FERGUSSON COLLEGE (AUTONOMOUS), PUNE

DEPARTMENT OF STATISTICS

TYBSC STATISTICS PROJECT 2023-24

Watt the Future Holds

ILLUMINATING INDIA'S ENERGY TAPESTRY

Deccan Education Society's

Fergusson College (Autonomous), Pune

Department of Statistics

T. Y. B. Sc.

Year 2023-24

STS3609: Statistics Practical III

CERTIFICATE

This is to certify that Mr./Ms. ______________________________________ Roll no.___________, has satisfactorily completed the project work entitled **"Watt the Future Holds: Illuminating India's Energy Tapestry''** during the academic year 2023 – 24 as per the rules and regulations laid down by FERGUSSON COLLEGE (Autonomous), Pune.

Place: Pune
Date: / / 2024

Ms Deepa Kulkarni	Dr Subhash Shende
Project Guide	Head of the Department,
	Vice Principal,
	Department of Statistics

Group Members:

Roll No	Name
Jai Lunkad	212230
Janhavi Tamhankar	212235
Ashokkumar Kushwaha	212237
Vedika Patil	212260
Sanket Brahmankar	212244
Leena Patil	212248

Acknowledgement

We would like to express our sincere gratitude to all those who have contributed to the successful completion of this project.

First and foremost, we would like to thank our esteemed project guide, Prof. Deepa Kulkarni, for her invaluable guidance and support throughout the entire project journey. Her insightful feedback and constant encouragement were a driving force for us. We are immensely grateful for her dedication and mentorship, without which this project would not have been possible.

We extend our heartfelt gratitude to the Department of Statistics, Fergusson College, for providing the necessary infrastructure, facilities, and resources that significantly aided our research and project development. We appreciate the department's commitment to fostering a supportive learning environment.

Finally, we would like to acknowledge and thank all the individuals who directly or indirectly contributed to this project. Their support, both seen and unseen, has been instrumental in our success.

Index

Sr. No	Topic	Page
1.	Introduction	5
2.	Abstract	6
3.	Motivation	7
4.	Objectives	8
5.	Methodology	9
6.	Exploratory Data Analysis	11
7.	Confirmatory Analysis: • Time Series Analysis • Regression Analysis • Correlation Analysis	15 15 34 48
8.	Conclusion	50
9.	Scope and Limitations	52
10.	References	53
11.	Appendix-1	54

Introduction

In today's digital age, electricity serves as the invisible backbone of society, powering everything from homes to industries. Understanding electricity consumption patterns becomes crucial for a nation like India, experiencing rapid development and population growth. This project delves into the heart of India's energy landscape, employing a two-pronged approach.

First, we scrutinize national-level electricity demand data. We want to analyse monthly data for various sources to identify seasonal variations, long-term trends, and unexpected fluctuations. By examining these relationships and leveraging forecasting models, we aim to predict future shifts in the country's electric pulse.

Secondly, we zoom in on the microcosm of Pune's households. A comprehensive survey will explore the tapestry of factors influencing household electricity bills. Through statistical analysis, we aim to isolate the most significant contributors to residential energy costs.

Ultimately, this research strives to illuminate a path towards a more energy-conscious and cost-effective future. By unravelling the complexities of electricity consumption in India, both nationally and within individual households, we hope to generate valuable insights that can inform future interventions and policy recommendations.

Abstract

Electricity is essential in our daily lives, powering everything from our homes to industries. It plays a vital role in shaping how we live, work, and connect, making it a cornerstone of modern society. From charging our phones to running factories, its impact is pervasive and indispensable. Without it, many aspects of our lives would come to a standstill, highlighting its crucial importance in our interconnected world.

This project explores the dynamic landscape of electricity consumption in India, with a keen focus on both macro and micro-level analysis. Through a meticulous examination of monthly electricity demand data, encompassing hydro, solar, wind, and total electricity consumption, the project aims to unveil time series patterns, identifying seasonality, long-term trends, and fluctuations. Additionally, by analysing correlations between different sources of electricity demand and the overarching total electricity demand, the project seeks to unravel the interplay of these variables, leveraging forecasting models to predict future trends and assessing the reliability of these forecasts.

Furthermore, the project delves into household electricity consumption in Pune through a comprehensive survey, aiming to unearth the factors influencing household electricity bills. From the physical attributes of homes to lifestyle choices and awareness levels regarding electricity consumption, various factors will be examined. Statistical techniques, including regression analysis, will be employed to pinpoint the most significant factors affecting household electricity bills, paving the way for informed interventions and policy recommendations to promote energy efficiency and cost savings in households.

Motivation

India, as one of the world's fastest-growing economies, faces the challenge of meeting the energy demands of its rapidly expanding population while minimizing environmental degradation. A project on electricity consumption and sustainable development addresses this urgent need for balanced growth that doesn't compromise the needs of future generations. There are many aspects of looking for renewable resources along with sustainable development in India:

1. **Environmental Concerns:**
   The reliance on fossil fuels for electricity generation contributes significantly to environmental challenges, including air pollution and climate change. By focusing on renewable energy sources such as solar, wind, and hydroelectric power, the project aims to reduce carbon emissions and mitigate the impact of climate change.
2. **Access to Energy:**
   Despite progress in electrification efforts, millions of people in India still need access to reliable electricity. By promoting sustainable development through renewable energy sources, the project aims to address this energy access gap. Off-grid and decentralized renewable energy solutions can provide power to remote and underserved communities, improving their quality of life and enabling socio-economic development.
3. **Energy Security:**
   Our country's dependence on imported fossil fuels makes its energy security vulnerable to global market fluctuations and geopolitical tensions. Embracing renewable energy sources enhances energy security by diversifying the energy mix and reducing reliance on imports, thus bolstering India's resilience to external disruptions.

Further narrowing down to Pune, we are interested in finding how much the population is aware of sustainable development and the changes they make to incorporate sustainability in their daily lives. The overall aspect of the project is to have more knowledge on how the supply and demand of electricity with the help of renewable resources is met in India. Additionally, we will also know how the general public (concentrated in Pune) is helping to build a sustainable future.

Objectives

1. Analyse historical energy production trends in India.

2. Investigate variations in electricity consumption across India.

3. Assess the proportion of electricity generated from renewable sources (Solar, Wind, and Hydro) in India's energy consumption.

4. Model and estimate domestic electricity consumption patterns in Pune city, considering factors such as House Area (in square feet), Household Size, Appliance Count, High-power Appliance Quantity, Presence and utilization of Electric Vehicles (EVs), Adoption of non-conventional energy sources, and Use of energy-efficient appliances.

5. Evaluate the awareness level regarding sustainable electricity usage within Pune households and examine its correlation with electricity consumption behaviour.

Methodology

The project consists of three major parts:

1. Analysis of Electricity Production and Consumption in India:

The initial phase of the project involves an in-depth examination of electricity production and consumption trends in India. This analysis is conducted utilizing secondary data obtained from the official government website, **GRID CONTROLLER OF INDIA LIMITED (formerly known as the Power System Operation Corporation Limited)**. Monthly reports issued by the Government of India serve as the primary source for data extraction.

- **Data Collection:**

  Monthly production and consumption data from various sources, including solar, wind, and hydro, as well as total production and consumption figures, are collected from the aforementioned reports.

- **Analysis Technique:**

  Time Series Analysis is employed to model the gathered data effectively. This analytical method enables the projection of future electricity consumption patterns, focusing on the contribution of renewable energy sources.

2. Modelling Domestic Electricity Consumption in Pune:

The subsequent phase of the project entails modelling domestic electricity consumption within the city of Pune, utilizing primary data obtained through a comprehensive survey.

- **Data Collection:**

  A structured survey is distributed among households in Pune to gather information about electricity consumption and its influencing factors. A random sample of respondents is used to estimate consumption patterns for the entire city.

- **Regression Analysis:**

  The collected data is subjected to Regression Analysis to discern the relationship between various independent variables and the dependent variable (electricity bill amount).

The following variables are considered:

Dependent Variable (Y): Electricity bill amount in different seasons.

Independent Variables (X):

X1: Area of House (in square feet)

X2: Number of occupants in the household

X3: Number of appliances in the household

X4: Existence and usage of Electric Vehicles (EVs) at home (Binary)

X5: Utilization of non-conventional energy sources (Binary)

X6: Deployment of energy-efficient appliances in the household (Binary)

3. **Awareness and Sustainability**

   In addition to conducting time series analysis and multiple regression for estimating domestic electricity consumption in Pune, this project incorporates **correlation analysis** to assess the relationship between awareness levels about sustainable electricity usage and actual electricity consumption.

- **Data Collection:**

  The awareness levels are determined based on responses to a set of 7 questions included in the survey, with responses measured on an ordinal scale ranging from 1 to 5.

- **Correlation analysis:**

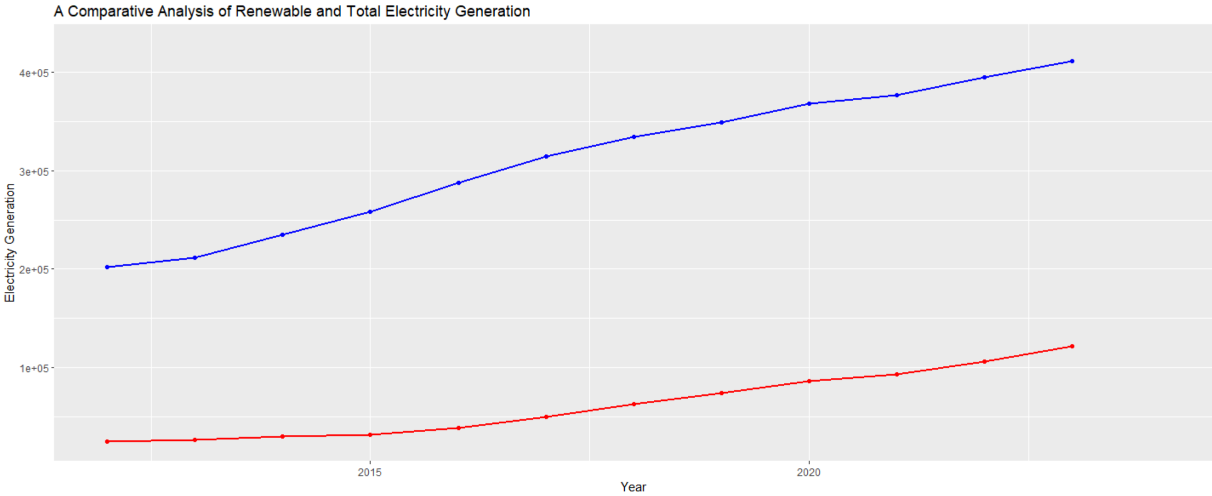
  Responses on the ordinal scale (1-5) are converted into numerical values to facilitate correlation analysis. **Spearman's Correlation coefficient** and **Kendall’s Tau Correlation coefficient** are calculated between the two defined variables.

Tools Used:

- **R Software:** The statistical analysis is predominantly conducted using R software due to its extensive capacity for statistical computing and analysis. The functionalities of R enable Regression Analysis, Time Series Analysis, and correlation assessment.
- **Microsoft Excel** is utilized for data storage and initial Exploratory Data Analysis (EDA).

Exploratory Data Analysis (EDA)

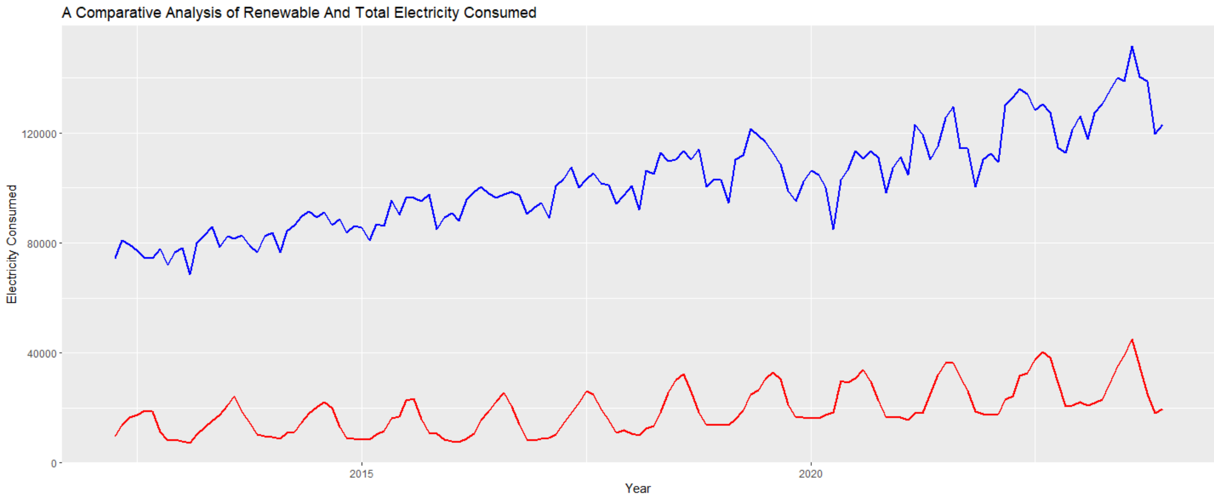
Energy Production in India



- The graph above shows the Electricity production in India between 2012 and 2023.

