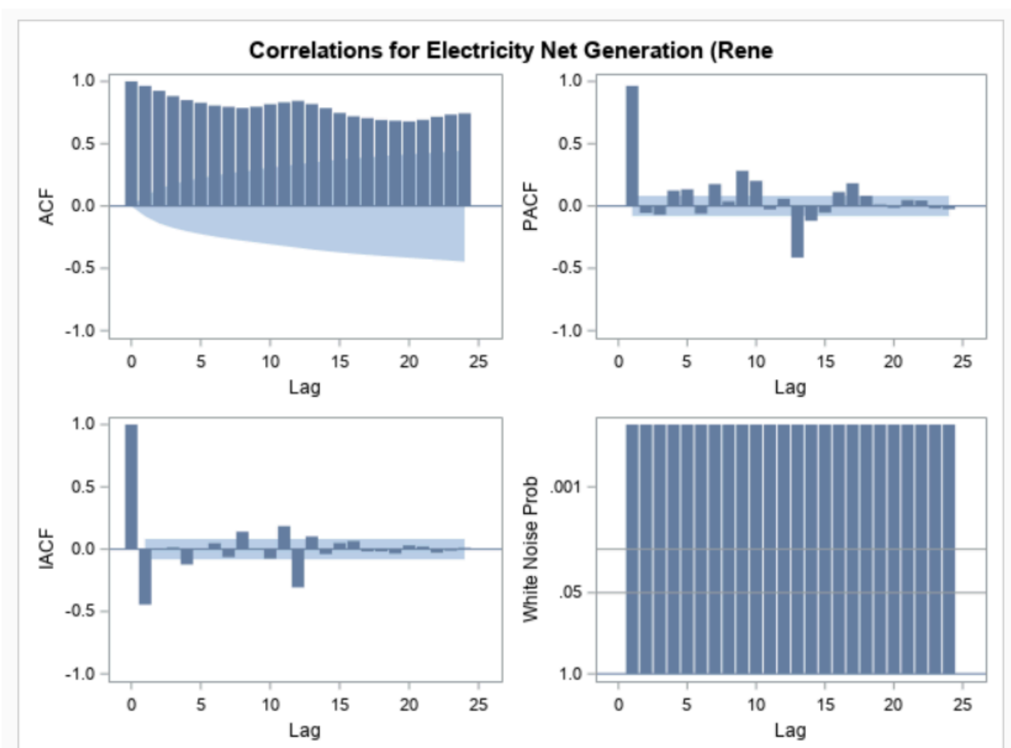
We performed basic decomposition analysis with the Total Renewable energy column as dependent variable and Month as time ID. Below are the results of decomposition analysis.


Trend Component for Electricity Net Generation (Rene
Trend
80000
60000
40000
20000
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
Month
Original
Trend


Seasonal Component for Electricity Net Generation (Rene
1.1
1.0
0.9
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
Month



Correlation Analysis f

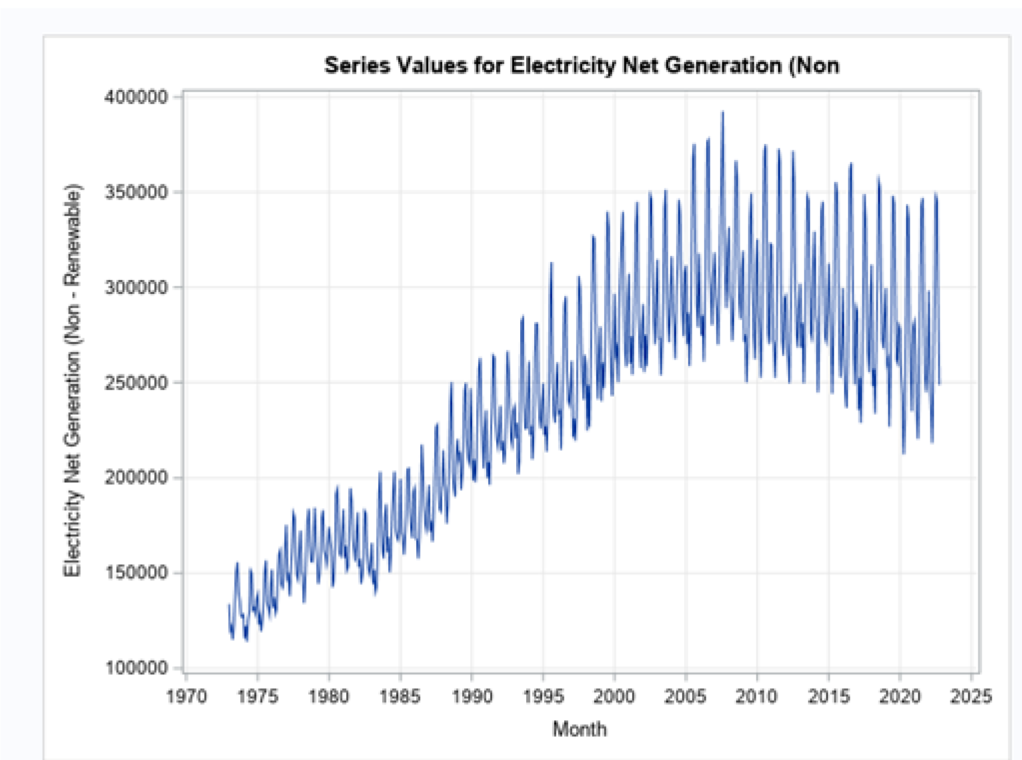


or Renewable Energy Source:

Upon analysis of the decomposition plots we observe that for renewable energy generation data points there is a seasonal and non-linear trend component. From the correlations plot we can see there is significant autocorrelation with all the previous lags. There is significant partial autocorrelation and significant inverse autocorrelation at some of the lags. Also we can see that the white noise probability test has failed, which means the errors are not yet distributed as white noise and there is still a significant signal that needs to be extracted from the noise.

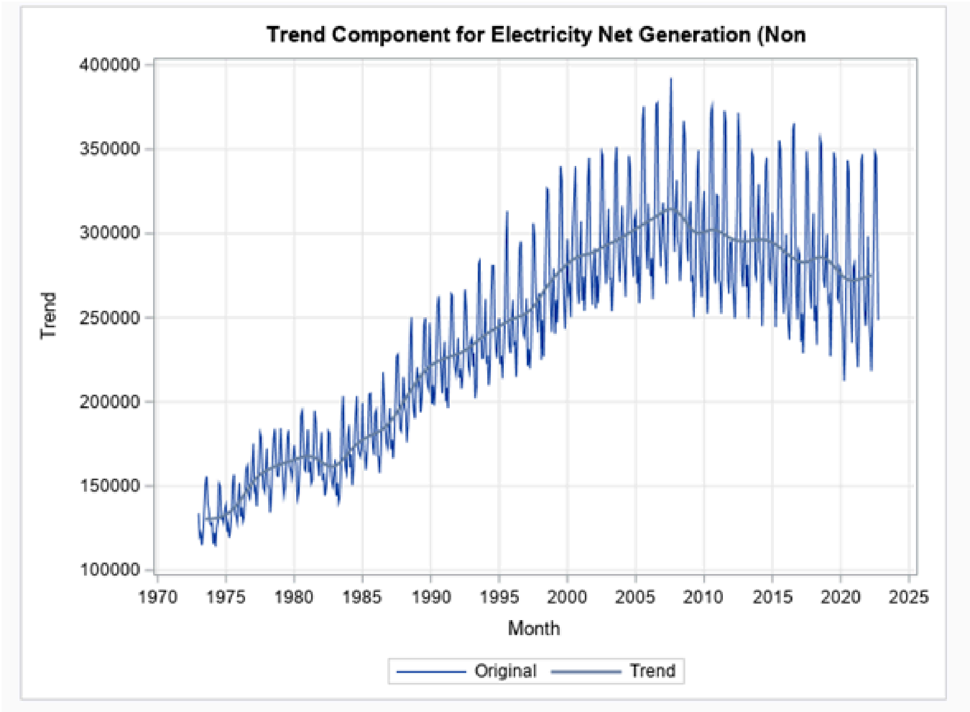
Electricity Generation From Non-Renewable Energy Sources:

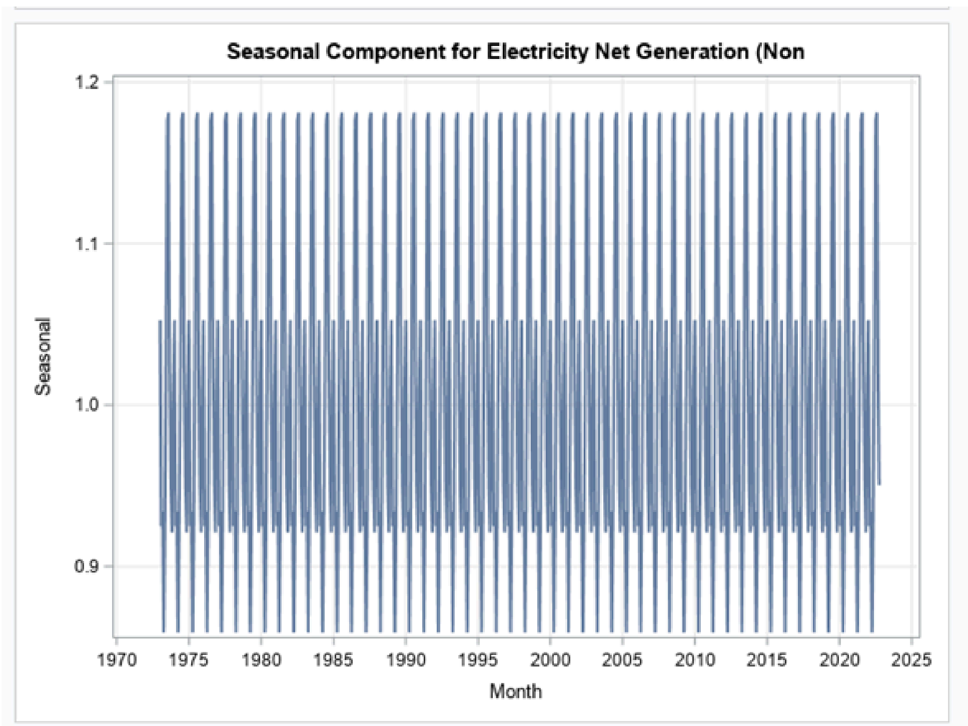
Series Value Graphs for Non-Renewable Sources:



Decomposition Analysis for Non-Renewable Sources:

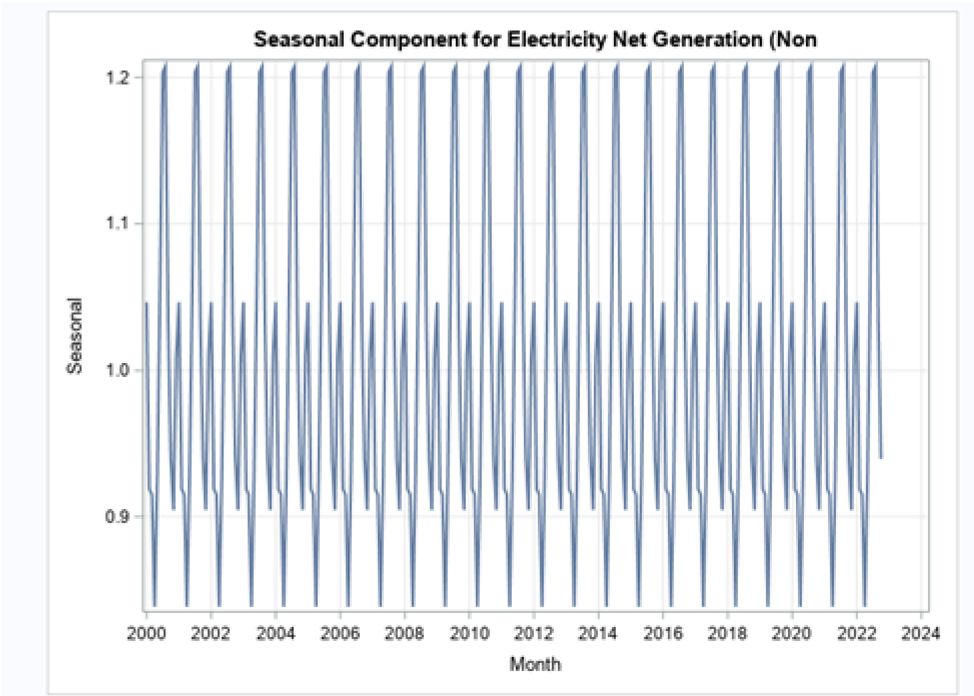
From the below trend component we observe that From 1970 to 2022, a trend in the net generation of power from non-renewable sources, which shows an increase in electricity production.

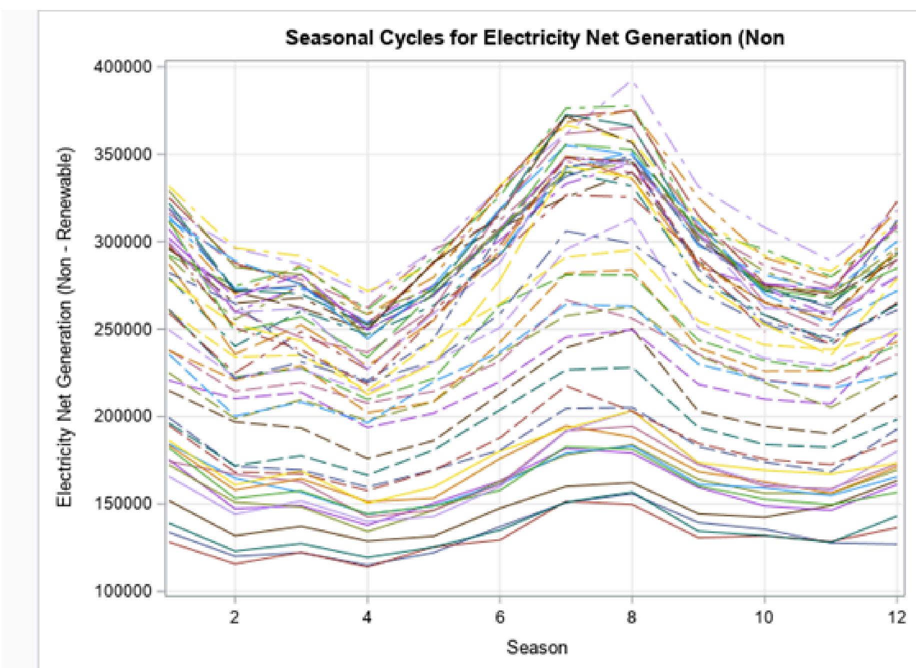




As per the above picture, we see the seasonal cycle is not clear, hence we have taken 22 years of data to check if the seasonality is clear.

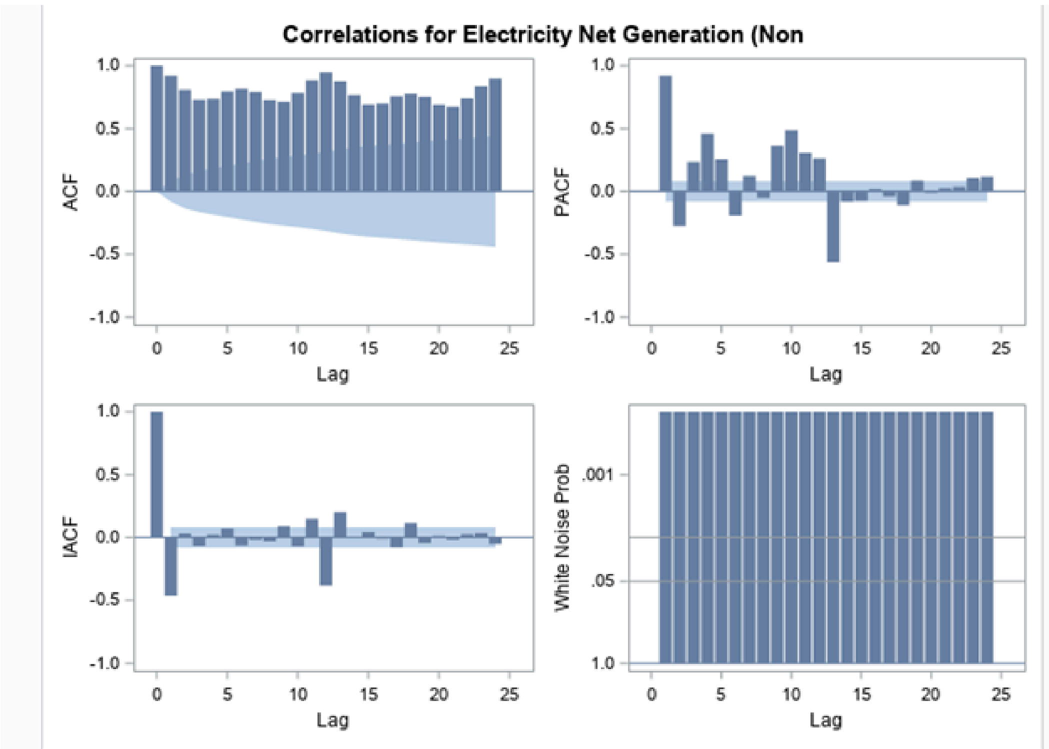
Below is the seasonality for past 22 years, there is a seasonality in the time series forecasting.





The pattern or changes that take place within a specific time or period are referred to as the seasonal cycle. The seasonal cycle of electricity Net generation is shown here, which enables us to spot trends and patterns for next predicted values. The screenshot below shows that seasonality has remained steady throughout time.

Correlations Analysis for Non-Renewable Sources:

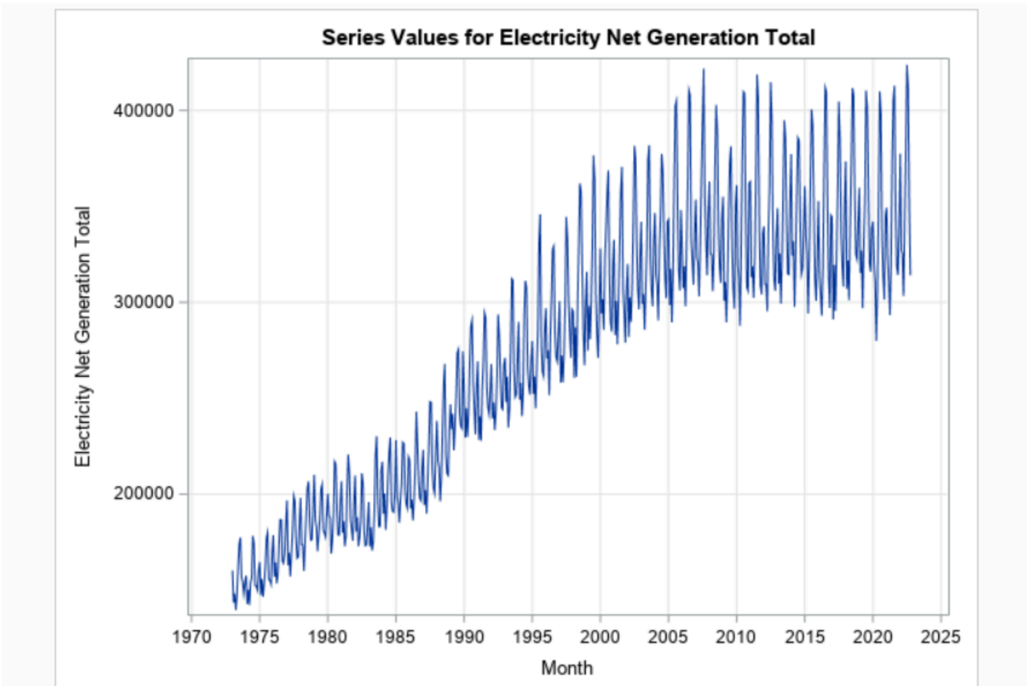


ACF and PACF correlation enables us to determine the relationship between a time series and its lag values. K is the lag, thus T-K. According to the graph below, there is a strong association between ACF and PACF, but not much of a noticeable lag.

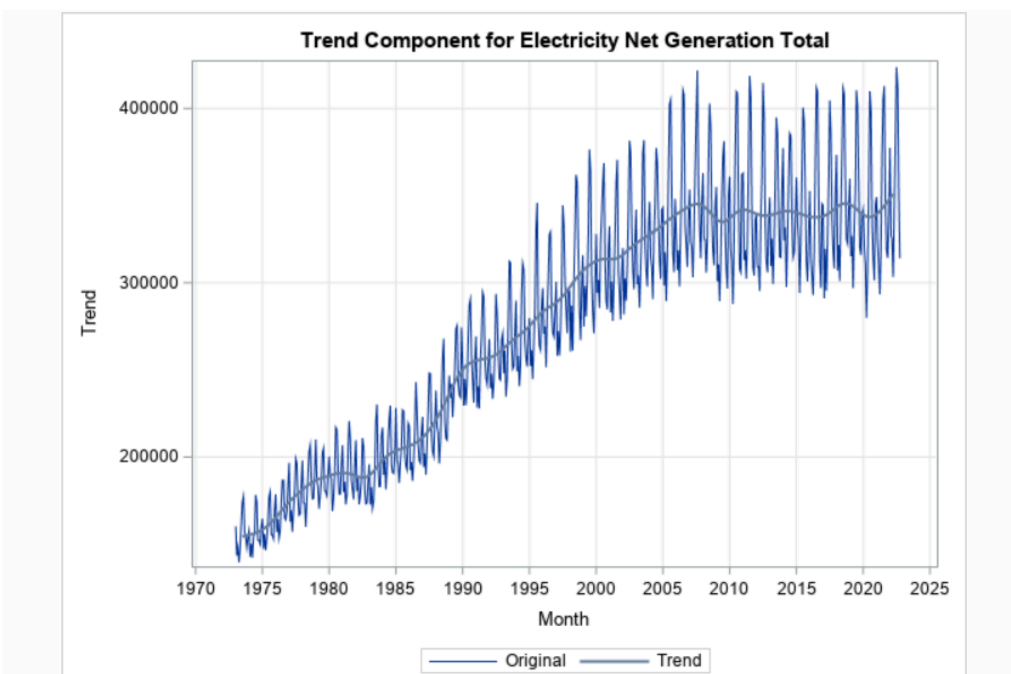
White noise has a mean of zero and a fixed variance, and each value in the time series is independently and identically distributed. According to Ljung-Box chi-square statistics, the null hypothesis, according to which a linear trend model would be insufficient for this series and there would be no white noise, is rejected.

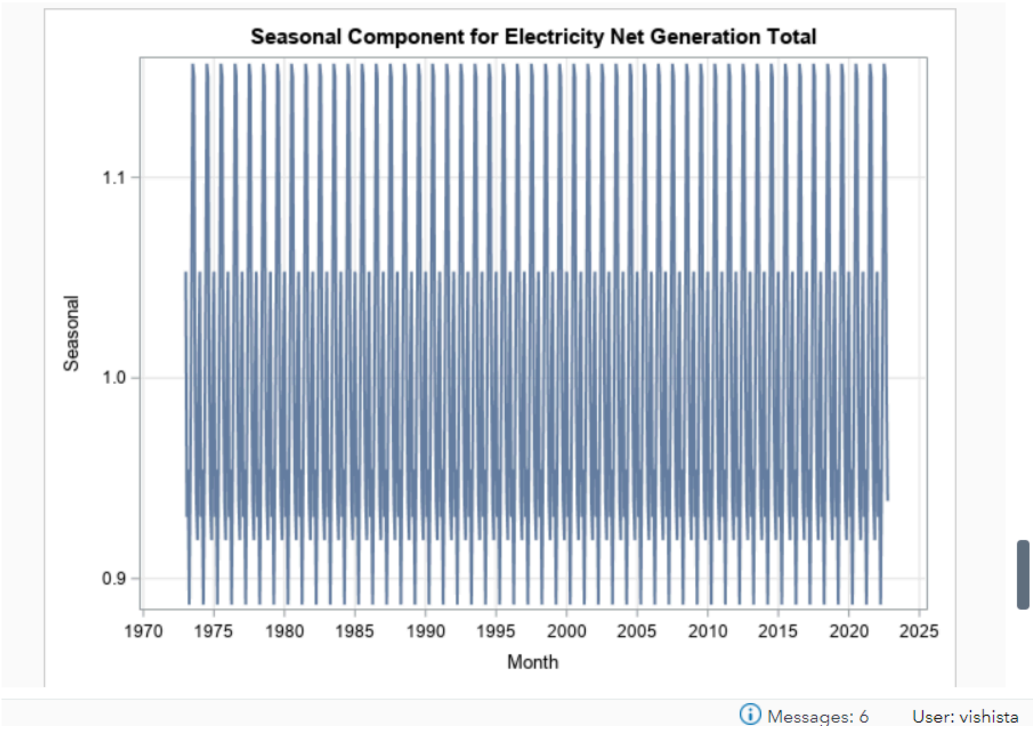
Total Electricity Generation:

Series Graphs for Total Electricity Generation:

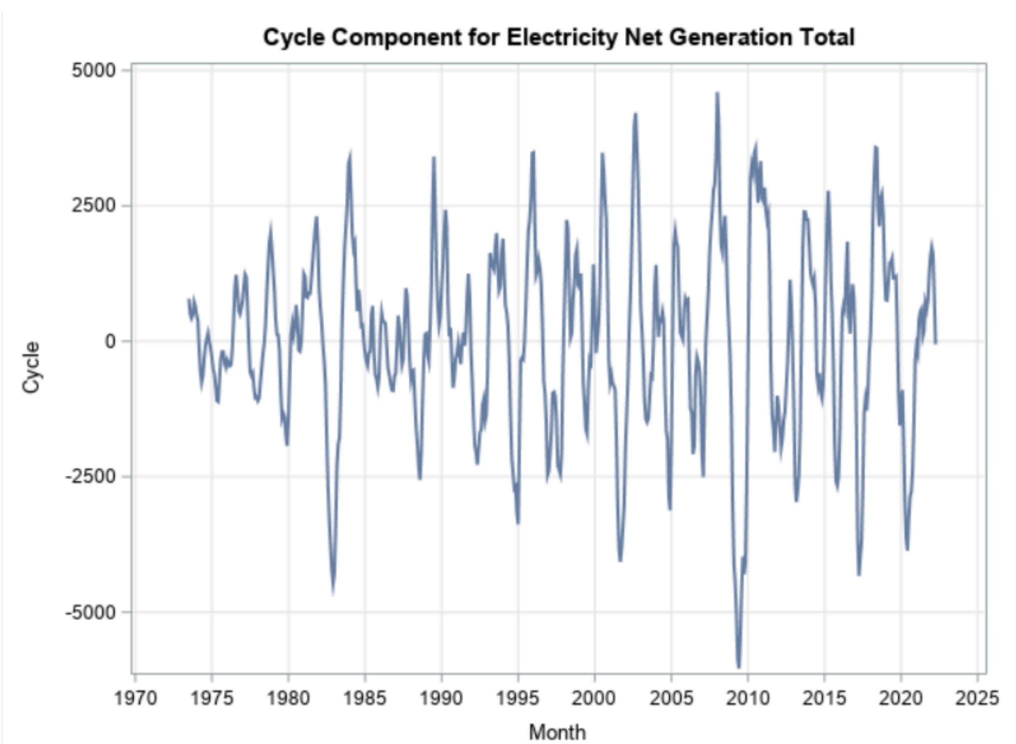


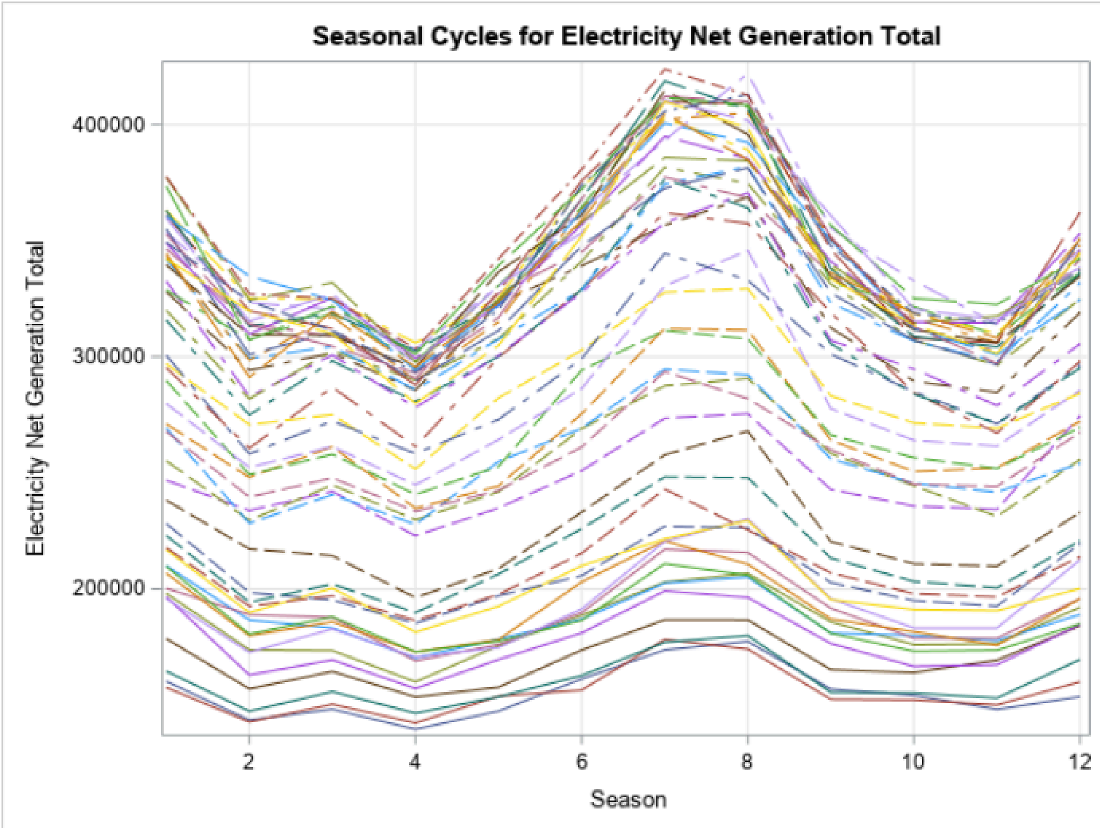
Decomposition Analysis for Total Electricity Generation:



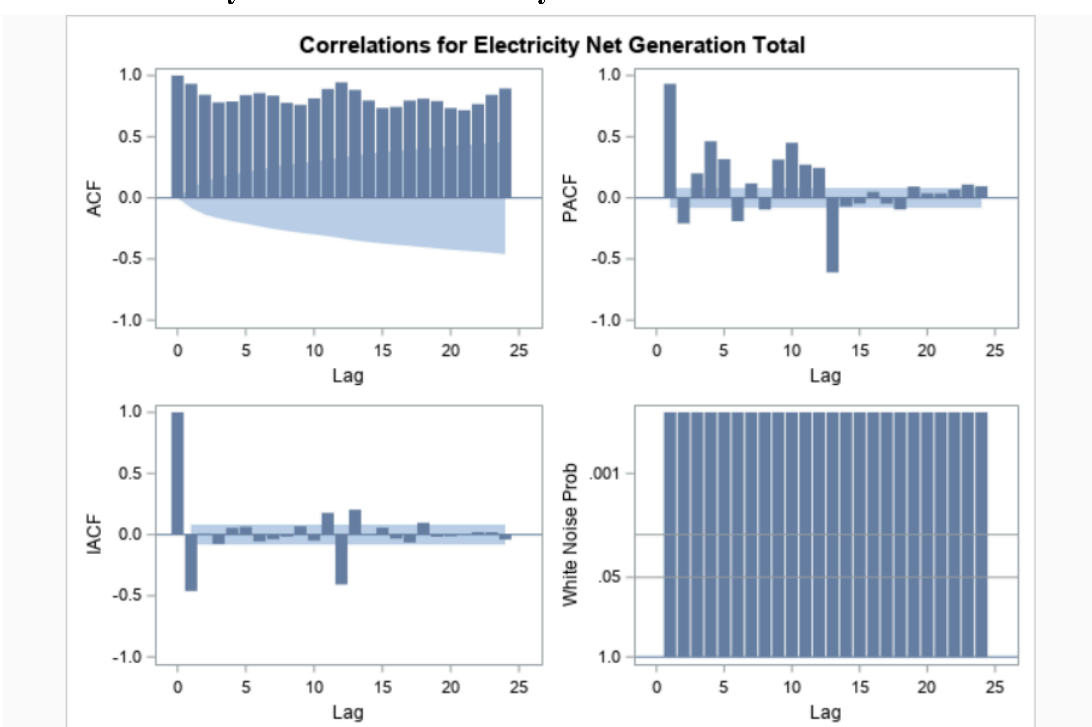


(CYCLE COMPONENT)



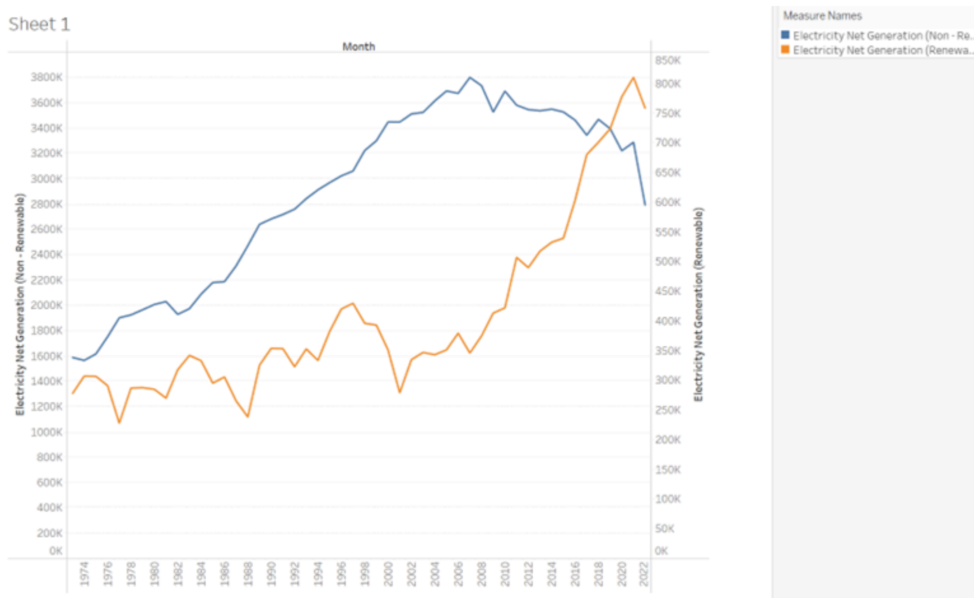


Correlation Analysis for Total Electricity Generation:



<u>**RENEWABLE ENERGY SOURCES AND NON-RENEWABLE ENERGY SOURCES**</u>

Below is the screenshot of the visualization of power generation from renewable resources and non-renewable resources from 1975 to 2022.



8

3. TIME SERIES MODELING AND FORECASTING

3.1. Non-Renewable Resources

Model Selection:

According to time series forecasting, we can see that the generation of electricity from non-renewable resources has a pattern, is seasonal, and seasonality is stable. For this, we must run the additive Winters model.

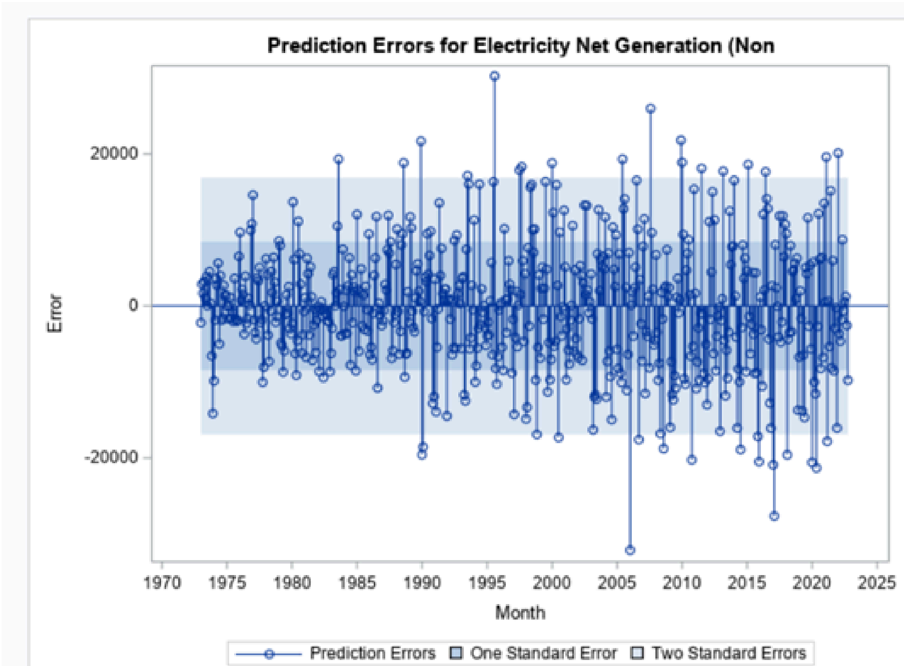
Winter additive model: When predicting time series with seasonal trends, the winter additive model is helpful. By breaking the series down into its constituent parts, it is possible to model the seasonal variations more precisely and to spot patterns and cycles in the data.

Winter Additive model:

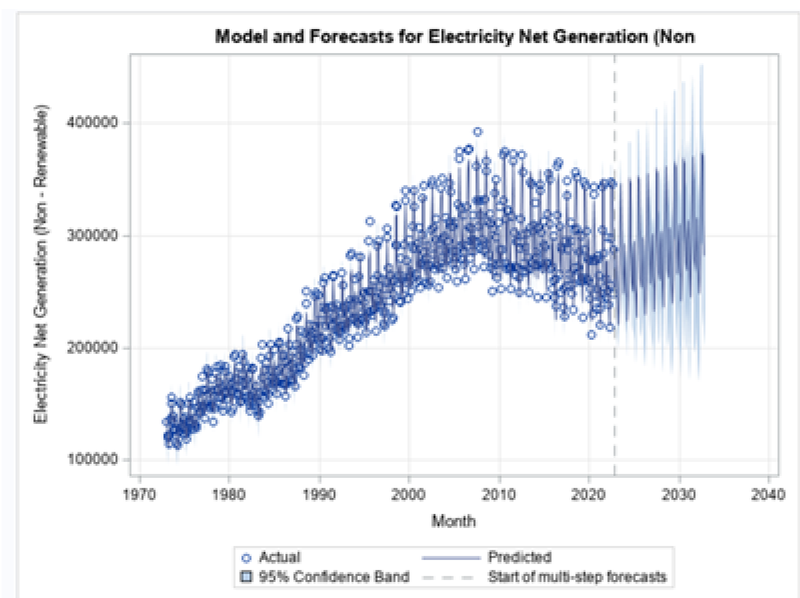
Forecast horizon: 120 and Holdout value as 60.

Input Data Set	
Name	WORK.PREPROCESSEDDATA
Label	
Time ID Variable	Month
Time Interval	MONTH
Length of Seasonal Cycle	12
Forecast Horizon	120
Forecast Back	60

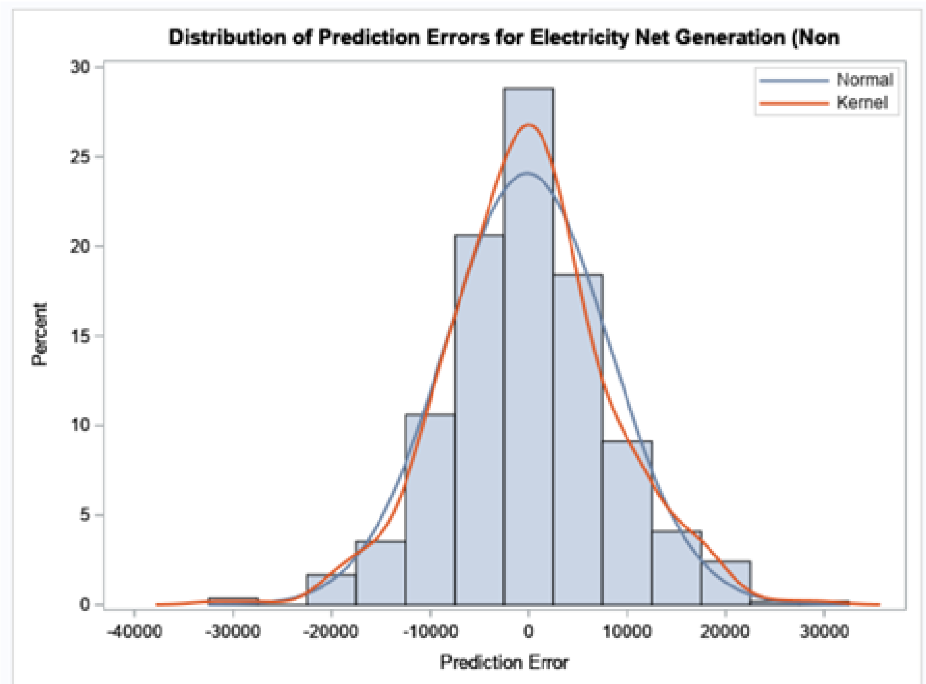
Variable Information	
Name	Electricity Net Generation (Non
Label	Electricity Net Generation (Non - Renewable)
First	JAN1973
Last	OCT2022
Number of Observations Read	598



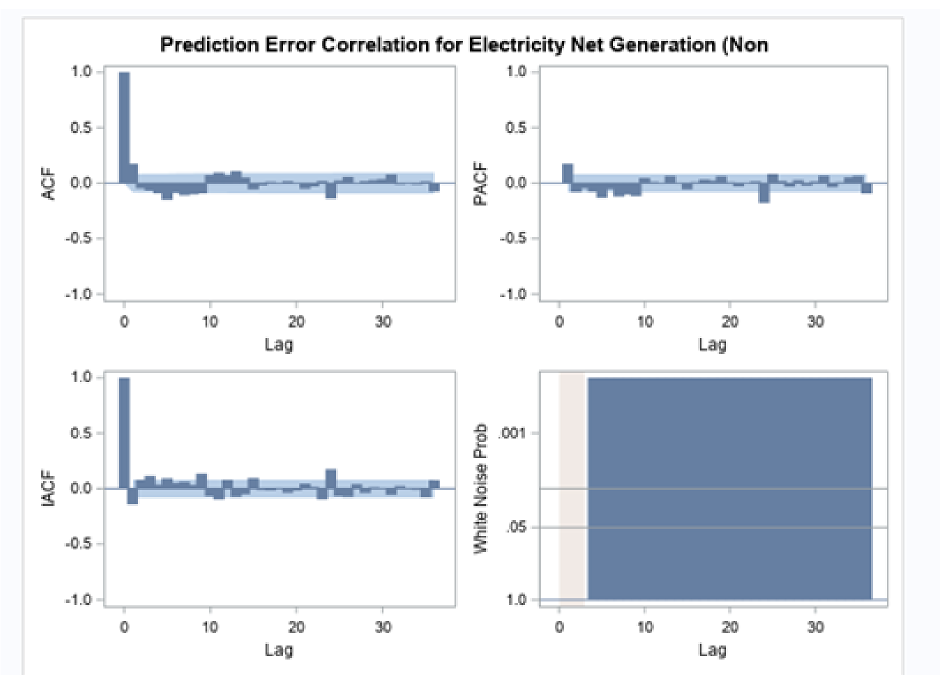
When we increase the number of forecasting periods to 120, which is equal to 10 years. The image of the 10-year forecast shows that power generation has decreased in recent years. The difference may be seen because there is a sizable drop when we compare the actual to the projected.



Errors are normally distributed as it is in a quadratic trend as per the below winter additive model.



There is no connection between errors from the first lag to the second lag, but there is white noise, therefore there is still a signal, thus we reject the null hypothesis, as shown below in the prediction error correlation for electricity.



We have taken 120 forecasting periods and the holdout value as 60. Below is the screenshot of the results which we have obtained.

Fit statistics:

	REGION	RMSE	MAPE	MAE
1	FIT	8278.6509857	2.6648502418	6344.2201682
2	FORECAST	19426.253114	6.14022688	16181.413054

MAE	RSQUARE	AIC	SBC
6344.2201682	0.9839597872	9713.0643932	9725.9279689
16181.413054	0.7423042403	1184.925682	1184.925682

AIC	**9713.064**
SBC	**9725.927**
MAPE	**2.664**
RMSE	**8278.650**

Stationary DATA analysis:

Dickey Filler Test(Unit test): Determine whether or not the data is stationary and whether a signal is there using the Dickey Fuller Test (unit test).

We can tell that the mean is not zero based on the time series descriptive analysis from the below screenshot.

Input Data Set	
Name	WORK.PREPROCESSEDDATA
Label	
Time ID Variable	Month
Time Interval	MONTH
Length of Seasonal Cycle	12

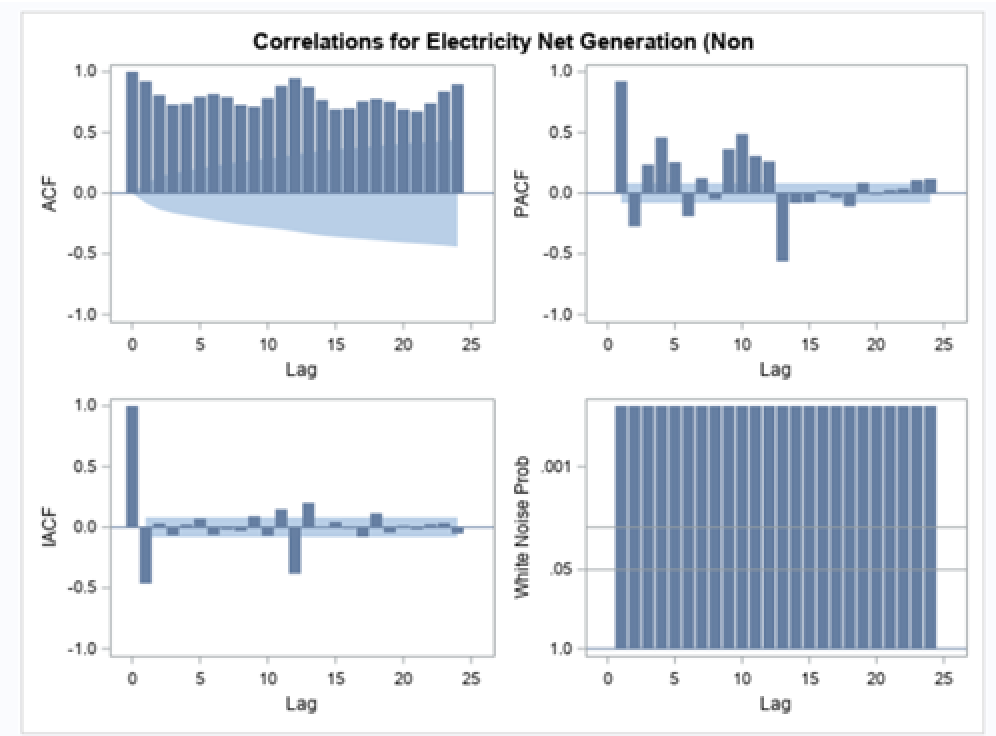
Variable Information	
Name	Electricity Net Generation (Non
Label	Electricity Net Generation (Non - Renewable)
First	JAN1973
Last	OCT2022
Number of Observations Read	598

Time Series Descriptive Statistics	
Variable	Electricity Net Generation (Non
Number of Observations	598
Number of Observations Used	598
Number of Missing Observations	0
Minimum	113965.4
Median	249634.6
Maximum	392542.4
Mean	240651.1
Standard Deviation	64476.47

As there is a trend in the series and the mean is non-zero, we proceed to calculate the Pr >Tau and Pr > F values and find that the values are less than 0.05, indicating that the dataset is stationary.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-2.6763	0.2612	-1.09	0.2509		
	1	-4.5032	0.1447	-1.44	0.1396		
	2	-2.1932	0.3089	-0.95	0.3045		
Single Mean	0	-46.5999	0.0018	-4.99	<.0001	12.48	0.0010
	1	-81.9110	0.0018	-6.48	<.0001	20.99	0.0010
	2	-48.7102	0.0018	-4.92	<.0001	12.13	0.0010
Trend	0	-138.671	0.0001	-8.74	<.0001	38.27	0.0010
	1	-318.398	0.0001	-12.46	<.0001	77.69	0.0010
	2	-241.989	0.0001	-9.91	<.0001	49.16	0.0010

We see there is a correlation for electricity power generation as per the below screenshot from first lag to the second however it still has the noise.



Seasonal decomposition/ adjustment shows the Trend cycle, Seasonal- Irregular cycle, Irregularity, and Seasonality adjusted cycle.

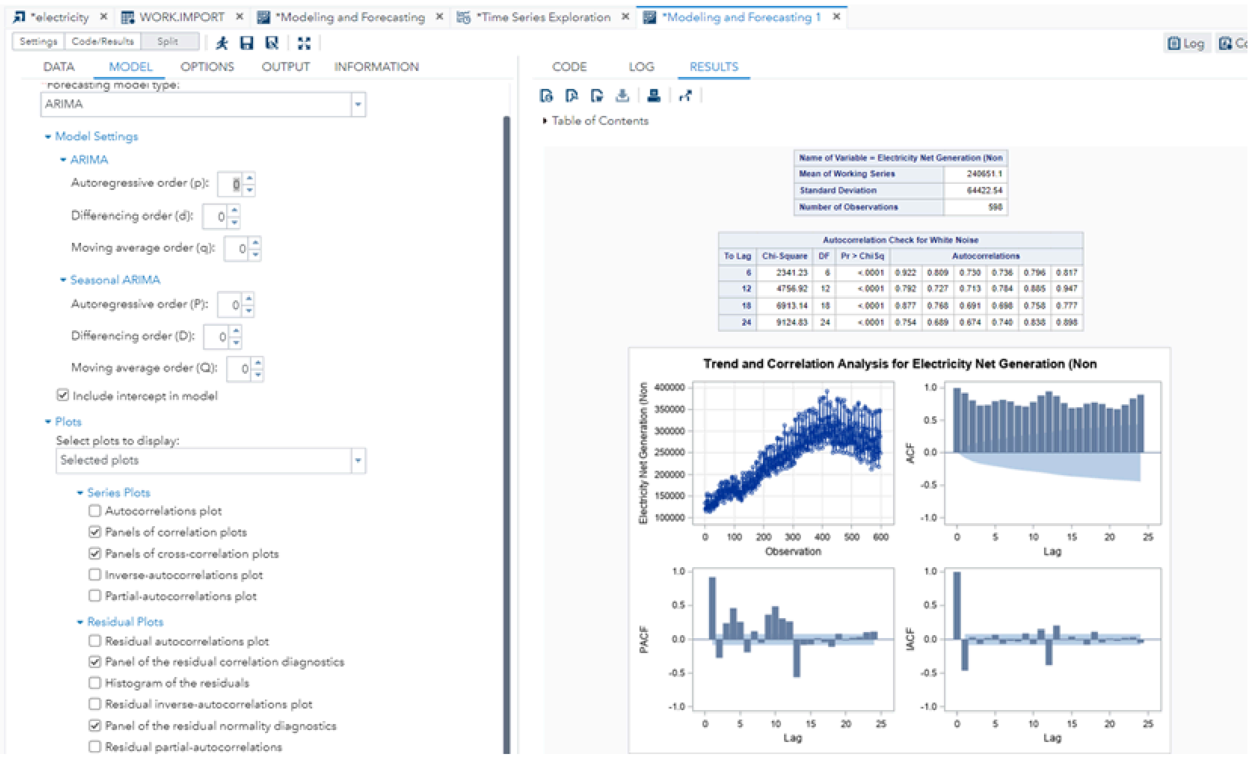
ARIMA MODEL: In many different fields, including finance, economics, and engineering, ARIMA models are frequently used for time series predictions and analysis. The algorithms can identify patterns in the data, such as trends, seasonality, and cycles, and can offer helpful information for making decisions.

As the dataset is stationary, we can now run ARIMA models.

The first step is to find if the data is the AR model or the AM model.

After running ARIMA, we see it is an AR model as the ACF is highly correlated however PACF is dying.

