ENERGY CONSUMPTION PREDICTION

*A project report submitted to ICT Academy of Kerala*

*in partial fulfillment of the requirements*

*for the certification of*

CERTIFIED SPECIALIST

IN

DATA SCIENCE & ANALYTICS

submitted by

Team 7

Members:

**Jomit George K**

**Abisa M**

**Ananthakrishnan U S**

**Christy Mariam Sunny**

**Shahana Shirin P T**



ICT ACADEMY OF KERALA

THIRUVANANTHAPURAM, KERALA, INDIA

Apr 2023

List of Figures

Figure 4.3.1 Energy consumption over time 16

Figure 4.3.2 Energy generation over time 16

Figure 4.3.3 Energy consumption of appliances distribution 17

Figure 4.3.4 Energy consumption on different weather conditions 17

Figure 4.3.5 Energy generation on different weather conditions 18

Figure 4.4.1 Heatmap showing the correlation between the columns. 20

Figure 5.2(a) Energy consumption by hour of the day 21

Figure 5.2(b) Energy Consumption by Dayparting 22

Figure 5.2(c) Energy consumption by day of the week 22

Figure 5.2(d) Energy consumption of appliances over hour of the day 23

Figure 5.2(e) Energy consumption of appliances over weekdays 23

Figure 5.2(f) Energy consumption of appliances over month 24

Figure 5.2(g) Energy consumption by season 24

Figure 6.1.1 Representation of House overall and rolling mean using moving average 27

Figure 6.1. 2. Illustration of power generated and rolling mean using moving average 27

Figure 6.2. Illustration of ACF and PACF 28

Figure 6.2.1 Representation of actual and forecasted values using ARIMA single forecasting 28

Figure 6.2.2 Illustration of test data and predicted values using ARIMA rolling forecasting 29

Figure 6.3.1 Illustration of test data and predicted values using LSTM 30

Figure 6.4.1 Illustration of test data and predicted values using Light GBM 31

Figure 6.5.1 Illustration of test data and predicted values using XGBoost. 31

Figure 6.6.1 Illustration of test data and predicted values using SARIMAX. 32

List of Abbreviations

1. Iot-Internet of Things

2. MA-Moving Average

3. ARIMA-AutoRegressive Integrated Moving Average

4. SARIMAX-Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors

5. LSTM-Long Short Term Memory

6. XGboost-Extreme Gradient Boosting

7. LightGBM-light gradient-boosting machine

8. MAE-Mean Absolute Error

9. MSE-Mean squared error

10. RMSE-Root mean squared error

11. NRMSE- Normalized root mean squared error

12. MAPE-Mean Absolute Percentage Error

13. R^2- R squared

14. kW-Kilowatt

15. ML-Machine Learning

16. CIoT- Cognitive Internet of Things

17. DT- Decision Tree

18. RF- Random Forest

19. kNN- k Nearest Neighbour

20. EDA- Exploratory Data Analysis

21. ADF- Augmented Dickey-Fuller

22. ACF- AutoCorrelation Function

23. PACF- Partial AutoCorrelation Function

24. RNN- Recurrent Neural Network

Table of Contents

**ABSTRACT**

**LIST OF FIGURES**

**LIST OF ABBREVIATIONS**

**1. PROBLEM DEFINITION**

1.1 Overview

1.2 Problem Statement

**2. INTRODUCTION**

**3. LITERATURE SURVEY**

3.1 Energy consumption prediction using machine learning; a review by Amir Mosavi and Abdullah Bahmani, 2019

3.2 R. A. Rashid, L. Chin, M. A. Sarijari, R. Sudirman and T. Ide, "Machine Learning for Smart Energy Monitoring of Home Appliances Using IoT", 2019

3.3 Energy consumption prediction by using machine learning for smart building: Case study in Malaysia”, by Mel Keytingan M. Shapi, Nor Azuana Ramli, Lilik J. Awalin, 2021

3.4 A machine-learning ensemble model for predicting energy consumption in smart homes”, by Ishaani Priyadarshini, Sandipan Sahu, Raghvendra Kumar, David Taniar,2022

**4. EXPLORATORY DATA ANALYSIS**

4.1 Null Value Analysis

4.2 Outlier Detection and Handling

4.3 Understanding Features and Target

4.4 Removing unnecessary columns (Correlation Analysis)

**5. DATA PREPROCESSING**

5.1 Missing Value Handling

5.2 Handling Time Series data

5.3 Encoding

5.4 Making time as index column

**6. ANALYSIS OF MACHINE LEARNING MODELS**

6.1 Moving Average

6.2 ARIMA

6.2.1 ARIMA Single Forecast

6.2.2 ARIMA Rolling Forecast

6.3 LSTM

6.4 LightGBM

6.5 XGBoost

6.6 SARIMAX

**7. RESULTS**

**8. CONCLUSION**

**References**

Abstract

Activity recognition in the smart home has emerged as a significant research area in recent years due to the multitude of human-centric applications. The Internet of Things (IoT) has unified various home appliances, such as air conditioning, alarm systems, lighting, heating, ventilation, telephone systems, TVs, etc., under one platform, enabling monitoring and remote control. Energy management is one of the IoT use cases that can enhance our comfort and security while promoting low energy consumption, allowing us to monitor energy being sent out or consumed. This project focuses mainly on utilizing past data to predict future energy consumption so that we can efficiently manage our daily appliance usage at home.

Also, the energy generated and consumed is influenced by various weather attributes, such as temperature, precipitation, etc. During winter, for instance, people tend to use heaters more frequently, while the use of air conditioning decreases significantly, and the reverse happens during summer.

So, in this project we aim to predict the changes in energy consumption patterns based on the weather conditions. We are planning to do the time-series analysis by focusing on models such as Moving Average, ARIMA, SARIMAX, XGBoost, LSTM and LightGBM Regressor and thereby predicting the future power consumption. Performance of the models will be evaluated using Mean Absolute Error (MAE), Mean squared error (MSE) and Root mean squared error (RMSE).

To achieve this, readings from a smart meter with a time span of 1 minute over 350 days of house appliances in kW, as well as weather conditions of a particular region, are analyzed. The dataset used comprises around 503911 rows and 32 columns.

**CHAPTER 1**

# **PROBLEM DEFINITION**

**1.1 Overview**

The "Smart Home Dataset with Weather Information" contains data related to the energy consumption of a household in the US, along with weather information. The dataset contains a total of 503910 data points covering the time period from January 2016 to December 2016.

The aim of the dataset is to predict future energy consumption using historical energy consumption and weather data.The dataset includes various features such as date/time, energy consumption, weather information,energy generation and other home-related information.

**1.2 Problem Statement**

To analyze a smart home dataset containing information about energy consumption and weather conditions, and to develop models that can predict changes in energy consumption patterns based on weather conditions.

**CHAPTER 2**

**INTRODUCTION**

The Internet of Things (IoT) has transformed the way we interact with technology and our daily lives. The integration of sensors, devices, and data analytics has made it possible to collect, process, and analyze large amounts of data in real-time. One of the most significant applications of IoT technology is in the field of smart homes. Smart homes are equipped with sensors and devices that can monitor and control various aspects of the home environment, including temperature, lighting, security, and energy usage. These devices generate a vast amount of data that can be used to reduce energy usage, improve efficiency, and reduce costs.

The dataset is a valuable resource for understanding the relationship between energy consumption and weather patterns, and for developing models to predict energy consumption. In this project, we aim to analyze this dataset and explore the factors affecting energy consumption in smart homes. Our analysis will involve data visualization, statistical analysis, and machine learning techniques. We will also develop models to predict future energy consumption.

This project can be used to reduce energy consumption in smart homes based on weather conditions. By predicting changes in energy consumption patterns, homeowners can adjust their energy usage accordingly, resulting in cost savings and reduced energy consumption. Overall, this project combines data exploration, preprocessing, time-series analysis, and machine learning techniques to analyze a smart home dataset and develop models that can predict changes in energy consumption patterns based on weather conditions. The insights and models developed in this project have practical applications for optimizing energy consumption in smart homes, and can be extended to other domains such as industrial automation and energy management.

The dataset contains a total of 503910 data points and 32 features, namely: time, use, gen, House overall, Dishwasher, Home office, Fridge, Wine cellar, Garage door, Barn, Well, Microwave, Living room, Solar, temperature, icon, humidity, visibility, summary, apparentTemperature, pressure, windSpeed, cloudCover, windBearing, precipIntensity, dewPoint, precipProbability, Furnace and Kitchen.