- The blue line represents non-renewable electricity generation. The red line represents renewable electricity generation. The graph shows a clear distinction between renewable and fossil fuel electricity generation over the years. The blue line, representing fossil fuel electricity generation, shows a steep increasing trend, whereas the red line, representing renewable electricity generation, appears relatively flat.

Energy Consumption in India

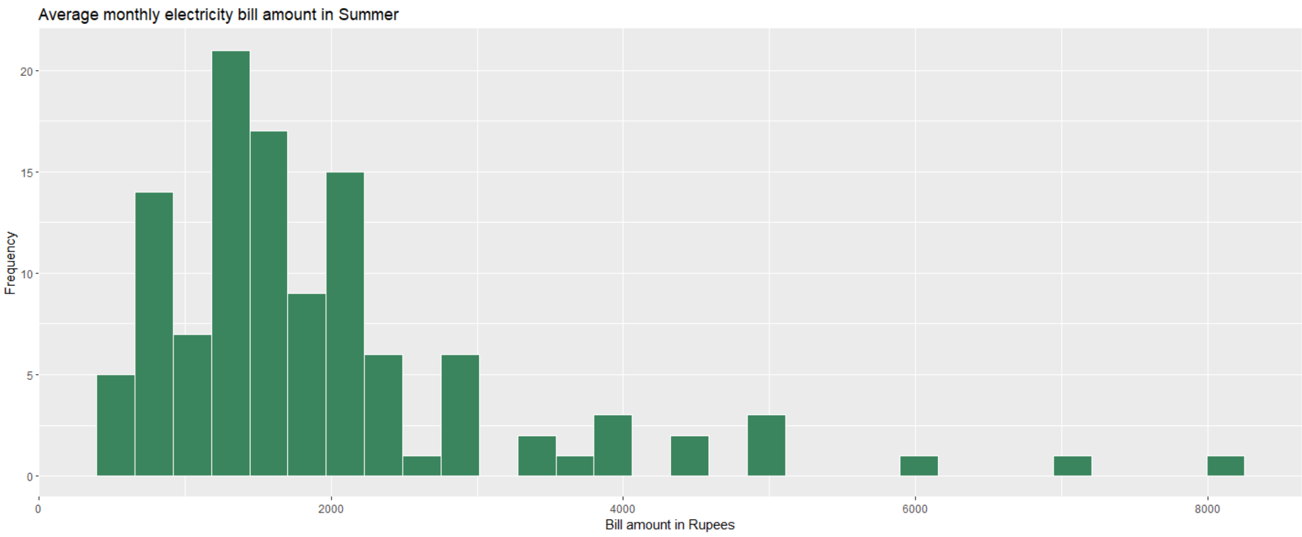


- The above graph shows the Electricity Consumption in India between 2012 and 2023.

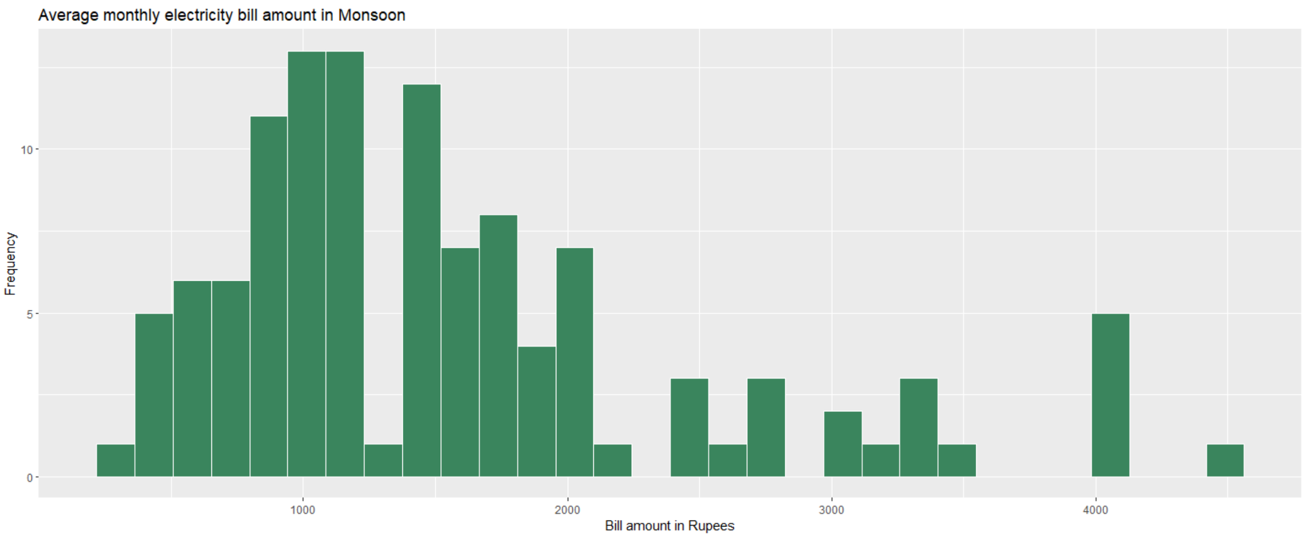
- The red line shows the amount of renewable energy consumed over time. It starts at a low point in 2012 and increases steadily to a higher value in 2023. This suggests that the consumption of renewable energy is increasing over time.

- The blue line shows the amount of total electricity consumed over time. It also started at a low point in 2012 and increased steadily to a higher value by 2022. However, the increase in total consumption appears to be steeper than the increase in renewable consumption. This suggests that the total consumption of electricity is growing faster than renewable consumption.

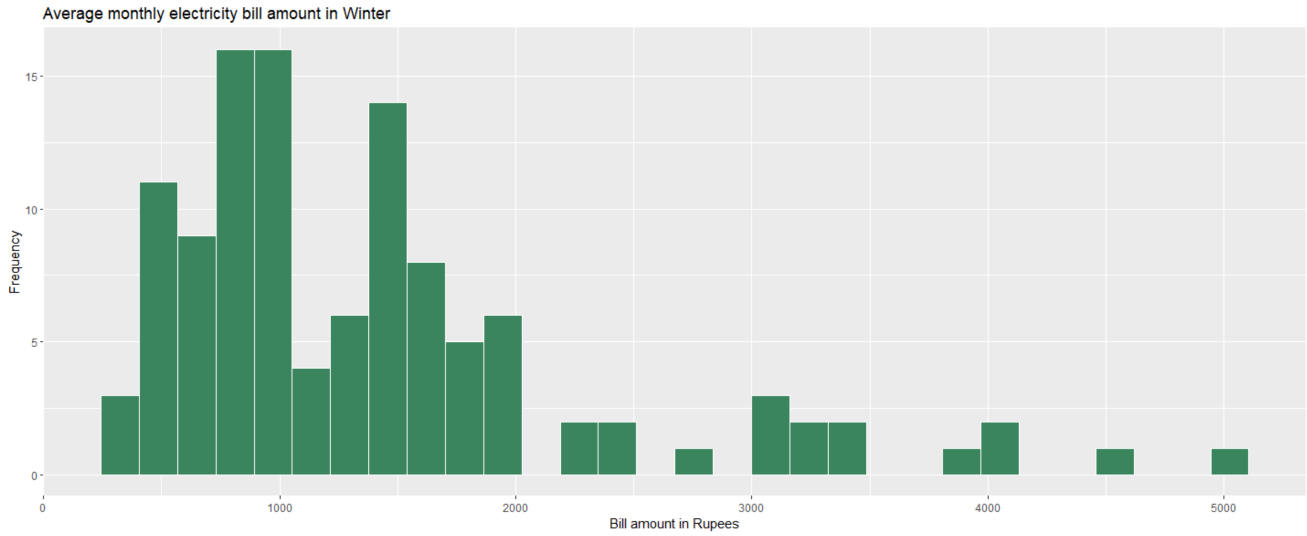
Electricity Consumption of Pune City



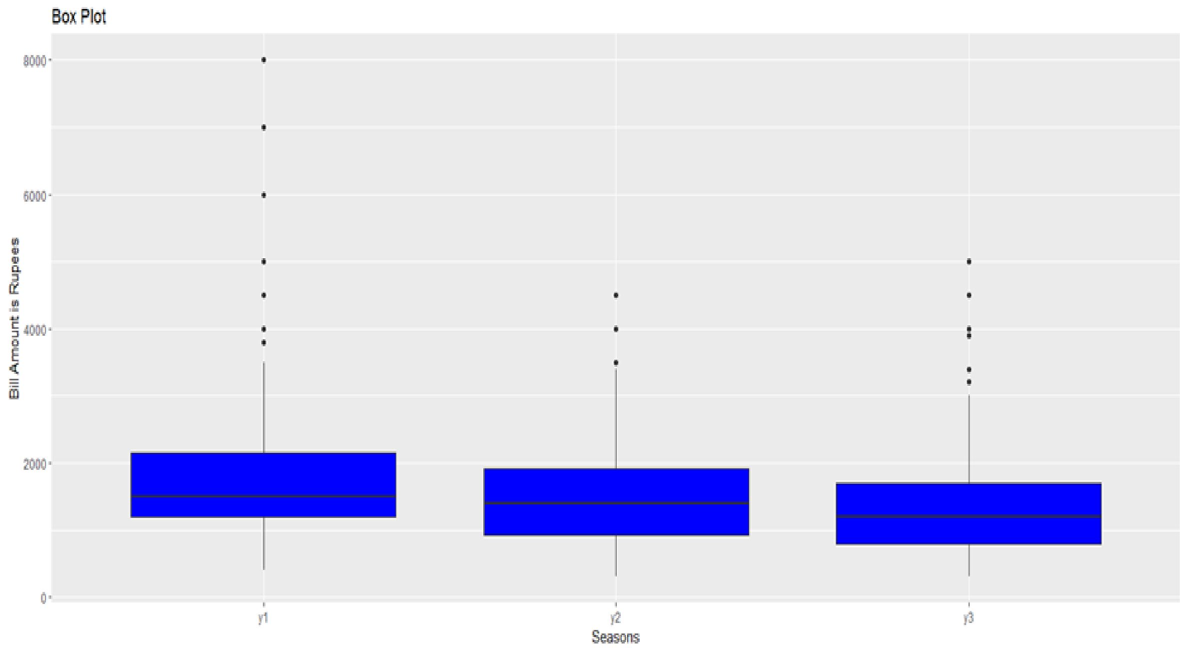
Electricity Consumption in Summer



Electricity Consumption in Monsoon



Electricity Consumption in Summer



Electricity Consumption in the three seasons

- The graphs above show the electricity consumption i.e. the average monthly electricity bill amount of Pune households, in three seasons- Summer, Monsoon and Winter.
- Overall, the histograms suggest that most electricity bills in Pune during any season fall within the 2000-4000 rupees range. However, there is a significant variation in bills, with some residents incurring considerably higher bills than others.

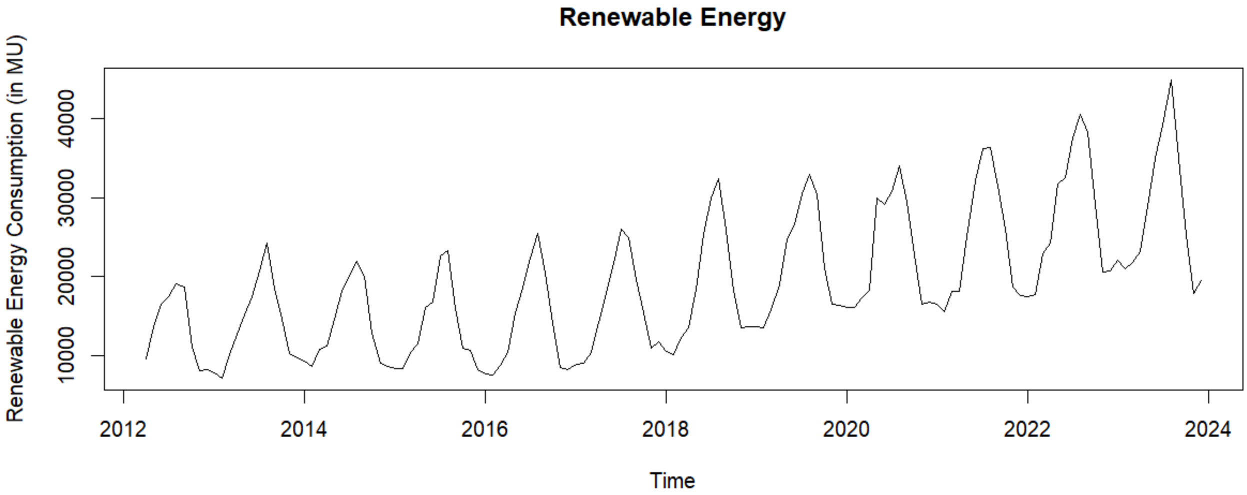
Confirmatory Analysis

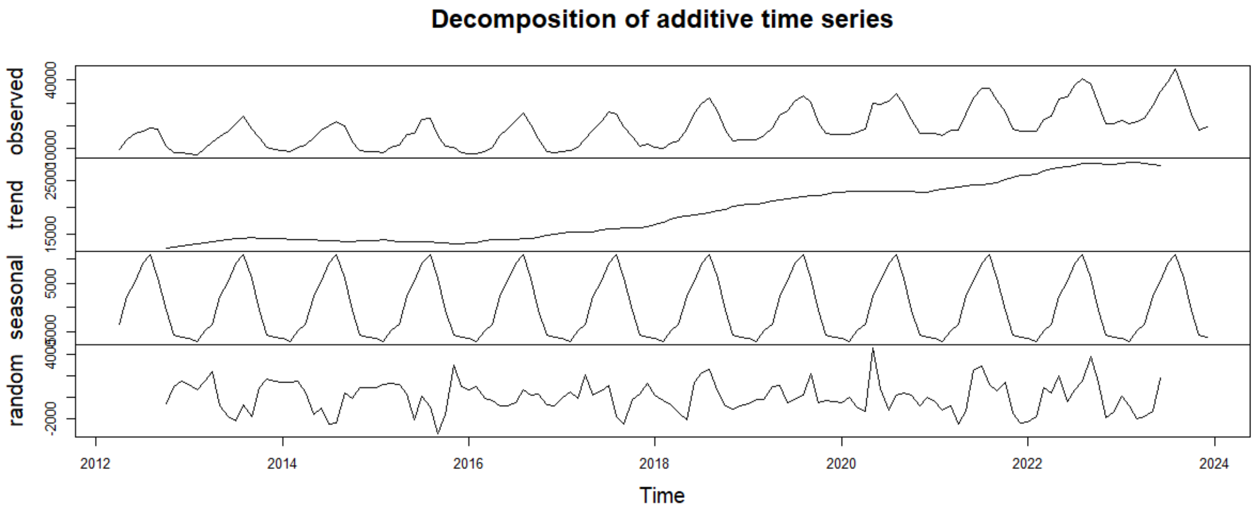
Time Series Analysis

Total Renewable Energy Consumption in India

To model and forecast the electricity consumed (in MU) in India, generated from renewable sources (Hydroelectricity, Solar and Wind energy).

