****

**ENERGY CONSUMPTION**

**TABLE OF CONTENTS**

**1.** Introduction

Problem statement

Dataset source

Data description

**2.** Data preprocessing

Time series exploration of important components

**3.** Time series exploration and modeling and forecasting

**4.** ESM Models

Additive Seasonal Exponential Smoothing Model

Winters Additive

Test For Stationarity

**5.** ARIMA Models

ARIMA(3,7)

ARIMA(2,1,0)

ARIMA(2,1,1)

**6.** Remaining Models

**7.** Model Comparison

**8.** Findings & Business Insights

**9.** Conclusion

**10.** References

Introduction:

Energy is the lifeline of any nation, powering everything from homes to industries. As nations progress, their energy needs evolve, necessitating careful planning and sustainable management. Romania with its rich history and dynamic future stands at a crossroads where understanding energy patterns is paramount to ensuring a sustainable future. This report delves deep into Romania's energy consumption patterns, offering insights and forecasts to guide the nation's energy policies.

As we strive for sustainability and efficiency, understanding the patterns and trends in energy utilization becomes increasingly crucial. This project centers on time series modeling of energy consumption, focusing on weekly data that provides a comprehensive snapshot of energy production and consumption dynamics.

As we continue to grapple with complex energy challenges, the ability to analyze and predict energy consumption patterns becomes increasingly vital. In this project, we embark on a journey to explore and model energy consumption using a weekly time series dataset. Our dataset comprises a range of energy sources, including hydroelectric, nuclear, wind, oil and gas, solar, biomass, and coal. By delving into this data, our goal is to extract meaningful insights, develop predictive models, and contribute to informed decision-making within the energy sector. This report provides an overview of our approach, methodologies, and key findings as we seek a deeper understanding of energy consumption trends and dynamics.

Dataset:

The dataset under examination encompasses various energy sources, including hydroelectric, nuclear, wind, oil and gas, solar, biomass, and coal. By delving into this rich dataset, we aim to uncover valuable insights, develop predictive models, and contribute to better-informed decision-making in the energy sector. This report outlines our approach, methodologies, and findings in this pursuit of understanding and forecasting energy consumption patterns.

Date: This column represents the date on which the data was recorded. In your example, the date is in the format "Month/Day/Year" (e.g., 1/7/2019), and it indicates the specific span in time for which the energy production and consumption data are reported.

Sum of Production: This column typically represents the total energy production during the specified time period. It is a cumulative figure that encompasses the combined output from all energy sources, both renewable (e.g., hydroelectric, wind, solar, biomass) and non-renewable (e.g., nuclear, oil and gas, coal).

Hydroelectric: This column represents the amount of energy generated from hydroelectric power sources during the specified time period. Hydroelectric power is derived from the flow of water, typically in rivers or dams, and is a form of renewable energy.

Consumption: This column represents the total energy consumption during the specified time period. It reflects the overall demand for energy across various sectors, including residential, commercial, and industrial.

Nuclear: This column represents the amount of energy generated from nuclear power sources during the specified time period. Nuclear energy is produced through nuclear reactions and is considered a non-renewable energy source.

Wind: This column represents the amount of energy generated from wind power sources during the specified time period. Wind energy is a renewable energy source that harnesses the kinetic energy of wind to generate electricity.

Oil and Gas: This column represents the amount of energy generated from oil and gas sources during the specified time period. These sources are non-renewable and involve the combustion of fossil fuels to produce energy.

Solar: This column represents the amount of energy generated from solar power sources during the specified time period. Solar energy is renewable and is harnessed from sunlight using solar panels or collectors.

Biomass: This column represents the amount of energy generated from biomass sources during the specified time period. Biomass energy is derived from organic materials such as wood, crop residues, and agricultural waste.

Coal: This column represents the amount of energy generated from coal sources during the specified time period. Coal is a non-renewable fossil fuel and has historically been a significant source of energy.

Data Preprocessing:

In our possession, we had data detailing Romania's hourly electricity consumption. However, for the depth of our analysis, we required a more aggregated, weekly perspective on energy use. To achieve this transformation from hourly to weekly data, we employed Excel's 'Group By' function. Subsequently, we harnessed the data into pivot tables, allowing us to effectively process and organize

the data into the desired weekly format.

Times Series Exploration

Input

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

For the Time Series Exploration, the dependent variable is Consumption, which is measured in kW. The Time ID is Date, and the interval is detected as weeks, as the data is in weekly format.

A screenshot of a data set

Description automatically generated

Output

The time interval for the dataset is week and the data is collected from Mon, 7 Jan 2019 through Mon, 13 Mar 2023. The total No.of observations are 219.

A graph with blue lines

Description automatically generated

A diagram of a distribution of values

Description automatically generated

A graph showing different colored lines

Description automatically generatedBased on the seasonal cycle graph, there is an increase in consumption every year during winters, and consumption decreases during summers. This pattern is observed consistently each year. Here, we can infer that there is seasonality in our data.