**CHAPTER 3**

**LITERATURE SURVEY**

# Smart building concept has been adapted more frequently as an initiative to create an intelligent space area by taking advantage of the rapid development of computational and communication architecture. General public understanding of the smart building concept rotates on the idea of automated processes. Other than automated control, a smart building also consists of an intelligent system which provides energy consumption forecasts as an energy efficiency initiative. Smart energy consumption forecasting is important, especially for buildings as buildings’ energy usage is increasing and almost reaches 40% of primary energy use in developed countries. Machine learning (ML) methods have recently contributed very well in the advancement of the prediction models used for energy consumption. Such models highly improve the accuracy, robustness, and precision and the generalization ability of the conventional time series forecasting tools.

## **3.1 Energy consumption prediction using machine learning; a review by Amir Mosavi and Abdullah Bahmani, 2019;**

This article reviews the state of the art of machine learning models used in the general application of energy consumption. A comprehensive review of the literature identifies the major ML methods, their application and a discussion on the evaluation of their effectiveness in energy consumption prediction. This paper further makes a conclusion on the trend and the effectiveness of the ML models. As a result, this research reports an outstanding rise in the accuracy and an ever increasing performance of the prediction technologies using the novel hybrid and ensemble prediction models.

## **3.2 R. A. Rashid, L. Chin, M. A. Sarijari, R. Sudirman and T. Ide, "Machine Learning for Smart Energy Monitoring of Home Appliances Using IoT", 2019;**

## This project proposes a smart energy monitoring system for home appliances incorporating Cognitive IoT (CIoT) which consists of three parts. A Raspberry Pi-based smart plug serving as the gateway, Google Colab as the training server and a dashboard using Matplotlib library where users may monitor the real-time energy consumption. The Tensorflow-based Long Short-term Memory (LSTM) model will forecast electricity bills and notify users if abnormal energy consumption of individual home appliances is detected. The completed LSTM model demonstrates a high accuracy of more than 80% with a low mean squared error and a high level of goodness of fit.

# 

## **3.3 “Energy consumption prediction by using machine learning for smart building: Case study in Malaysia”, by Mel Keytingan M. Shapi, Nor Azuana Ramli, Lilik J. Awalin, 2021;**

# This research aims to develop a predictive model for energy consumption in Microsoft Azure cloud-based machine learning platform, by focusing on real-life application in Malaysia. Three methodologies which are [Support Vector Machine](https://www.sciencedirect.com/topics/engineering/support-vector-machine), [Artificial Neural Network](https://www.sciencedirect.com/topics/engineering/artificial-neural-network), and k-Nearest Neighbour are proposed for the algorithm of the predictive model. The performance of each of the methods is compared based on [RMSE](https://www.sciencedirect.com/topics/engineering/root-mean-square-error), NRMSE, and MAPE metrics.

## **3.4 “A machine-learning ensemble model for predicting energy consumption in smart homes”, by Ishaani Priyadarshini, Sandipan Sahu, Raghvendra Kumar, David Taniar,2022;**

This paper proposes an overall analysis of energy consumption in smart homes by deploying [machine learning](https://www.sciencedirect.com/topics/computer-science/machine-learning) models. They relied on [machine learning techniques](https://www.sciencedirect.com/topics/engineering/machine-learning-technique), like [Decision Trees](https://www.sciencedirect.com/topics/computer-science/decision-trees) (DT), [Random Forest](https://www.sciencedirect.com/topics/computer-science/random-decision-forest) (RF), eXtreme Gradient Boosting (XGBoost), and k-Nearest Neighbor (KNN) for predicting the [power consumption](https://www.sciencedirect.com/topics/engineering/electric-power-utilization) of multiple datasets. The authors also proposed a DT-RF-XGBoost-based Ensemble Model for analyzing the consumption and comparing it with the baseline algorithms. The evaluation parameters used in the study are [Mean Square Error](https://www.sciencedirect.com/topics/engineering/mean-square-error) (MSE), R-squared (R2,), [Root Mean Square Error](https://www.sciencedirect.com/topics/engineering/root-mean-square-error) (RMSE), and [Mean Absolute Error](https://www.sciencedirect.com/topics/engineering/mean-absolute-error) (MAE), respectively.

From the literature survey, it is clear that machine learning (ML) techniques have great potential in energy consumption prediction. Different ML techniques, including ARIMA (Auto-Regressive Integrated Moving Average),SARIMA (Seasonal ARIMA), Artificial neural networks, and LSTM (Long and Short-Term Memory network) have been used to forecast energy consumption patterns. The choice of ML technique depends on the nature of the data and the specific problem at hand. Future research can focus on exploring new ML techniques or hybrid approaches to improve the accuracy and efficiency of energy consumption forecasting.

**CHAPTER 4**

**EXPLORATORY DATA ANALYSIS**

Exploratory Data Analysis is very important in machine learning. It helps you gather insights and make better sense of the data, and removes irregularities and unnecessary values from data. It also helps you prepare your dataset for analysis and allows a machine learning model to predict better by giving you more accurate results.

The Dataset has around 503910 observations in it with 32 columns. It has both categorical and numerical columns. The categorical columns are time, icon, summary and cloud cover and the numerical columns are use, gen, House overall, Dishwasher,Home office, Fridge, Wine cellar, Garage door, Barn, Well, Microwave, Living room, Solar, temperature, humidity, visibility, apparentTemperature, pressure, windSpeed, windBearing, precipIntensity, dewPoint, precipProbability, Furnace and Kitchen.

The types of Graphs used for visualization of the dataset are Bar chart, Line graph, Scatter Plot ,Box plot & Heat map. Matplotlib, Seaborn and Holoviews are the three Python libraries that have been used here to generate the graph for visualization.

**4.1 Null Value Analysis**

At the beginning of this EDA, null values were investigated and we have found that there are null values in every column except in time, kitchen and furnace. Also, the last row of the dataset contained no values and hence they were replaced.

**4.2 Outlier Detection**

The boxplots for the majority of the features were drawn and found that every column has outliers. But, we neglected it since the outliers were insignificant.

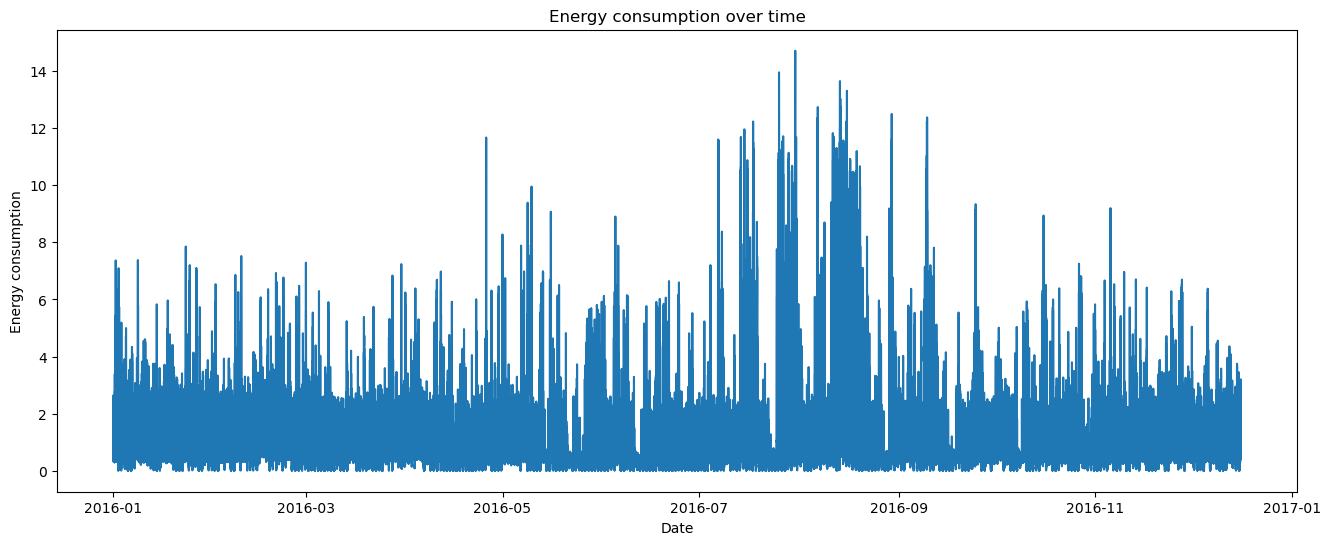
**4.3 Understanding Features and Target**

The features column in our data can be divided as weather data and energy data. We have decided to take overall\_usage as the target variable.Also, we removed unit(kW) from dataset column names.There are multiple columns in Furnace and Kitchen. To make it simple, we aggregated them into new columns by summing them up.