Time Series Plot:



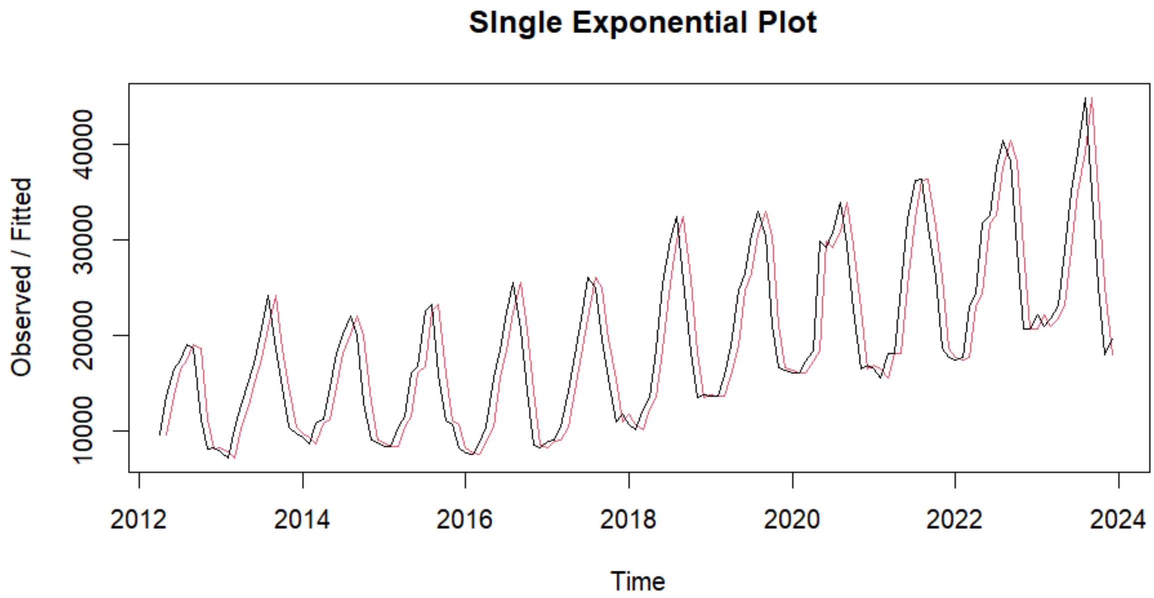


- We may observe a slightly increasing trend, there is a seasonal component in the data, and the irregularities seem sufficiently random.
- The electricity supply from renewable sources tends to increase from the beginning of the year. It peaks in August and then starts to decline.
- We consider the three main sources of renewable electricity- Hydro, Solar and Wind. Thus, the behaviour of supply depends on the factors that affect these three sources, each of which is explained in the later sections.

Exponential Smoothing to Forecast Total Renewable Energy Consumption

Using single, double and triple exponential smoothing on the data, we get the following results:

1. **Single Exponential Smoothing:**

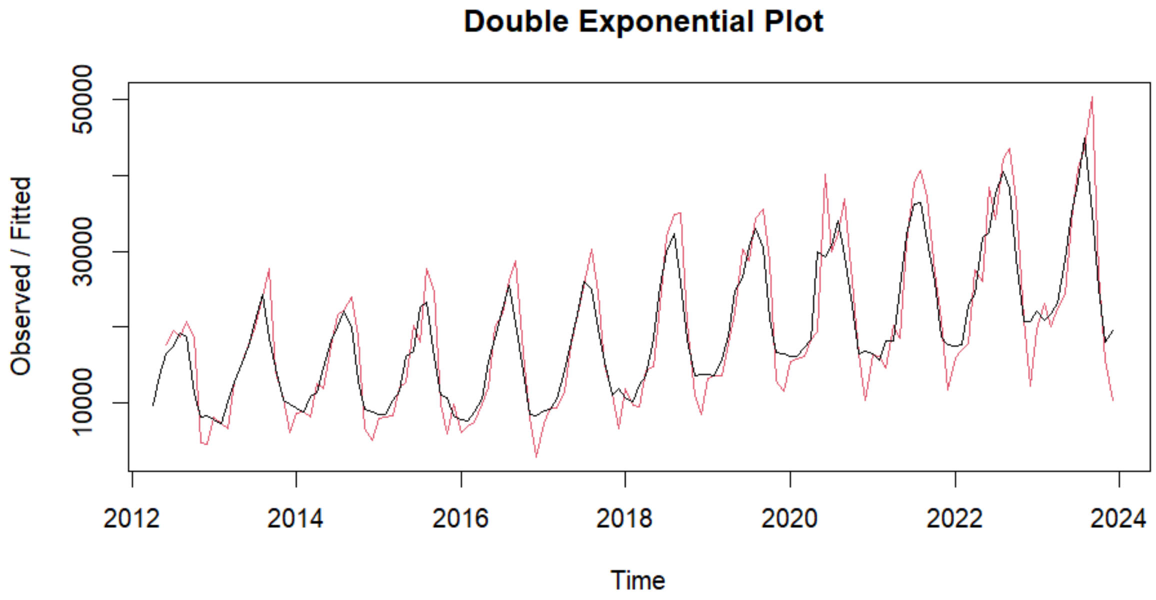


- Smoothing parameters:

  α = **0.9999339**

- Root Mean Square Error = **48055.83**

2. Double Exponential Smoothing



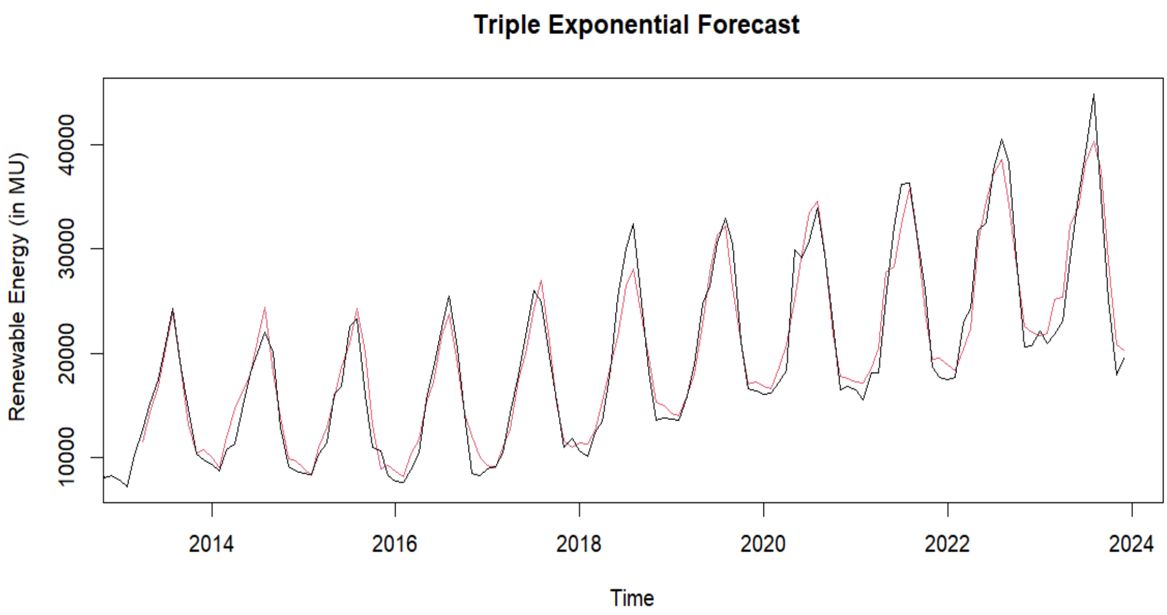
- Smoothing parameters:

  α = **1**

  β = **0.8665156**

- Root Mean Square Error = **46397.12**

3. Triple Exponential Smoothing



- Smoothing parameters:

  α = **46397.12**

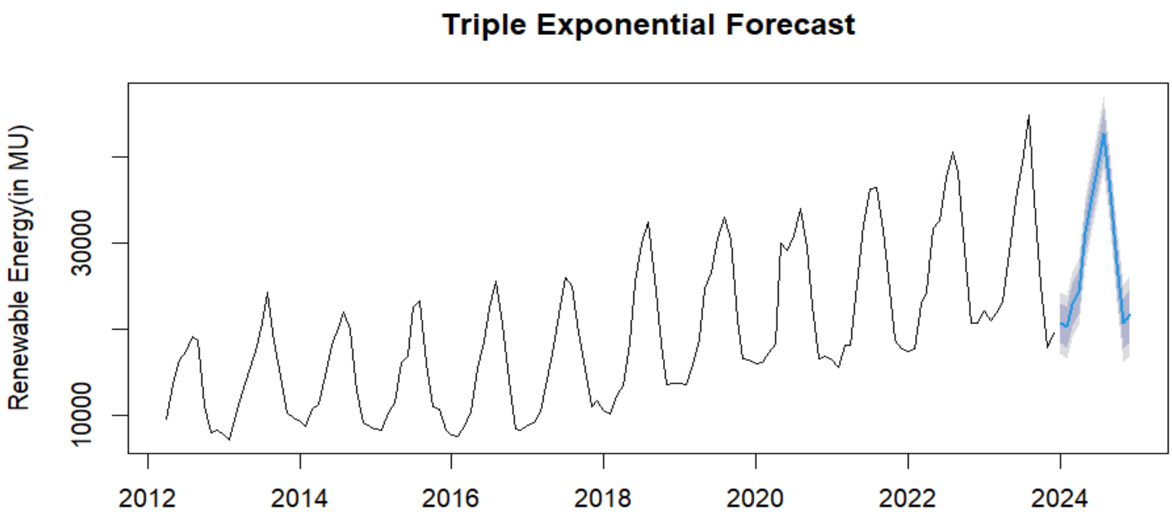
  β = **0.005199607**

  γ = **0.5170413**

- Root Mean Square Error = **1690.307**

➔ **As the RMSE value for triple exponential smoothing is the lowest, it is the most appropriate for forecasting.**

Forecasting for 2024 with Triple Exponential Smoothing:

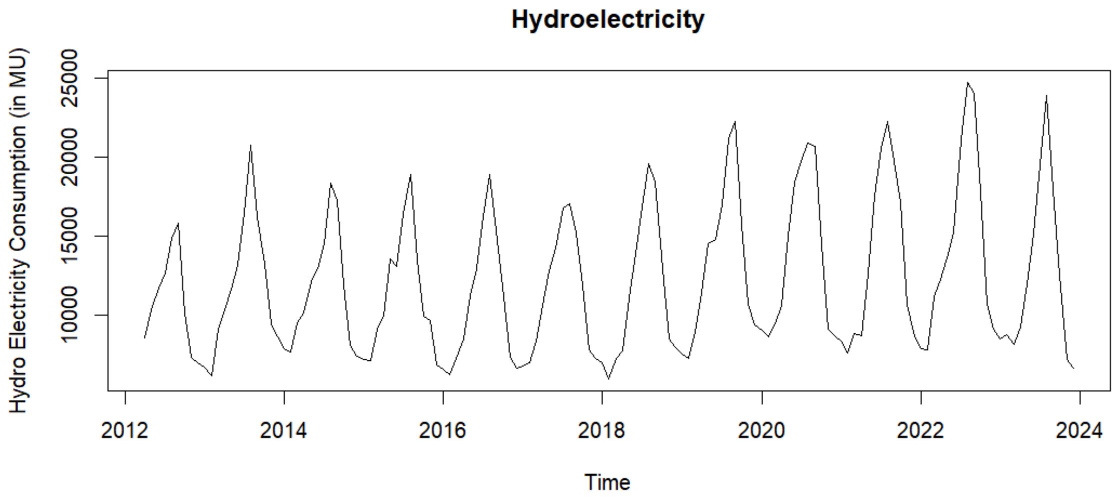


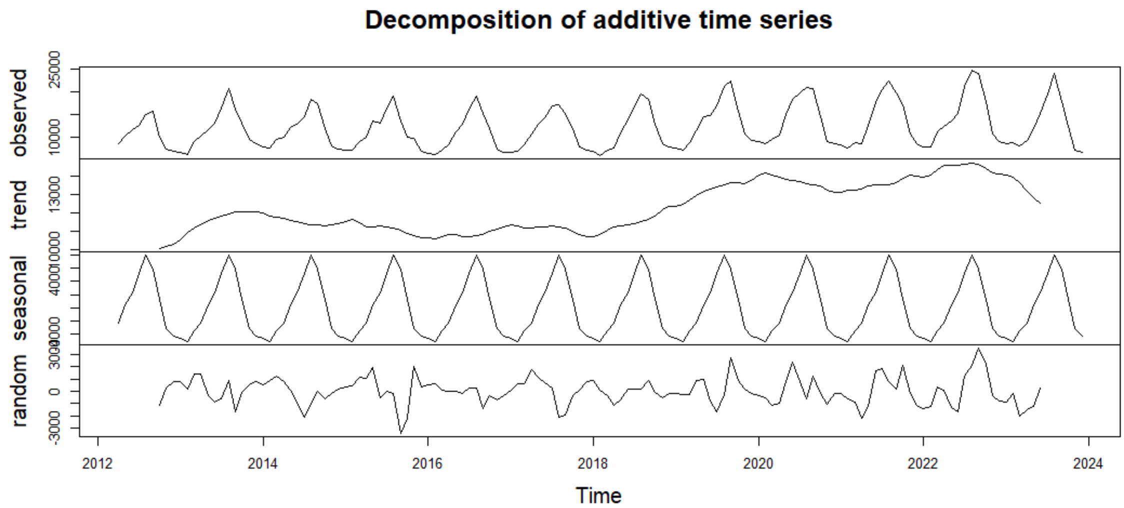
Sr. No.	Time	Forecasted Renewable Electricity consumption (in MU)
1	January 2024	20720.25
2	February 2024	20258.87
3	March 2024	22809.12
4	April 2024	24283.38
5	May 2024	31262.40
6	June 2024	35472.20
7	July 2024	39468.71
8	August 2024	42700.79
9	September 2024	35955.10
10	October 2024	27515.85
11	November 2024	20677.85
12	December 2024	21658.89

Hydroelectricity Consumption in India

To model and forecast the electricity consumed (in MU) in India that was produced from hydropower.

Time Series Plot:





- We may observe a slight increasing trend, there is a seasonal component in the data, and the irregularities seem fairly random.

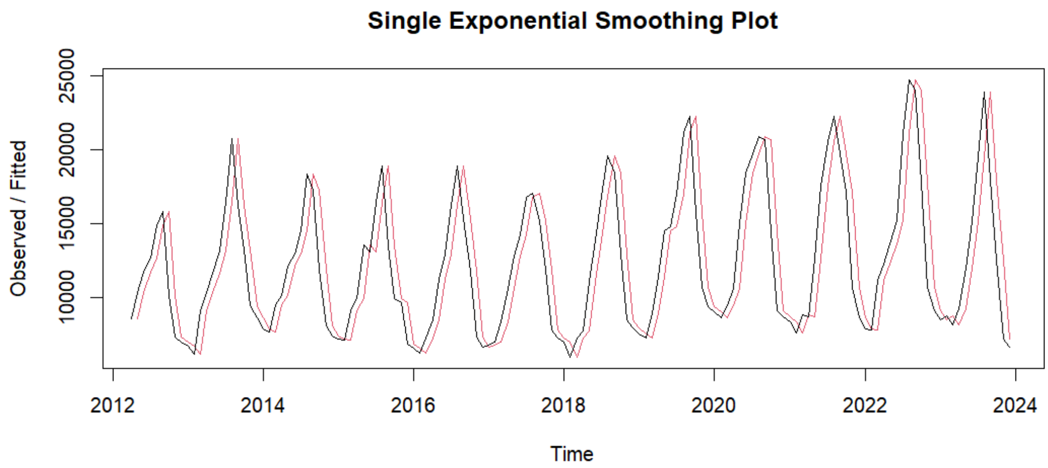
- The supply of hydroelectricity tends to increase from the beginning of the year, it peaks around August and then starts to decline.

This is caused by the rainfall patterns in the country as monsoons in India persist between June and September, leading to increased water flow. Also, after the monsoons, during September towards January of the next year, the conservation of stored water to ensure reliable power generation during the dry seasons is prioritized.

Using Exponential Smoothing to Forecast Hydroelectricity Supply

Using single, double and triple exponential smoothing on the data, we get the following results:

1. Single Exponential Smoothing:

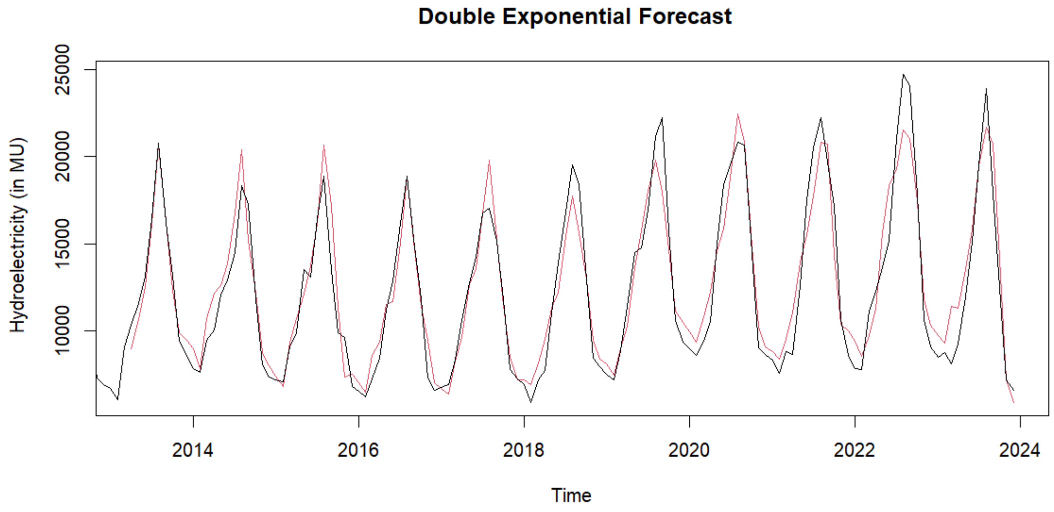


- Smoothing parameters:

  α = **0.9999339**

- Root Mean Square Error = **33973.36**

2. Double Exponential Smoothing:



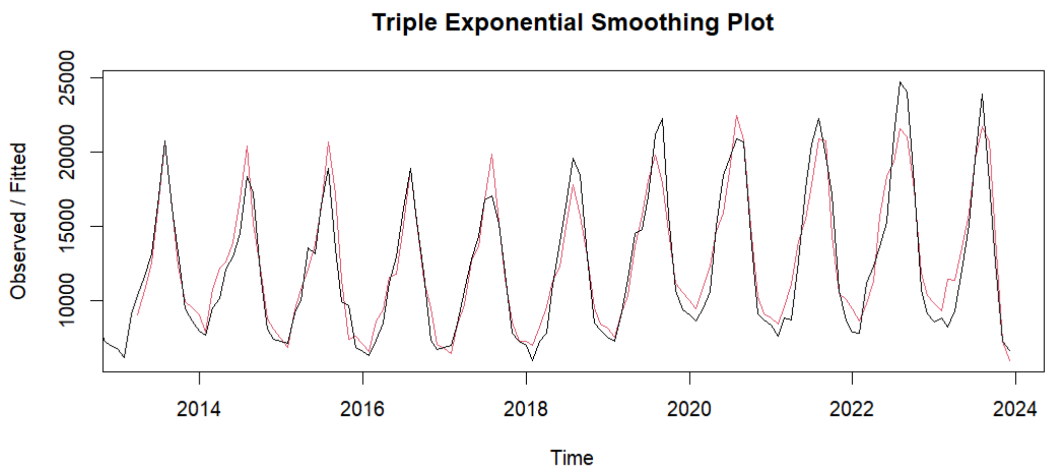
- Smoothing parameters:

  α = **1**

  β = **0.04263834**

- Root Mean Square Error = **35102.55**

3. Triple Exponential Smoothing:



- Smoothing parameters:

  α = **0.3065658**

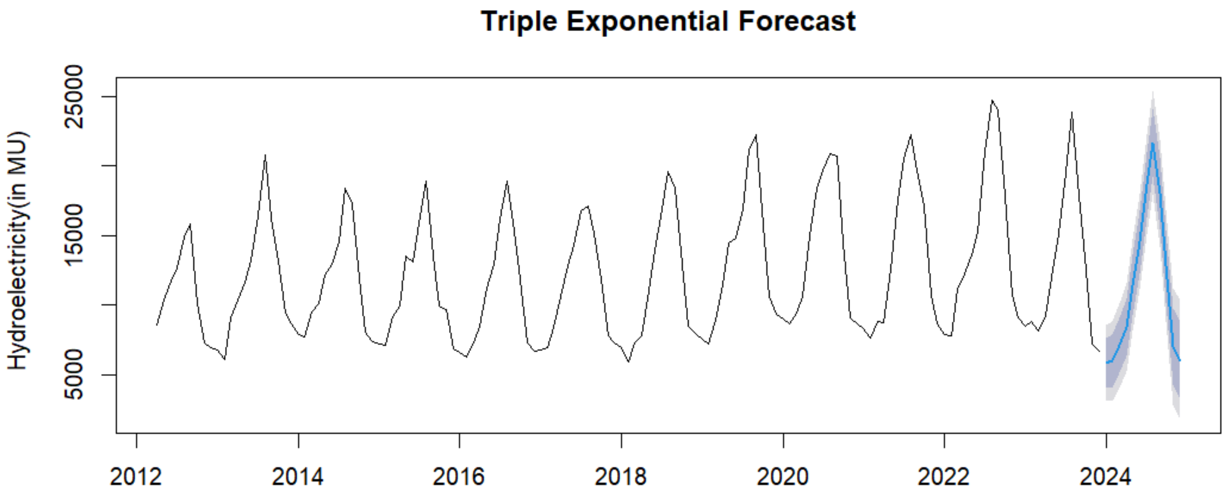
  β = **0.02984244**

  γ = **0.5672972**

- Root Mean Square Error = **16053.74**

➔ **As the RMSE value for triple exponential smoothing is the lowest, it is the most appropriate for forecasting.**

Forecasting for 2024 with Triple Exponential Smoothing:

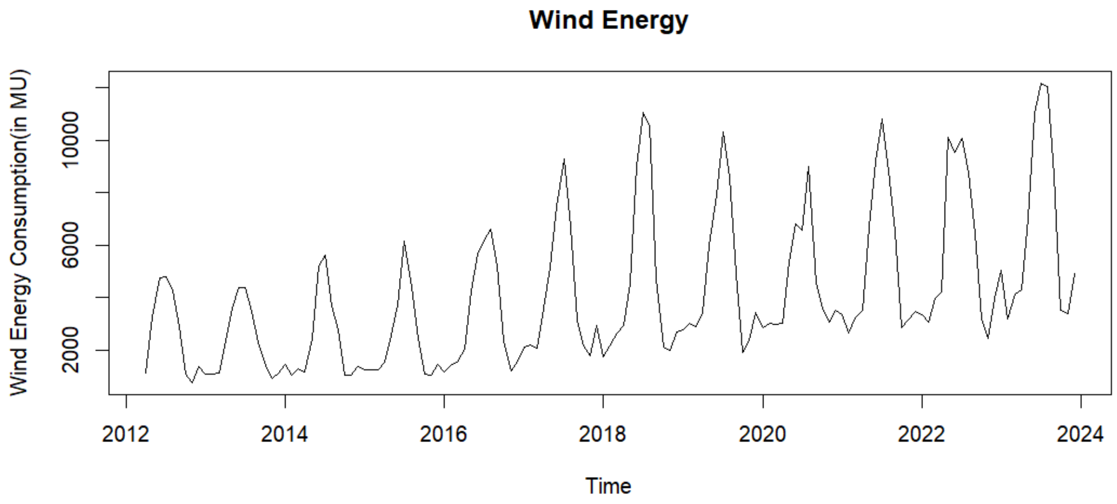


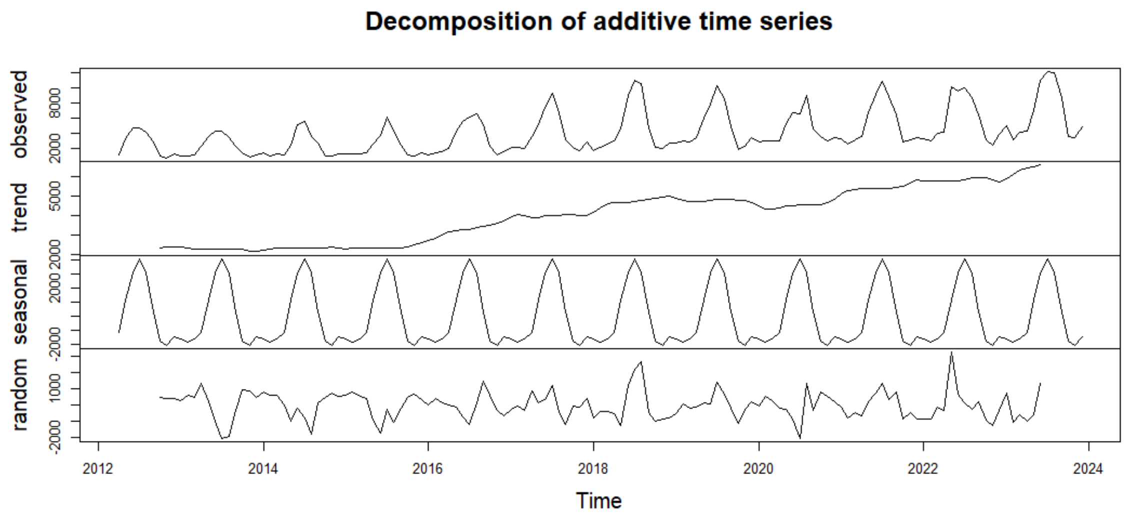
Sr. No.	Time	Forecasted Hydroelectricity consumption (in MU)
1	January 2024	5812.159
2	February 2024	5948.097
3	March 2024	7066.672
4	April 2024	8390.581
5	May 2024	11404.900
6	June 2024	14515.881
7	July 2024	18619.929
8	August 2024	21675.639
9	September 2024	18037.442
10	October 2024	12806.033
11	November 2024	7014.314
12	December 2024	6018.573

Wind Energy Consumption in India

To model and forecast the electricity consumed (in MU) in India that was produced from wind power.