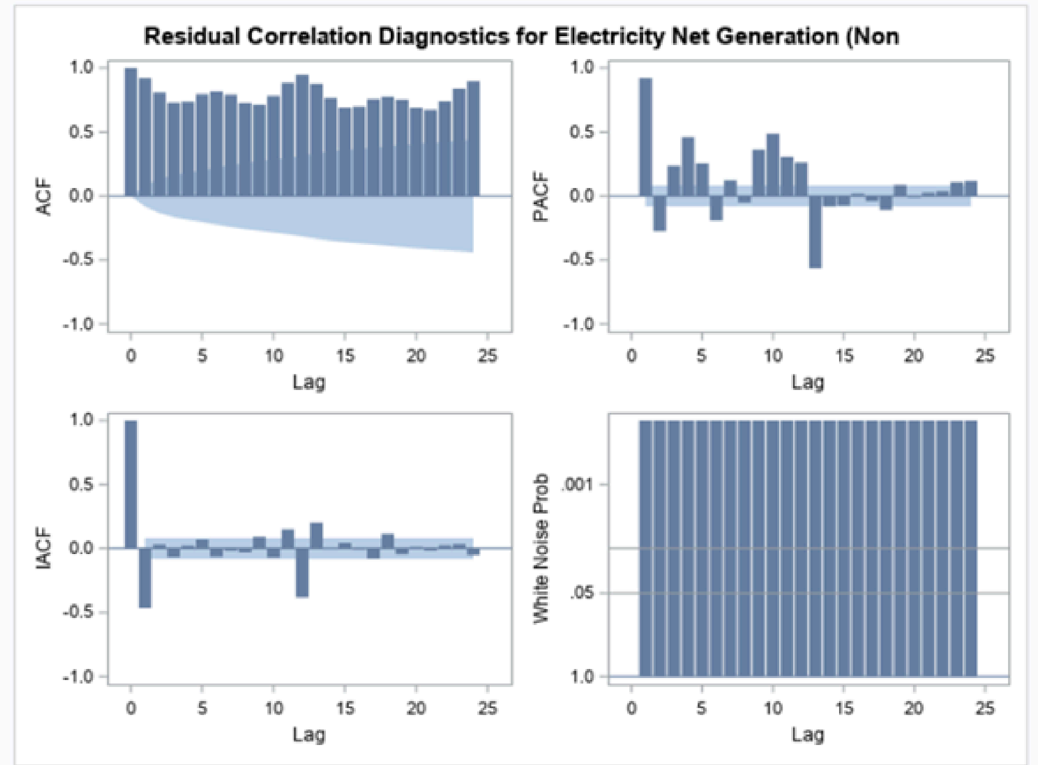
ARIMA(0,0,0)

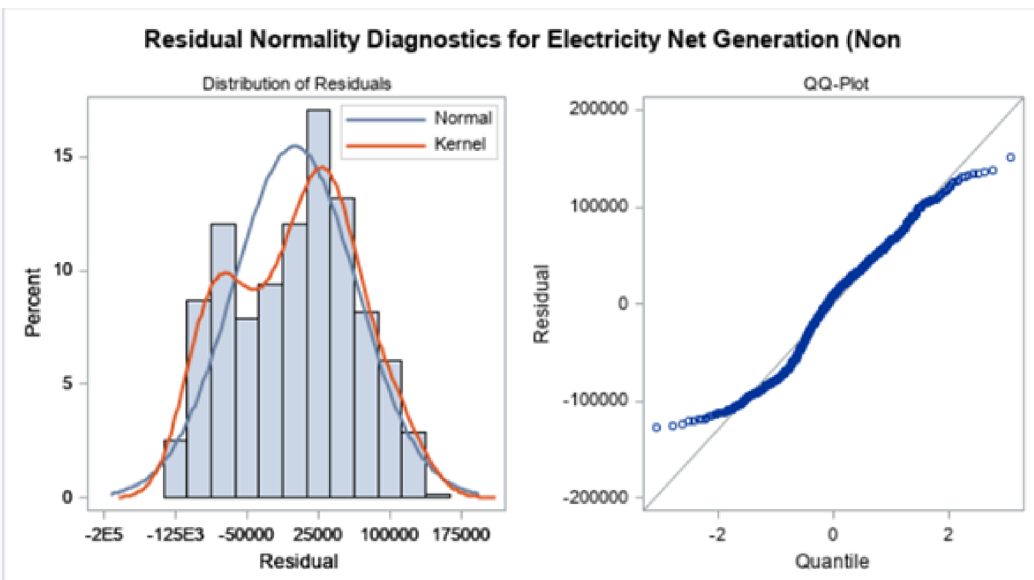


Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	240651.1	2636.6	91.27	<.0001	0

Constant Estimate	240651.1
Variance Estimate	4.1572E9
Std Error Estimate	64476.47
AIC	14942.62
SBC	14947.01
Number of Residuals	598

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2341.23	6	<.0001	0.922	0.809	0.730	0.736	0.796	0.817
12	4756.92	12	<.0001	0.792	0.727	0.713	0.784	0.885	0.947
18	6913.14	18	<.0001	0.877	0.768	0.691	0.698	0.758	0.777
24	9124.83	24	<.0001	0.754	0.689	0.674	0.740	0.838	0.898
30	9999.99	30	<.0001	0.830	0.725	0.651	0.659	0.716	0.736
36	9999.99	36	<.0001	0.713	0.650	0.634	0.698	0.793	0.852
42	9999.99	42	<.0001	0.786	0.684	0.612	0.617	0.672	0.691
48	9999.99	48	<.0001	0.668	0.605	0.590	0.653	0.746	0.806


Residual Correlation Diagnostics for Electricity Net Generation (Non
ACF
PACF
IACF
White Noise Prob
Lag
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
5
10
15
20
25


Residual Normality Diagnostics for Electricity Net Generation (Non
Distribution of Residuals
Normal
Kernel
Percent
Residual
-2E5
-125E3
-50000
25000
100000
175000
QQ-Plot
200000
100000
0
-100000
-200000
Quantile
-2
0
2

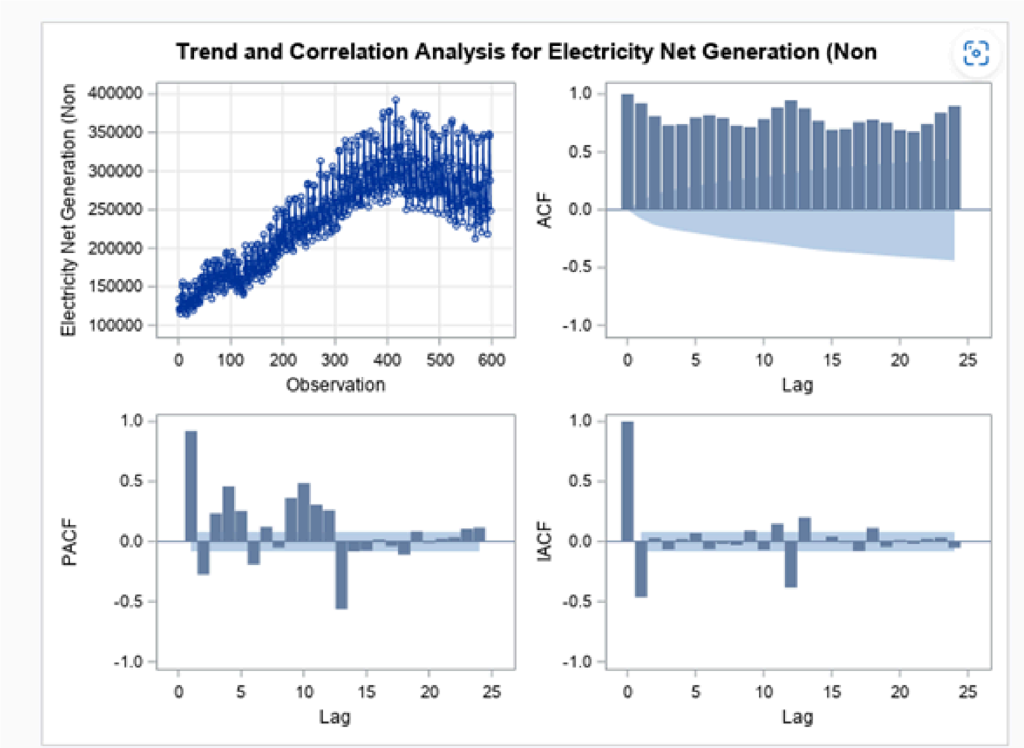
Model for variable Electricity Net Generation (Non	
Estimated Mean	240651.1

Forecasts for variable Electricity Net Generation (Non				
Obs	Forecast	Std Error	95% Confidence Limits	
599	240651.1	64476.47	114279.5	367022.7
600	240651.1	64476.47	114279.5	367022.7
601	240651.1	64476.47	114279.5	367022.7
602	240651.1	64476.47	114279.5	367022.7
603	240651.1	64476.47	114279.5	367022.7
604	240651.1	64476.47	114279.5	367022.7
605	240651.1	64476.47	114279.5	367022.7
606	240651.1	64476.47	114279.5	367022.7
607	240651.1	64476.47	114279.5	367022.7
608	240651.1	64476.47	114279.5	367022.7
609	240651.1	64476.47	114279.5	367022.7
610	240651.1	64476.47	114279.5	367022.7

Tried with ARIMA (12,0,12), however, it is the most complex model:

Name of Variable = Electricity Net Generation (Non	
Mean of Working Series	240651.1
Standard Deviation	64422.54
Number of Observations	598

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2341.23	6	<.0001	0.922	0.809	0.730	0.736	0.796	0.817
12	4756.92	12	<.0001	0.792	0.727	0.713	0.784	0.885	0.947
18	6913.14	18	<.0001	0.877	0.768	0.691	0.698	0.758	0.777
24	9124.83	24	<.0001	0.754	0.689	0.674	0.740	0.838	0.898

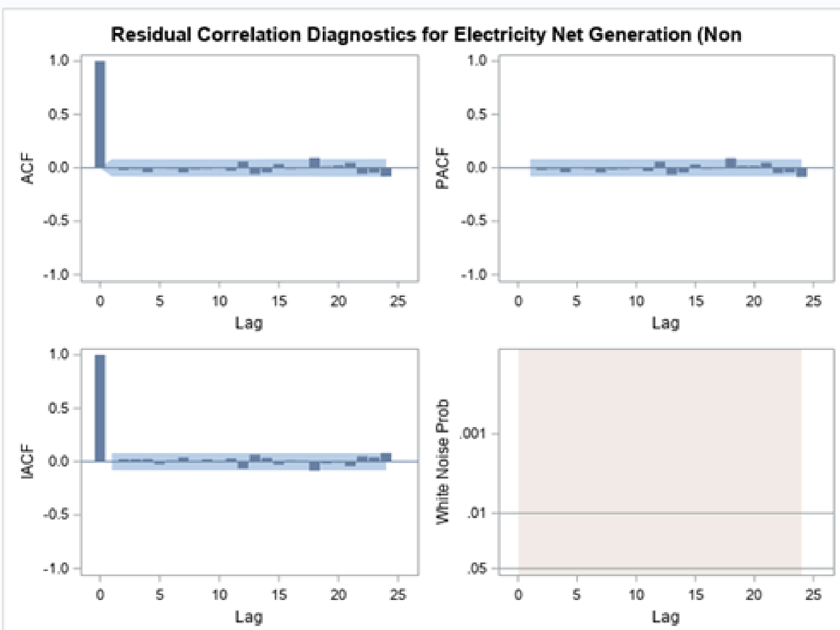

Trend and Correlation Analysis for Electricity Net Generation (Non
Electricity Net Generation (Non
Observation
ACF
PACF
IACF
Lag

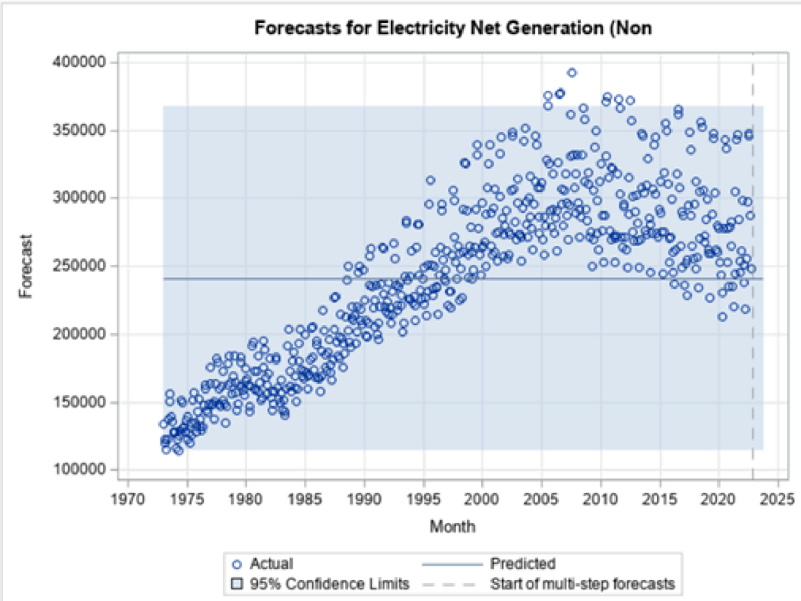
ARIMA Estimation Optimization Summary	
Estimation Method	Maximum Likelihood
Parameters Estimated	25
Termination Criteria	Maximum Relative Change in Estimates
Iteration Stopping Value	0.001
Criteria Value	9.43E-13
Maximum Absolute Value of Gradient	6.5746E9
R-Square Change from Last Iteration	0.128962
Objective Function	Log Gaussian Likelihood
Objective Function Value	-6256.6
Marquardt's Lambda Coefficient	1E12
Numerical Derivative Perturbation Delta	0.001
Iterations	17
Warning Message	Estimates may not have converged.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	209662.8	50966.1	4.11	<.0001	0
MA1,1	-0.29192	0.08614	-3.39	0.0007	1
MA1,2	-0.17636	0.08346	-2.11	0.0346	2
MA1,3	-0.53208	0.04908	-10.84	<.0001	3
MA1,4	-0.17953	0.08864	-2.03	0.0428	4
MA1,5	0.07712	0.08336	0.93	0.3549	5
MA1,6	-0.50558	0.04957	-10.20	<.0001	6
MA1,7	-0.01634	0.08763	-0.19	0.8520	7
MA1,8	0.07551	0.08321	0.91	0.3641	8
MA1,9	-0.37767	0.05197	-7.27	<.0001	9
MA1,10	-0.17946	0.08468	-2.12	0.0341	10
MA1,11	-0.06678	0.07775	-0.86	0.3903	11
MA1,12	0.15772	0.04367	3.61	0.0003	12
AR1,1	0.24486	0.06967	3.51	0.0004	1
AR1,2	0.07334	0.10449	0.70	0.4827	2
AR1,3	-0.29286	0.04188	-6.99	<.0001	3
AR1,4	0.18247	0.07316	2.49	0.0126	4
AR1,5	0.14442	0.10582	1.36	0.1723	5
AR1,6	-0.35399	0.03855	-9.18	<.0001	6
AR1,7	0.24927	0.07048	3.54	0.0004	7
AR1,8	0.07699	0.10448	0.74	0.4612	8
AR1,9	-0.28961	0.04084	-7.09	<.0001	9
AR1,10	0.17878	0.07334	2.44	0.0148	10
AR1,11	0.13812	0.10544	1.31	0.1902	11
AR1,12	0.63674	0.03816	16.68	<.0001	12

Constant Estimate	2406.004
Variance Estimate	71469196
Std Error Estimate	8453.946
AIC	12563.19
SBC	12673.03
Number of Residuals	598

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	.	0	.	0.011	-0.008	0.002	-0.030	0.014	0.003
12	.	0	.	-0.033	-0.001	0.003	0.011	-0.014	0.068
18	.	0	.	-0.049	-0.032	0.046	0.002	0.011	0.105
24	.	0	.	0.023	0.032	0.058	-0.046	-0.036	-0.069
30	38.55	6	<.0001	-0.058	0.007	0.066	0.030	-0.021	0.125
36	46.87	12	<.0001	0.062	0.011	0.094	0.001	-0.020	0.003
42	64.71	18	<.0001	-0.025	0.006	0.131	-0.006	-0.044	0.089
48	80.43	24	<.0001	-0.046	-0.023	0.037	-0.002	-0.016	0.141





Outlier Detection Summary	
Maximum number searched	5
Number found	3
Significance used	0.05

Outlier Details				
Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
306	Shift	51630.9	137.09	<.0001

We have tried to generate all the possible correlations however we tried with SARIMA, which is a statistical model used to analyze and forecast time-series data that exhibit seasonality. Extension of ARIMA and used to analyze nonseasonal time series.

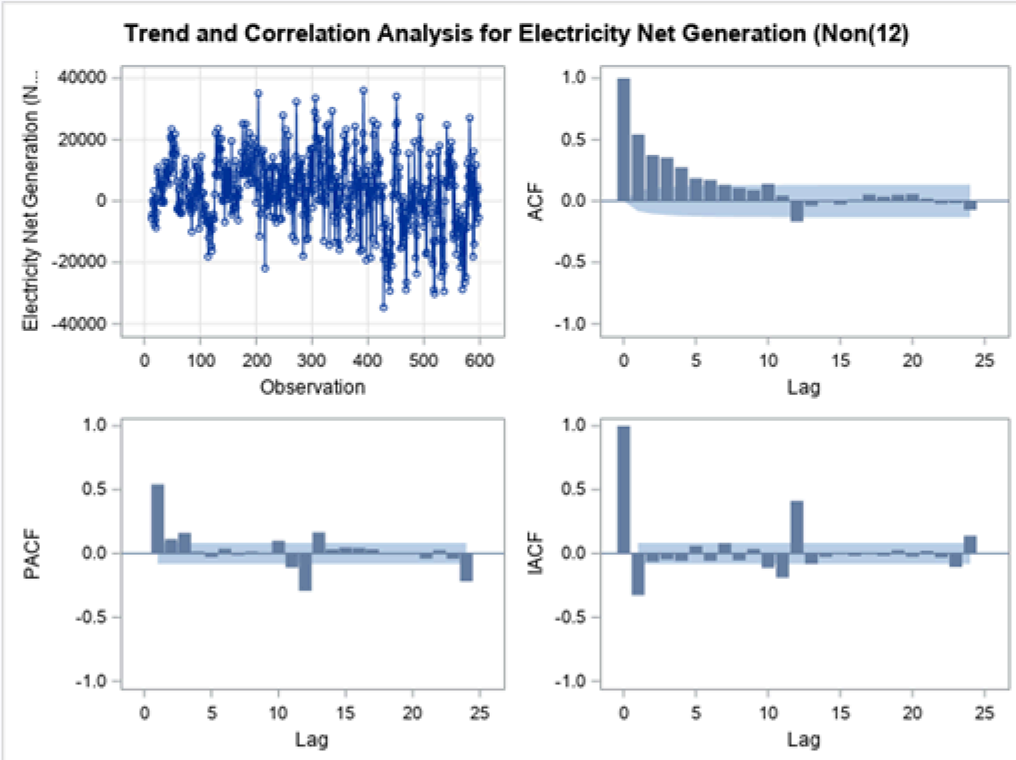
SARIMA is used when the data exhibits seasonality such as monthly, quarterly, or yearly patterns. It is used to predict time series based on patterns and trends observed in the past.

After running SARIMA with (1,1,1) and ARIMA(1,0,0) showing the data which we have for residuals, it shows the model is a good fit for the data and is effectively capturing the underlying structure of the time series. The presence of white noise in the residuals indicates that the model is capturing all the systematic patterns in the data and any remaining variability is random and unpredictable.

And as the ACF shows significant autocorrelation residuals at any lag, it suggests that the model is effectively removing any serial correlation in the data and PACF indicates the residuals are not significantly correlated with their own lagged values beyond the first lag. This model has captured all the relevant information in the data and there is no remaining structure to be modeled.

Name of Variable = Electricity Net Generation (Non	
Period(s) of Differencing	12
Mean of Working Series	2916.453
Standard Deviation	11695.05
Number of Observations	586
Observation(s) eliminated by differencing	12

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	412.58	6	<.0001	0.542	0.374	0.354	0.275	0.186	0.172
12	464.29	12	<.0001	0.132	0.106	0.089	0.140	0.045	-0.168
18	468.26	18	<.0001	-0.039	-0.006	-0.028	0.010	0.053	0.036
24	475.74	24	<.0001	0.051	0.057	0.020	-0.022	-0.017	-0.073

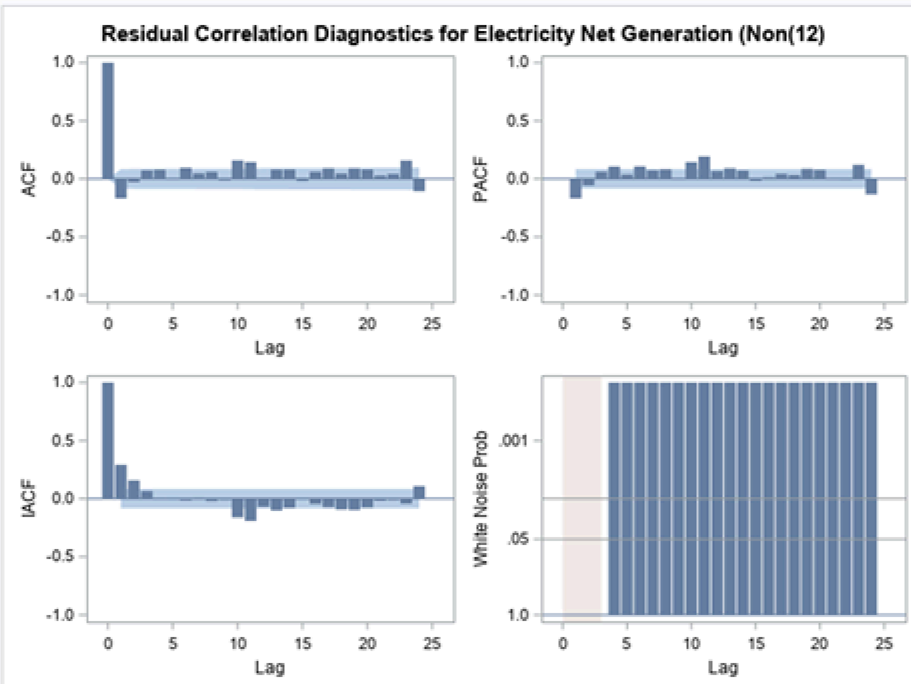


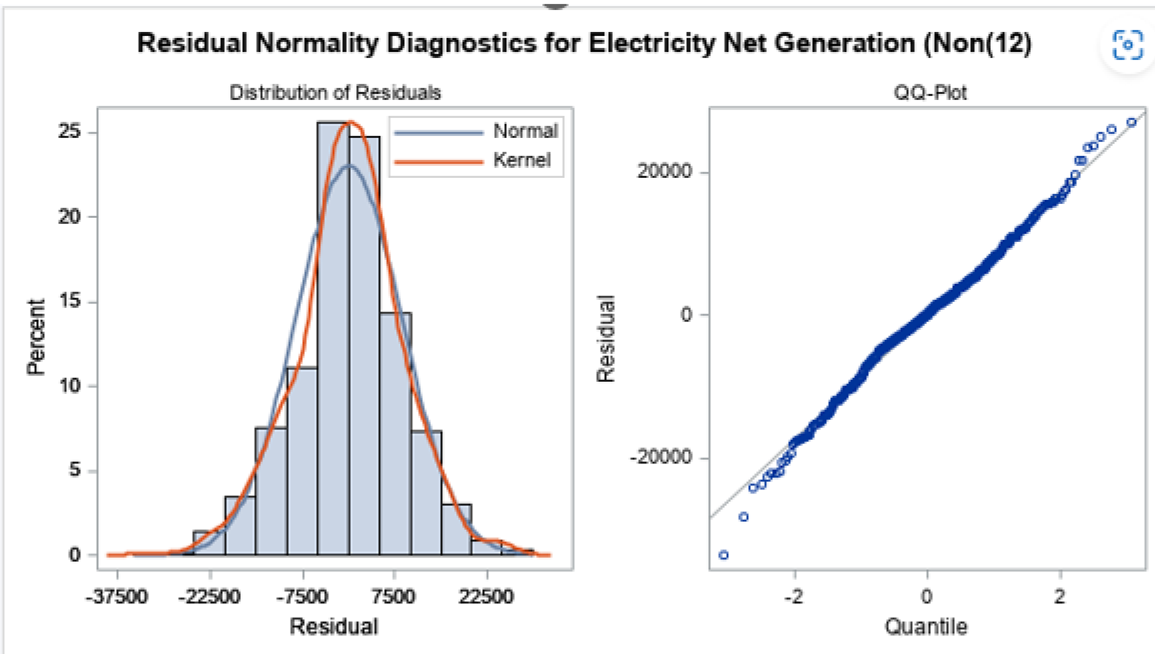
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	3072.3	579.48087	5.30	<.0001	0
MA1,1	0.70887	0.05219	13.58	<.0001	12
AR1,1	0.79528	0.02850	27.90	<.0001	1
AR2,1	0.07945	0.06518	1.22	0.2228	12

Constant Estimate	578.9711
Variance Estimate	75650279
Std Error Estimate	8697.717
AIC	12305.96
SBC	12323.45
Number of Residuals	586

Correlations of Parameter Estimates				
Parameter	**MU**	**MA1,1**	**AR1,1**	**AR2,1**
MU	1.000	0.113	0.037	0.078
MA1,1	0.113	1.000	0.400	0.742
AR1,1	0.037	0.400	1.000	0.123
AR2,1	0.078	0.742	0.123	1.000

Autocorrelation Check of Residuals									
To Lag	**Chi-Square**	**DF**	**Pr > ChiSq**	**Autocorrelations**					
6	29.72	3	<.0001	-0.169	-0.027	0.074	0.080	-0.002	0.098
12	60.83	9	<.0001	0.047	0.060	-0.012	0.159	0.144	-0.001
18	77.33	15	<.0001	0.081	0.081	-0.018	0.060	0.090	0.046
24	110.01	21	<.0001	0.090	0.081	0.028	0.041	0.158	-0.107
30	125.84	27	<.0001	0.084	0.084	0.018	0.066	0.062	0.056
36	148.90	33	<.0001	0.139	0.009	0.065	0.035	0.102	-0.042
42	166.94	39	<.0001	0.078	0.039	0.118	0.005	0.036	0.076
48	175.57	45	<.0001	0.045	0.021	0.035	0.045	0.029	0.083



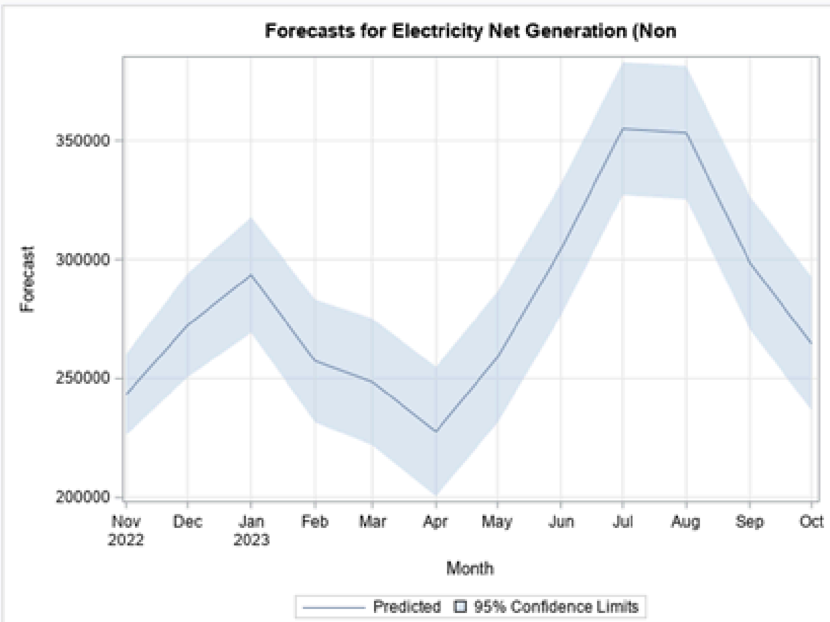


Model for variable Electricity Net Generation (Non	
Estimated Mean	3072.26
Period(s) of Differencing	12

Autoregressive Factors	
Factor 1:	1 - 0.79528 B**(1)
Factor 2:	1 - 0.07945 B**(12)

Moving Average Factors	
Factor 1:	1 - 0.70887 B**(12)

Forecasts for variable Electricity Net Generation (Non				
Obs	Forecast	Std Error	95% Confidence Limits	
599	243253.6	8697.72	226206.4	260300.8
600	272356.5	11112.93	250575.5	294137.4
601	293601.5	12399.97	269298.0	317905.0
602	257389.6	13149.12	231617.8	283161.4
603	248596.8	13601.66	221938.0	275255.5
604	227641.5	13880.26	200436.7	254846.3
605	258941.6	14053.62	231397.1	286486.2
606	304368.0	14162.17	276610.7	332125.3
607	355113.9	14230.39	327222.8	383004.9
608	353385.5	14273.38	325410.2	381360.8
609	298687.0	14300.50	270658.5	326715.4
610	264673.0	14317.62	236611.0	292735.0



Outlier Detection Summary	
Maximum number searched	5
Number found	5
Significance used	0.05

Outlier Details				
Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
204	Additive	27351.3	26.87	<.0001
397	Additive	-26866.5	26.38	<.0001
542	Shift	-24840.3	21.05	<.0001
589	Additive	24109.4	17.38	<.0001
272	Additive	20932.0	16.02	<.0001

However, later we tried to run the model with SARIMA(1,1,1) and ARIMA(1,1,1) . Below is the graph and there is no white noise as well.