Invalid values in the cloudCover column because it must be float64-type but it is object-type. Maybe this is a data collection mistake.So we replaced these invalid values with the next valid value.

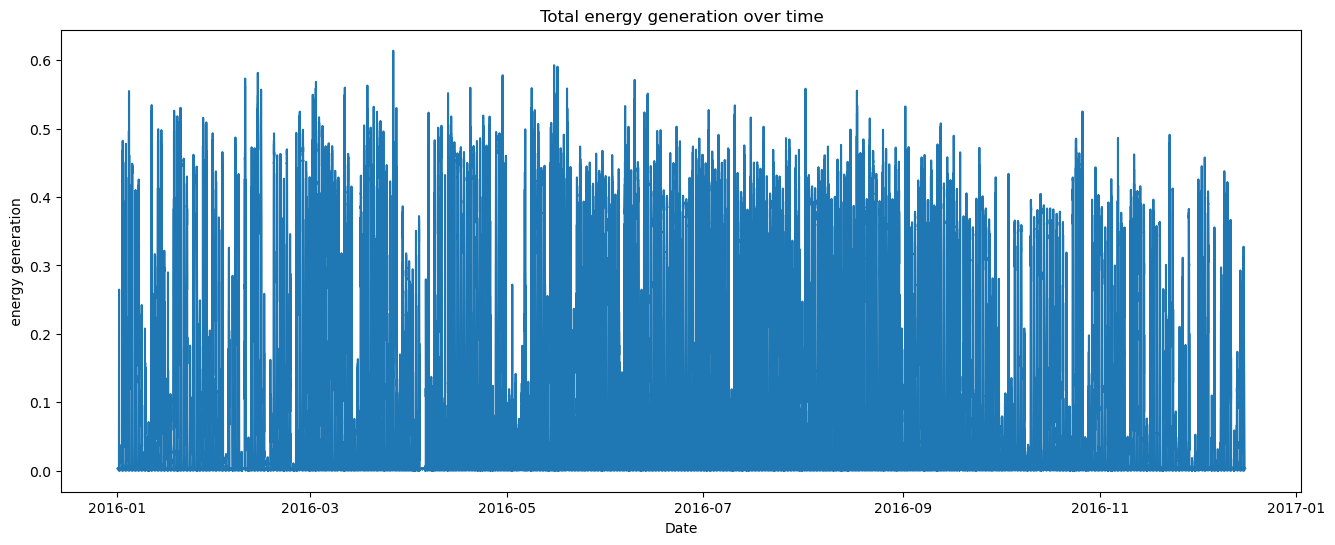
After converting the unix timestamp, we realized the time-step is in increments of seconds though the data was collected with a time-span of 1 minute. To utilize datetime information such as year, month and day in EDA and modeling phase, we extracted them from the time column. Also we separated the Hour variable into night, morning, afternoon and evening based on its number.

**4.3.1 Energy consumption over time**



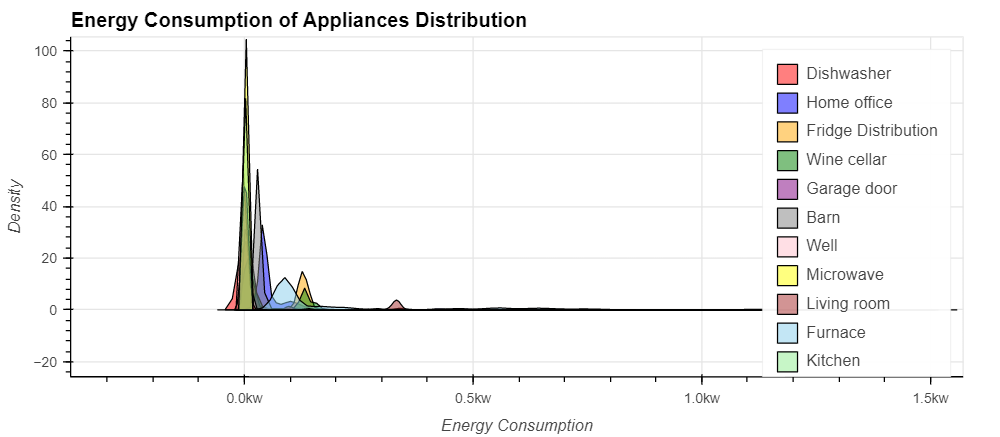
From this we can see that energy consumption has increased from July to September. In the rest of the months we cannot see any pattern.

**4.3.2 Energy generation over time**



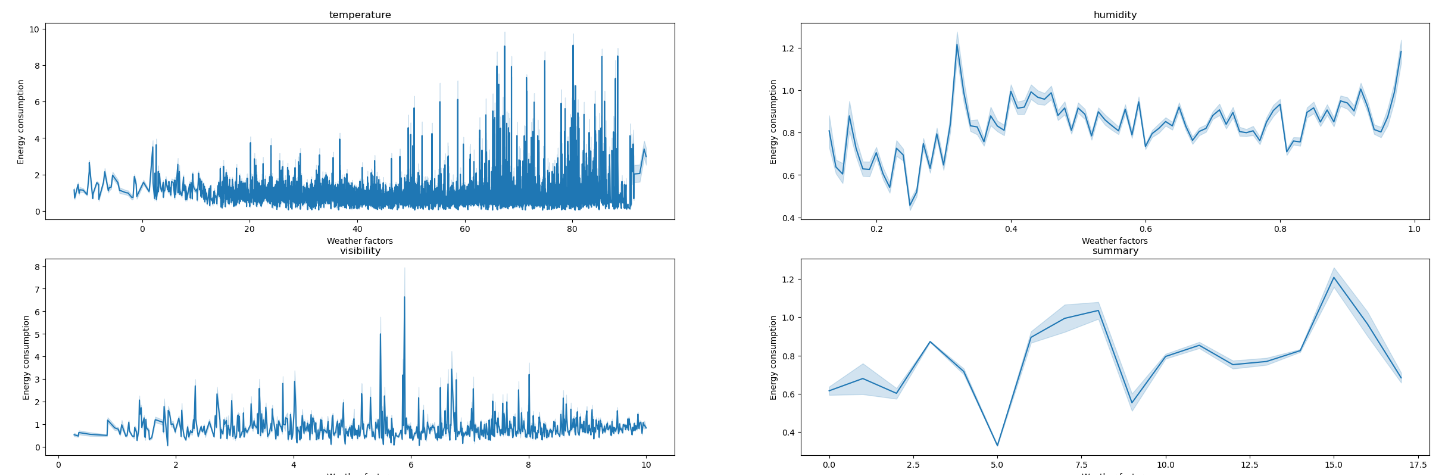
No visible pattern can be seen in the energy generated over time.

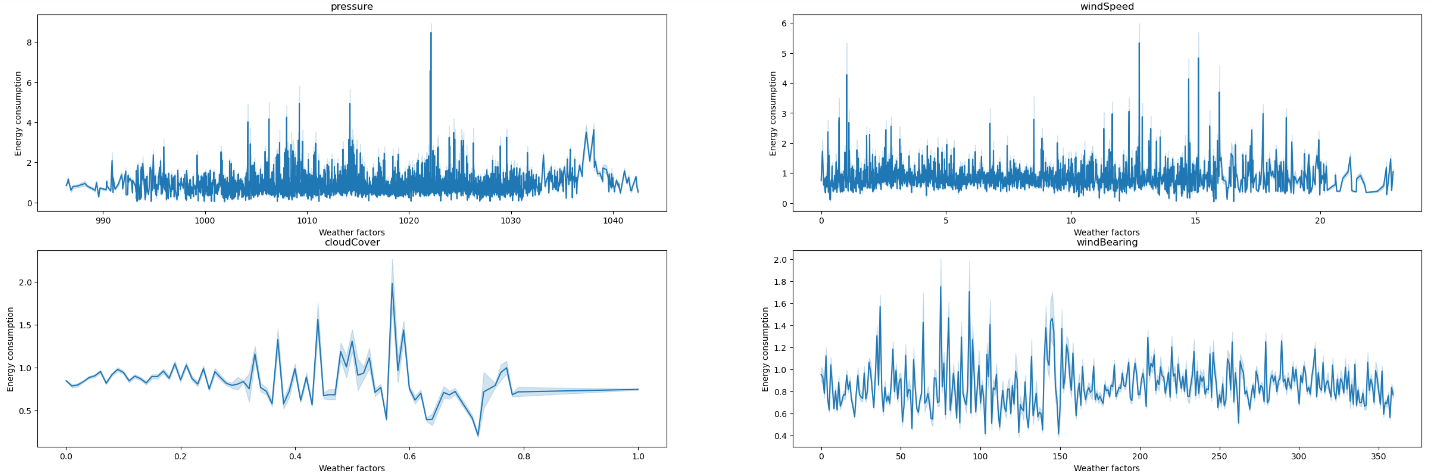
**4.3.3 Energy consumption of appliances distribution**