Time series Plot:





- We can observe a slightly increasing trend, there is a seasonal component in the data, and the irregularities seem fairly random.

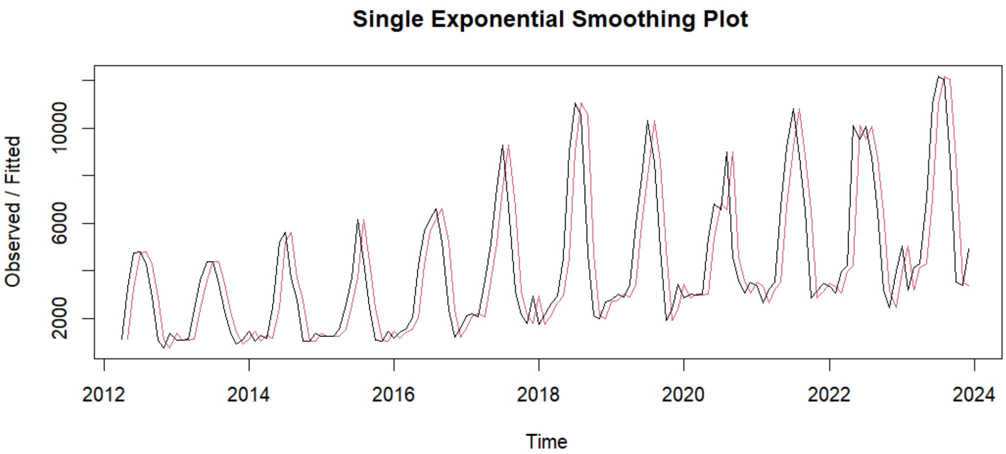
- The wind energy supply tends to increase from the start of the year, peaks in June and then starts to decline.

- India experiences seasonal changes in wind patterns due to the monsoon season. During the summer months, especially from March to May, India experiences the southwest pre-monsoon winds, which brings strong and consistent winds to many parts of the country, particularly along the western coast and in certain inland regions. This leads to an increase in wind energy production during these months. The combined effect of strong pre-monsoon winds and relatively less rain might lead to the highest wind speeds and power generation in June.
- As the monsoon progresses and transitions into the post-monsoon season (August onwards), the wind speeds generally weaken. This reduction in wind speed translates to lower power generation from wind farms.

Using Exponential Smoothing to Forecast Wind Energy Supply

Using single, double and triple exponential smoothing on the data, we get the following results:

1. Single Exponential Smoothing:

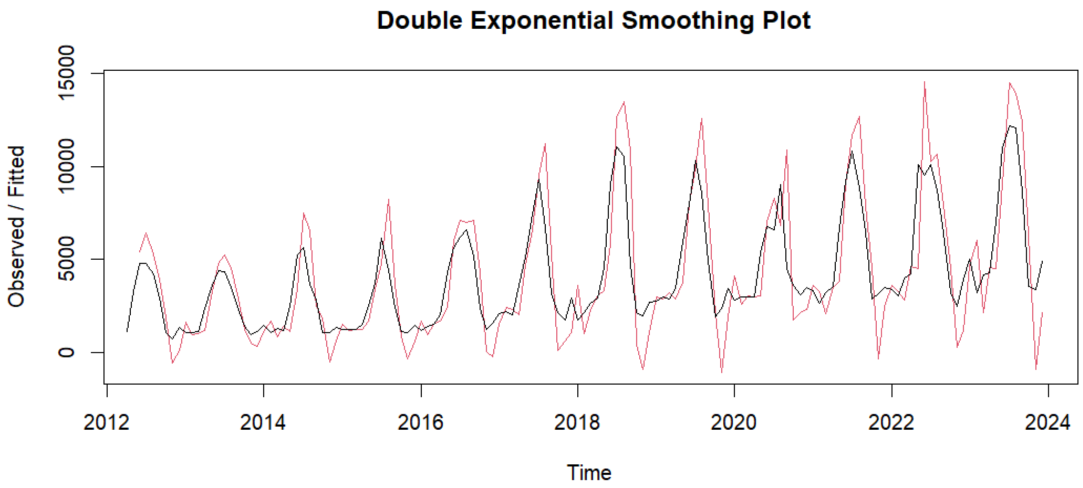


- Smoothing parameters:

  α = **0.9999339**

- Root Mean Square Error = **20691.21**

2. Double Exponential Smoothing:



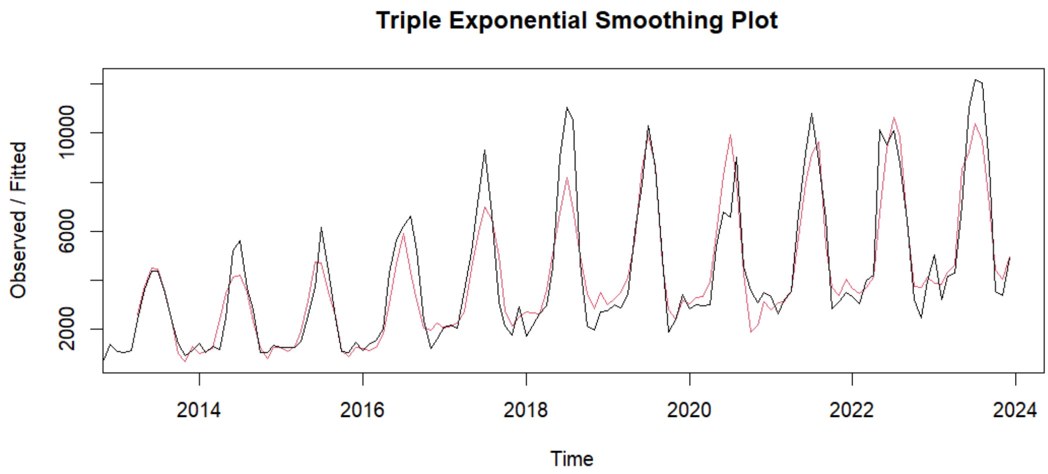
- Smoothing parameters:

  α = **1**

  β = **0.7355735**

- Root Mean Square Error = **22300.99**

3. Triple Exponential Smoothing:



- Smoothing parameters:

  α = **0.1176794**

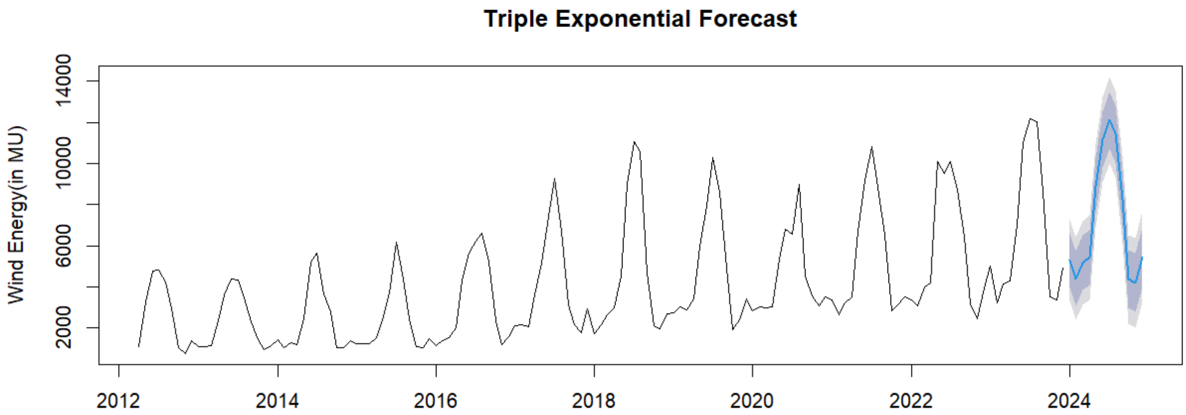
  β = **0.02922646**

  γ = **0.4594893**

- Root Mean Square Error = **11613.38**

→ **As the RMSE value for triple exponential smoothing is the lowest, it is the most appropriate for forecasting.**

Forecasting for 2024 with Triple Exponential Smoothing:

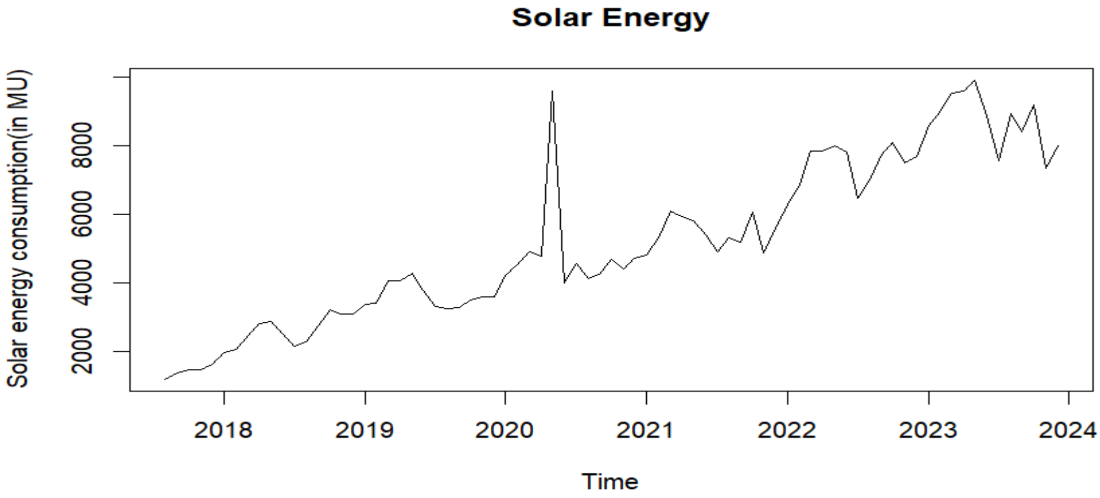


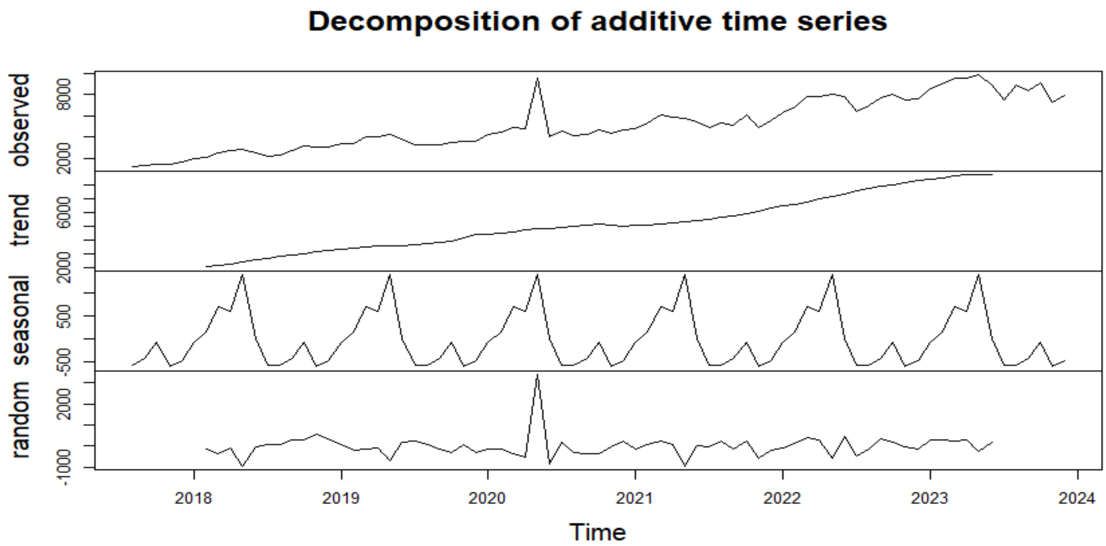
Sr. No.	Time	Forecasted Hydroelectricity consumption (in MU)
1	January 2024	5314.386
2	February 2024	4408.902
3	March 2024	5177.087
4	April 2024	5432.358
5	May 2024	8908.189
6	June 2024	11153.862
7	July 2024	12114.180
8	August 2024	11432.677
9	September 2024	8238.109
10	October 2024	4381.860
11	November 2024	4217.725
12	December 2024	5465.047

Solar Energy Consumption in India

To model and forecast the electricity consumed (in MU) in India that was produced from wind power.