Romania, which is a cold Country, there is usually an increase in electricity usage during winter months due to heating requirements.

A graph of a graph of a graph

Description automatically generated with medium confidence

Considering the visual representation of ACF, PACF, and IACF, the ACF is slowly dying out, while PACF shows significant lags at lag=3. IACF exhibits significant lags at lag=2 and 3, which represents the values correlated with the previous lag values. Based on this, we can make observations, such as it may be an AR model. The q values may go up to 7 as ACF graph suggests, and the p values can go up to 3 according to the PACF graph.

A graph of a number of probabilities

Description automatically generatedThe white noise test fails, which indicates that the series is not distributed as white noise. There is a presence of signal in the data which can be derived from Time series Modelling.

A graph showing the number of the year

Description automatically generated with medium confidence

A graph with blue lines

Description automatically generated

The above graph shows that there is no trend component in our dataset.

**Test for stationarity:**

A screenshot of a table

Description automatically generated

To test the stationarity of the data, ADF Unit Root Tests are conducted. Since we do not have a trend component in our data, we are only considering the values of a single mean here. It can be observed that the p-value and Tau values indicate weak stationarity for the Augmented order 0. To make the data stationary, we may need to apply 1st differencing.

In time series analysis and forecasting, having stationary data is preferred because it signifies that the statistical characteristics of the data, such as mean and variance, remain constant over time.

**MODELING AND FORECASTING**

**EXPONENTIAL SMOOTHING MODELS**

**A screenshot of a computer

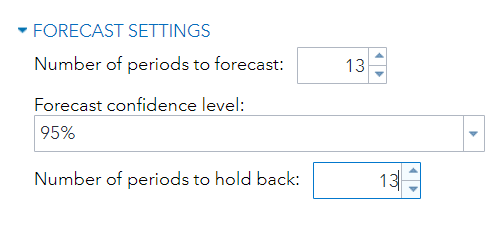
Description automatically generatedA screenshot of a computer

Description automatically generatedADDITIVE SEASONAL EXPONENTIAL SMOOTHING MODEL**

Consumption is our dependent variable used in the Additive Seasonal Exponential Smoothing Model; Time ID is Date. Interval is week as we have a weekly data format.

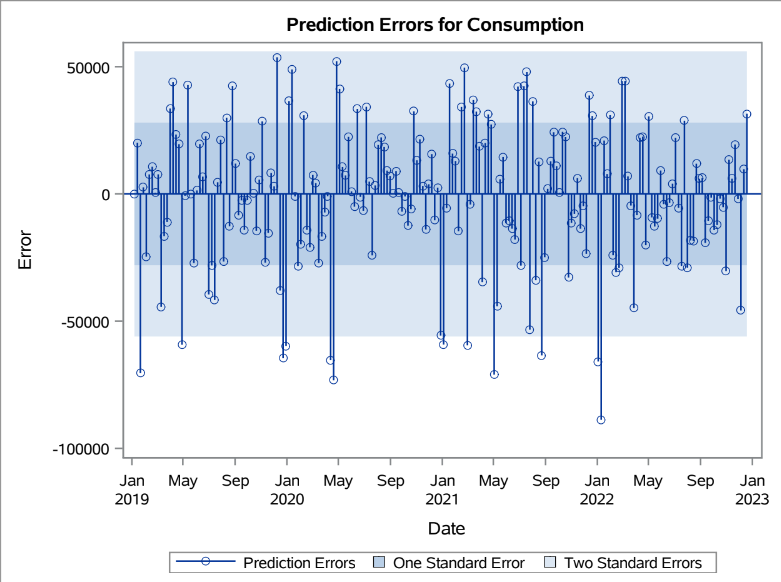
**A white rectangular object with black text

Description automatically generated**

****

We are trying to hold back and forecast for a period of 13 weeks hence the holdout sample is 13 and the period of forecast is 13

**A graph of a graph of a graph

Description automatically generated with medium confidence** From the prediction error plot, we can see that the errors are in the interval of one standard error and two standard errors.

Based on the visual examination of ACF, PACF of Prediction Error Correlation for Consumption, we have observed some significant lags in Lag 1, lag 2 of ACF and Lag 0 and Lag 1 of PACF. Rest of the lags are not so significant. This shows that the prediction errors are not much correlated/ dependent with their previous values. However, we could see that residuals are not white noise, which indicates that there is still some amount of data that can be modelled.

So, this is not a significant model for our data.

A graph of a graph showing the forecasts for consumption

Description automatically generated with medium confidence

By observing the above forecast graph, we could see that the predicted line is fitting the actual values although some of them are deviating.

|  |  |  |  |
| --- | --- | --- | --- |
|  | MAPE | AIC | SBC |
| Actual | 1.9738 | 4377.03 | 4383.76 |
| Forecasted | 5.0467 | 135.35 | 135.35 |

By using the Additive Seasonal Exponential Smoothening Model, we are able to attain a MAPE of 5.04, AIC of 135.35 and SBC of 135.35 for the forecasted model.