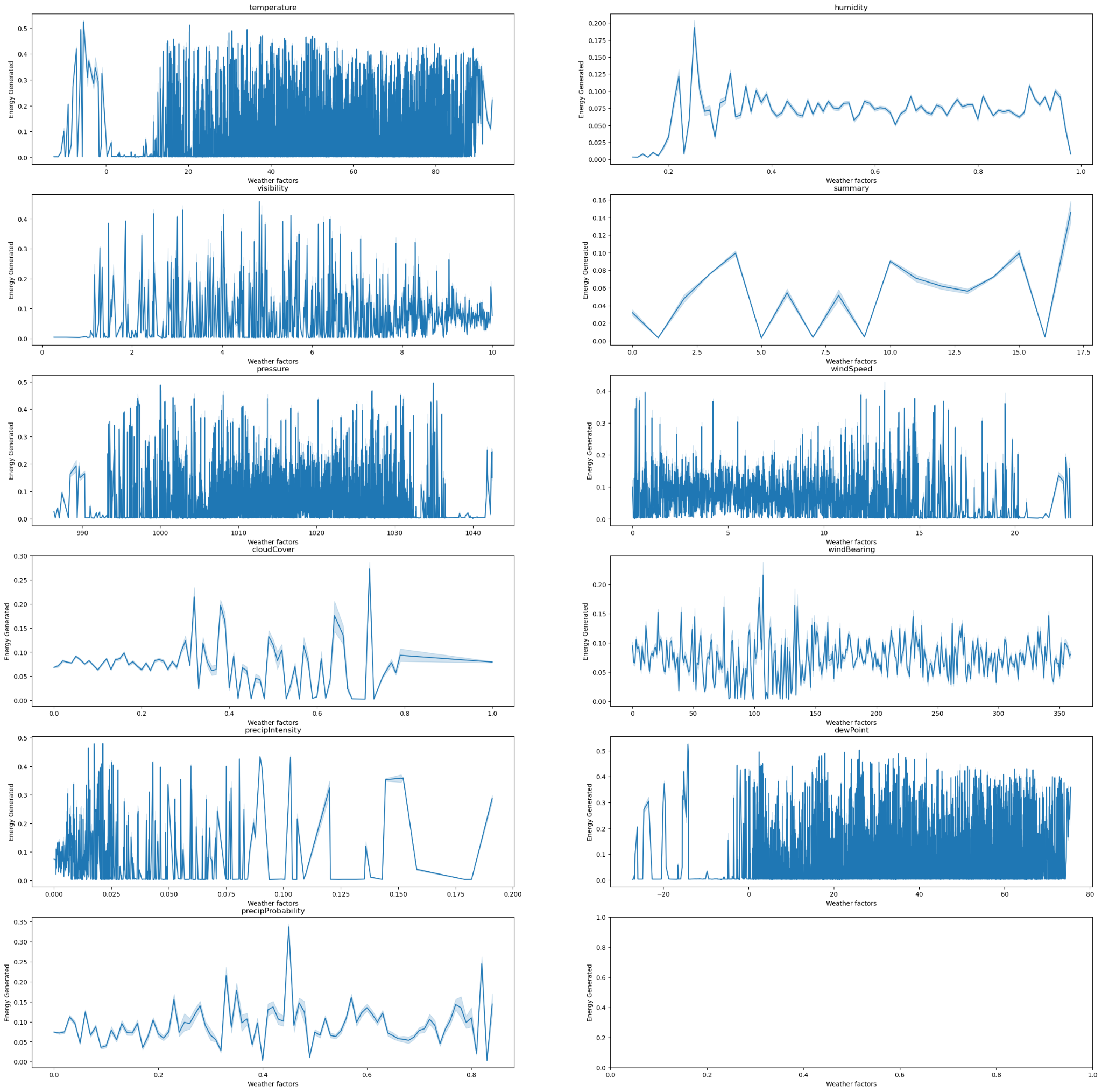


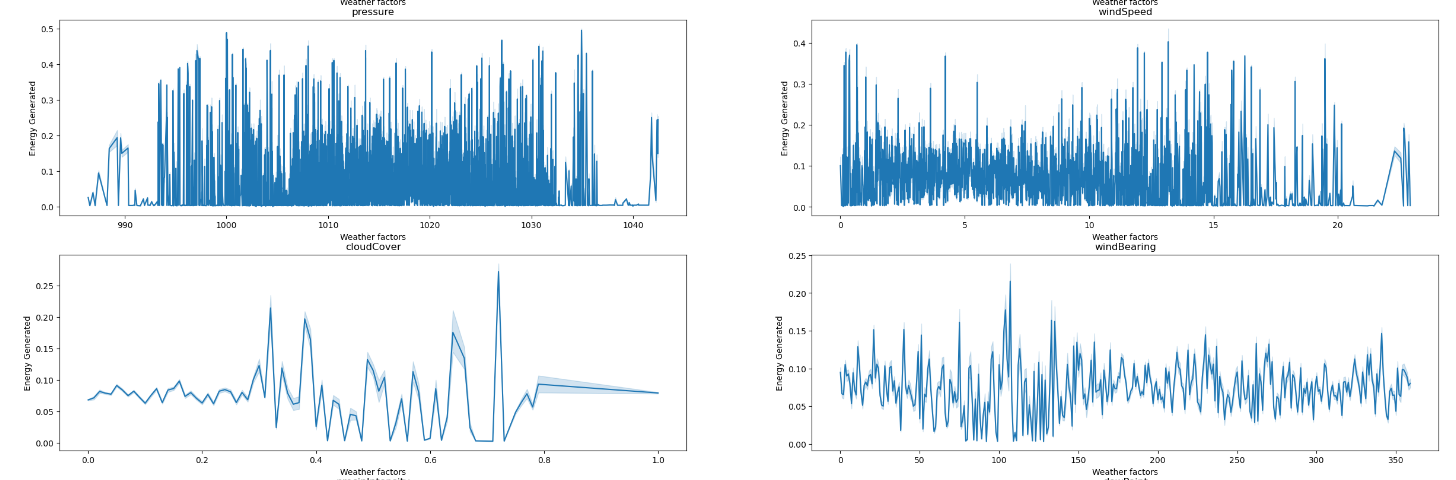
From this graph, we can see that microwave consumes the most energy

**4.3.4 Energy consumption on different weather conditions**





**4.3.5 Energy generation on different weather conditions**



**4.4 Removing unnecessary columns ( Correlation Analysis)**

We checked the correlation between different columns . We got the inference that some of the columns were highly correlated. From the fields with correlation greater than 0.99 one of the fields was dropped from the dataframe. There are two furnace values and 3 kitchen columns . They were summed up as a single furnace and single kitchen and rest were dropped.