Name of Variable = Electricity Net Generation (Non	
Period(s) of Differencing	1,12
Mean of Working Series	0.40646
Standard Deviation	11188.79
Number of Observations	585
Observation(s) eliminated by differencing	13

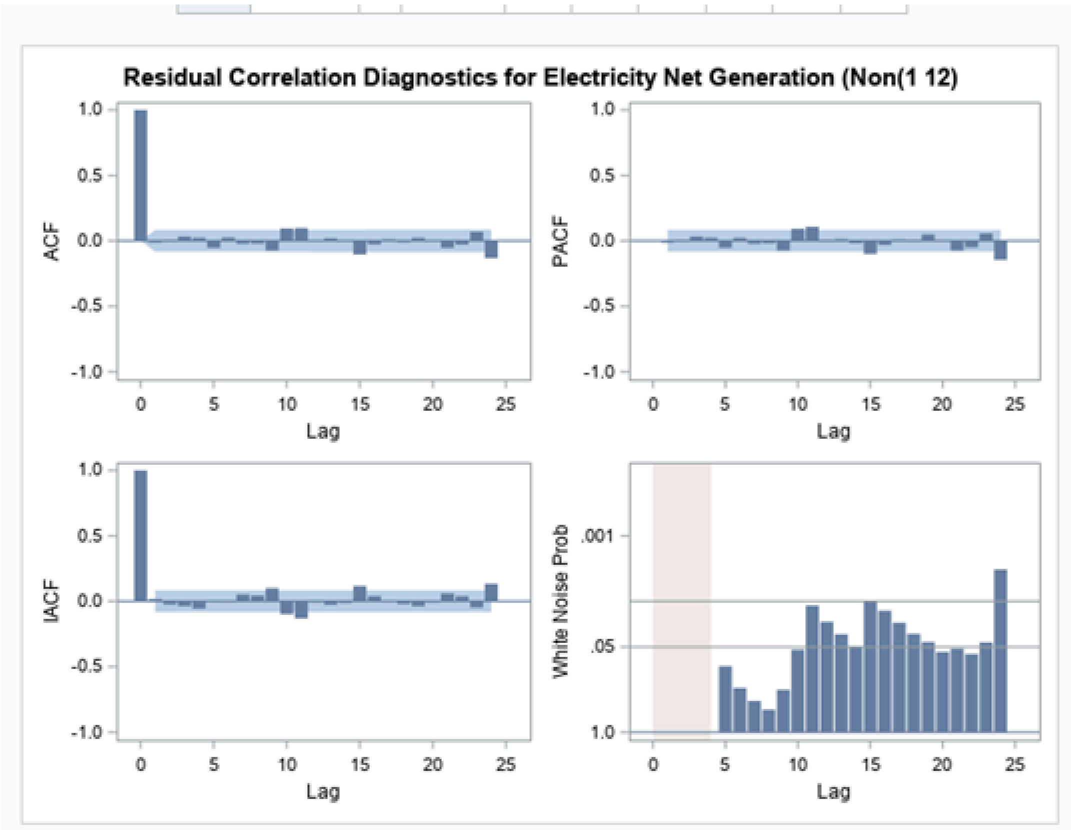
Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	81.22	6	<.0001	-0.316	-0.162	0.065	0.010	-0.082	0.029
12	193.79	12	<.0001	-0.016	-0.012	-0.070	0.156	0.133	-0.375
18	208.80	18	<.0001	0.106	0.060	-0.068	-0.002	0.065	-0.037
24	224.54	24	<.0001	0.009	0.048	0.005	-0.053	0.065	-0.128

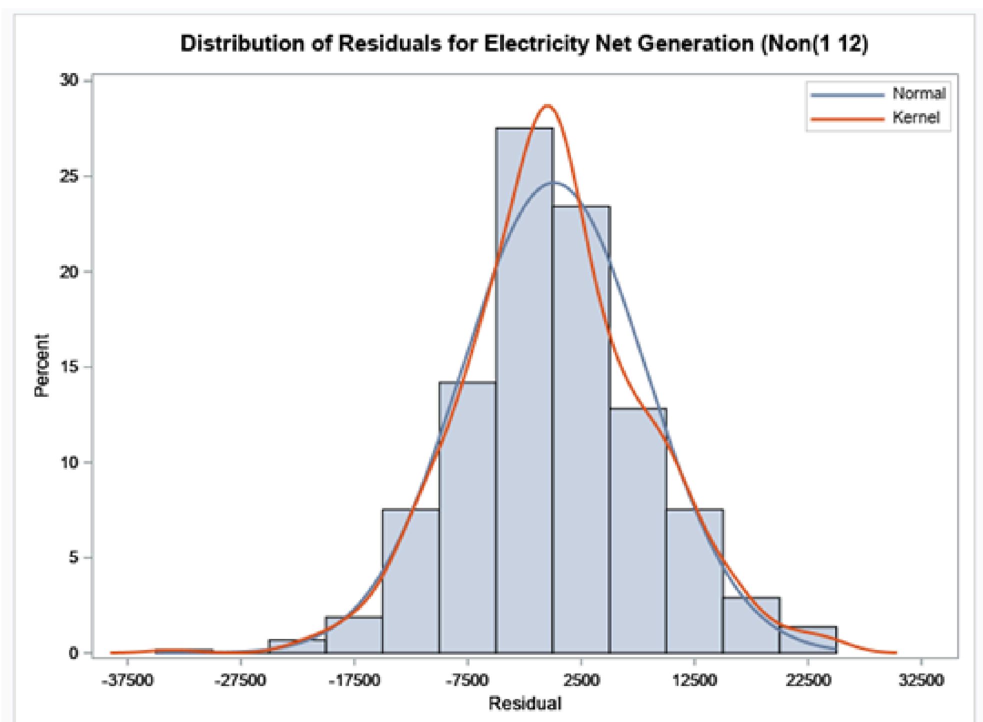
Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	-13.95422	15.19305	-0.92	0.3584	0
MA1,1	0.90774	0.02332	38.92	<.0001	1
MA2,1	0.77660	0.03756	20.68	<.0001	12
AR1,1	0.46311	0.04787	9.67	<.0001	1
AR2,1	0.07780	0.05734	1.36	0.1748	12

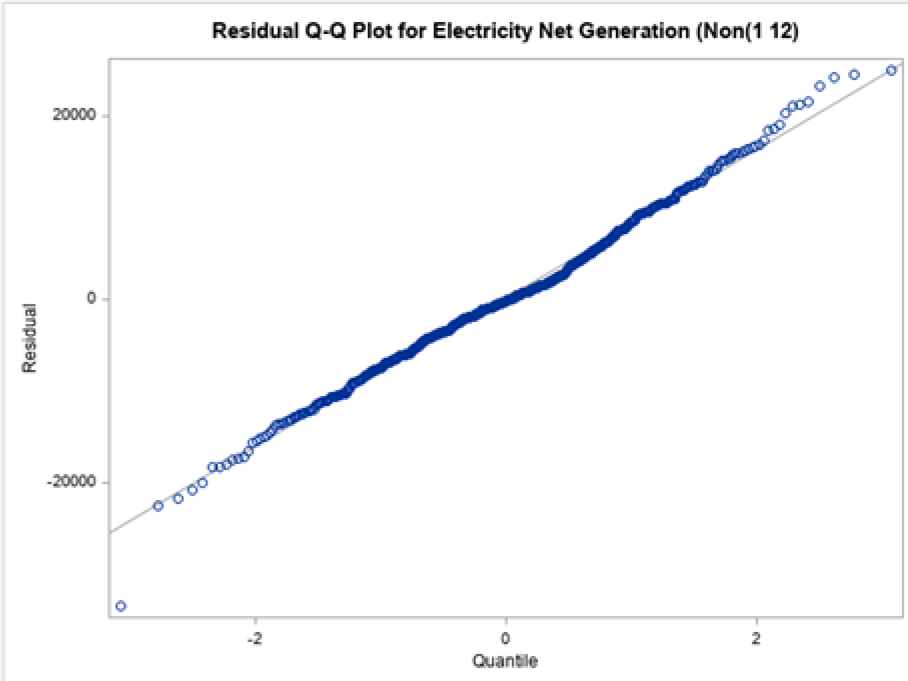
Constant Estimate	-6.90894
Variance Estimate	65855560
Std Error Estimate	8115.144
AIC	12207.94
SBC	12229.8
Number of Residuals	585

Correlations of Parameter Estimates					
Parameter	MU	MA1,1	MA2,1	AR1,1	AR2,1
MU	1.000	-0.029	-0.009	-0.022	0.000
MA1,1	-0.029	1.000	0.044	0.639	0.141
MA2,1	-0.009	0.044	1.000	0.070	0.670
AR1,1	-0.022	0.639	0.070	1.000	0.057
AR2,1	0.000	0.141	0.670	0.057	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	3.11	2	0.2110	-0.012	0.008	0.033	0.023	-0.053	0.025
12	18.08	8	0.0206	-0.022	-0.021	-0.074	0.095	0.098	0.003
18	25.32	14	0.0315	0.019	0.001	-0.104	-0.026	0.011	-0.010
24	41.44	20	0.0033	0.023	0.001	-0.053	-0.028	0.069	-0.132
30	44.09	26	0.0148	0.002	0.005	-0.058	-0.002	0.016	0.027
36	53.55	32	0.0098	0.093	-0.027	0.006	-0.006	0.054	-0.053
42	57.20	38	0.0235	0.033	-0.003	0.045	-0.042	-0.012	0.028
48	62.39	44	0.0353	-0.002	-0.037	-0.026	0.002	0.012	0.077


Residual Correlation Diagnostics for Electricity Net Generation (Non(1 12)
ACF
PACF
IACF
White Noise Prob
Lag

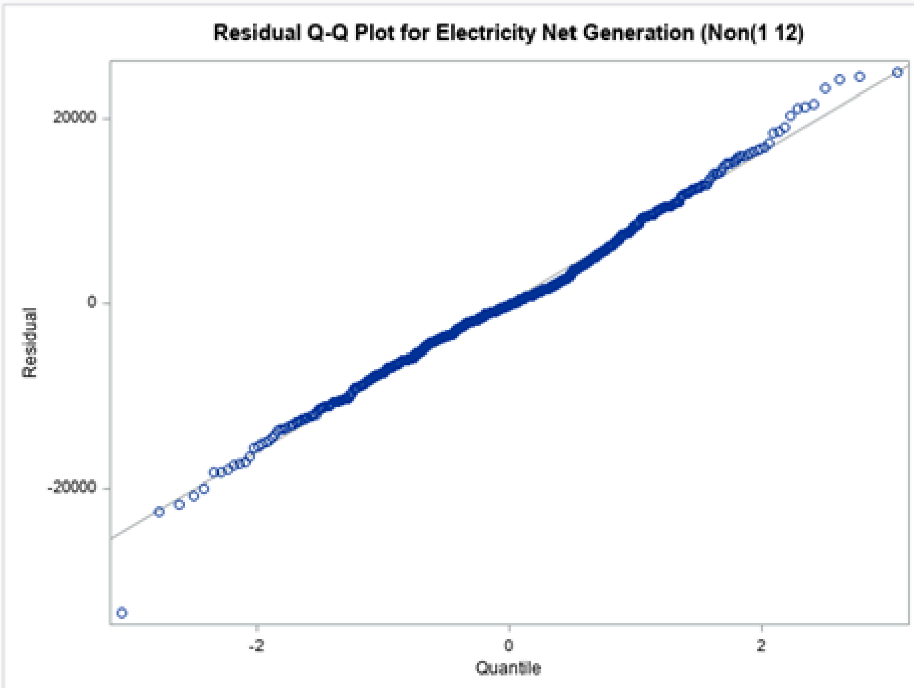

Distribution of Residuals for Electricity Net Generation (Non(1 12)
Normal
Kernel
Percent
Residual



Model for variable Electricity Net Generation (Non	
Estimated Mean	-13.9542
Period(s) of Differencing	1,12

Autoregressive Factors	
Factor 1:	1 - 0.46311 B**(1)
Factor 2:	1 - 0.0778 B**(12)

Moving Average Factors	
Factor 1:	1 - 0.90774 B**(1)
Factor 2:	1 - 0.7766 B**(12)

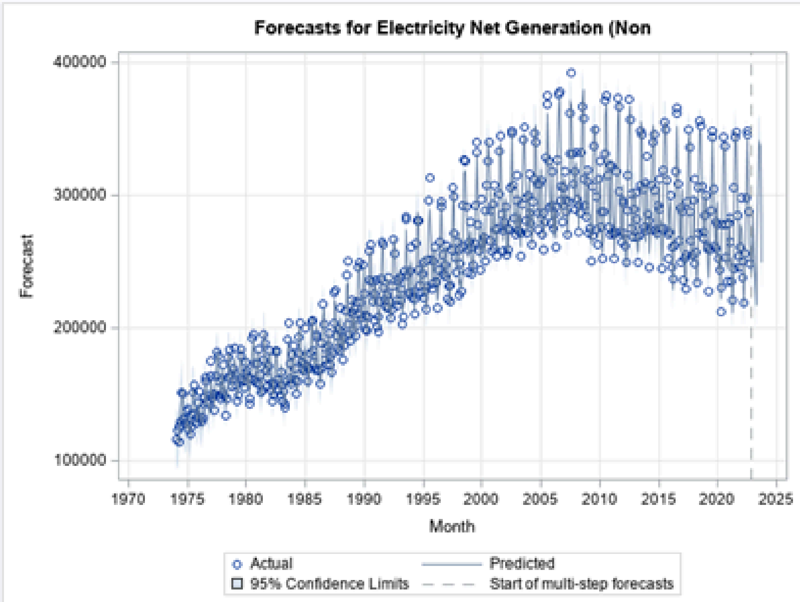


Model for variable Electricity Net Generation (Non	
Estimated Mean	-13.9542
Period(s) of Differencing	1,12

Autoregressive Factors	
Factor 1:	1 - 0.46311 B**(1)
Factor 2:	1 - 0.0778 B**(12)

Moving Average Factors	
Factor 1:	1 - 0.90774 B**(1)
Factor 2:	1 - 0.7766 B**(12)

Forecasts for variable Electricity Net Generation (Non				
Obs	Forecast	Std Error	95% Confidence Limits	
599	241518.1	8115.14	225612.7	257423.5
600	269151.4	9282.67	250957.7	287345.1
601	286728.4	9706.20	267704.6	305752.2
602	248420.8	9922.82	228972.4	267869.1
603	239628.3	10068.01	219895.3	259361.2
604	216904.3	10184.76	196942.5	236866.0
605	246529.1	10288.99	226363.0	266695.1
606	290155.3	10387.23	269796.7	310513.9
607	339638.8	10482.32	319093.8	360183.7
608	337267.9	10575.55	316540.2	357995.6
609	283104.4	10667.49	262196.5	304012.3
610	249243.1	10758.44	228156.9	270329.2



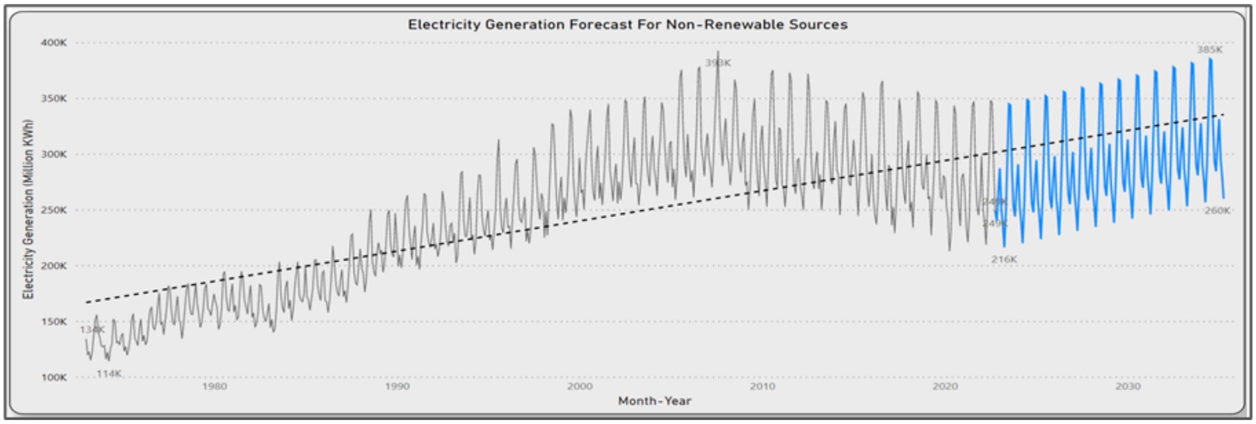
Outlier Detection Summary	
Maximum number searched	5
Number found	5
Significance used	0.05

Outlier Details

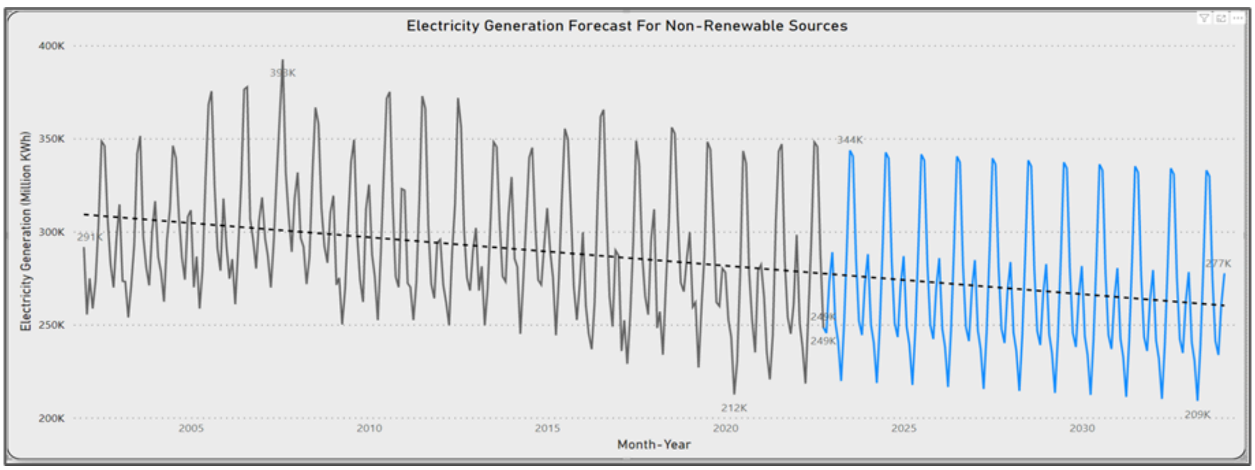
Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
204	Additive	27484.4	21.18	<.0001
397	Additive	-27457.9	22.00	<.0001
578	Additive	22305.7	14.25	0.0002
272	Additive	21284.2	13.77	0.0002
416	Additive	20827.4	13.21	0.0003

For ARIMA modeling, I have modeled the data based on the white noise, As there was still signal and the white noise was not going away, So I have tried to model based on that and in the end, SARIMA(1,1,1) and ARIMA(1,1,1) worked much better and even the AIC and SBC values were also better. The corelation between the lags was also significantly good.

For non-renewable energy, SARIMA(1,1,1) and ARIMA(1,1,1) worked better based on the AIC,SBC values. Below is the screenshot of forecasting from 1978 to 2022.

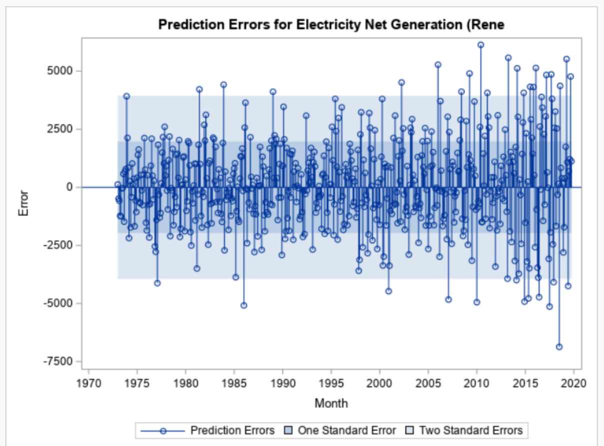


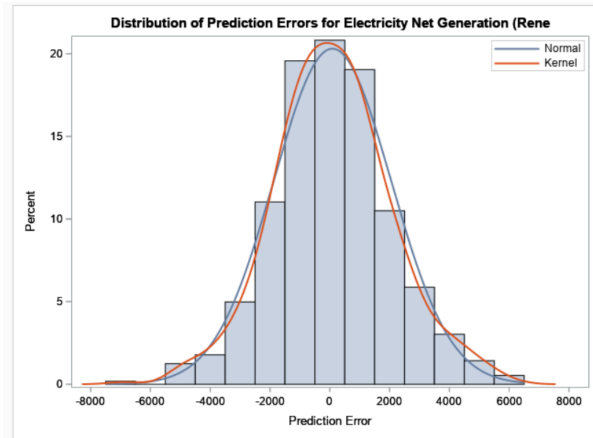
If we consider, last 20 years data, there is a decreasing trend for non-renewable energy.



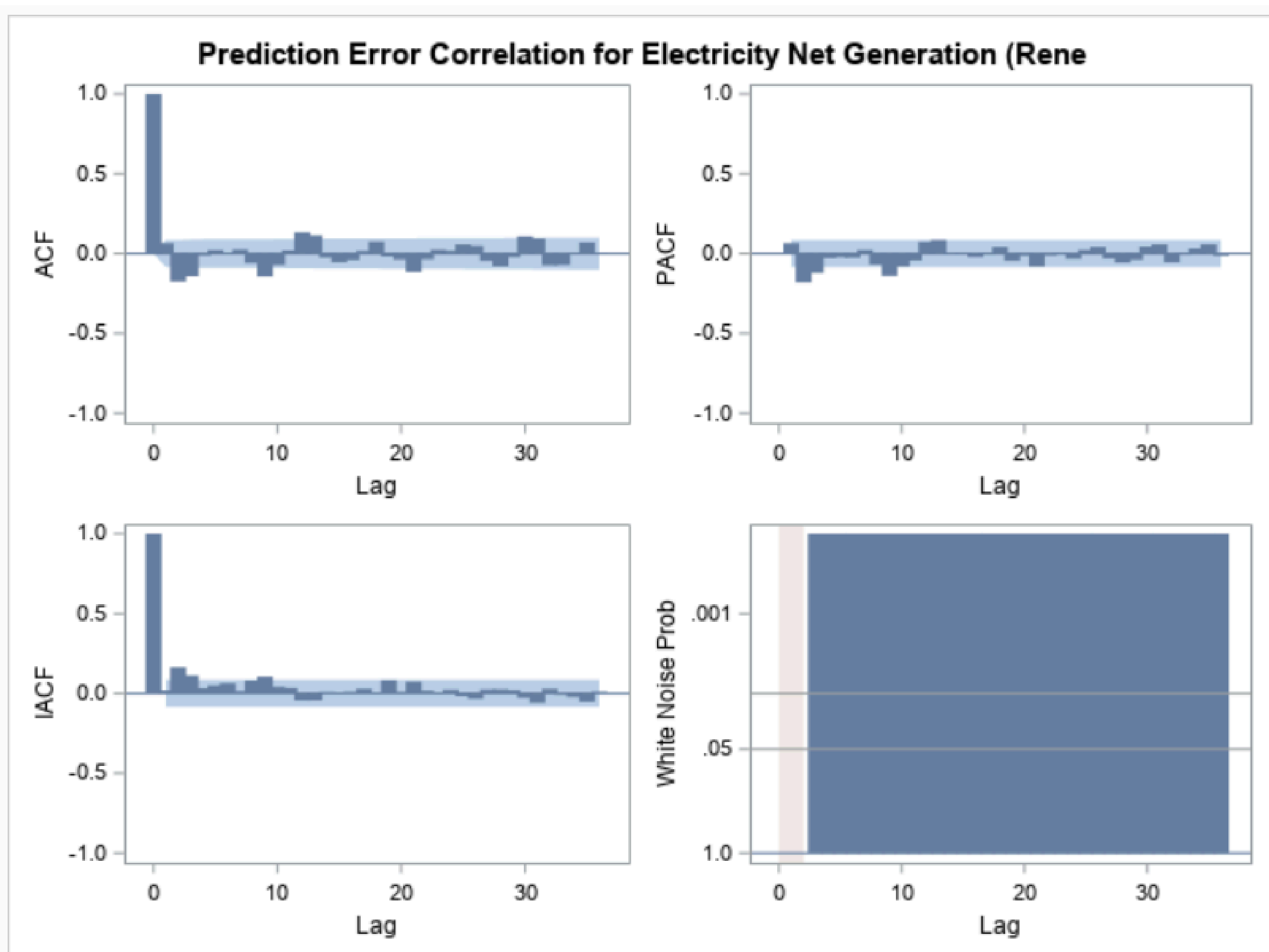
3.2. Renewable Resources

Below are the results of the Multiplicative Seasonal Exponential Smoothing model (MSESM).





Upon analysis of the results of the Multiplicative Seasonal model, it was found that the errors are distributed normally and the majority of errors are not correlated to any lags.



Also the model couldn't extract a significant amount of signal from the noise and hence the white noise probability test has failed.

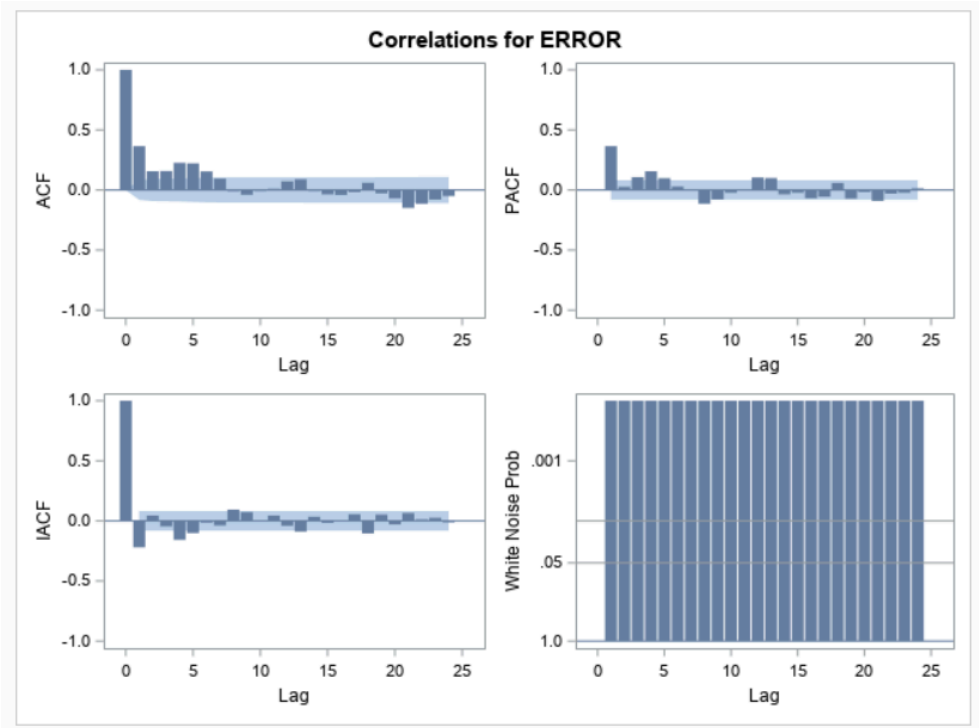
After generating the forecasted values for 36 months with a holdout sample of 36 months. The errors of the output of the Multiplicative Seasonal Exponential Smoothing model (MSESM) were fed to the ARIMA model for extracting significant signal from noise.