Time Series Plot:



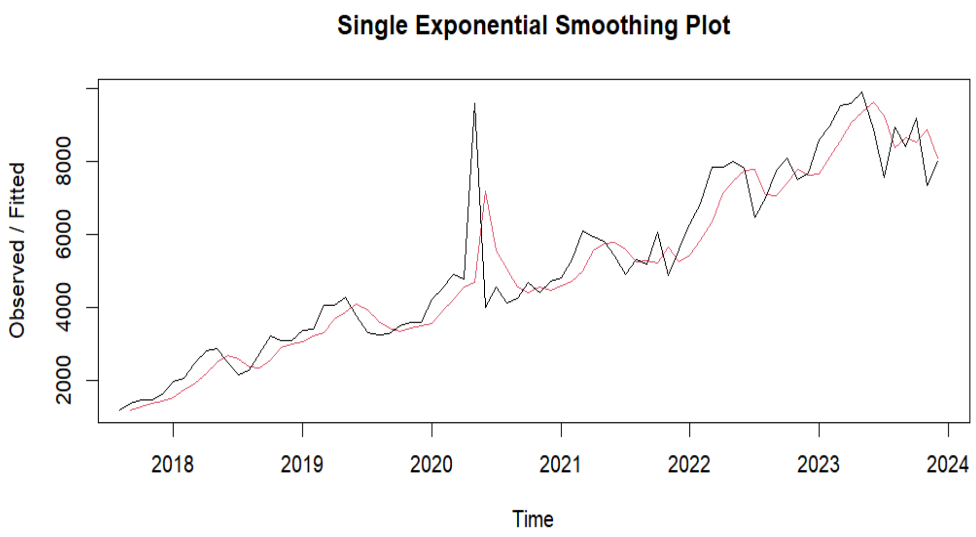


- We can observe a slight increasing trend, a seasonal component may be present in the data, and the irregularities seem fairly random.
- India's solar power dips after May due to a shift in the sun's angle, reducing direct sunlight captured by panels. Additionally, the monsoon season (June onwards) brings increased cloud cover, further hindering solar energy generation. While dust storms can also play a minor role, advancements in solar panel technology are helping mitigate some of these seasonal variations.

Using Exponential smoothing to forecast Solar Energy Supply

Using single, double and triple exponential smoothing on the data, we get the following results:

1. **Single Exponential Smoothing:**

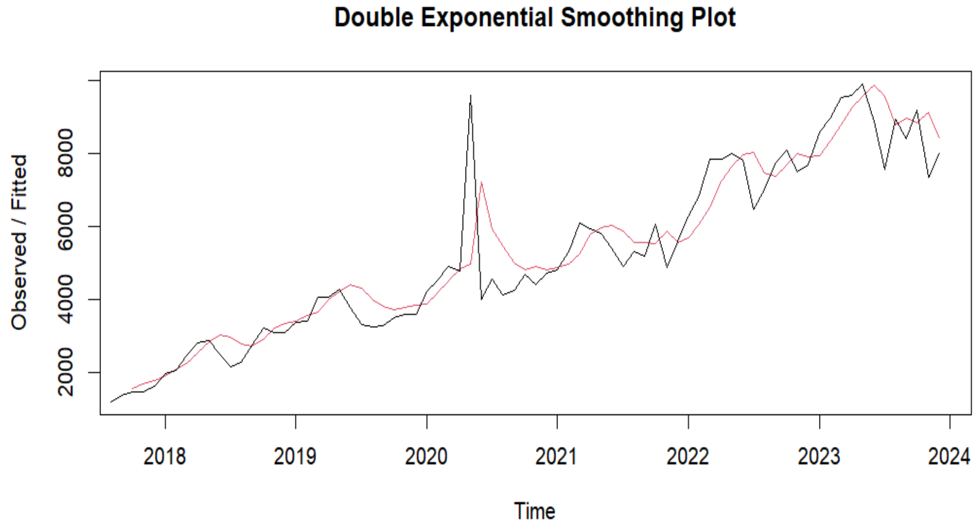


- Smoothing parameters:

  α = **0.5098932**

- Root Mean Square Error = **7911.696**

2. Double Exponential Smoothing:



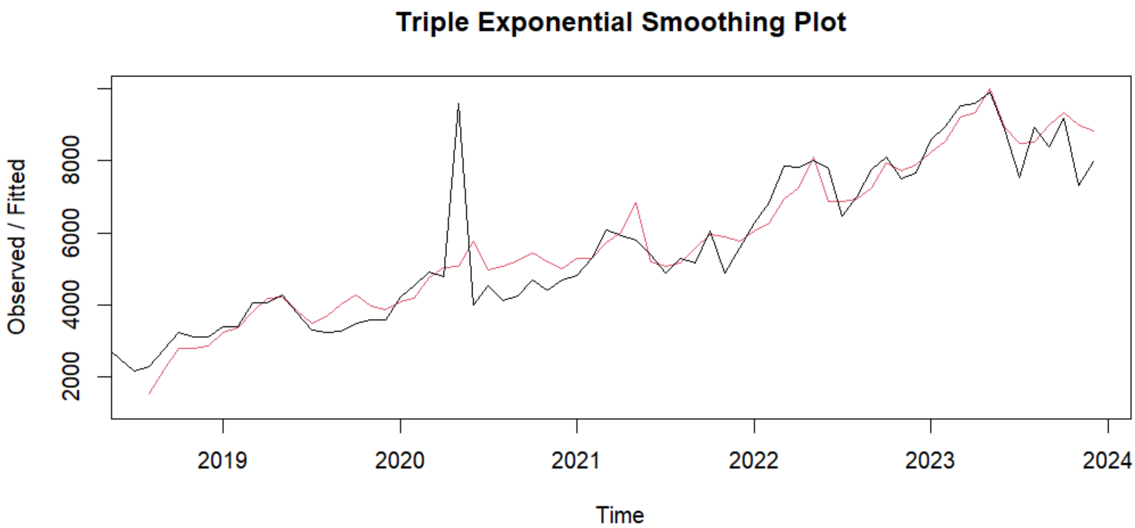
- Smoothing parameters:

  α = **0.4482388**

  β = **0.01781607**

- Root Mean Square Error = **7859.637**

3. Triple Exponential Smoothing:



- Smoothing parameters:

  α = **0.2402498**

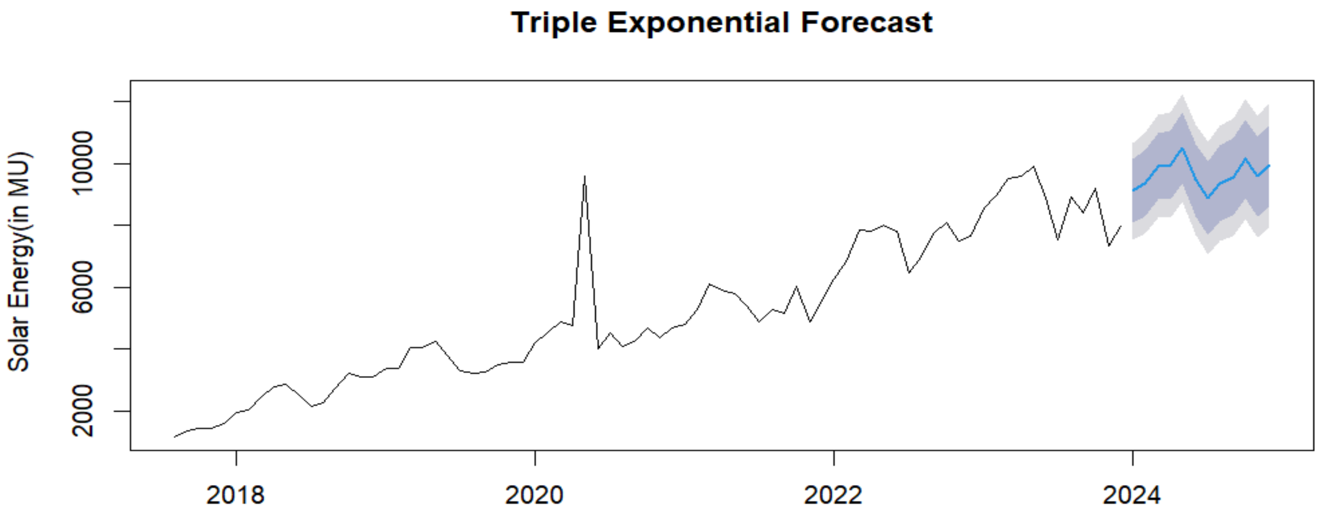
  β = **0**

  γ = **0.213544**

- Root Mean Square Error = **6428.832**

→ **As the RMSE value for triple exponential smoothing is the lowest, it is the most appropriate for forecasting.**

Forecasting for 2024 with Triple Exponential Smoothing:



Sr. No.	Time	Forecasted Solar energy consumption (in MU)
1	January 2024	9106.112
2	February 2024	9355.783
3	March 2024	9907.751
4	April 2024	9952.507
5	May 2024	10492.544
6	June 2024	9458.969
7	July 2024	8876.101
8	August 2024	9348.672
9	September 2024	9560.066
10	October 2024	10133.017
11	November 2024	9573.538
12	December 2024	9934.911

Multiple Linear Regression

In general, the response Y may be related to **k** regressors or predictor variables.

Following is the multiple regression model:

$$\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon}$$

The parameters β_j, j = 0, 1…, k, are called the regression coefficients. This model describes a hyperplane in the k-dimensional space of the regressor variables X_j. The parameter β_j represents the expected change in the response y per unit change in X_j when all remaining regressor variables xi ($i \neq j$) are held constant. For this reason, the parameters β_j, j = 1, 2…, k, are often called **partial regression coefficients**.

Modelling Household Electricity Consumption of Pune City

To model the electricity consumption of Pune's households, primary data is collected through a survey for the following variables, and an adequate random sample is selected.

Three models are fit, one for each of the seasons.

The following variables are considered:

Response/Dependent Variable (Y): Electricity bill amount in different seasons.

Y_1: Average monthly electricity bill amount during Summer

Y_2: Average monthly electricity bill amount during Monsoon

Y_3: Average monthly electricity bill amount during Winter

Regressors/Independent Variables (X):

X_1: Area of House (in square feet)

X_2: Number of occupants in the household

X_3: Number of appliances in the household

X_4: Existence and charging of Electric Vehicles (EVs) at home (Binary)*

X_5: Utilization of non-conventional energy sources (Binary)*

X_6: Deployment of energy-efficient appliances in the household (Binary)*

**Note: The variables X_4 and X_5 would be measured by the following rule:*

X_4: Existence and usage of Electric Vehicles (EVs) at home		Existence of EVs	
		Yes (1)	No (0)
Charged at Home	Yes (1)	X_4= 1	X_4=0
	No (0)	X_4=0	X_4=0

X_5: Utilization of non-conventional energy sources		Existence of Sources	
		Yes (1)	No (0)
Energy generated from these sources at home	Yes (1)	X6= 1	X6=0
	No (0)	X6=0	X6=0

**Further as the sample size is large enough, the binary variables are converted into Standard Normal Variates using the Central Limit Theorem:

As X_4, X_5, X_6 are all Binomial random variables and sample size(n) is large,

$$\frac{x_i - \mu_i}{\sigma_i} \sim N(0,1)$$

where μ_i: Mean of x_i

σ_i: Standard Deviation of x_i

Log-Transform of Response variables to meet Normality Assumption

The histograms of the response variables suggest that $\mathbf{Y_1}$, $\mathbf{Y_2}$ and $\mathbf{Y_3}$ seem positively skewed, suggesting that these variables might be lognormal.

So, if $\mathbf{Y_1}$, $\mathbf{Y_2}$ and $\mathbf{Y_3}$ are lognormal variates then their natural logs should be normally distributed.

Checking the normality of response variables using Shapiro-Wilk Test for normality:

- **The hypothesis to be tested:**

  H_0 : The sample comes from normal distribution

  Vs

  H_1 : The sample does not come from normal distribution

- **Decision criterion:**

  We reject H_0 if the $p\text{-}value < 0.05$, otherwise we accept H_0.

Variable	Y_1	Y_2	Y_3
P-value	1.64e-11	2.013e-08	8.136e-10

Variable	$ln(Y_1)$	$ln(Y_2)$	$ln(Y_3)$
P-value	0.181	0.2368	0.2267

- **Decision and Conclusion:**

  As P- value<0.05 for Y_1, Y_2 and Y_3, thus we may conclude that they are not normally distributed.