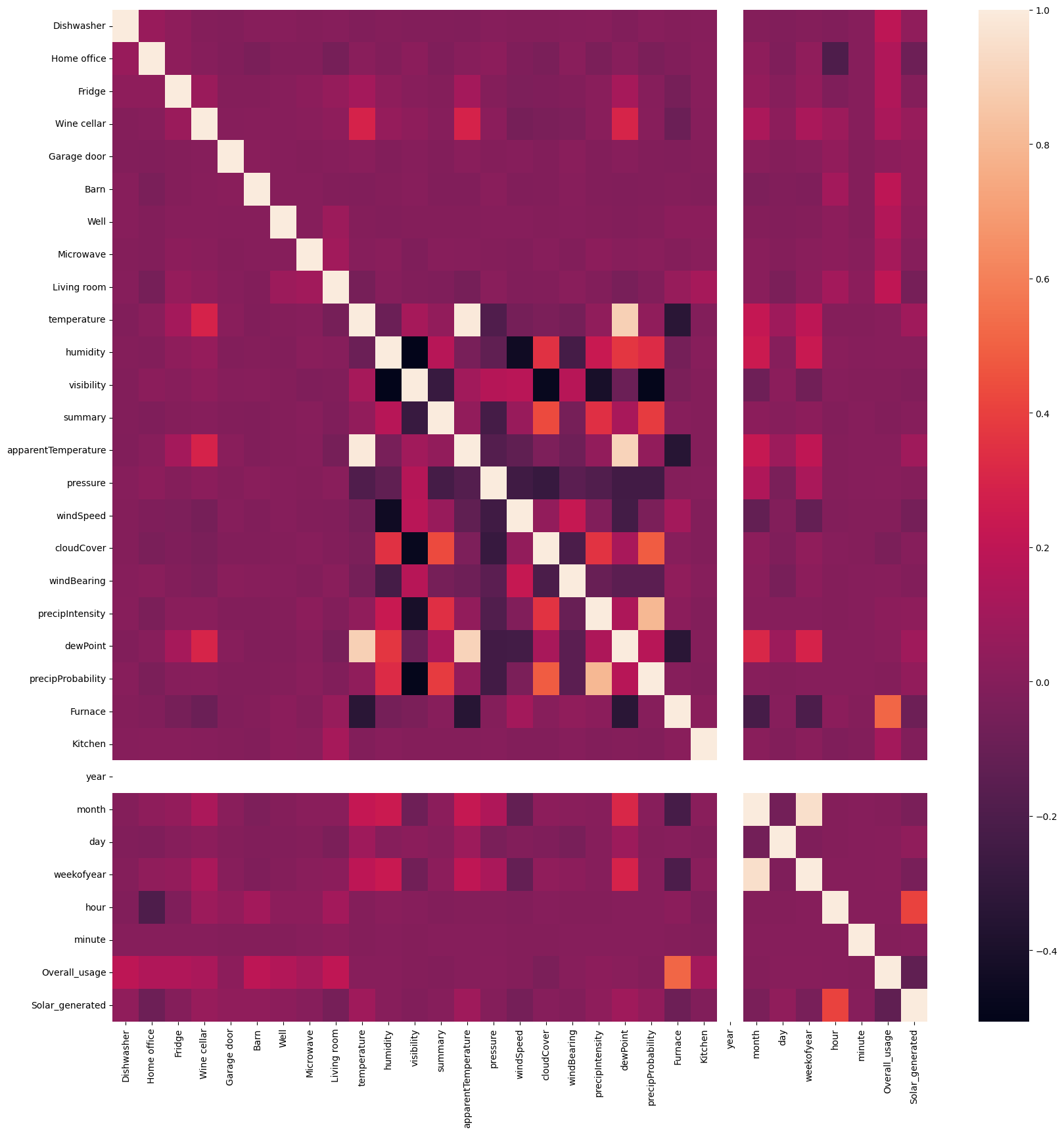


Figure 4.4.1 Heatmap showing the correlation between the columns.

**CHAPTER 5**

**PREPROCESSING**

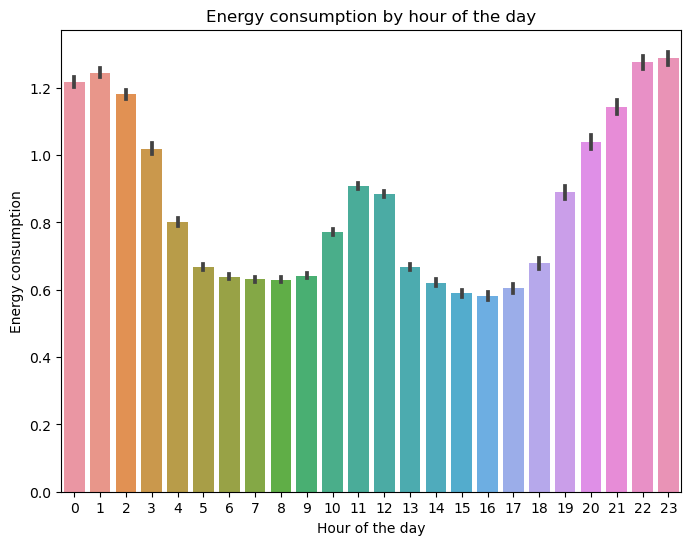
**5.1 Handling missing values**

Certain values in the cloud cover column were of string type , so they were converted to null values and filled with median.

**5.2 Handling Time Series data**

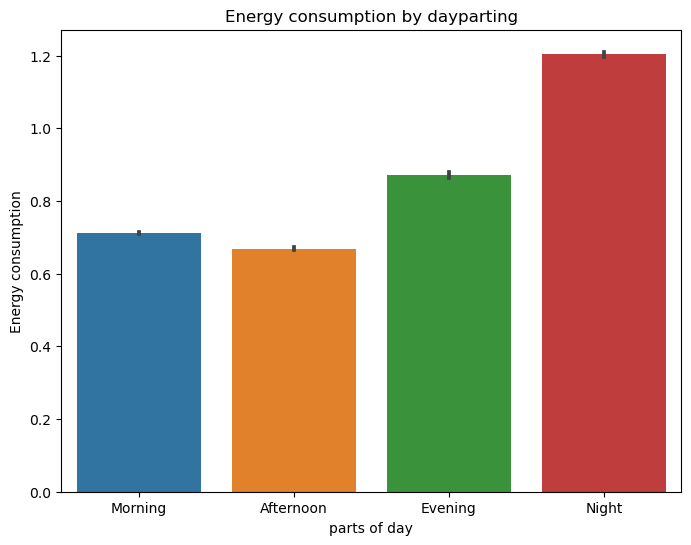
We converted time series data to a different frequency, i.e, from seconds to minutes. The time column was then split into hourly, daily, weekly and monthly data. The hours of day is again splitted into morning, afternoon, evening and night, also the months were grouped as winter, spring, summer and autumn seasons We plotted graphs showing the relationship between energy consumption and different time frequencies.

1. **Energy consumption by hour of the day**



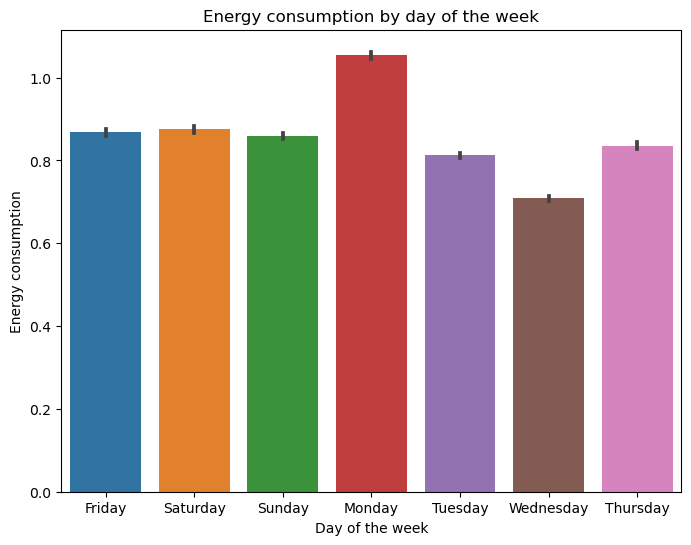
From this graph, we can see that energy consumption peaks during early morning, noon and late night hours. The off peak hours are during morning and afternoon/evening.

# **Energy Consumption by Dayparting**



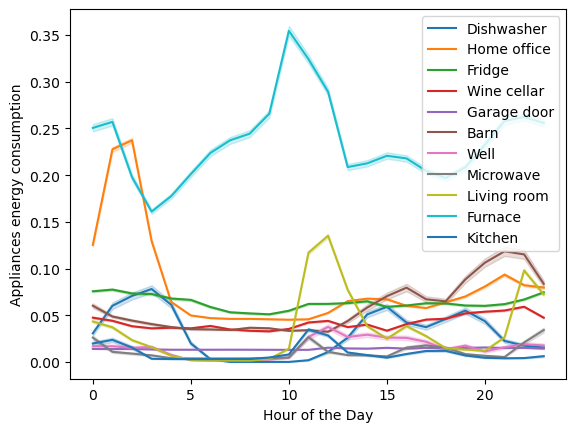
From this graph we can see that energy consumption increases from morning to night. And it's highest at night.

**c) Energy consumption by day of the week**



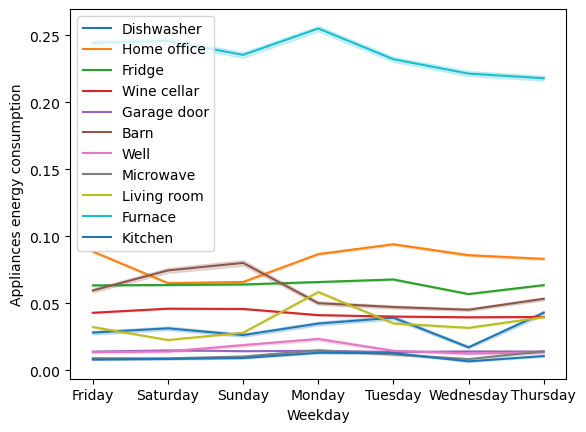
In general,it can be seen that energy consumption is highest on Mondays and lowest on Wednesdays.