Before ARIMA modeling a stationarity test was performed on the output of MSESM. Below are the results of the stationarity test.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-373.215	0.0001	-16.32	<.0001		
	1	-347.906	0.0001	-12.95	<.0001		
	2	-258.941	0.0001	-10.03	<.0001		
Single Mean	0	-376.015	0.0001	-16.40	<.0001	134.57	0.0010
	1	-352.781	0.0001	-13.04	<.0001	85.07	0.0010
	2	-264.292	0.0001	-10.12	<.0001	51.24	0.0010
Trend	0	-382.125	0.0001	-16.60	<.0001	137.88	0.0010
	1	-363.405	0.0001	-13.25	<.0001	87.86	0.0010
	2	-275.972	0.0001	-10.33	<.0001	53.36	0.0010

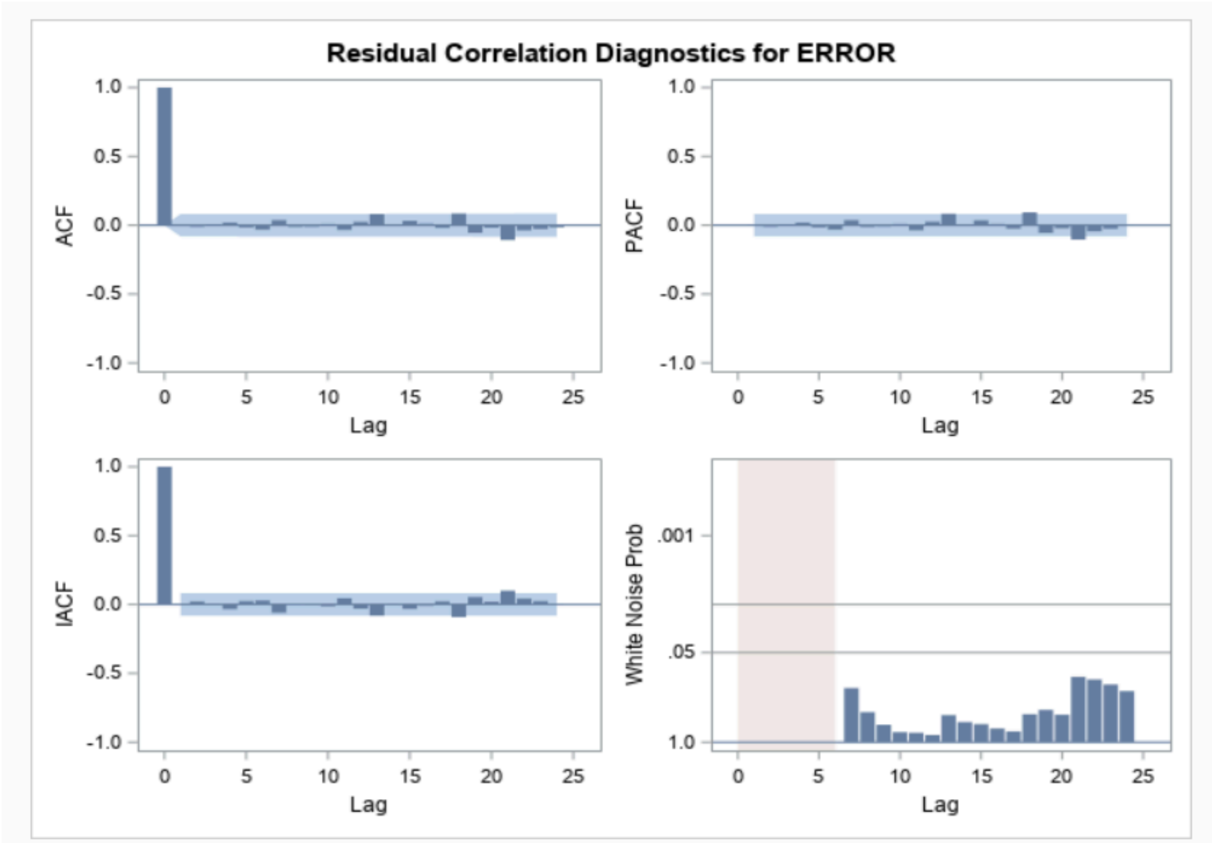
From the results of the Dickey-Fuller Unit Root Test it is confirmed that the model is stationary and we can proceed with ARIMA modeling.

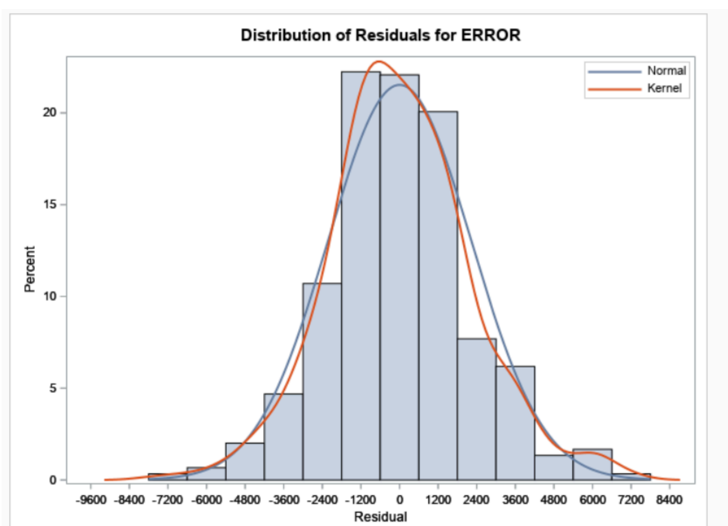
MSESM error autocorrelation plot:



From the error correlations plot of the Multiplicative Seasonal model the ACF, PACF and IACF seem to be having sinusoidal nature. Hence we go ahead and try to fit a seasonal ARIMA.

Upon various trial and error, it was found that SARIMA(3,0,3)(0,0,0) was the best fit for the errors. Below are the results of SARIMA(3,0,3)(0,0,0).





From the correlation plot it is clear that the errors are now distributed as white noise with no significant correlation with any of the lags. Also, the distribution is normal so the data points are independent.

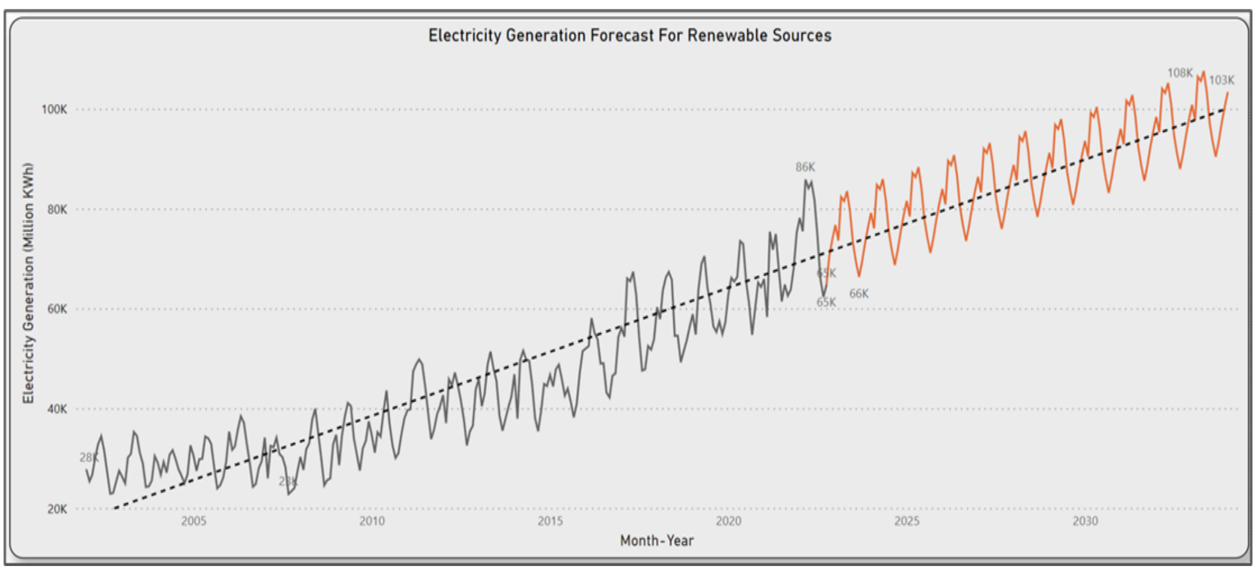
So, the final model for the renewable energy dataset was a combination of Seasonal Multiplicative Exponential Smoothing model and Seasonal ARIMA model. The out forecasted values of the total renewable energy generation from 2022– 2034 was the addition of forecasted values from SMESM and forecasted errors from SARIMA(3,0,3).

	Month	Electricity Net Generation (Rene
601	JAN2023	69562
602	FEB2023	63986
603	MAR2023	72801
604	APR2023	73132
605	MAY2023	74715
606	JUN2023	72300
607	JUL2023	67256
608	AUG2023	61121
609	SEP2023	55324
610	OCT2023	57244
611	NOV2023	59494
612	DEC2023	65358
613	JAN2024	69562
614	FEB2024	63986
615	MAR2024	72801
616	APR2024	73132
617	MAY2024	74715

Total rows: 634 Total columns: 6

	ERROR	FORECAST
601	.	242.78089206
602	.	241.02213865
603	.	240.87511477
604	.	241.98774875
605	.	243.16103386
606	.	243.49293619
607	.	242.97244973
608	.	242.23355853
609	.	241.89397145
610	.	242.09479439
611	.	242.53026878
612	.	242.80536593
613	.	242.75854994
614	.	242.51671902
615	.	242.31738672
616	.	242.29834714

Forecast:

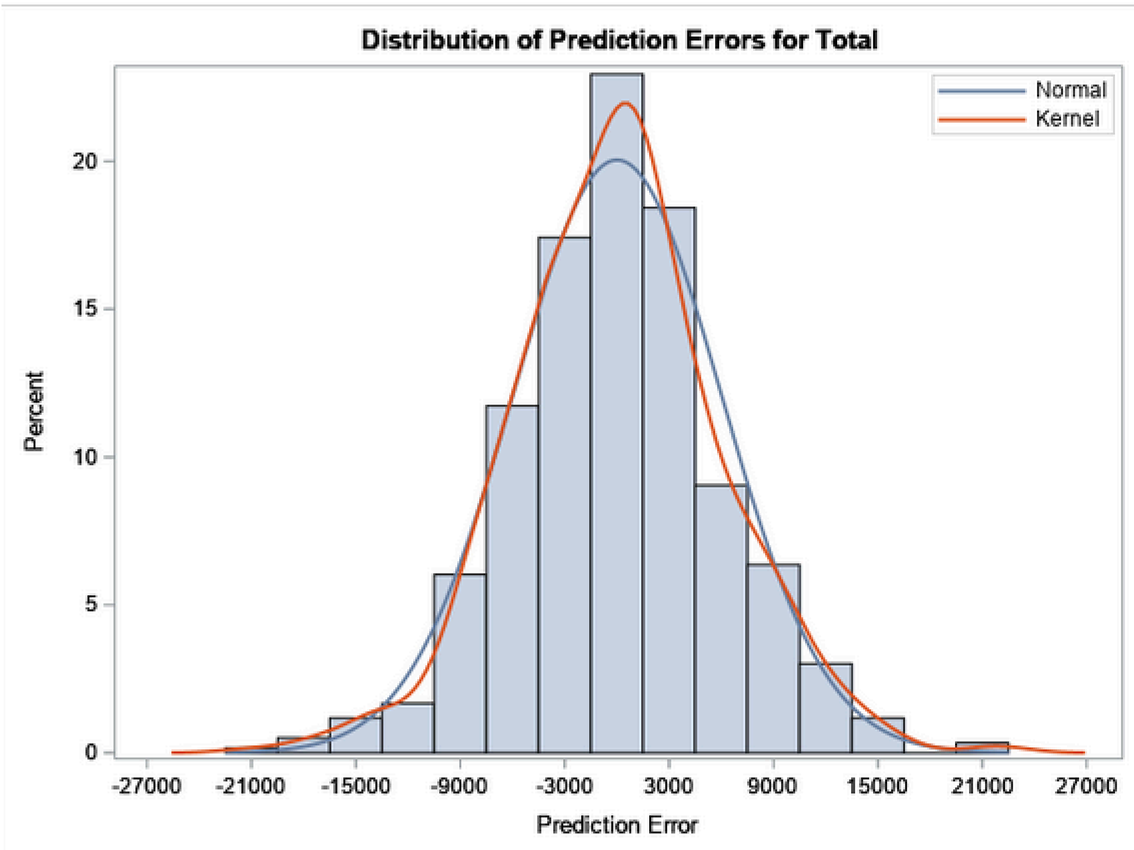


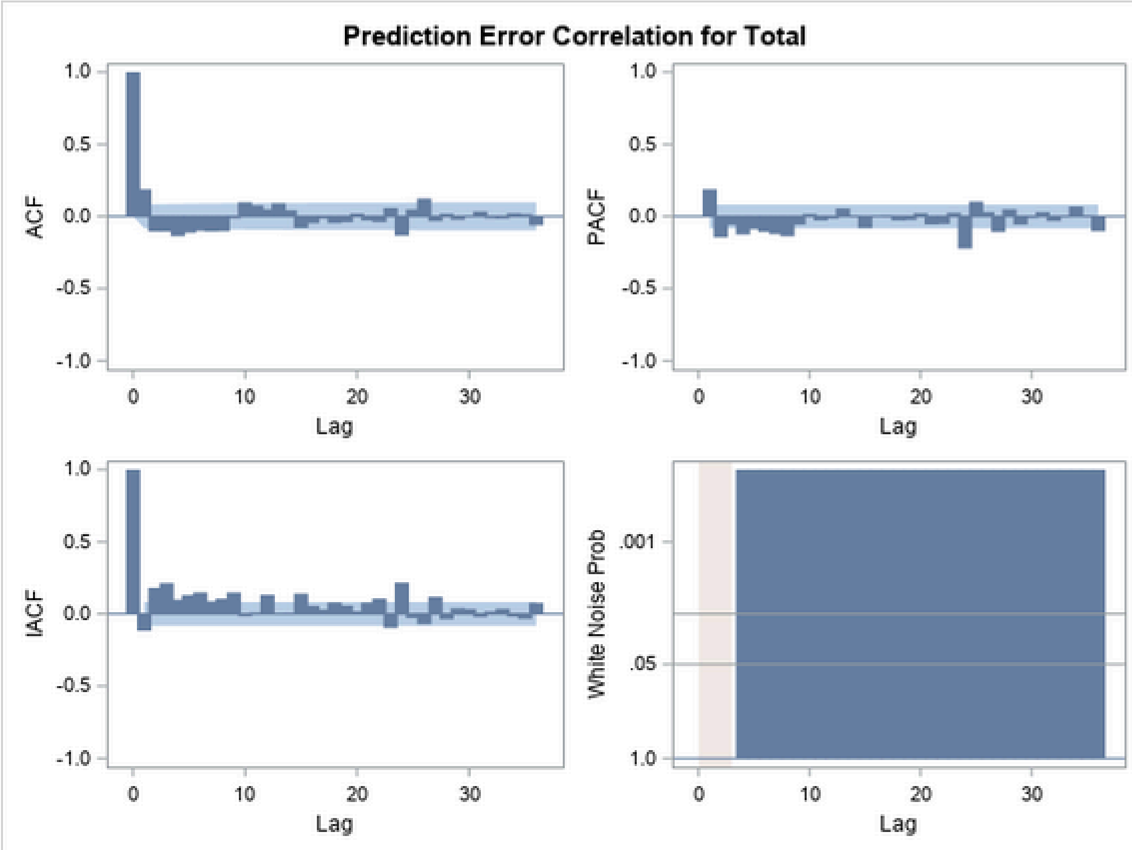
3.3. ELECTRICITY END USE:

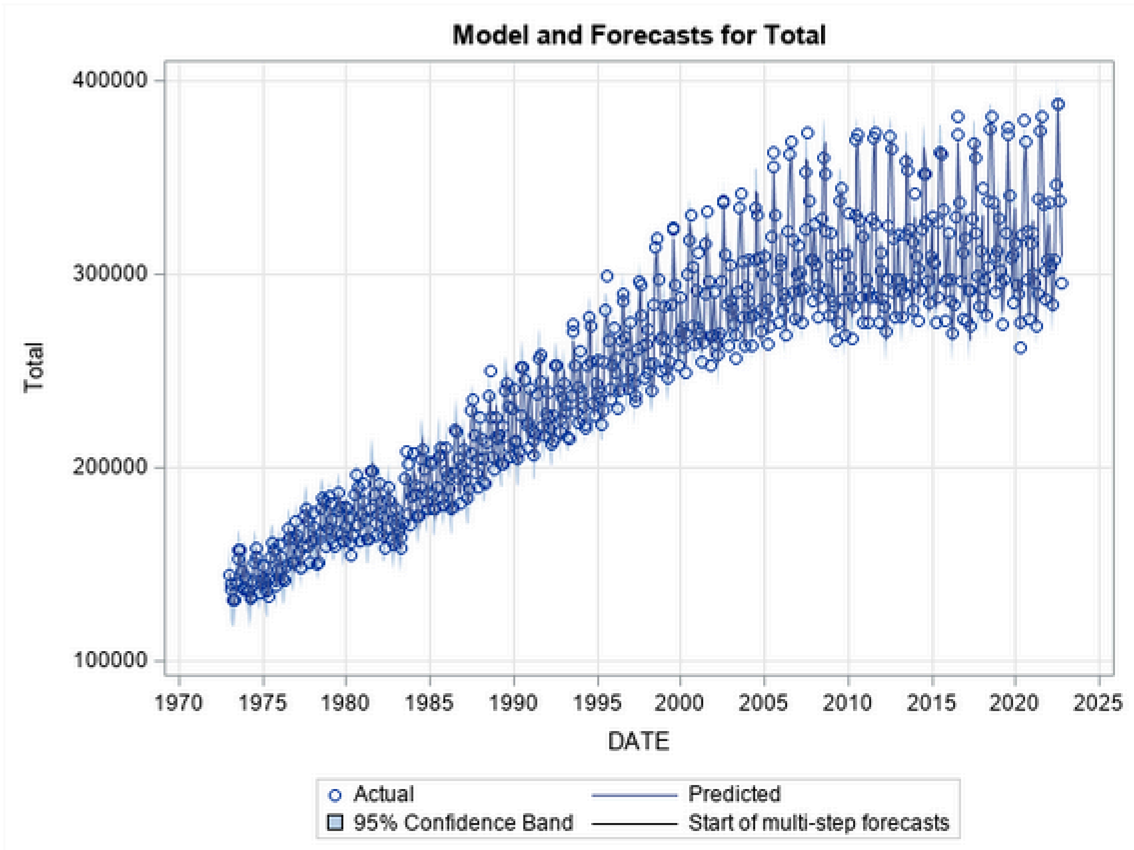
As we observed both trend and seasonality in this data, we will first try to fit winter's multiplicative model to capture both trend and seasonality.

Winters Multiplicative Model:

After modeling the data using the winters multiplicative model, the results are shown below.




Prediction Error Correlation for Total
ACF
PACF
IACF
White Noise Prob
Lag
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
10
20
30



Winters Multiplicative Model Fit Statistics:

	MSE	RMSE	MAPE	MAE
1	35590461.343	5965.7741613	1.8016241103	4588.4093792

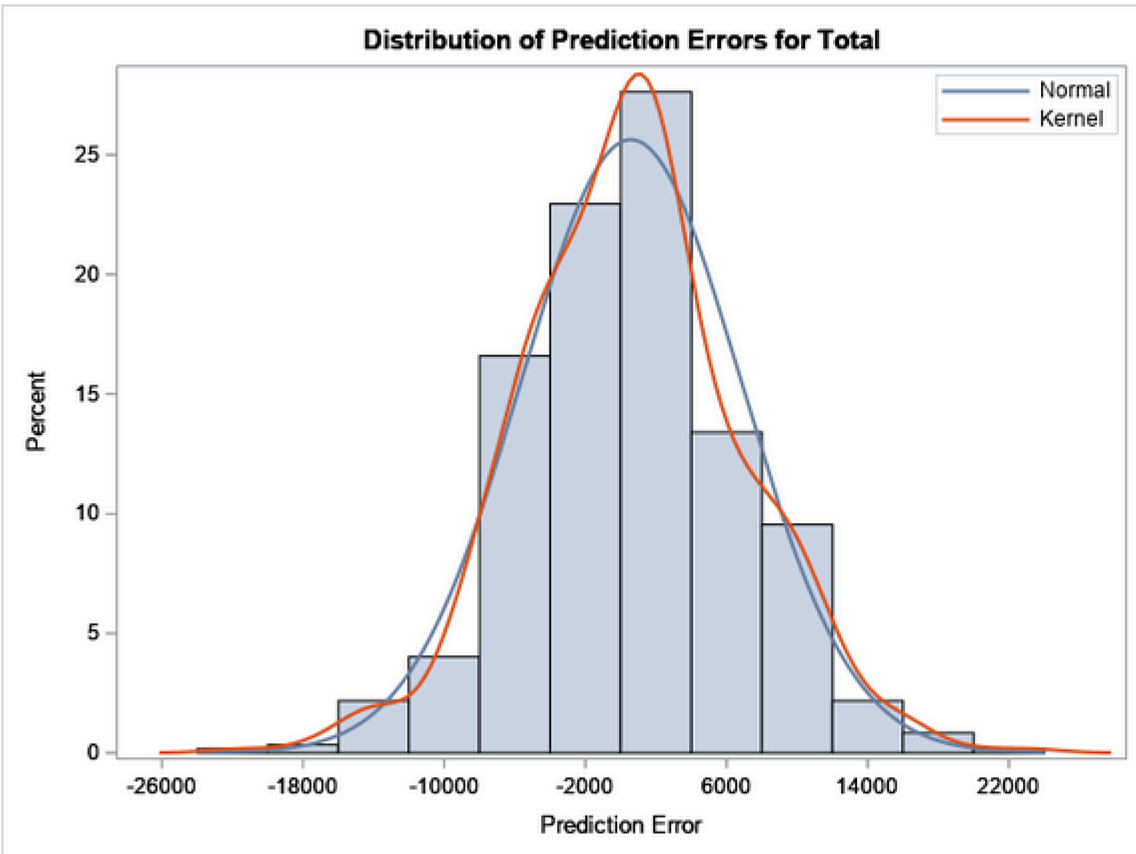
Winters Multiplicative Model Parameter Estimates:

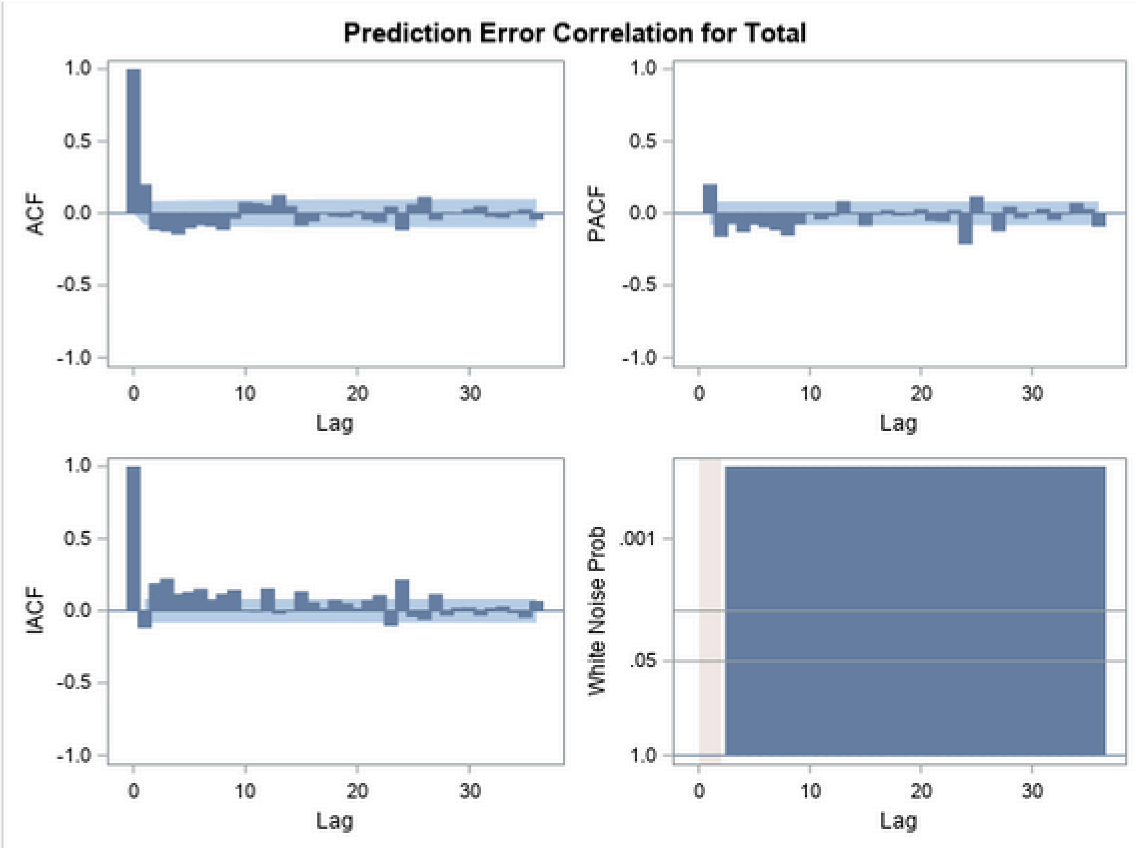
	MODEL	_PARM_	_EST_	_PVALUE_
1	WINTERS	LEVEL	0.5078640817	3.027399E-84
2	WINTERS	TREND	0.001	0.8517258576
3	WINTERS	SEASON	0.4850739162	2.718494E-33

From the model fit statistics, we can see that the model has performed so well as the MAPE and MAE are low. But from the parameter estimates, we can see that the p value of trend parameter estimate is not significant as it is greater than 0.05. This means that the trend component is not playing a major role in forecasting the data.