  On the other hand, as P- value>0.05 for natural logs of Y_1, Y_2 and Y_3, thus $ln(Y_1)$, $ln(Y_2)$ **and** $ln(Y_3)$ are normally distributed.

As a consequence of the above result, we use the log transformation of the response variables to fit the regression model.

Consider:

$$Y_1^* = ln(Y_1)$$

$$Y_2^* = ln(Y_2)$$

$$Y_3^* = ln(Y_3)$$

Multiple Regression Model of Household Electricity Consumption in Summer

Complete Multiple Regression Model:

$$Y_1^* = 6.3489898 + (0.0007298)X_1 + (-0.0188885)X_2 + (0.0285456)X_3 + (0.0824444)X_4 + (0.0385655)X_5 + (0.0352542)X_6$$

This is a preliminary model which might not be optimal.

Backward Elimination Method of Variable Selection:

We use the backwards elimination method to find the best fit model to remove the *unnecessary* variables.

Coefficients	Estimate	Std. Error	t value	P- Value
(Intercept)	6.297e+00	8.279e-02	76.067	**< 2e-16** ***
X_1	1.5222	0.1499	10.152	**< 2e-16** ***
X_3	228.2366	72.7923	3.135	**2.55e-05** ***
X_4	259.8187	76.2894	3.406	**0.00299** **
X_6	4.271e-02	2.737e-02	1.561	**0.12146**

Thus, after the backward elimination method, we are only left with three regressors- X_1, X_3, X_4 and X_6.

The best-fit model obtained is:

$$Y_1^* = 6.2971937 + (0.0007008\ ) X_1 + (0.0282784) X_3 + (0.0874918) X_4 + (0.0427100)X_6$$

Verifying the model by checking the Significance of regressors:

- **The hypothesis to be tested:**

$$\mathbf{H_0 : \beta_i = 0 \quad Vs \quad H_1 : \beta_i \; ; \; i=1,2,...6}$$

- **Decision Criterion:**

  We reject H_0 if the *p-value < 0.05*, otherwise we accept H_0.

Coefficients	**Estimate**	**Std. Error**	**t value**	**P- Value**
(Intercept)	6.349e+00	8.948e-02	70.953	**< 2e-16 *****
X_1	7.298e-04	8.225e-05	8.873	**1.68e-14 *****
X_2	-1.889e-02	1.734e-02	-1.090	**0.27832**
X_3	2.855e-02	6.430e-03	4.440	**2.18e-05 *****
X_4	8.244e-02	2.904e-02	2.839	**0.00541 ****
X_5	3.857e-02	3.076e-02	1.254	**0.21269**
X_6	3.525e-02	2.774e-02	1.271	**0.20645**

- **Decision and Conclusion**

  As P- value<0.05 for X_1, X_3, X_4 and X_6, there is a linear relation between the Electricity Bill in summer and only the **Area of the House, Number of appliances in the household, Existence and charging of Electric Vehicles (EVs) at home, Deployment of energy-efficient appliances in the household.**

Therefore, the best fit model obtained is:

$$\boldsymbol{Y_1^* = 6.2971937 + (0.0007008)\, X_1 + (0.0282784)\, X_3 + (0.0874918)\, X_4 + (0.0427100)\, X_6}$$

With a coefficient of determination: $\mathbf{R^2 = 0.7752}$

Multicollinearity:

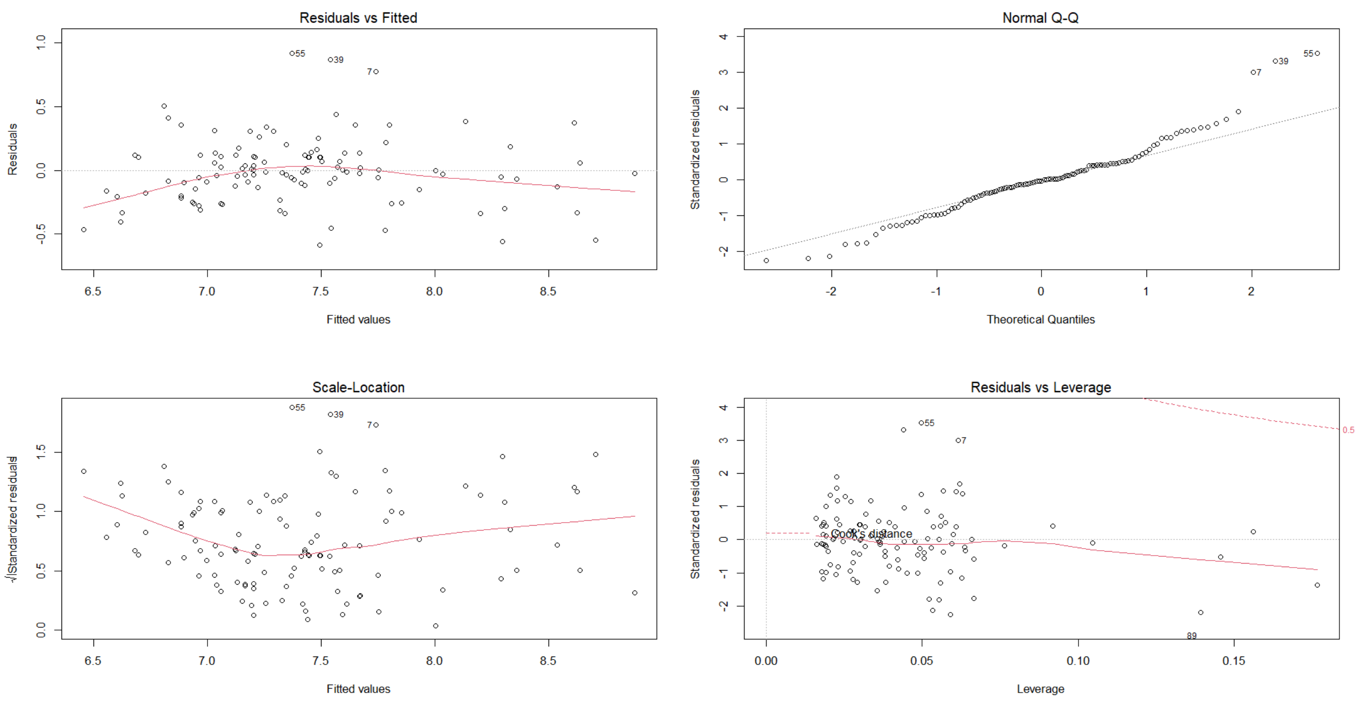
VIF: The VIF for each term in the model measures the combined effect of the dependencies among the regressors on the variance of that term. One or more large VIFs indicate multicollinearity. Practical experience indicates that if any of the VIFs exceeds 5 or 10, it is an indication that the associated regression coefficients are poorly estimated because of multicollinearity.

X_1	X_3	X_4	X_6
1.581171	**1.154029**	**1.322094**	**1.192720**

Therefore, as the VIF values for all variables are close to 1, there is no multicollinearity in our model.

Diagnostic Checking:

Diagnostic plots for our model:



From the above diagnostic plots, we can observe that all the assumptions of linear regression are satisfied, thus our model is adequate.

Multiple Regression Model of Household Electricity Consumption in Monsoon

Complete Multiple Regression Model:

$$Y_2^* = 6.1440296 + (0.0006769)X_1 + (-0.0074799)X_2 + (0.0303099)X_3 + (0.0990281)X_4 + (0.0163887)X_5 + (0.0318239)X_6$$

This is a preliminary model which might not be optimal.

Backward Elimination Method of Variable Selection:

We use the backwards elimination method to find the best-fit model to remove the *unnecessary* variables.

Coefficients	**Estimate**	**Std. Error**	**t value**	**P- Value**
(Intercept)	6.095e+00	8.203e-02	74.297	**< 2e-16 *** **
X_1	6.931e-04	5.573e-05	12.436	**< 2e-16 *** **
X_3	2.948e-02	6.578e-03	4.481	**1.82e-05 *** **
X_4	1.017e-01	2.958e-02	3.438	**0.000827 *** **

Thus, after the backward elimination method, we are only left with three regressors- X_1, X_3, and X_4.

The best-fit model obtained is:

$$Y_2^* = 6.0948395 + (0.0006931)\,X_1 + (0.0294751)\,X_3 + (0.1016741)\,X_4$$

Verifying the model by checking the Significance of regressors:

- **The hypothesis to be tested:**

$$H_0 : \beta_i = 0 \quad Vs \quad H_1 : \beta_i \; ; i=1,2,...6$$

- **Decision Criterion:**

  We reject H_0 if the *p-value< 0.05*, otherwise we accept H_0.

Coefficients	Estimate	Std. Error	t value	P- Value
(Intercept)	6.144e+00	9.249e-02	66.432	**< 2e-16 *****
X_1	6.768e-04	8.501e-05	7.962	**1.83e-12 *****
X_2	-7.480e-03	1.792e-02	-0.417	**0.408**
X_3	3.031e-02	6.646e-03	4.561	**1.35e-05 *****
X_4	9.903e-02	3.001e-02	3.299	**0.00131 ****
X_5	1.639e-02	3.180e-02	0.515	**0.60731**
X_6	3.182e-02	2.867e-02	1.110	**0.26943**

- **Decision and Conclusion**

  As p-value < 0.05 for X_1, X_3, and X_4, there is a linear relation between the Electricity Bill in summer and only the **Area of the House, the Number of appliances in the household and the Existence and charging of Electric Vehicles (EVs) at home.**

Therefore, the best-fit model obtained is:

$$Y_2^* = 6.0948395 + (0.0006931)\, X_1 + (0.0294751)\, X_3 + (0.1016741)\, X_4$$

With a coefficient of determination: $\mathbf{R^2 = 0.7596}$

Multicollinearity:

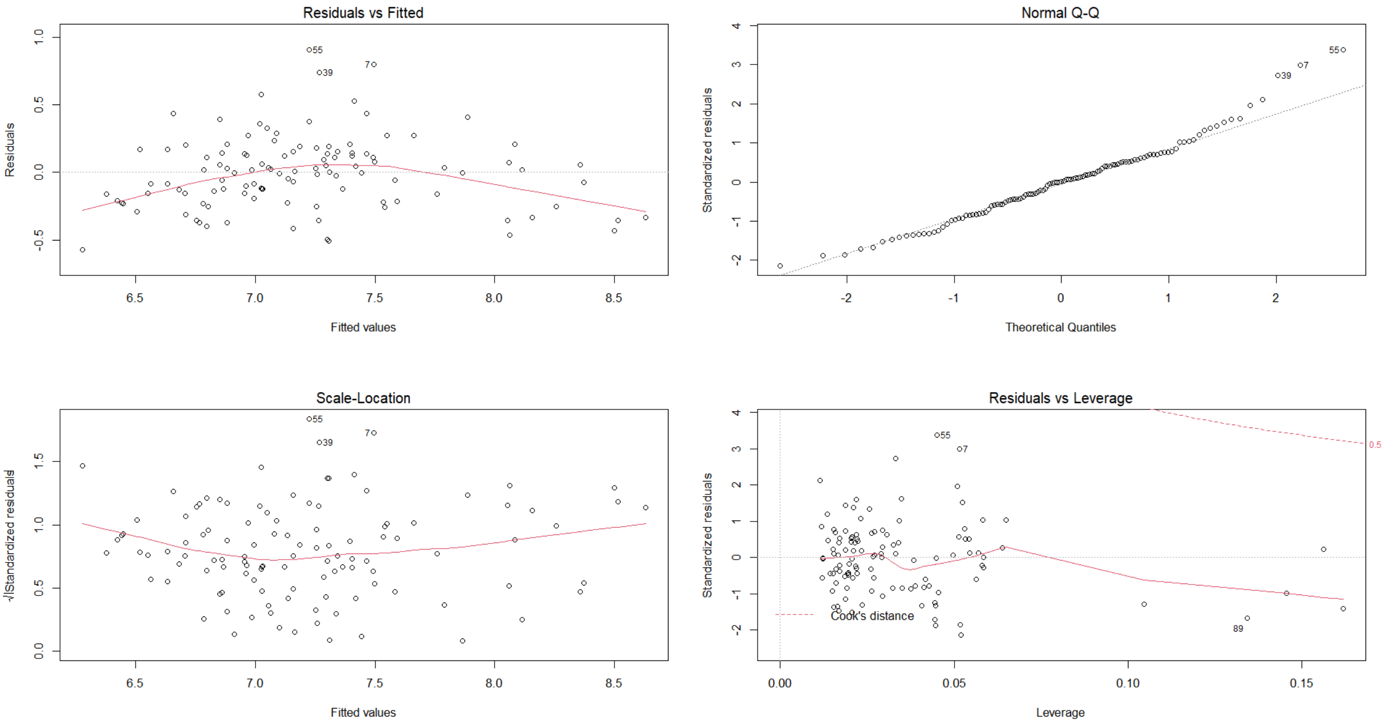
VIF:

X_1	X_3	X_4
1.375249	1.145217	1.321849

Therefore, as the VIF values for all variables are close to 1, there is no multicollinearity in our model.