A graph of a distribution of a number of error

Description automatically generated with medium confidence

From the above analysis, we can infer that the Exponential Smoothing Models are not working out in our case. We are now trying ARMA models

**BEST MODEL:**

As our data is weakly stationary, trying out with 1st order differencing.

**ARIMA (2,1,0)**

A screenshot of a computer

Description automatically generated

After exploring Exponential smoothing models and ARMA models, now we have transitioned to ARIMA models with 1st order differencing as our data is weakly stationary. In this model we used Auto-Regressive order p=2, d=1, Moving Average q=0.

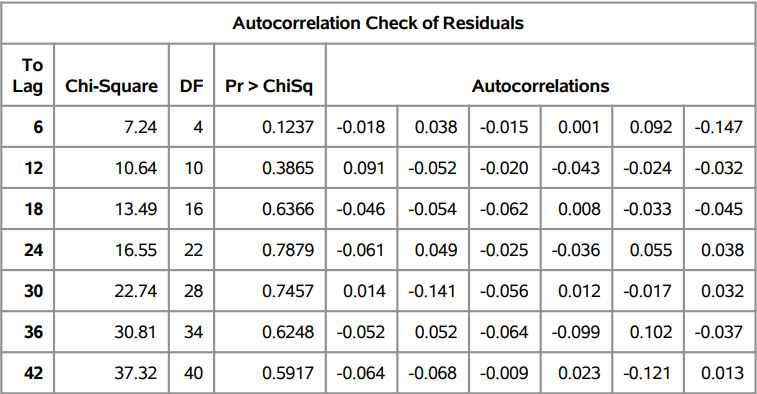
A screenshot of a table

Description automatically generated

A group of graphs with numbers

Description automatically generated with medium confidenceThe table displays the outcomes of an autocorrelation analysis, including chi-square statistics, degrees of freedom (DF), p-values (Pr > ChiSq), and autocorrelation coefficients at various lag orders. The significant chi- square values and p- values < 0.0001 indicate the presence of Auto correlation in the data up to lag 6 only. Since there is not much auto correlation, this indicates that the model can capture almost all the patterns and dynamics in the data.

Here, the ACF, PACF and IACF values show a good correlation.

 A screenshot of a computer

Description automatically generated

The table suggests the presence of no significant autocorrelation in the residuals at multiple lags, indicating that the model may adequately capture all the patterns and dynamics in the data.

A group of graphs with numbers

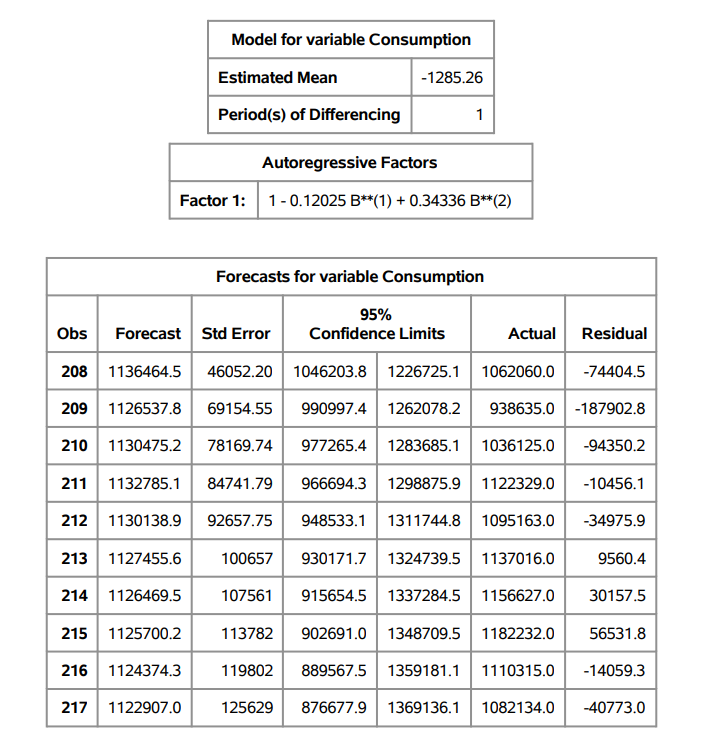
Description automatically generated

A graph of normality and normality

Description automatically generatedWe can observe that the residuals are white noise, and ACF, PACF and IACF graphs are significant here.

From the Residuals normality plot, we can see that the residuals are normally distributed.

A graph of a graph with numbers and a line

Description automatically generated with medium confidence

A screenshot of a table

Description automatically generated

From the Forecasts for consumption table, most of the forecasted values are within the 95% confidence interval range.

Write mape, aic, sbc table here

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **AIC** | **SBC** | **MAPE** | **RMSE** | **MAE** |
| 5303.4586 | 5313.6121 | 2.92 | 47193.76 | 32709.99 |

The ARIMA(2,1,0) has a better MAPE values and lower AIC and SBC.