**d) Energy consumption of appliances over hour of the day**



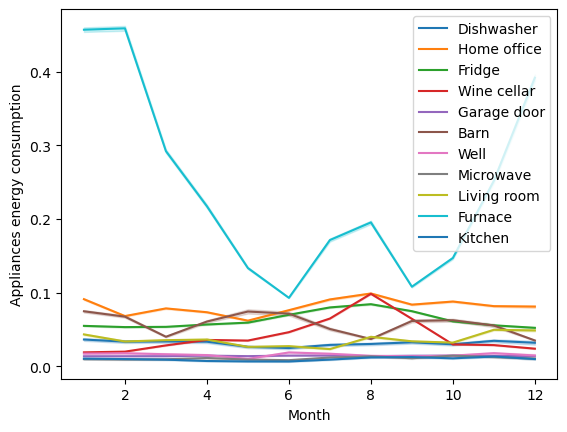
From this graph, it can be seen that furnaces consume the most energy and also overall appliances show peak consumption during morning hours.

**e) Energy consumption of appliances over weekdays**

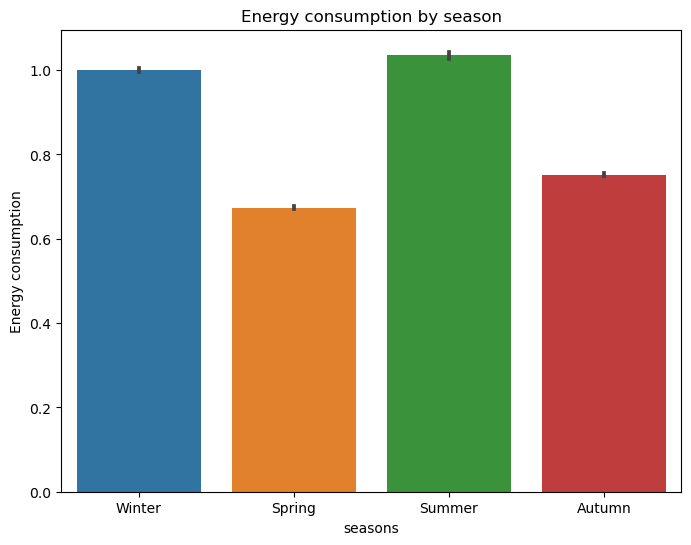


From this graph it can be seen that out of all appliances, the furnace consumes the most energy and its highest on mondays.

**f) Energy consumption of appliances over month**

From this graph it can be seen that, energy usage of furnaces is highest in the months of December, January, February, July, and August. It might be because of the drop in temperature.

**g) Energy consumption by season**



From this graph we can say that the consumption is higher in winter and summer than spring and autumn.

**5.3 Encoding**

We did Label encoding to the summary column to make it numerical.

**5.4 Making time as index column**

In order to facilitate working on different time series analysis models we have made time column as index column.

**CHAPTER 6**

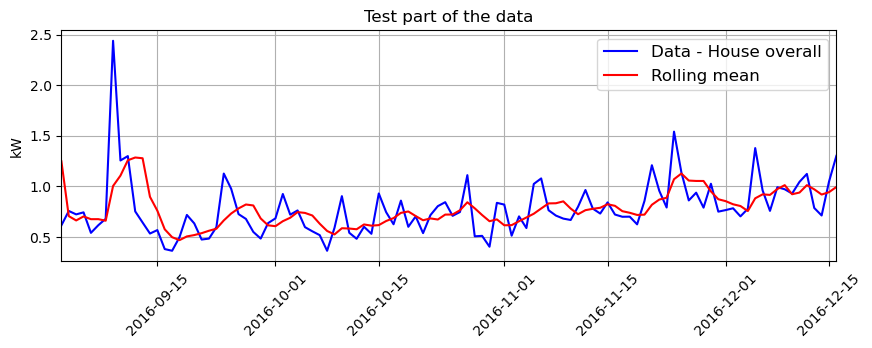
**ANALYSIS OF MACHINE LEARNING MODELS**

We focused on the analysis of machine learning models for predicting anomalies and future energy consumption in a smart home dataset. We explored various machine learning techniques including Moving Average,ARIMA, SARIMA, LSTM and LightGBM Regressor.

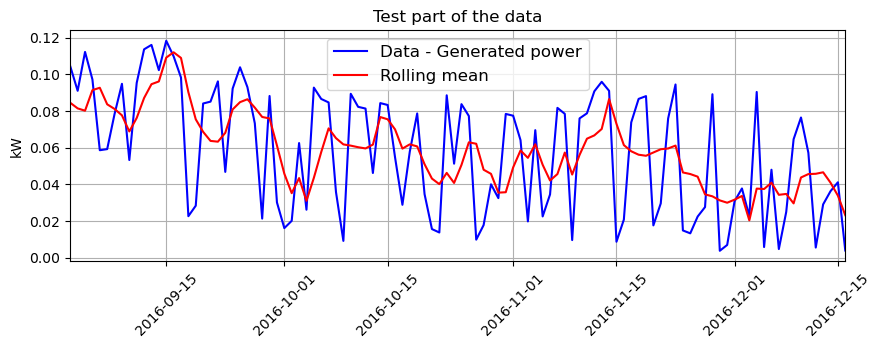
The main objective was to detect anomalies in power consumption of appliances and predict their future power consumption based on weather conditions. We evaluated the performance of these models using MSE,RMSE,MAE and R^2 Square as the evaluation metrics.

**6.1**  **Moving Average**

Moving Average is a simple model that calculates the average of a specified window of historical data to make predictions. It provides a baseline for comparison with other models. Here, we separately resampled the energy consumption and solar generation columns to calculate the average for each day. We calculated three different moving averages with rolling averages over a window of 5 days, 20 days, and 30 days. The first graph is representing the relationship between the overall house energy consumption and its corresponding rolling mean, obtained using the moving average approach. Analyzing the plot reveals performance metrics, including a MSE of 0.057,RMSE of 0.240, MAE of 0.157, and an R-squared score of 0.250. Moving on to the second graph, it illustrates the total power generated and its associated rolling mean using the moving average technique. Analyzing this plot yields performance metrics of MSE: 0.001, RMSE: 0.027, MAE: 0.023, and an R-squared score of 0.338.



**Graph 6.1.1**.Representation of House overall and rolling mean using moving average

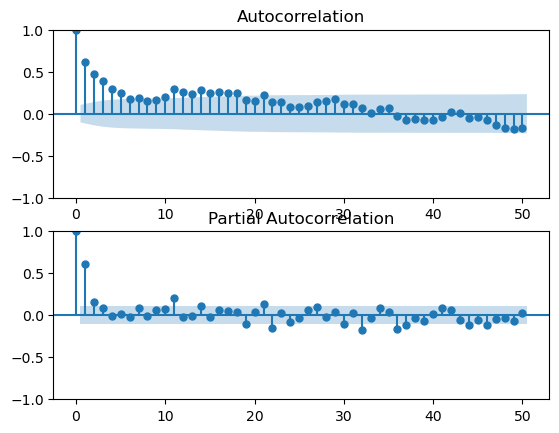


**Graph 6.1. 2**. Illustration of power generated and rolling mean using moving average

**6.2 ARIMA**

ARIMA (Autoregressive Integrated Moving Average) is a widely used time series forecasting model. It combines autoregressive (AR) and moving average (MA) components with differencing (I) to handle non-stationary data. ARIMA models are effective for capturing trends, seasonality, and long-term dependencies in time series data.

First we did the Augmented Dickey-Fuller test on the target column and printed the ADF statistic, p-value, and critical values. The p value was 0.036030 which is less than the critical value of 0.05 so the data is stationary. Then we plotted the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the target column.

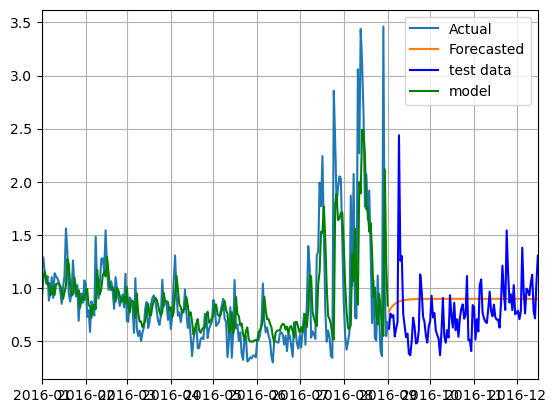
****

**Graph 6.2**. Illustration of ACF and PACF

**6.2.1 ARIMA Single forecast**

ARIMA Single Forecasting is a technique used to predict the value of a future data point in a time series using the ARIMA model. It involves fitting an ARIMA model to the historical data and then using this model to generate a forecast for a single future observation.