Diagnostic Checking:

Diagnostic plots for our model:



From the above diagnostic plots, we can observe that all the assumptions of linear regression are satisfied, thus our model is adequate.

Multiple Regression Model of Household Electricity Consumption in Winter

Complete Multiple Regression Model:

$$Y_3^* = 5.9563410 + (0.0007408)X_1 + (-0.0024377\ )X_2 + (\ 0.0242238)X_3 + (0.1124401)X_4 + (0.0250122)X_5 + (\ 0.0116217)X_6$$

This is a preliminary model which might not be optimal.

Backward Elimination Method of Variable Selection:

We use the backwards elimination method to find the best fit model to remove the *unnecessary* variables.

Coefficients	**Estimate**	**Std. Error**	**t value**	**P- Value**
(Intercept)	5.919e+00	8.722e-02	67.855	**< 2e-16 *****
X_1	7.653e-04	5.926e-05	12.914	**< 2e-16 *****
X_3	2.367e-02	6.995e-03	3.384	**0.000989 *****
X_4	1.159e-01	3.145e-02	3.686	**0.000354 *****

Thus, after the backward elimination method, we are only left with three regressors- X_1, X_3, and X_4.

The best-fit model obtained is:

$$Y_3^* = 5.9186546 + (\ 0.0007653)\ X_1 + (\ 0.0236684\ )\ X_3 + (0.1159253)\ X_4$$

Verifying the model by checking the Significance of regressors:

- **The hypothesis to be tested:**

$$H_0 : \beta_i = 0 \quad Vs \quad H_1 : \beta_i \; ; \; i=1,2,...6$$

- **Decision Criterion:**

  We reject H_0 if the *p-value< 0.05*, otherwise we accept H_0.

Coefficients	**Estimate**	**Std. Error**	**t value**	**P- Value**
(Intercept)	5.956e+00	9.885e-02	60.259	**< 2e-16 *****
X_1	7.408e-04	9.085e-05	8.154	**6.86e-13 *****
X_2	-2.438e-03	1.915e-02	-0.127	**0.898943**
X_3	2.422e-02	7.103e-03	3.411	**0.000666 *****
X_4	1.124e-01	3.208e-02	3.505	**0.00131 ****
X_5	2.501e-02	3.398e-02	0.736	**0.463313**
X_6	1.162e-02	3.064e-02	0.379	**0.705207**

- **Decision and Conclusion**

  As p-value < 0.05 for X_1, X_3, and X_4, there is a linear relation between the Electricity Bill in summer and only the **Area of the House, Number of appliances in the household and Existence and charging of Electric Vehicles (EVs) at home.**

Therefore, the best-fit model obtained is:

$$Y_3^* = 5.9186546 + (0.0007653) X_1 + (0.0236684) X_3 + (0.1159253) X_4$$

With coefficient of determination: $\mathbf{R^2 = 0.7613}$

Multicollinearity:

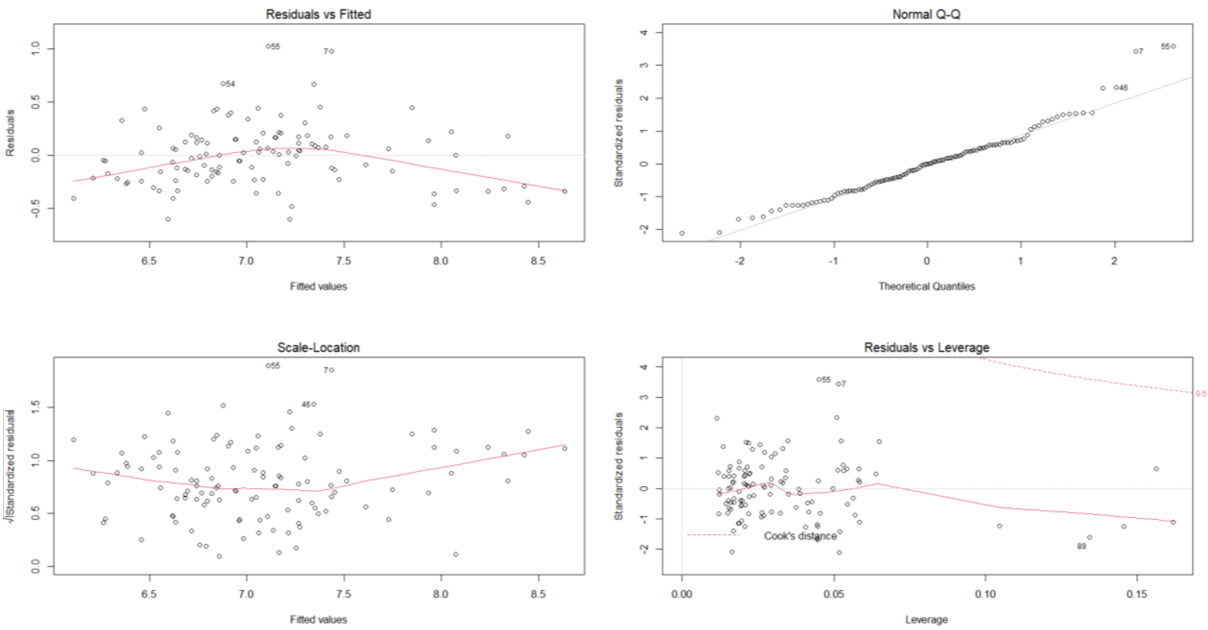
VIF:

X_1	X_3	X_4
1.375249	**1.145217**	**1.321849**

Therefore, as the VIF values for all variables are close to 1, there is no multicollinearity in our model.

Diagnostic Checking:

Diagnostic plots for our model:



From the above diagnostic plots, we can observe that all the assumptions of linear regression are satisfied, thus our model is adequate.

A confidence interval for household electricity consumption in Pune

The obtained multiple regression model can be written in the form:

$$Y = \beta \underline{X}$$

Thus, the estimated value of response variable, say Y_0 for particular values of regressors ($\underline{X_0}$= X1, X2…, Xn) can be defined as

$$\hat{Y}_0 = \hat{\beta} \underline{X_0}$$

A 100(1-α) percent confidence interval for these estimates can be given by:

$$(\hat{Y}_0 - t_{\alpha,n-p}\sqrt{\hat{\sigma}^2(1 + {X_0}'(X'X)^{-1}X_0)}\,, \hat{Y}_0 + t_{\alpha,n-p}\sqrt{\hat{\sigma}^2(1 + {X_0}'(X'X)^{-1}X_0)})$$

From our data, the 95% confidence interval is:

$7.153448 \le Y_1^* \le 7.671032$

$6.942798 \le Y_2^* \le 7.477703$

$6.790203 \le Y_3^* \le 7.361897$

Thus the 95% confidence interval for Average Electricity Bill Amount of a household in Pune (in Rs.) is:

During Summer: ***$1278.506 \le Y_1 \le 2145.294$***

During monsoon: ***$1035.664 \le Y_2 \le 1768.174$***

During Winter: ***$889.0944 \le Y_3 \le 1574.8213$***

Thus the 95% confidence interval for the Average Electricity consumed by a household in Pune (in Units):

During Summer: ***$109.1807 \le Z_1 \le 183.2019$***

During monsoon: ***$88.44266 \le Z_2 \le 150.99692$***

During Winter: ***$75.92608 \le Z_3 \le 134.48516$***

*Where Z is the Average Electricity consumed

Correlation Analysis

Correlation between Energy Consumption and Awareness about Sustainable Practices

To check if there is enough awareness among people in Pune and if it affects their Electricity consumption, the awareness levels are determined based on responses to a set of questions included in the survey, with responses measured on an ordinal scale ranging from 1 to 5.

To find the association between electricity consumption and awareness levels, as the awareness data is on an ordinal scale, **Spearman's Correlation coefficient** and **Kendall's Tau Correlation coefficient were used.**

Kendall's Tau and Spearman's rank correlation coefficient assess statistical associations based on the ranks of the data. Ranking data is carried out on the variables that are separately put in order and numbered.

A value of ± 1 indicates a perfect degree of association between the two variables. As the correlation coefficient value goes towards 0, the relationship between the two variables will be weaker. The direction of the relationship is indicated by the sign of the coefficient; a + sign indicates a positive relationship and a – sign indicates a negative relationship.

Let,

Y: Average monthly Electricity bill of the household.

AS: The awareness score of individuals

Here,

There is a set of seven questions asked in the survey, the responses were measured on an ordinal scale ranging from 1 to 5 for each. So, the maximum and minimum scores possible are 35 and 7 respectively. Thus, the scores were converted into percentages and then further to an ordinal scale.

So, 0-10 % → 1, 11-20 % → 2 and so on.

For our data, the correlation between Y and AS is:

- Spearman’s Correlation Coefficient = -0.01595184
- Kendall’s Tau Coefficient = -0.02089948

These values of the correlation coefficients suggest that

1. There is almost no association between electricity consumption and people's awareness of sustainable electricity usage.
2. The little association that might be there is negative.

This might be because even when there is awareness, people lack the proactiveness to adopt these practices in everyday life. Also, People might be aware of the need for sustainability but lack specific knowledge about impactful actions in their daily lives.

Conclusion

1. Analysis of Energy Consumption in India

In India, the total consumption of energy (from renewable as well as non-renewable sources) has shown an increase in the last decade. The increase in total consumption appears to be steeper than the increase in consumption of energy from renewable sources. This suggests that the total consumption of electricity is growing faster than renewable consumption.

The Total Renewable Energy Consumption:

India's total renewable energy supply is steadily increasing. However, this rise is accompanied by fascinating seasonal variations depending on the specific renewable resource.

Hydroelectricity:

The supply of hydroelectricity depends largely on monsoons. As rainfall peaks in August, hydropower generation surges. But, with an eye towards long-term sustainability, water is then conserved for drier months, leading to a decline in output.

Wind Energy:

India's wind power follows a rhythm dictated by the monsoon cycle. Pre-monsoon winds, particularly strong in June, propel wind energy production to its zenith. However, as the monsoon progresses, wind speeds weaken, causing a decrease in a generation.

Solar Energy:

Solar energy production is the highest during the summer seasons. But then, as the sun's angle shifts and the monsoon brings cloud cover, solar energy dips. While dust storms can also play a temporary dampener, advancements in technology are helping to lessen the impact of these seasonal fluctuations.

2. Modelling Household Electricity Consumption of Pune City

Pune’s household Electricity consumption depends on the square footage area of the household, Number of appliances in the household and Existence and charging of Electric Vehicles (EVs) at home. Factors such as the Number of Occupants in the household and the Deployment of energy-efficient appliances in the household were found to have no significant effect.

In addition to this, only during the summer, deployment of energy-efficient appliances in the household was also found to have an effect.

3. Correlation between Energy Consumption and Awareness about Sustainable Practices

There seems to be almost no association between electricity consumption and people's awareness of sustainable electricity usage. This might be because even when people are aware, they lack the proactiveness to adopt these practices in everyday life. People might understand the importance of sustainable practices but lack the motivation or resources to implement them. Also, People might be aware of the need for sustainability but lack specific knowledge about impactful actions in their daily lives.

Scope and Limitations

- More historical data would allow for time series modelling, providing more precise forecasts.

- Considering geographic and economic factors would refine our forecasts for a more nuanced view.

- This project could be further expanded by conducting the same analysis at a regional level to explore the diversity in consumption patterns in India.

- The modelling of household electricity consumption would be carried out with greater accuracy with a larger sample size.

- The assessment of awareness among the public that was carried out was relatively vague, and a more sophisticated test would have given us a more accurate measure that potentially could have been incorporated into the model itself for a more holistic view.

References

Links:

1. Grid Controller of India Limited- https://posoco.in/en/
2. Journal of Electrical Systems and Information Technology- https://jesit.springeropen.com/articles/10.1186/s43067-023-00104-2
3. Analysing and Projecting Indian Electricity Demand to 2030 - Thomas Spencer, Fellow, TERI, & Associate Fellow, IDDRI Aayushi Awasthy, Associate Fellow, TERI- https://www.teriin.org/sites/default/files/2019-02/Analysing%20and%20Projecting%20Indian%20Electricity%20Demand%20to%202030.pdf
4. Energy Consumption Analysis Based on Energy Efficiency Approach: A Case of Suburban Area- https://www.researchgate.net/publication/312310279_Energy_Consumption_Analysis_Based_on_Energy_Efficiency_Approach_A_Case_of_Suburban_Area

Books:

1. *Introduction to Linear Regression Analysis- Book by Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining*
2. *The analysis of time series- Book by Christopher Chatfield*

Appendix-1

Questions about awareness of Sustainable Lifestyle Practices in the survey;

1. How aware are you of the energy consumption of different household appliances?

2. How often do you consciously turn off lights and appliances when not in use?

3. How likely are you to invest in energy-efficient upgrades for your home? (irrespective of affordability)

4. How aware are you about the concept of carbon footprint and its effect on electricity consumption?

5. To what extent do you believe individuals can make a significant impact on reducing electricity consumption and promoting sustainability?

6. How much do you believe cultural and societal norms influence electricity consumption patterns in your household and community?

7. Overall, how satisfied are you with your household's current efforts towards sustainable electricity consumption?

Responses assessed on a 5-point Likert scale:

1	2	3	4	5
Not at all	**Somewhat**	**Moderately**	**Mostly**	**Completely**