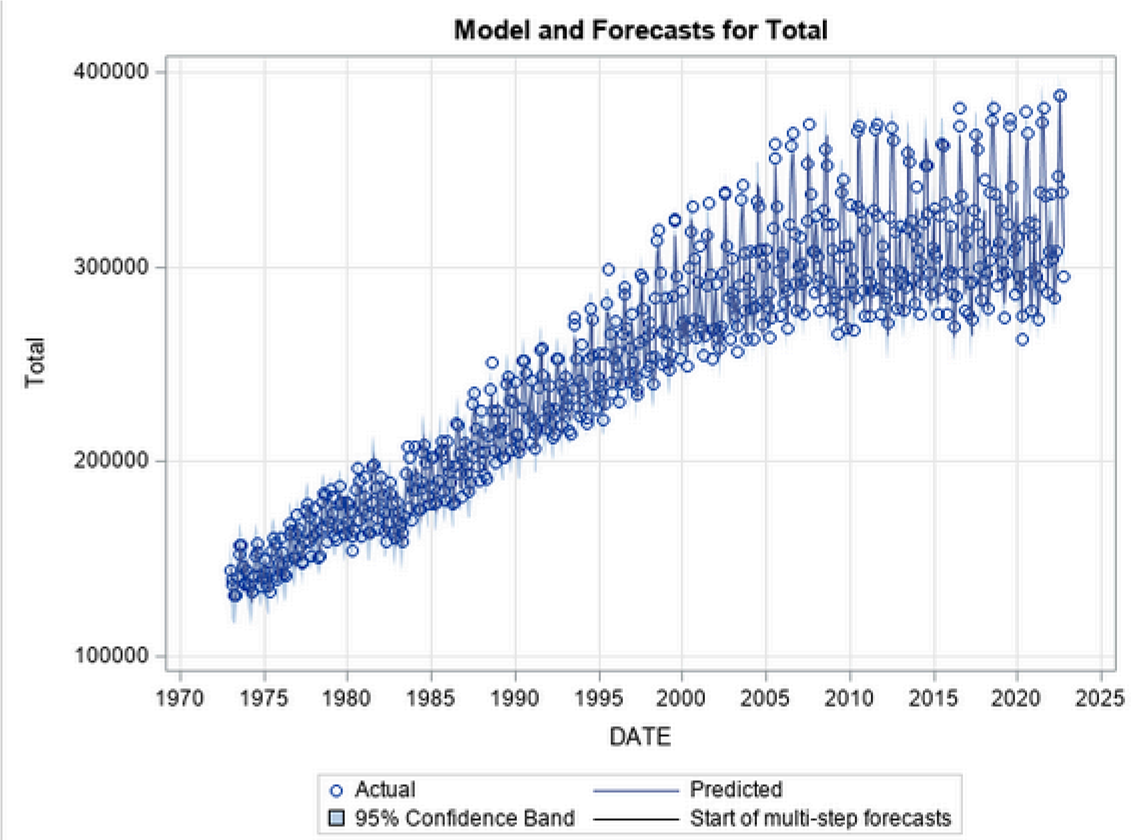
Hence we will try to fit the Additive Seasonal Model to the data.

Additive Seasonal Model:





Prediction Error Correlation for Total
ACF
PACF
IACF
White Noise Prob
Lag
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
10
20
30



Additive Seasonal Model Fit Statistics:

	MODEL	_PARM_	_EST_	_PVALUE_
1	SEASONAL	LEVEL	0.5026667751	3.9442E-86
2	SEASONAL	SEASON	0.600684614	1.536565E-37

Additive Seasonal Model Parameter Estimates:

	MSE	RMSE	MAPE	MAE
1	39019768.946	6246.5805803	1.9021764153	4836.8183457

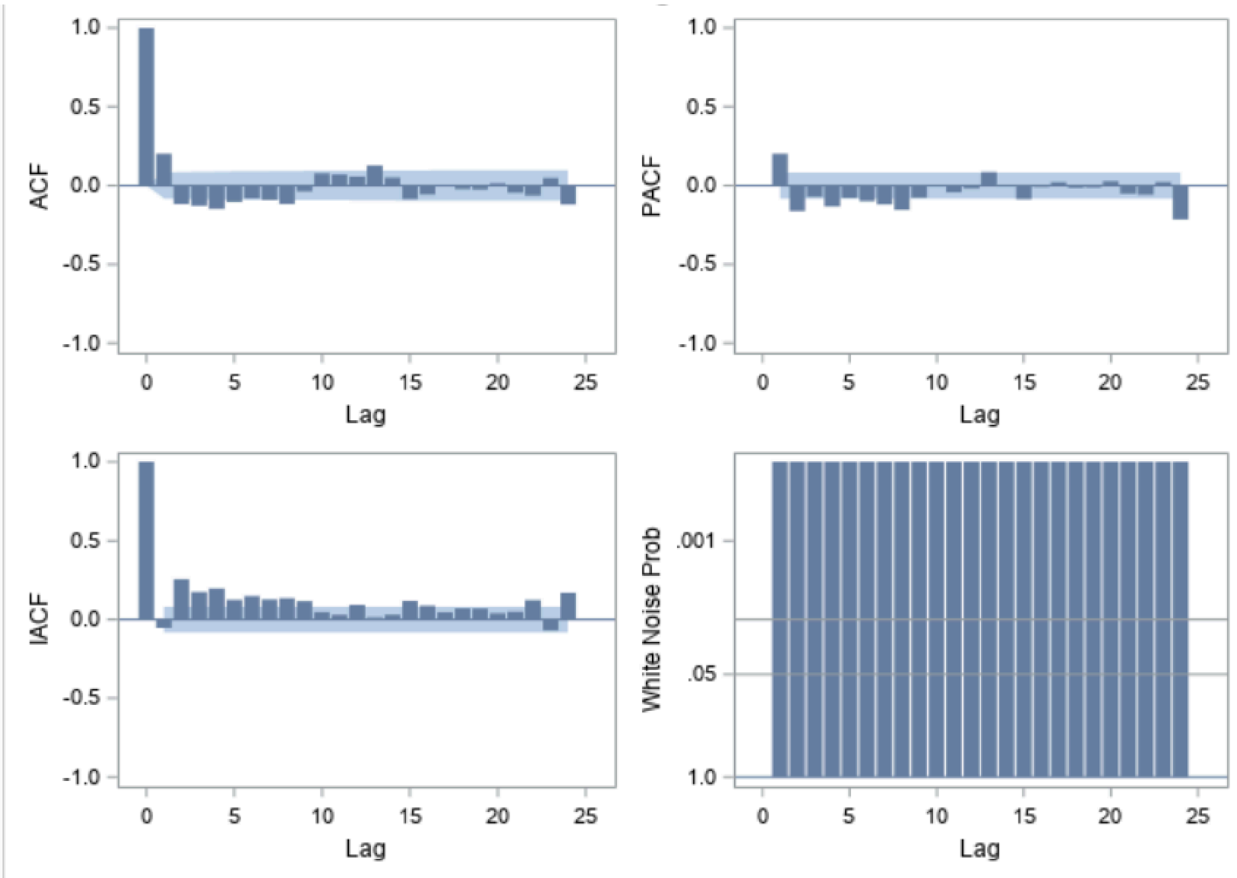
Modeling Errors From Additive Seasonal Model:

To perform ARIMA Modeling, we first need to check stationarity of the data.

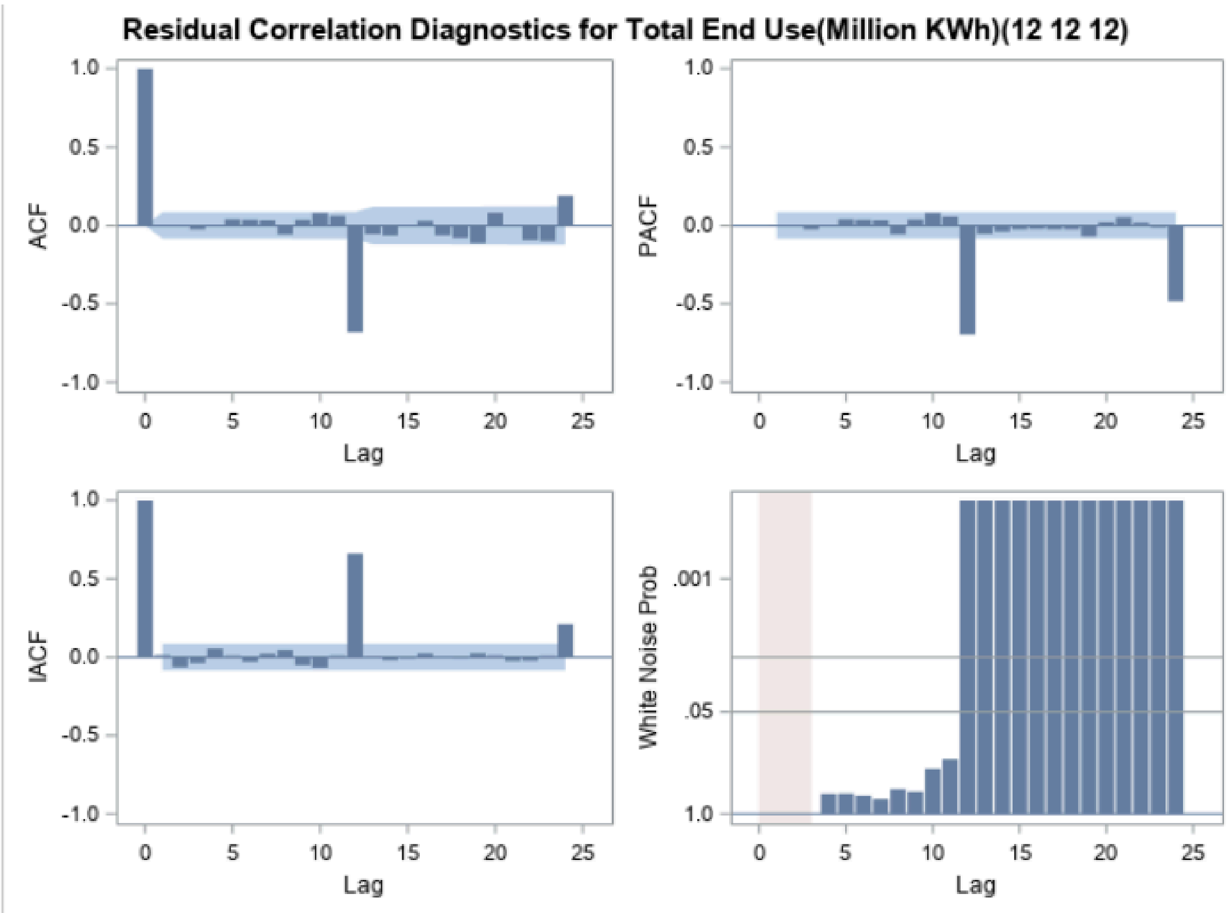
Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-1.3373	0.4187	-0.70	0.4159		
	1	-3.1875	0.2199	-1.18	0.2176		
	2	-0.7667	0.5153	-0.46	0.5155		
	3	0.2254	0.7367	0.23	0.7531		
Single Mean	0	-29.0444	0.0018	-3.93	0.0021	7.78	0.0010
	1	-60.3502	0.0018	-5.56	<.0001	15.48	0.0010
	2	-24.5692	0.0038	-3.54	0.0075	6.37	0.0024
	3	-9.0236	0.1693	-2.22	0.1991	2.78	0.3560
Trend	0	-162.728	0.0001	-9.60	<.0001	46.14	0.0010
	1	-534.970	0.0001	-16.15	<.0001	130.39	0.0010
	2	-310.566	0.0001	-10.99	<.0001	60.45	0.0010
	3	-107.370	0.0001	-6.69	<.0001	22.42	0.0010

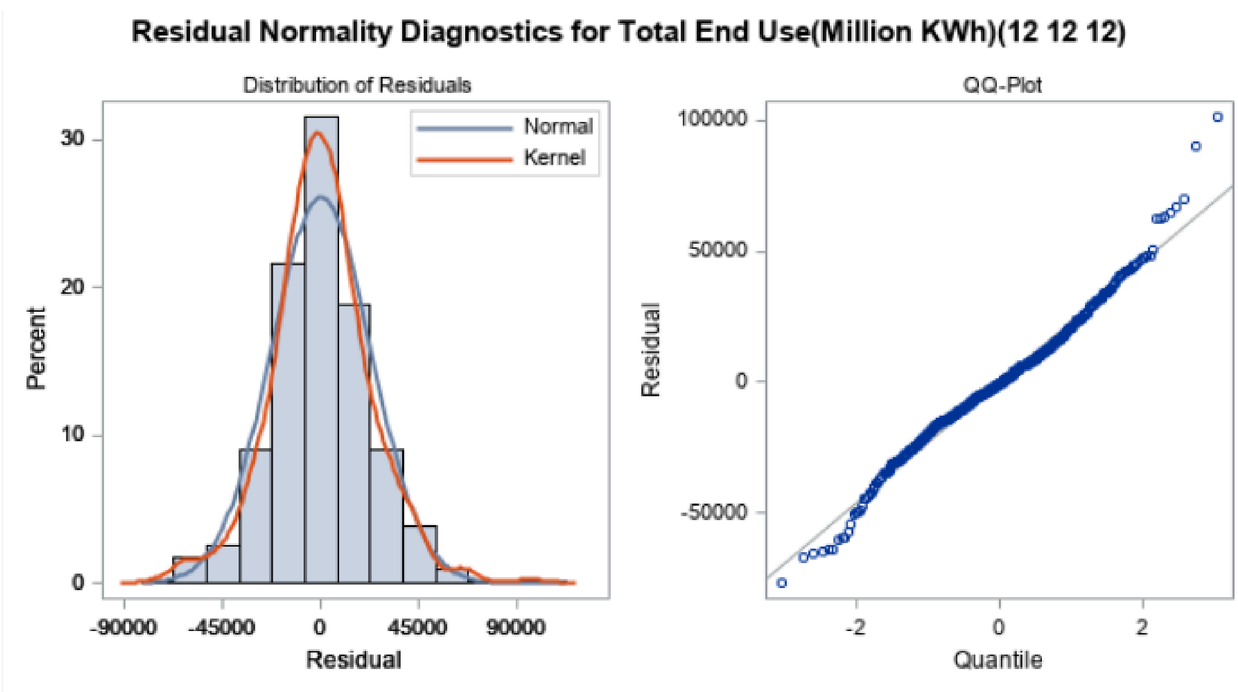
We can conclude that the data is stationary.

Correlation Analysis:



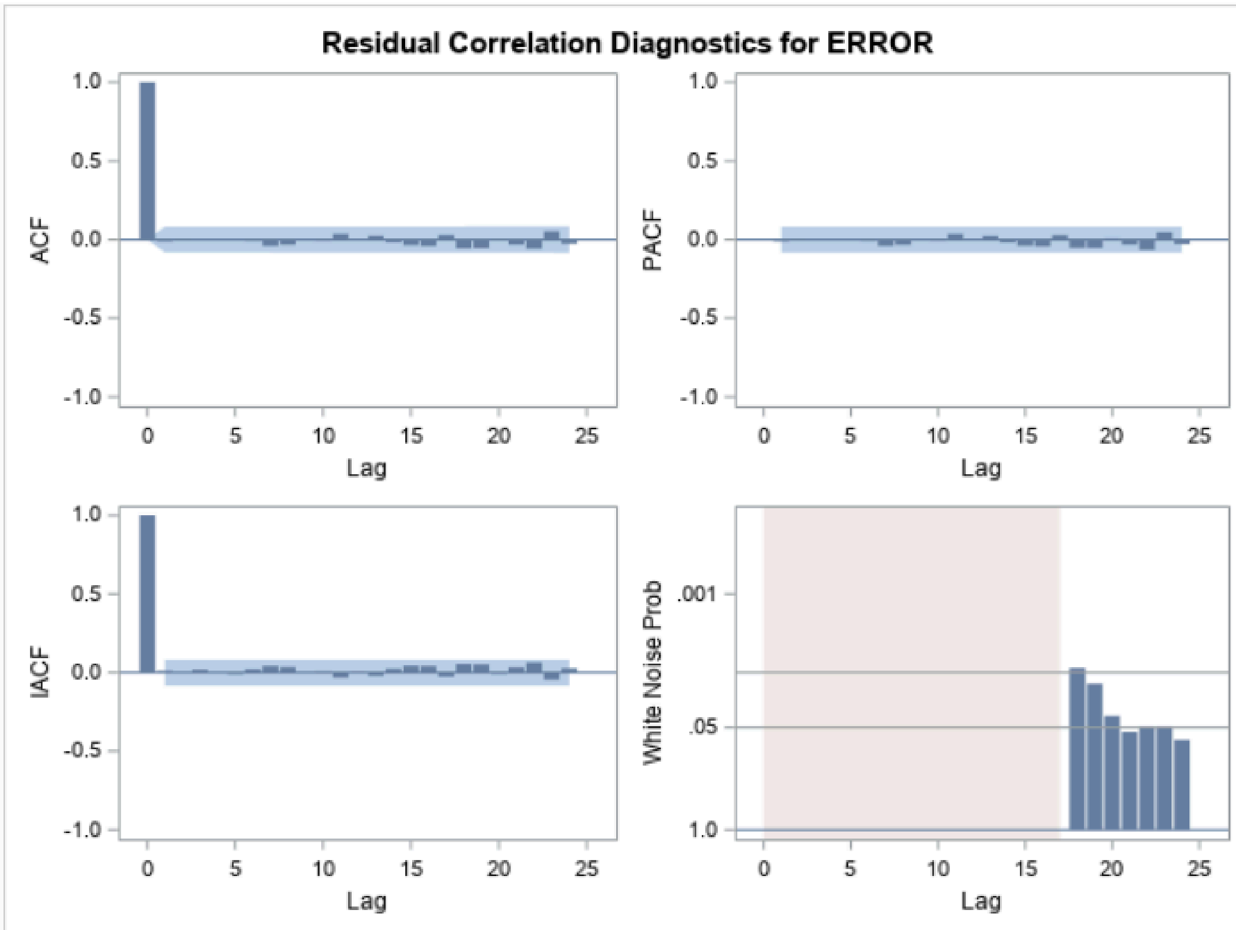
Fitting ARIMA (4,0,4):

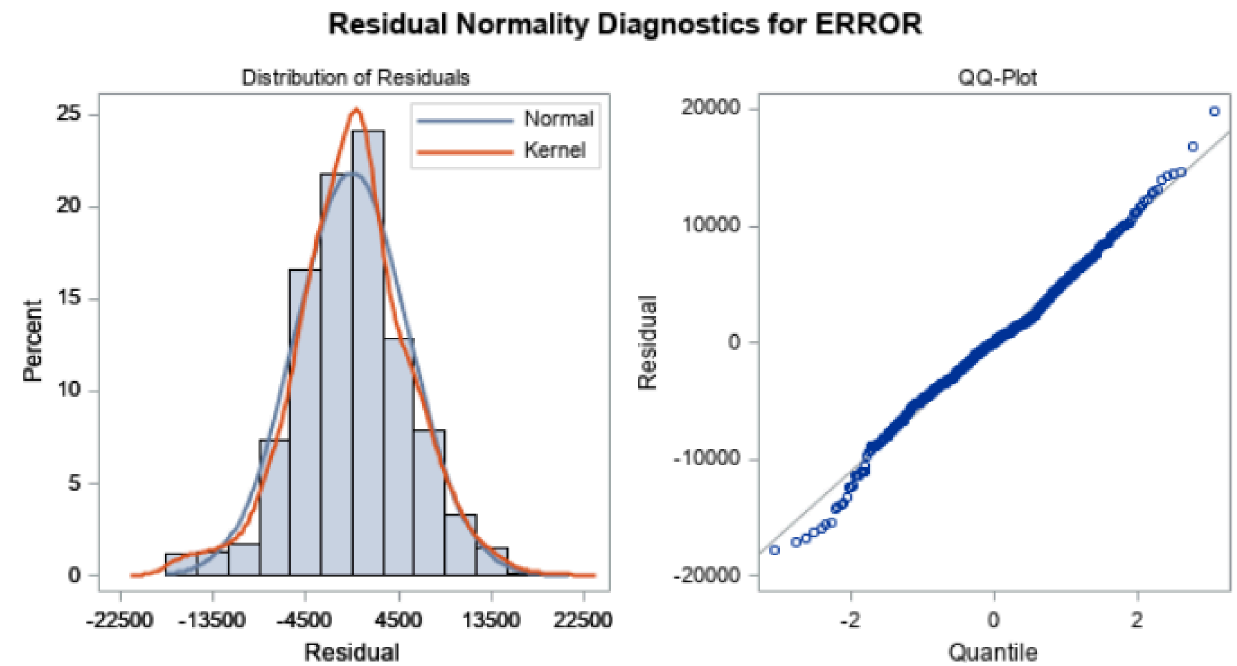

Residual Correlation Diagnostics for Total End Use(Million KWh)(12 12 12)
ACF
PACF
IACF
White Noise Prob
Lag
1.0
0.5
0.0
-0.5
-1.0
.001
.05
0
5
10
15
20
25


Residual Normality Diagnostics for Total End Use(Million KWh)(12 12 12)
Distribution of Residuals
Normal
Kernel
Percent
Residual
-90000
-45000
0
45000
90000
QQ-Plot
100000
50000
-50000
Quantile
-2
2

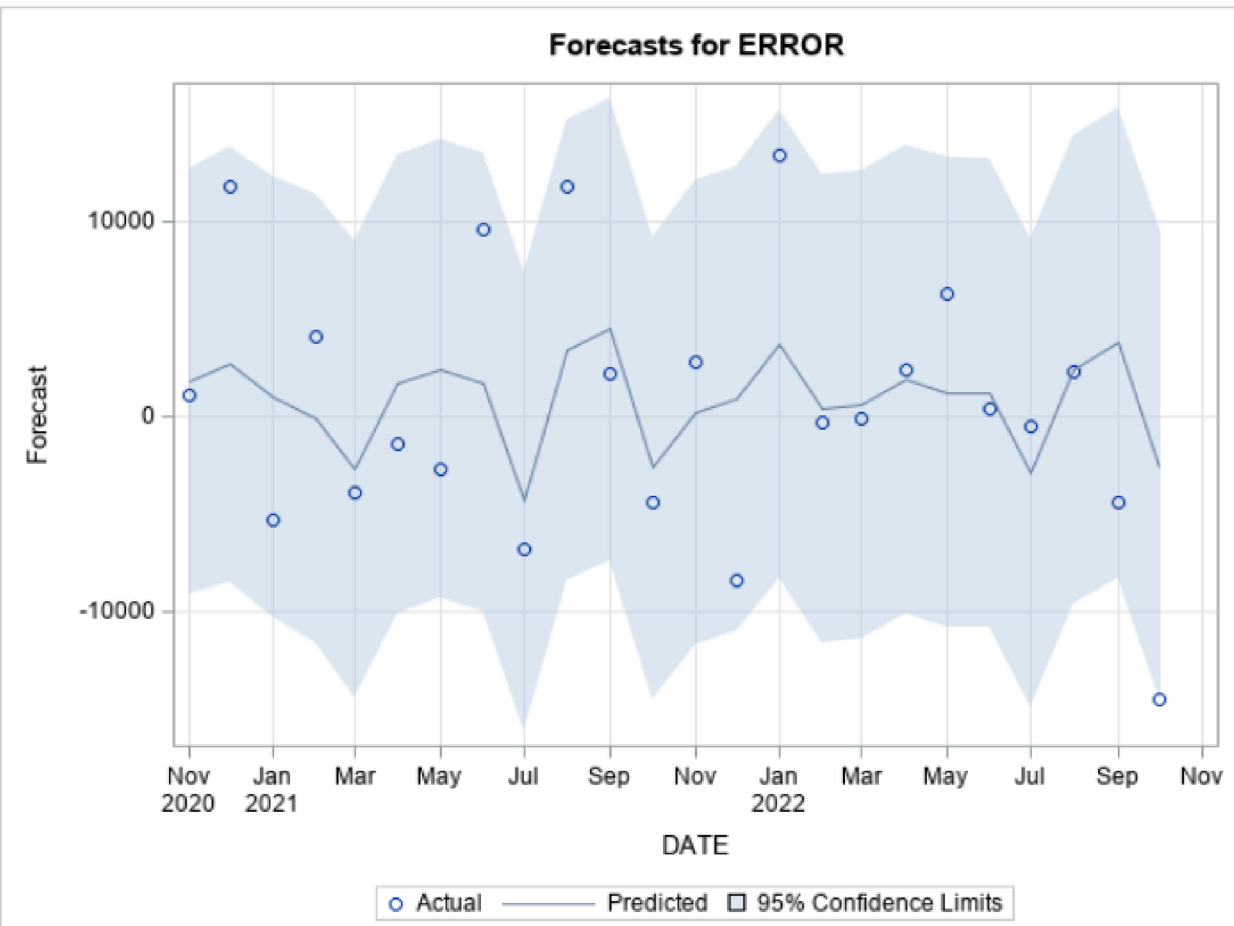
From the residual plots we can see that the residuals are almost normally distributed. But from the correlation and white noise test, we can see that there is still a significant signal present in the residuals.

Fitting ARIMA (8,0,8) Model:





Residual Normality Diagnostics for ERROR
Distribution of Residuals
Normal
Kernel
Percent
Residual
-22500
-13500
-4500
4500
13500
22500
QQ-Plot
Residual
Quantile

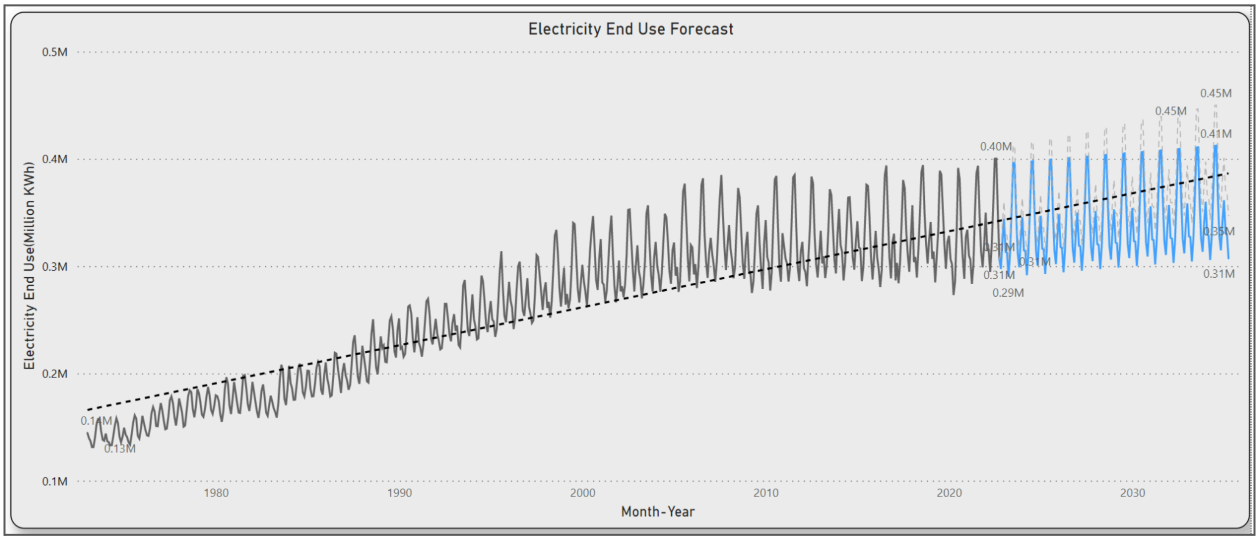


Forecasts for ERROR
Forecast
DATE
Nov 2020
Jan 2021
Mar
May
Jul
Sep
Nov
Jan 2022
Mar
May
Jul
Sep
Nov
Actual
Predicted
95% Confidence Limits

The correlation plots and white noise test indicate that the ARIMA(8,0,8) model has almost nearly matched the error data, however the forecast plot reveals that the forecasted values fall far short of the actual values. From the model statistics, we can note that the MAPE is larger than 100%, indicating that the errors are "far greater" than the actual values, implying a forecast accuracy of zero or a forecast that is significantly incorrect.