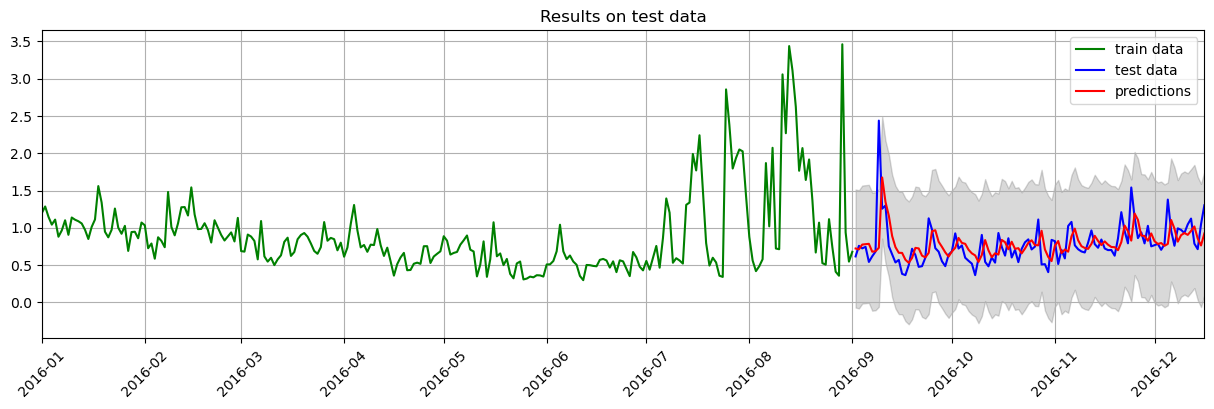
The time series data was then split into a training set and a testing set. The training set contains 70% of the data, and the testing set contains the remaining 30%. Analyzing the plot reveals performance metrics, including the MSE of 0.089, RMSE of 0.298, MAE of 0.231 and R^2 score: -0.160



**Graph 6.2.1** Representation of actual and forecasted values using ARIMA single forecasting

**6.2.2 ARIMA Rolling forecast**

ARIMA rolling forecast is a technique used to make sequential predictions on a time series by iteratively updating the model with new observations. It involves splitting the time series data into a training set and a test set. The model is initially trained on the training set, and then it predicts the next value in the test set. This predicted value is then added to the history, and the process is repeated for the next time step. By updating the model with each new observation, the rolling forecast takes into account the evolving nature of the time series and provides updated predictions at each step. The calculated values for these metrics are as follows: MSE: 0.06930, RMSE: 0.263, MAE: 0.175, MAPE: 0.309 and R^2 score: 0.095.The graph 6.2.2 is representing the test data and predicted values using ARIMA rolling forecasting.



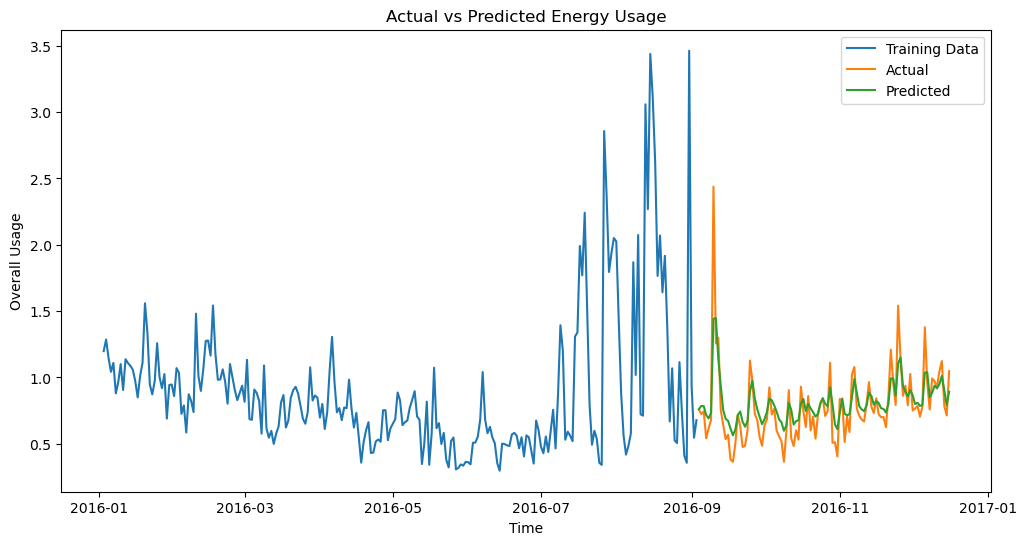
**Graph 6.2.2** Illustration of test data and predicted values using ARIMA rolling forecasting

**6.3** **LSTM**

LSTM (Long Short-Term Memory) is a type of recurrent neural network (RNN) architecture that is widely used for sequential data processing, such as time series data, text data, and speech recognition. LSTM networks are designed to address the limitations of traditional RNNs in capturing and retaining long-term dependencies in sequences.

The time series data was then split into a training set and a testing set. The training set contains 70% of the data, and the testing set contains the remaining 30%. Analyzing the plot reveals performance metrics, including the MSE of 0.118, RMSE of 0.343, MAE of 0.246 and R^2 score: 0.260.

The graph 6.3.1 is representing the test data and predicted values using LSTM.

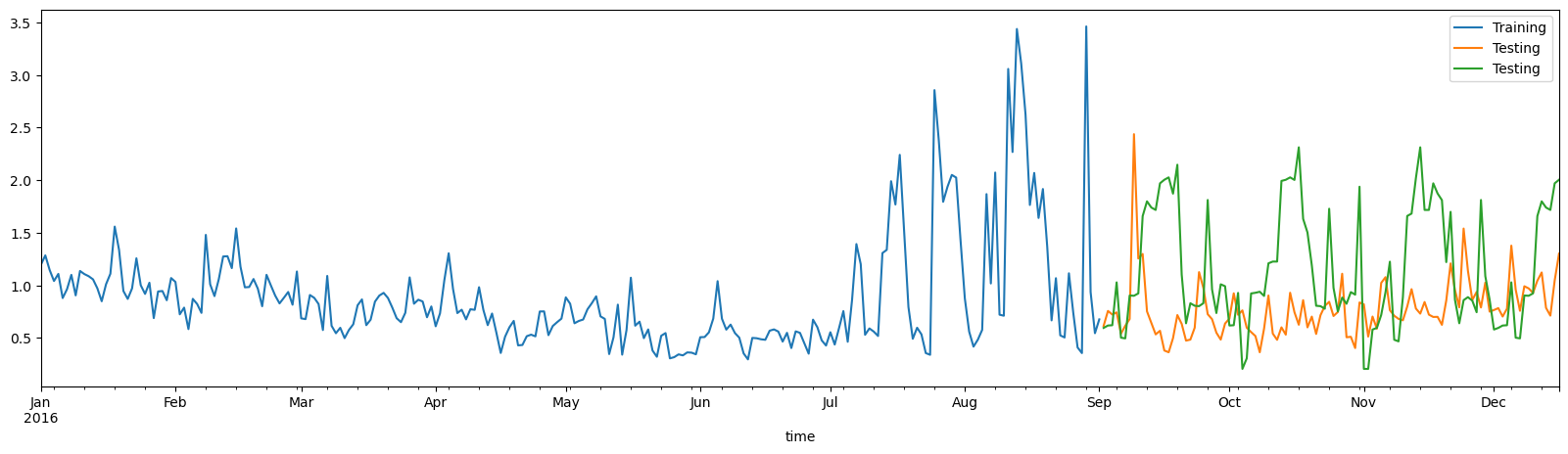
**Graph 6.3.1** Illustration of test data and predicted values using LSTM

**6.4** **Light GBM**

LightGBM (Light Gradient Boosting Machine) is a gradient boosting framework that is designed to be efficient and highly performant. It is widely used for both classification and regression tasks. In the context of regression, LightGBM can be utilized as a LightGBM Regressor model.

The time series data was then split into a training set and a testing set. The training set contains 70% of the data, and the testing set contains the remaining 30%. Analyzing the plot reveals performance metrics, including the MSE of 0.524, RMSE of 0.724, MAE of 0.554 and R^2 score: -5.844

The graph 6.4.1 is representing the test data and predicted values using Light GBM.