To model the total electricity end-use data, therefore, additive seasonal modeling is recommended. The MAPE value of 1.98 for additive seasonal modeling corresponds to an almost 98% forecast accuracy.

Total End Use Forecasting:



From the above forecast, we can conclude that the electricity demand is increasing with a short positive slope. We can assume that the forecast is also following the same seasonality as the previous years.

3.4. TOTAL ELECTRICITY GENERATION

Model Selection:

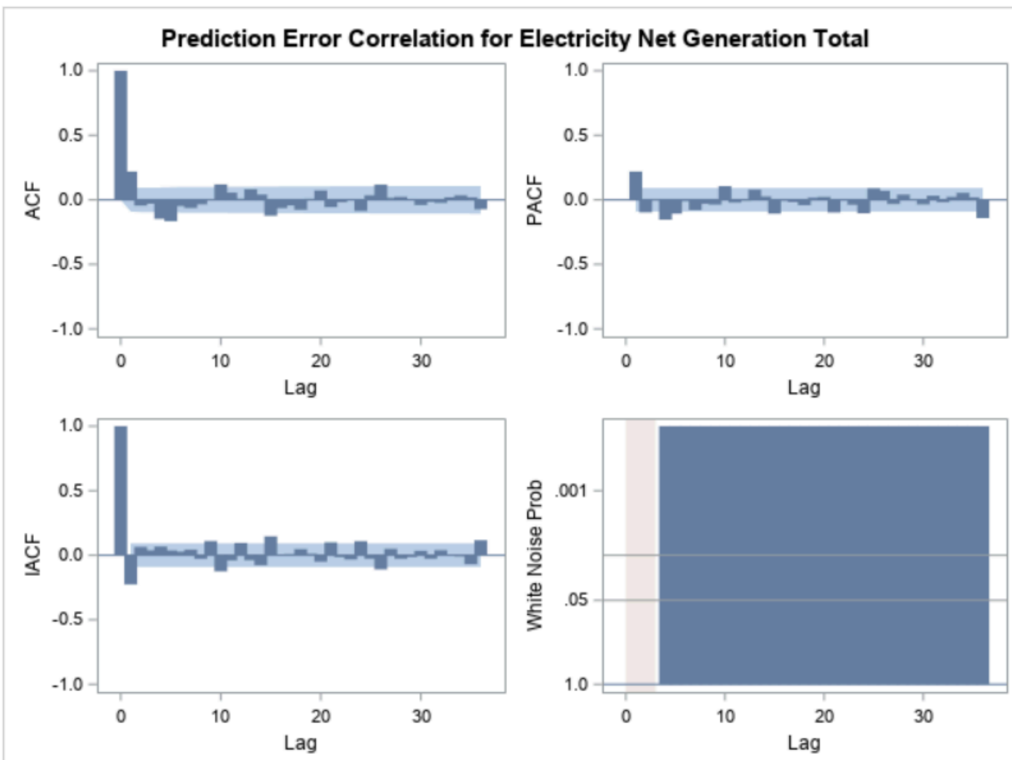
Considering the trend and seasonality for the total energy generation both trend and seasonality are evident so we can go ahead by performing the winters additive and the multiplicative smoothing models as per the conditions

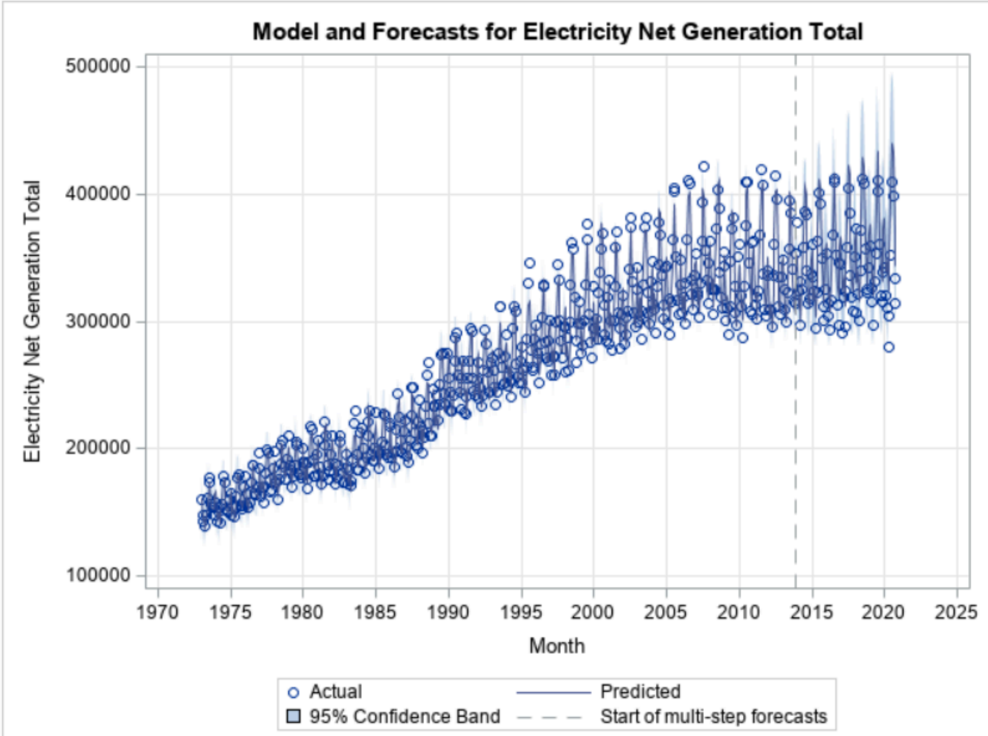
Winters Multiplicative Model:

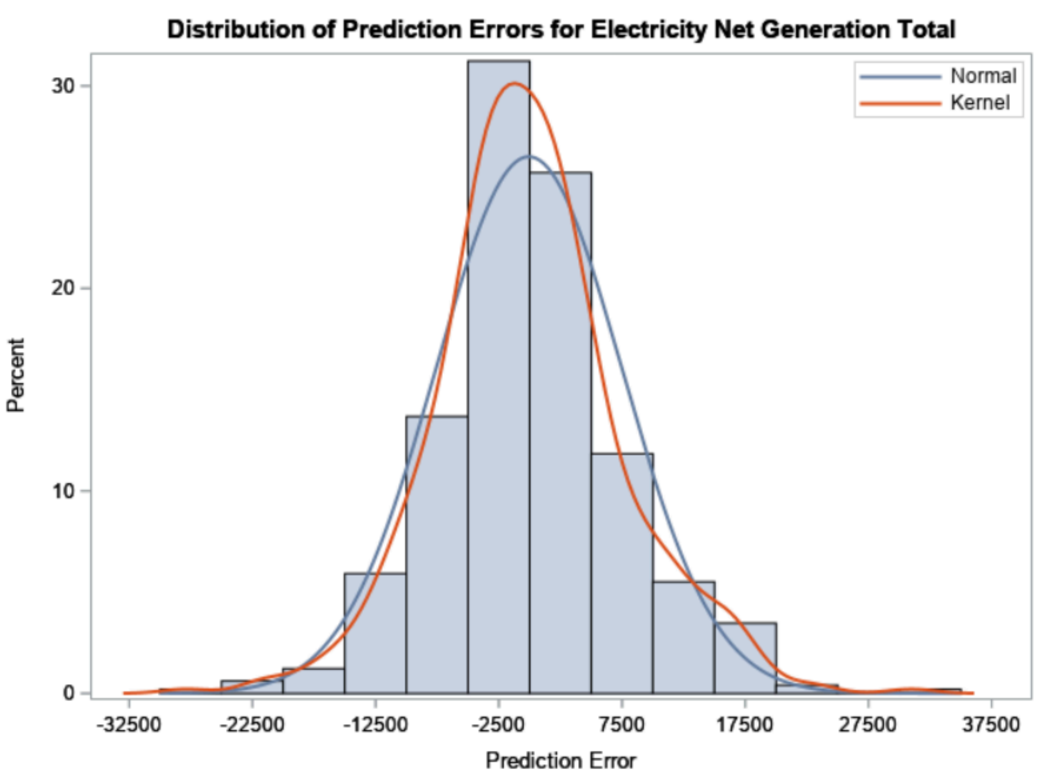
The analysis reveals the presence of both trend and seasonality in the data, despite their lack of stability. Nonetheless, we conducted model performance tests to identify the best fit. We observed stable correlations but found that the error is not disturbed as white noise, and there exists a signal in the errors that needs to be captured.

The forecast plots indicate that the predictions are well-contained within the confidence interval. To assess the accuracy of the model, we calculated the Mean Absolute Percentage Error (MAPE) value from the statistical output dataset and derived the accuracy percentage by subtracting the MAPE from 100.

Best Model Accuracy Percentage = 95%







RMSE	MAPE	MAE	AIC	SBC
7518.0851128	2.1435874321	5686.9208023	8752.5654104	8765.1486265
20171.431333	5.0127372145	16806.953587	1665.2197954	1665.2197954

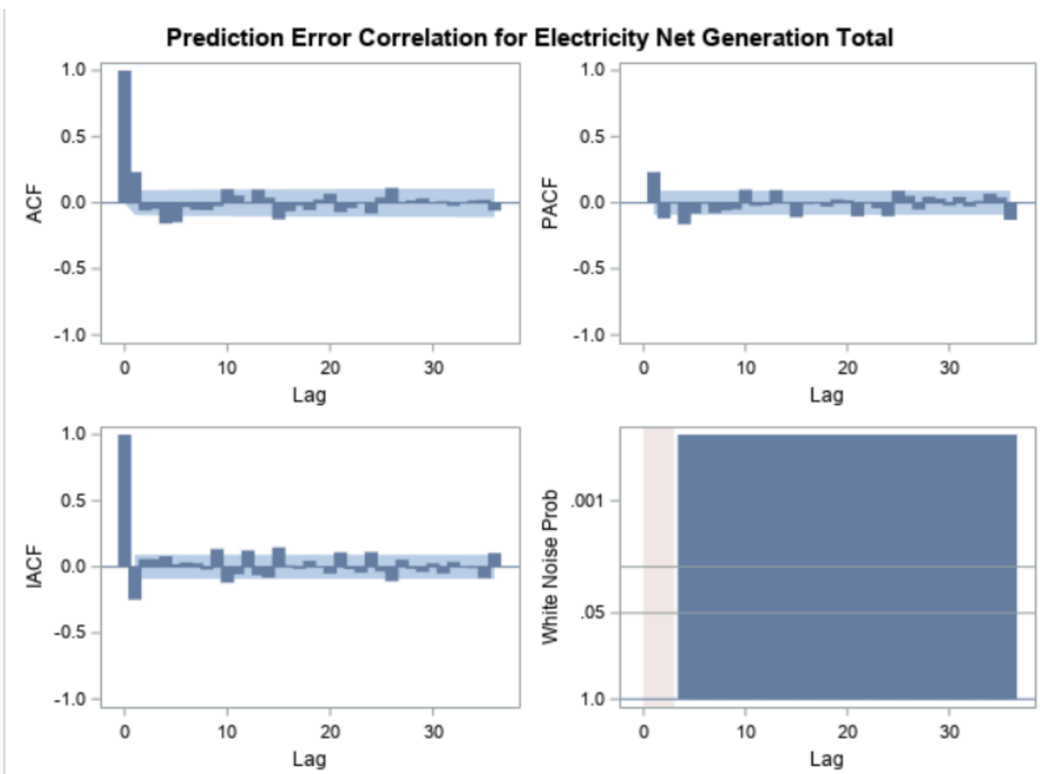
Winters Additive Model:

If there are both trends and seasonality even though it is not stable we can do the model performance and test the best fit.

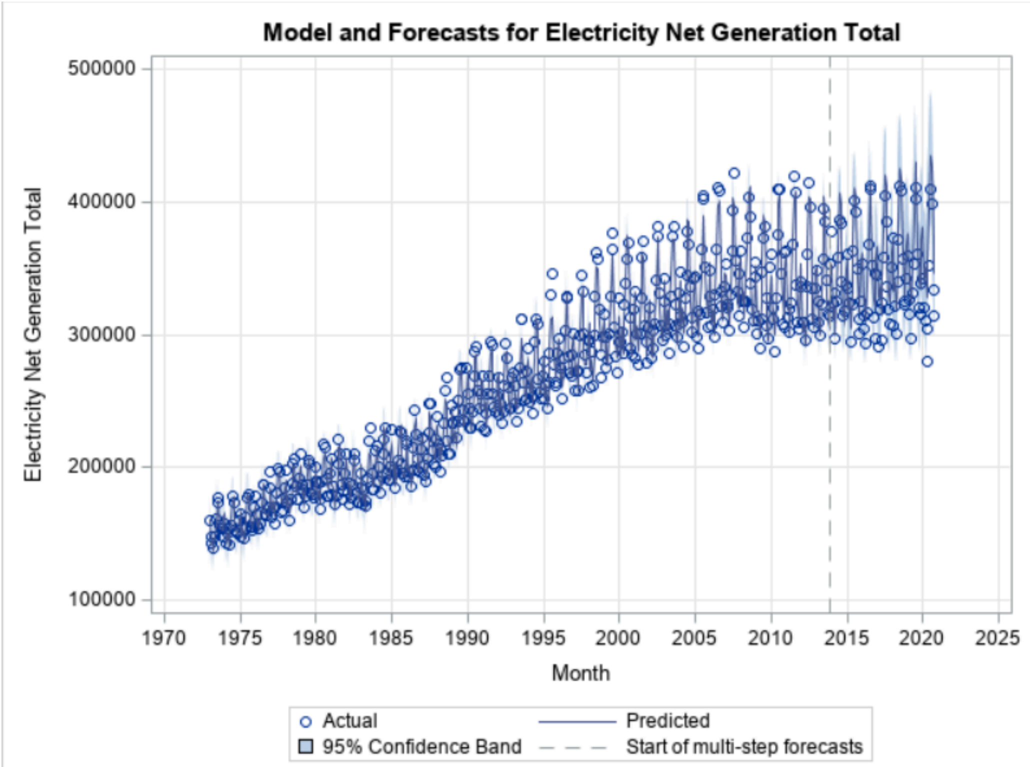
There are correlations which are stable and the error is not distributed as white noise and still signals the errors to capture. The forecast plots are merging and present inside the confidence interval.

The accuracy percentage is achieved by looking at the MAPE value in the statistics output dataset by subtracting it from 100.

Accuracy percentage = 95%

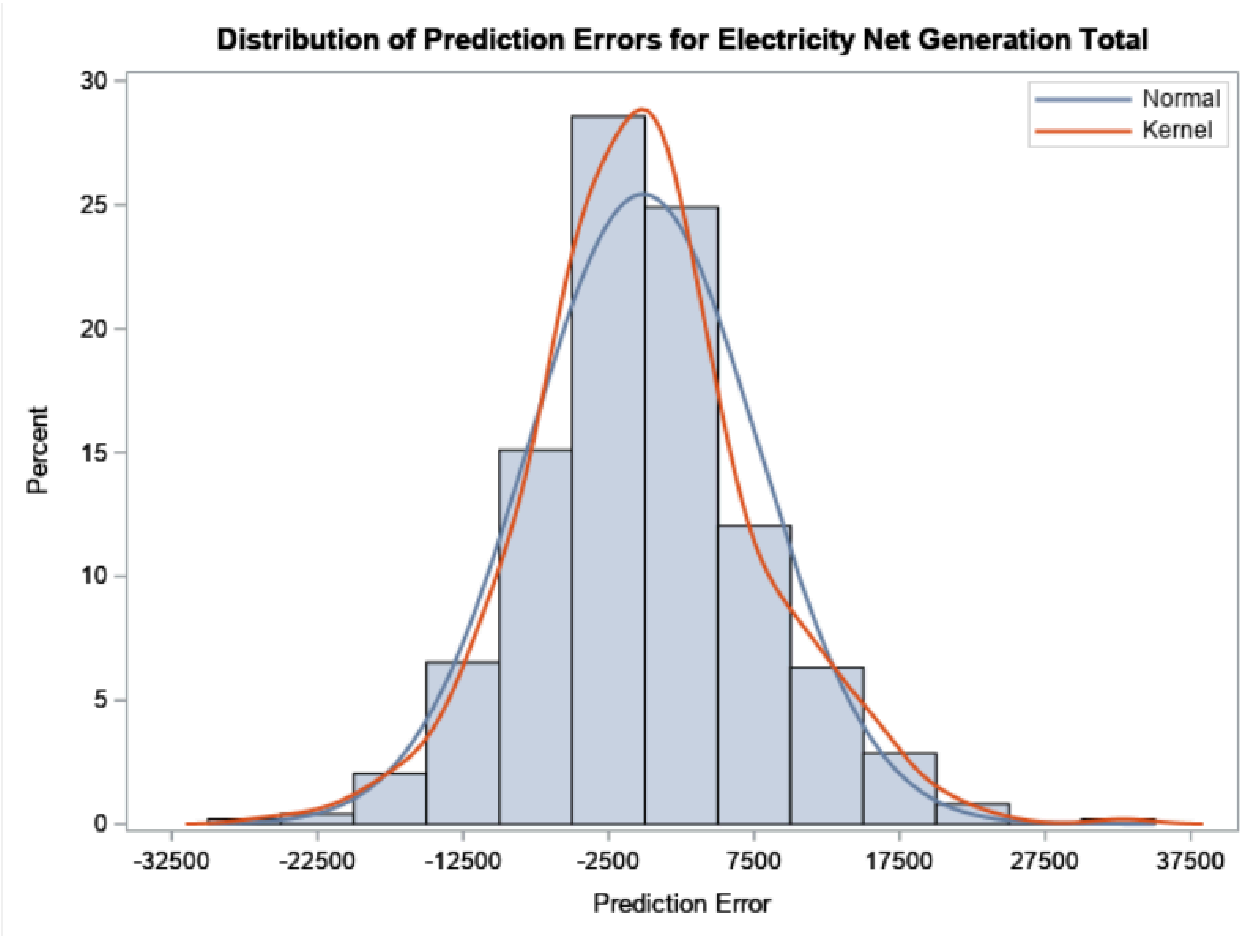


The ACF plot is decreasing and there is a significant lag 0 in the PACF plot and a significant lag 1 in the IACF plot and the error is not distributed as white noise in the plot.



RMSE	MAPE	MAE	AIC	SBC
7837.4112443	2.240177902	5956.0562316	8793.3305832	8805.9137994
20177.770829	5.0538205093	16806.396671	1665.2725863	1665.2725863

The forecast plots are merging inside the confidence interval. The actual and predicted values are stable and the energy generated in total has a slight increase as the predicted values have a bigger difference and there are less errors with an accuracy of 95%. The normality spike is under the value of residuals and the error is minimal.



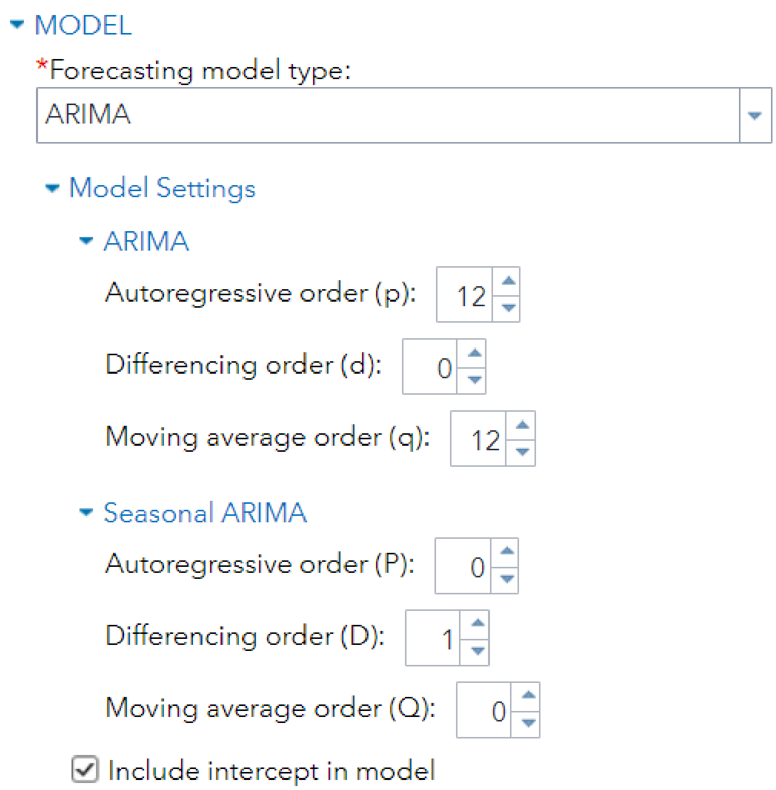
Test for Stationarity:

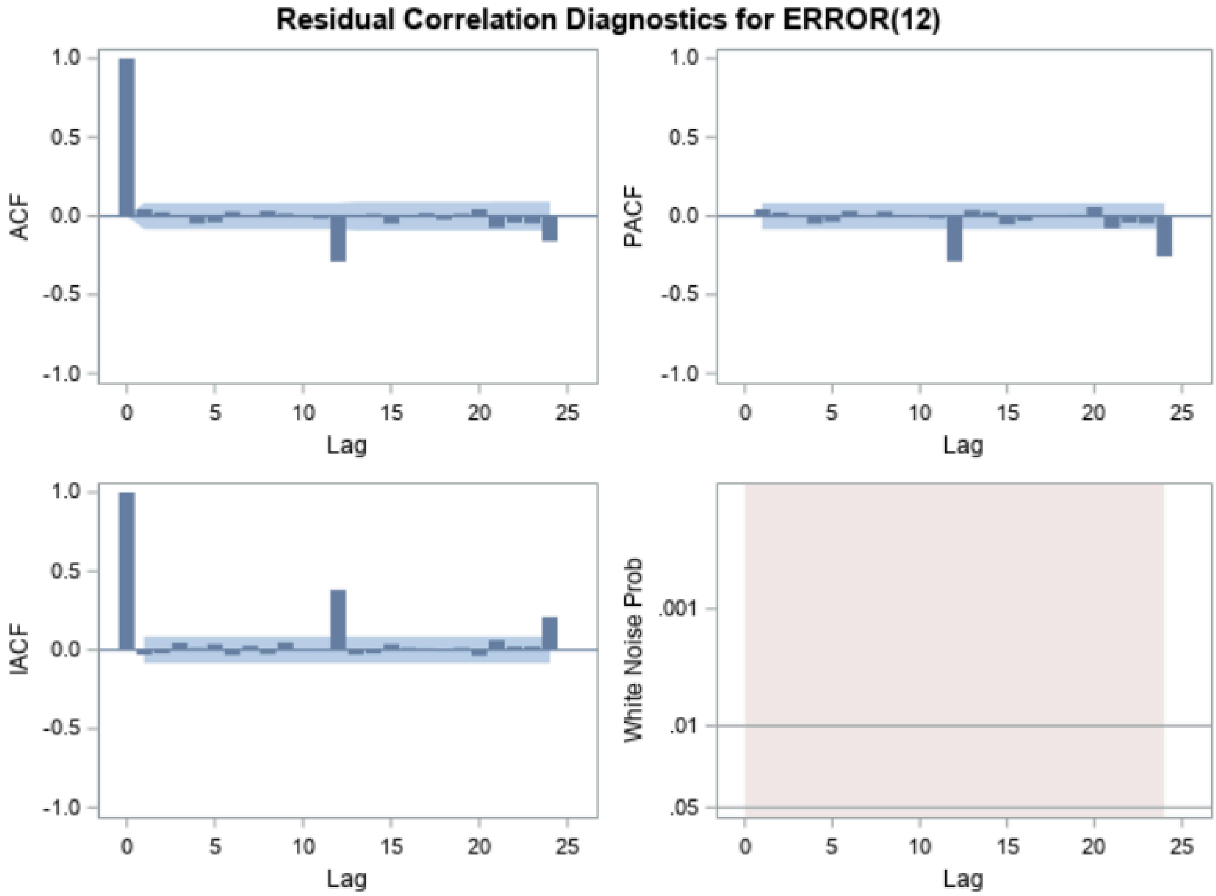
Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-2.0080	0.3297	-0.91	0.3243		
	1	-2.9980	0.2342	-1.14	0.2329		
	2	-1.5287	0.3910	-0.75	0.3927		
	3	-0.0070	0.6812	-0.01	0.6810		
Single Mean	0	-39.8332	0.0018	-4.60	0.0002	10.61	0.0010
	1	-60.6737	0.0018	-5.59	<.0001	15.63	0.0010
	2	-38.7902	0.0018	-4.40	0.0004	9.74	0.0010
	3	-13.2860	0.0599	-2.66	0.0830	3.74	0.1114
Trend	0	-187.841	0.0001	-10.45	<.0001	54.63	0.0010
	1	-425.264	0.0001	-14.43	<.0001	104.09	0.0010
	2	-437.905	0.0001	-12.52	<.0001	78.42	0.0010
	3	-140.310	0.0001	-7.48	<.0001	27.98	0.0010

There is no stationarity after doing the ADF test even though the winter multiplicative model is best fitting for the above given dataset which has a minimum error and since it fails the white noise test we can further go on with ARIMA modeling for the error analysis and reduce the white noise.

Below mentioned are the plots and residuals for the ARIMA for errors and there is white noise in the test and the forecast values are also merging within the 95% confidence bonds.

ARIMA Model(12,0,12):

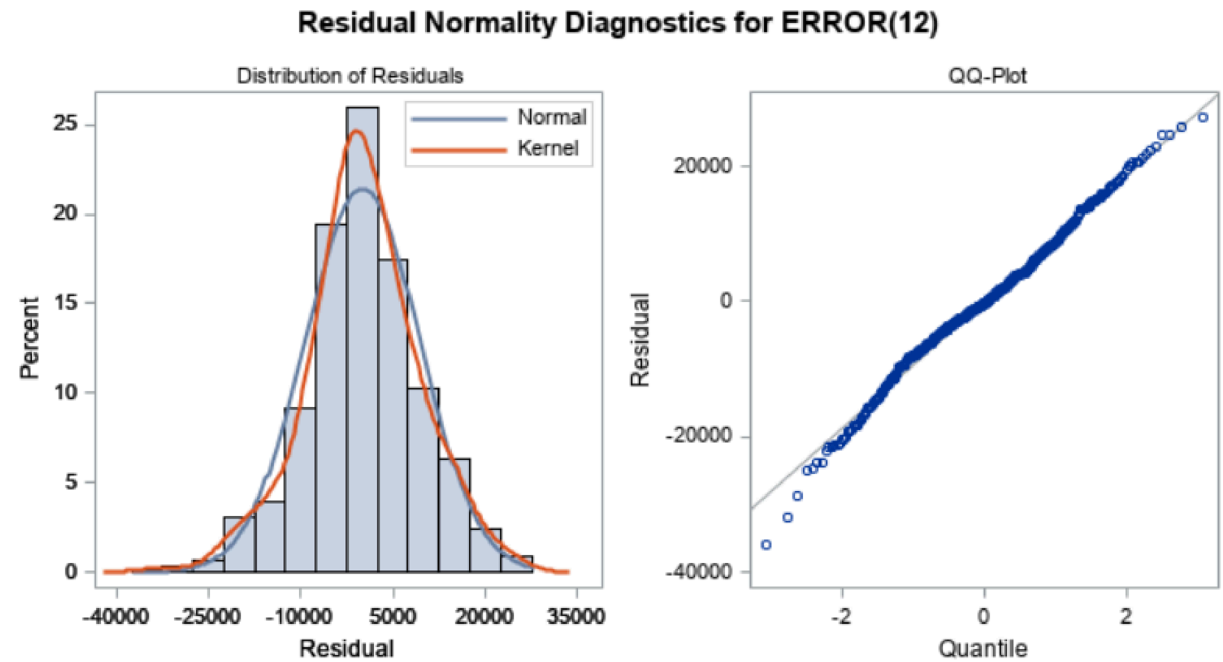


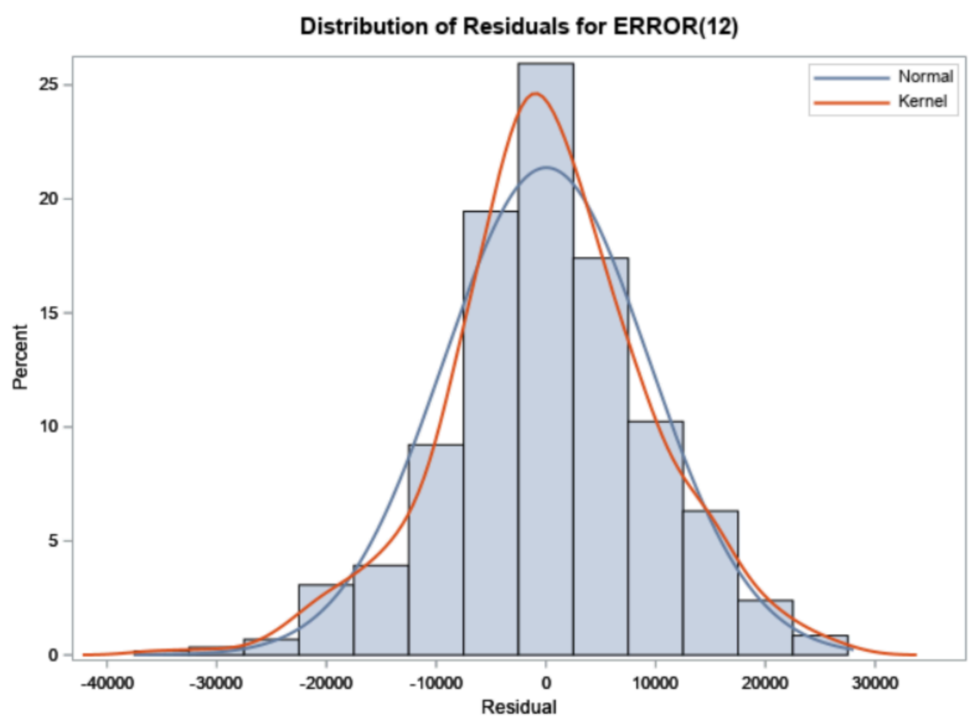


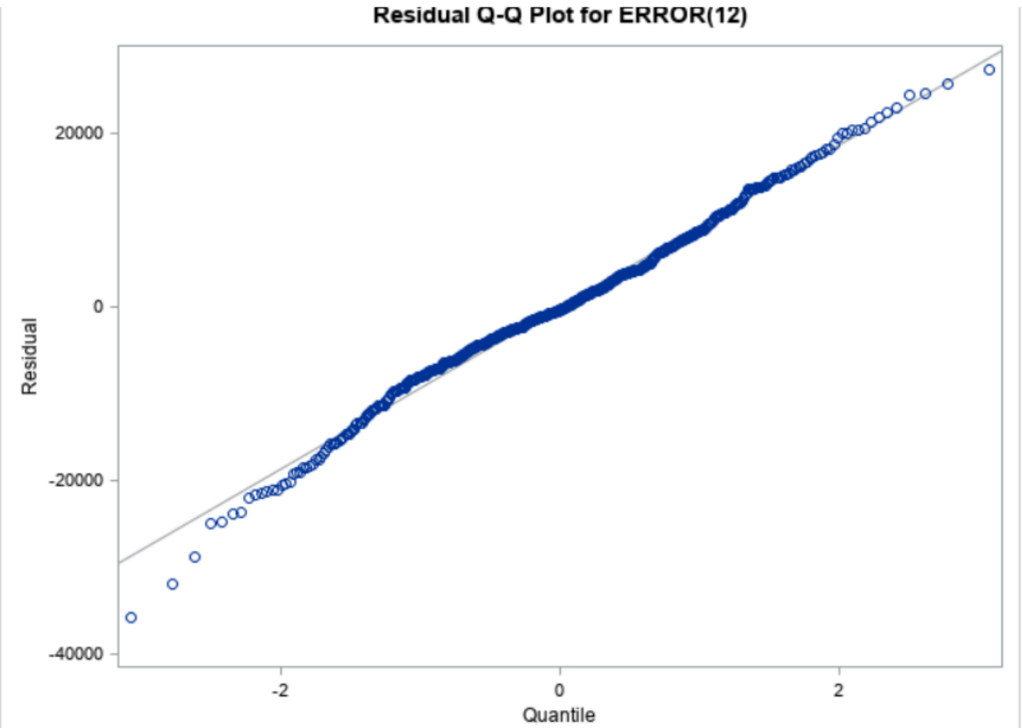
After conducting the ARIMA for the forecast values of the best fit model to reduce the errors distribution the autocorrelation values are significant at lag 12 so that can be given as the autoregressive order for the ARIMA and the partial autocorrelation function has a lag at 12 which is the moving average order.

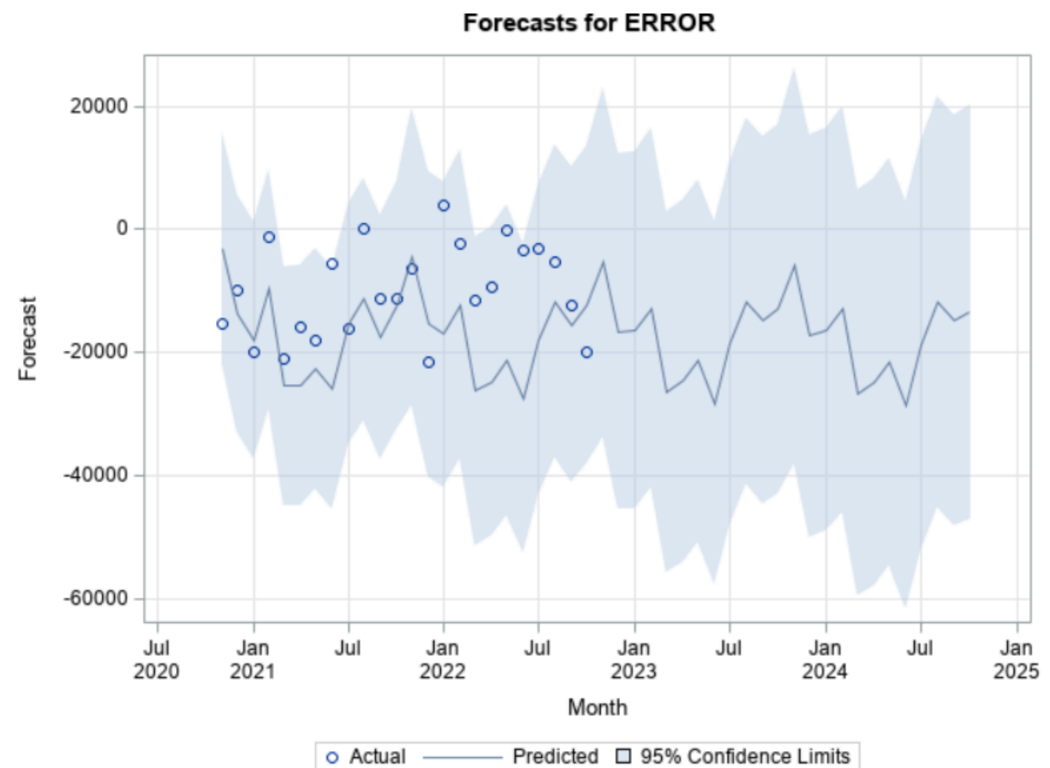
The errors in the signal have been completely neglected and the forecast plots can be achieved by seasonal differencing order 1 to reduce the seasonality.

The ARIMA modeling for error prediction has a normality in the error interval and merges with residual values of the plot.

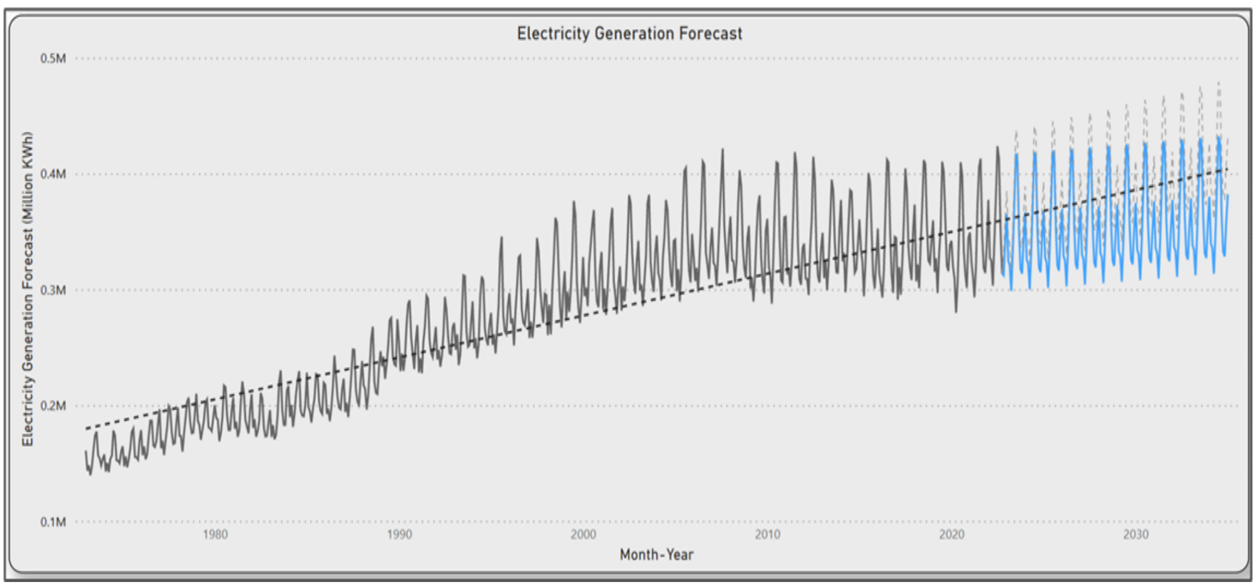

Residual Normality Diagnostics for ERROR(12)
Distribution of Residuals
Normal
Kernel
Percent
Residual
QQ-Plot
Quantile


Distribution of Residuals for ERROR(12)
Normal
Kernel
Percent
Residual


Residual Q-Q Plot for ERROR(12)
Residual
20000
0
-20000
-40000
-2
0
2
Quantile


Forecasts for ERROR
Forecast
20000
0
-20000
-40000
-60000
Jul 2020
Jan 2021
Jul
Jan 2022
Jul
Jan 2023
Jul
Jan 2024
Jul
Jan 2025
Month
Actual
Predicted
95% Confidence Limits

	TYPE	_STAT_	_VALUE_
1	ML	AIC	12436.140933
2	ML	SBC	12545.473928
3	ML	LOGLIK	-6193.070466
4	ML	SSE	50947644866
5	ML	NUMRESID	586
6	ML	NPARMS	25
7	ML	NDIFS	12
8	ML	ERRORVAR	90815766.25
9	ML	MU	-221.1132914
10	ML	CONV	1
11	ML	NITER	10



4. Model Comparison:

Model comparison for Non renewable resources based on AIC, SBC values. For forecasting for non-renewable resources, SARIMA(1,1,1) and ARIMA(1,1,1) worked much better than any other models.

Variable	Models	AIC	SBC	MAPE
Non-renewable	Winter Additive modeling	9713.064	9725.927	2.66
Non-renewable	ARIMA(0,0,0)(errors from winter additive model)	14942.62	14947.01	
Non-renewable	ARIMA(12,0,12)(based on the white noise and residuals)	12563.19	12673.03	116.67%
Non-renewable	SARIMA(1,1,1),ARIMA(1,1,1)(based on the complex model and to take care of the trend)	12305.95	12323.45	466.13%

To take care of the seasonal and trend component we first try to fit exponential smoothing models to our dataset. Below are the output statistics of three exponential smoothing models, namely, winters additive, winters multiplicative and seasonal multiplicative.

Model comparison for Renewable resources based on accuracy statistics values MAE, MAPE, RMSE

Model	RMSE	MAPE	MAE

Winters Additive Model	12534.339032	17.958613986	10457.988706
Winters Multiplicative Model	10594.982333	15.215972329	10594.982333
Multiplicative Seasonal Model	**6573.172352**	**8.0146217295**	**6573.172352**

Based on the accuracy statistics on the holdout sample for the three ESM models the Multiplicative Seasonal Exponential Smoothing model was found to fit the best.

Model Comparison for Total Electricity Generated:

Model	RMSE	MAPE	MAE	AIC	SBC
Winters Multiplicative Model	**20013.431333**	**3.0127372145**	**16806.953587**	**1558.2197954**	**1620.2197954**
Winters Additive Model	20177.770829	5.0538205093	16806.396671	1665.2725863	1665.2725863

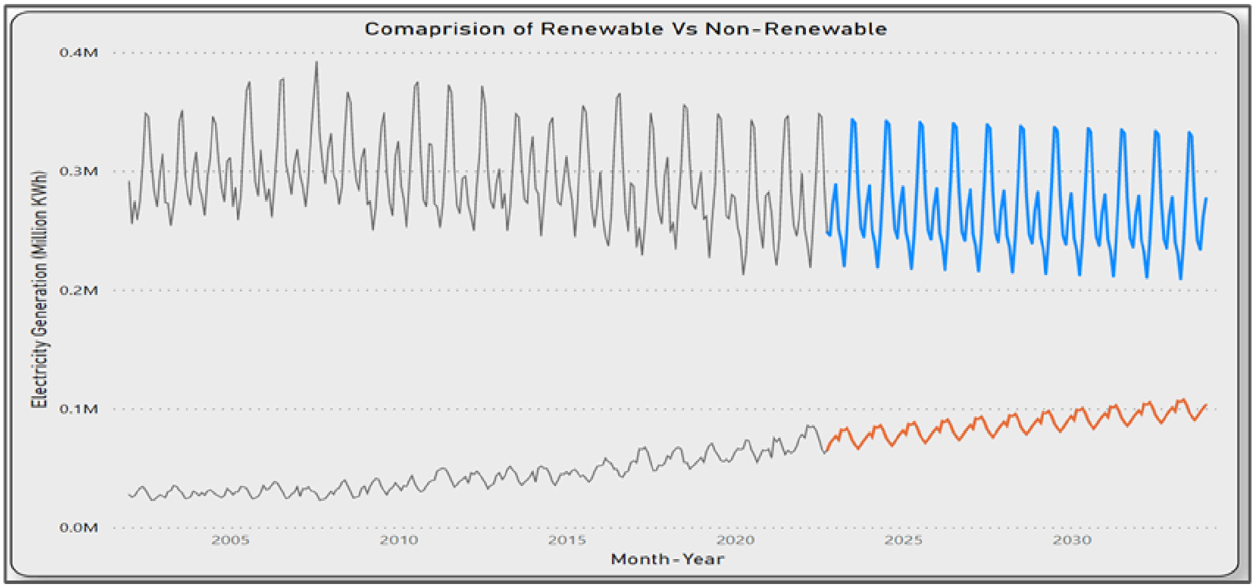
Model Comparison for Electricity End Use:

Model	**RMSE**	**MAPE**	**MAE**

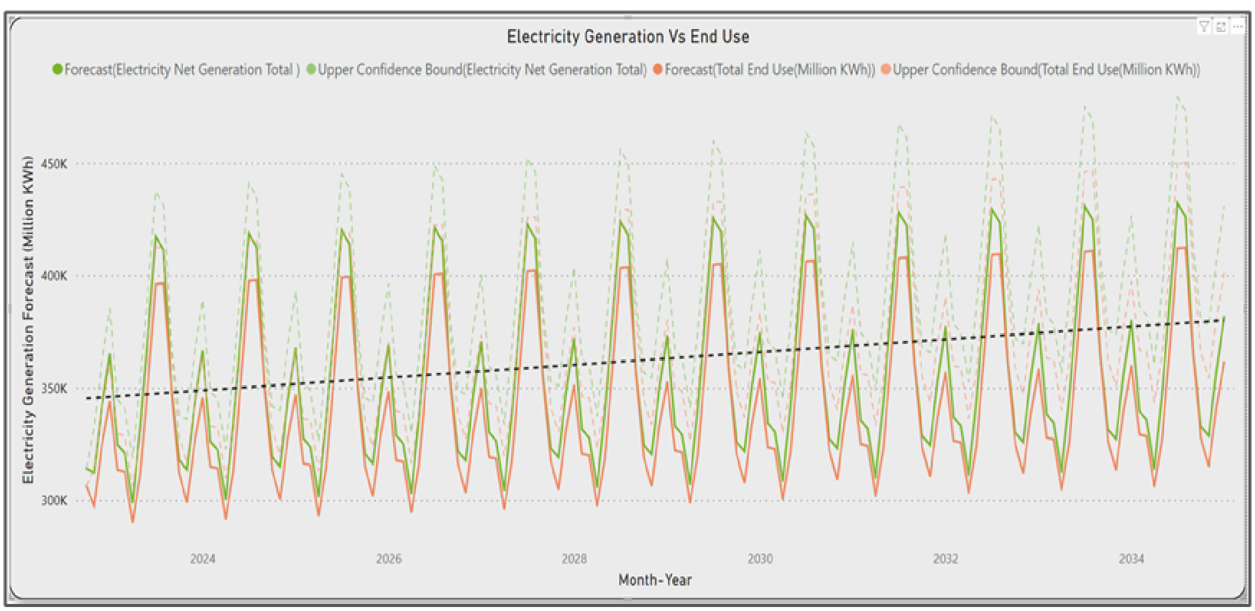
Winters Multiplicative Model	5965.77	2.2	4988.40
Additive Seasonal Model	6246.5	1.9	4836.8
ARIMA(4,0,4) on Residuals from Additive Seasonal Model	6224.23	5242.82	177.06
ARIMA(8,0,8) on Residuals from	5742.45	4333.99	152.06

5. Findings

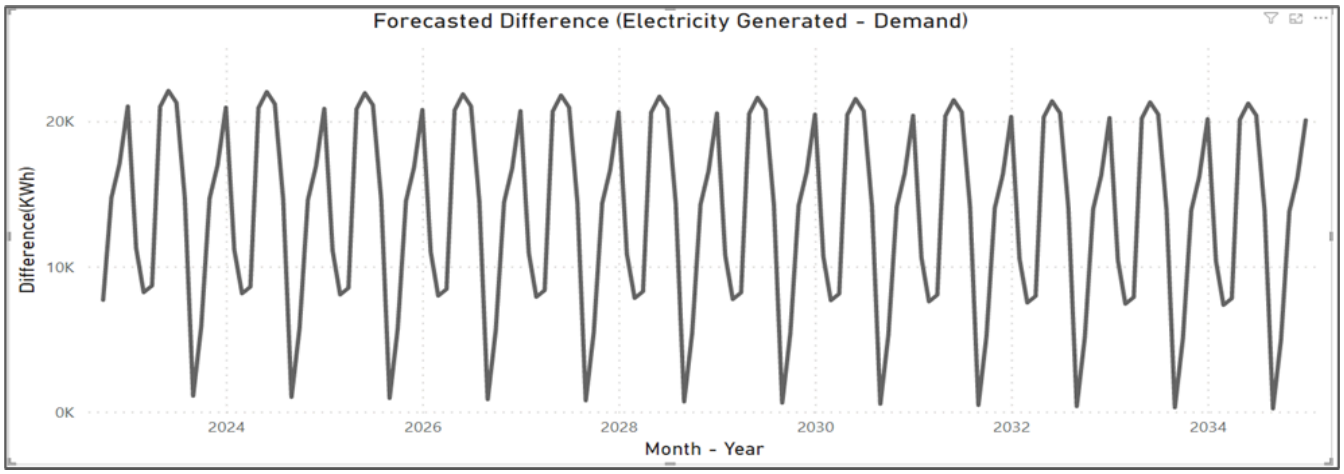
- From the non-renewable models, we may observe a non-renewable trend from 1978 to 2022. The generation of electricity from non-renewable resources has increased, but if we look at the data over the past 20 years, it has declined. The decrease in nonrenewable electricity production is due to the increase in renewable resources. Since the 2000s, solar panels and other renewable energy sources have gained popularity, which has had a stronger influence. Moreover, the government supported solar energy production with subsidies and other regulations.



- The findings of the project suggest that energy generation is likely to meet future electricity demands, but there is a risk of blackout periods if appropriate measures are not taken. Through a comparison of the forecasted electricity demand and generation, we can identify potential energy shortages and take necessary actions to meet the demand. When the forecasted demand exceeds the forecasted generation, this may indicate the possibility of a shortage. Furthermore, the upper 95% level CF values of electricity generation forecast need to be generated to meet all electricity demand needs and avoid the risk of blackout periods.



- The electricity demand is increasing with a small slope of trend, and more energy is being wasted during summer months. Our analysis shows that electricity wastage is consistently high during April to July, which coincides with the summer season. This information can be particularly useful for electricity providers as it can help them manage their energy production and distribution.



- From the trend component of time series exploration of renewable energy resources we can see from 2002 there is a steep increase in energy generation through renewable energy sources. This can be attributed to the fact of increased harnessing of solar power due to large scale installation of solar panels.

Business Value:

Lowering Costs: Firstly, the forecasted energy generation and consumption can help businesses plan their production and operation schedules to optimize their energy use, leading to potential cost savings. The use of renewable sources of energy can also help to reduce costs by reducing the reliance on expensive non-renewable sources of energy, especially during peak demand periods. By forecasting the upper 95% level of capacity factor values, the business can ensure that they have sufficient electricity generation capacity to meet their peak demand needs, which can help avoid costly blackouts.

For example, the company may be able to schedule production during off-peak hours to take advantage of lower electricity rates, or adjust their inventory levels to reduce waste and storage costs.

Increasing revenue: Secondly, the forecasted trends in energy generation through non-renewable sources and electricity generation through renewable sources can help businesses make informed decisions about their future investments in energy infrastructure. By investing in renewable sources of energy, businesses can reduce their carbon footprint and potentially attract environmentally conscious customers, which can lead to increased revenue.

Lowering risk: Finally, the additive seasonal modeling and ARIMA models used in this project can help businesses make more accurate and informed decisions about their energy usage and generation. This can help businesses to reduce the risk associated with energy price volatility and ensure that they have sufficient energy supply to meet their production needs. In addition, the use of these models can help businesses to optimize their energy usage patterns, which can lead to cost savings and improve operational efficiency.

In summary, the forecasting models used in this project can help businesses to reduce costs, increase revenue, and mitigate risks associated with energy usage and generation.

Conclusions:

In conclusion, our forecasting project revealed several important insights about energy generation and consumption in the future.

1. Energy generation is expected to meet electricity demand, but there is a risk of blackout periods if measures are not taken to ensure reliability.
2. Electricity demand is increasing gradually, but not at a steep rate.
3. More energy is wasted during the summer months, from March to July, which suggests that measures should be taken to improve energy efficiency during this time.

4. The use of renewable energy sources is increasing linearly, while the use of non-renewable sources for electricity generation is decreasing.
5. Exponential smoothing models with ARIMA models on errors from ESM models were found to be effective for forecasting electricity generation from renewable and non-renewable sources.
6. Additive seasonal models were found to be effective for forecasting electricity end use.
7. Upper 95% confidence level CF values for electricity generation should be used to ensure that electricity demand is met reliably.
8. Our findings can help businesses and policymakers make better decisions to reduce costs, increase revenue, and lower the risk associated with energy generation and consumption.
9. For businesses, the insights can be used to optimize their energy consumption patterns, reduce wastage, and improve efficiency.
10. For policymakers, the findings can be used to develop policies that promote the use of renewable energy sources and reduce reliance on non-renewable sources, leading to a more sustainable future.

References:

1. *AlRashidi, M. R., & El-Naggar, K. M. (2015). Forecasting of primary energy consumption using an improved grey model. Energy, 93, 2282-2292.*
2. *Arora, A., & Singh, J. P. (2019). Forecasting electricity consumption in India using artificial neural networks. Energy Reports, 5, 1321-1328.*
3. *Azevedo, I. L., Morgan, M. G., & Apt, J. (2013). Public acceptance of climate change technologies: wind and solar energy, carbon capture and storage, and nuclear power. Proceedings of the National Academy of Sciences, 110(3), 9209-9214.*

4. *Balli, F., & Guerineau, S. (2017). Forecasting electricity consumption: A review of the state-of-the-art with a look into the future. Renewable and Sustainable Energy Reviews, 75, 86-97.*
5. *Bhandari, R., & Sharma, J. (2018). Forecasting electricity consumption of Nepal using ARIMA models. Sustainable Energy Technologies and Assessments, 29, 48-55.*
6. *Cai, X., Wang, J., & Zhang, X. (2017). A hybrid model for electricity demand forecasting with applications to China. Energy Economics, 63, 174-181.*
7. *Chen, S. T., Kuo, H. C., & Chen, C. L. (2015). Forecasting Taiwan's electricity consumption using a novel grey-based hybrid model. Energy, 93, 2293-2302.*
8. *Cheng, Y., & Zeng, Y. (2019). A comparative analysis of energy consumption and economic development in China and the United States. Energy Economics, 84, 104501.*
9. *Ghodsi, M. R., & Kargarsharifabad, H. (2016). A comparative analysis of energy consumption forecasting methods for buildings. Renewable and Sustainable Energy Reviews, 58, 165-177.*
10. *Giri, R. K., & Dash, P. K. (2018). Electricity demand forecasting using artificial neural network models: A review. Renewable and Sustainable Energy Reviews, 81, 399-417.*
11. *Huang, S. L., Chen, Y. M., & Wang, Y. L. (2016). An integrated model for forecasting short-term electricity consumption in Taiwan. Energy, 115, 1154-1164.*
12. *Liu, Y., Li, Y., & Qian, Y. (2018). Electricity consumption forecasting using deep learning approaches: A review. Renewable and Sustainable Energy Reviews, 81, 249-258.*
13. *Meng, J., & Li, X. (2017). A comparison study of grey model, ARIMA model and GM-ARIMA model for predicting China's energy demand. Energy, 135, 645-654.*
14. *Mishra, A., & Rana, S. (2016). Electricity demand forecasting using seasonal ARIMA model with interventions. Energy, 112, 435-447.*
15. *Mo, J., Lee, W. J., & Kim, T. H. (2017). Short-term forecasting of electricity consumption in South Korea using an artificial neural network model with residual modification. Applied Energy, 185, 1783-1793.*
16. *Moro, S., Haji, A., & Babaie, M. (2018). A comparative study of electricity consumption forecasting using machine learning models*