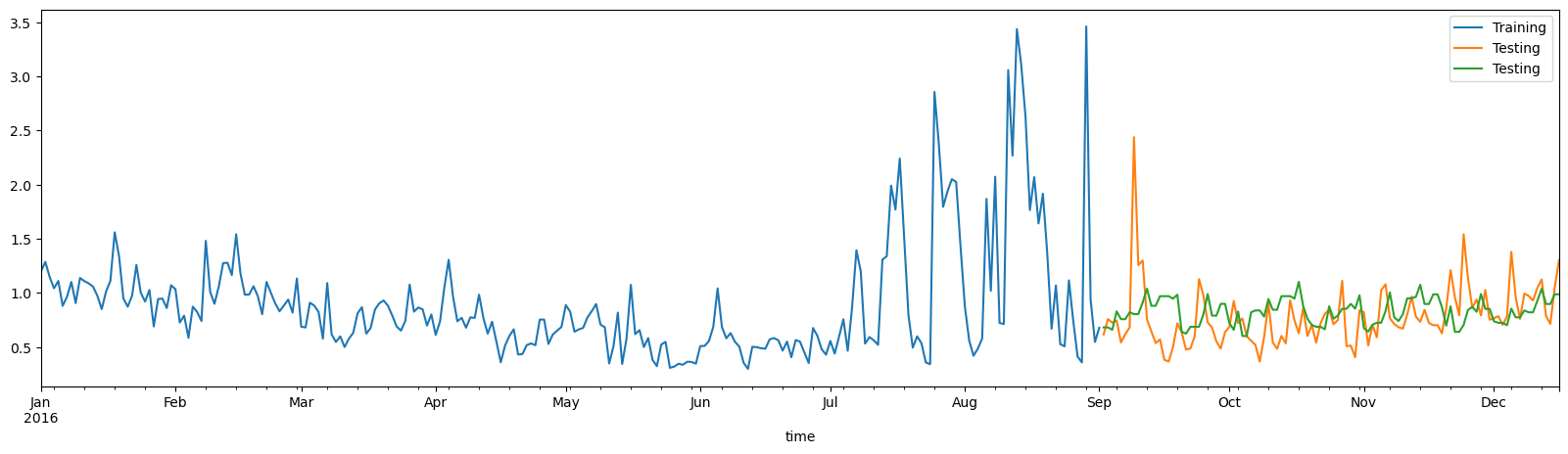
**Graph 6.4.1** Illustration of test data and predicted values using Light GBM

**6.5** **XGBoost**

XGBoost (eXtreme Gradient Boosting) is another popular gradient boosting framework widely used for both classification and regression tasks. It is known for its scalability, speed, and state-of-the-art performance.

The time series data was then split into a training set and a testing set. The training set contains 70% of the data, and the testing set contains the remaining 30%. Analyzing the plot reveals performance metrics, including the MSE of 0.091, RMSE of 0.301, MAE of 0.218 and R^2 score: -0.184

The graph 6.5.1 is representing the test data and predicted values using XGBoost.

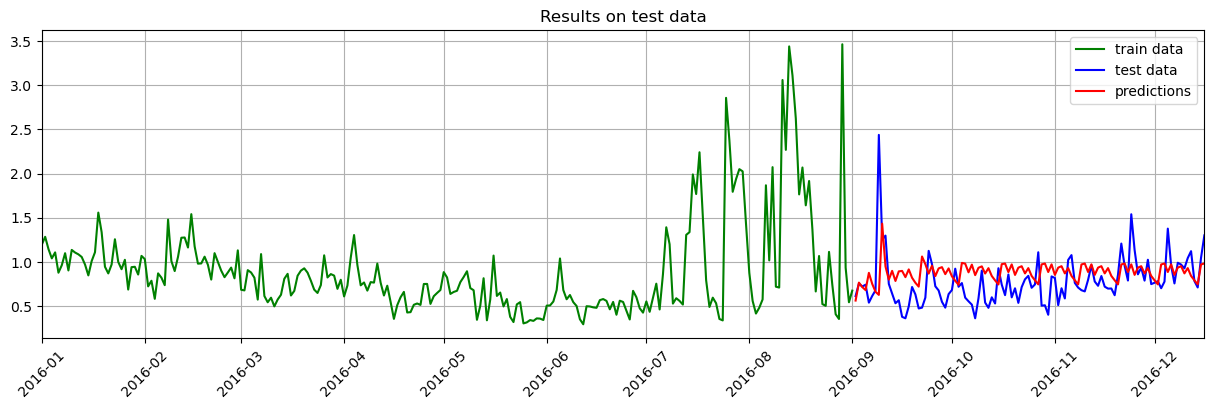


**Graph 6.5.1** Illustration of test data and predicted values using XGBoost.

**6.6** **SARIMAX**

SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors) is a popular statistical model used for time series forecasting. It is an extension of the ARIMA model that takes into account seasonality and external regressors.

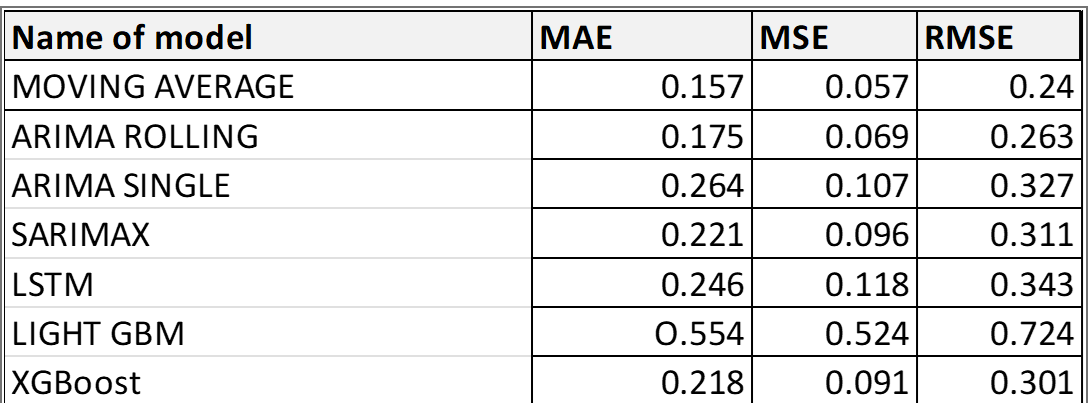
The time series data was then split into a training set and a testing set. The training set contains 70% of the data, and the testing set contains the remaining 30%. Analyzing the plot reveals performance metrics, including the MSE of 0.096, RMSE of 0.311, MAE of 0.221.The graph 6.6.1 is representing the test data and predicted values using SARIMAX.



**Graph 6.6.1** Illustration of test data and predicted values using SARIMAX.

**CHAPTER 7**

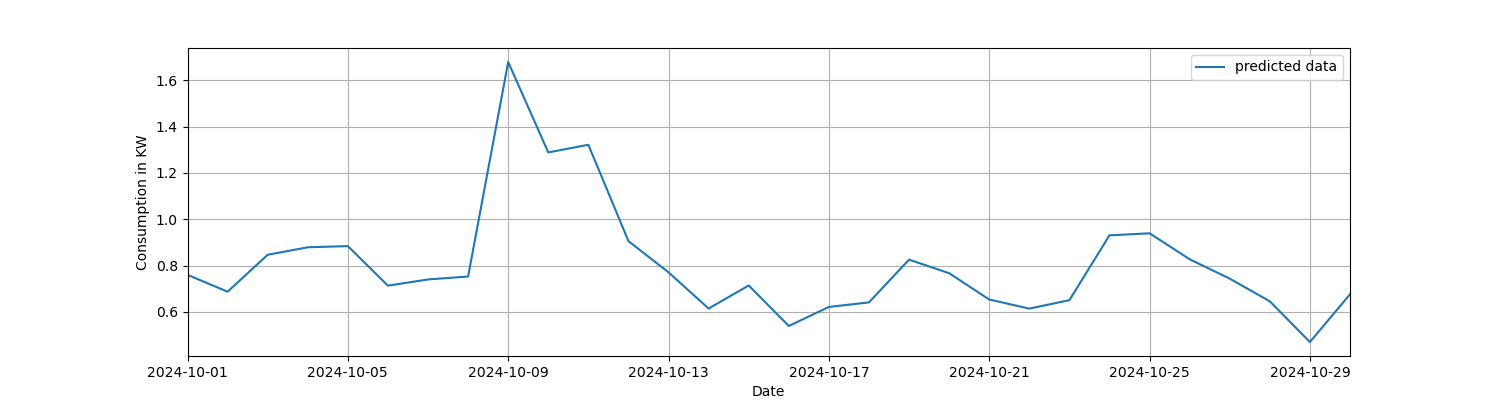
**RESULTS**

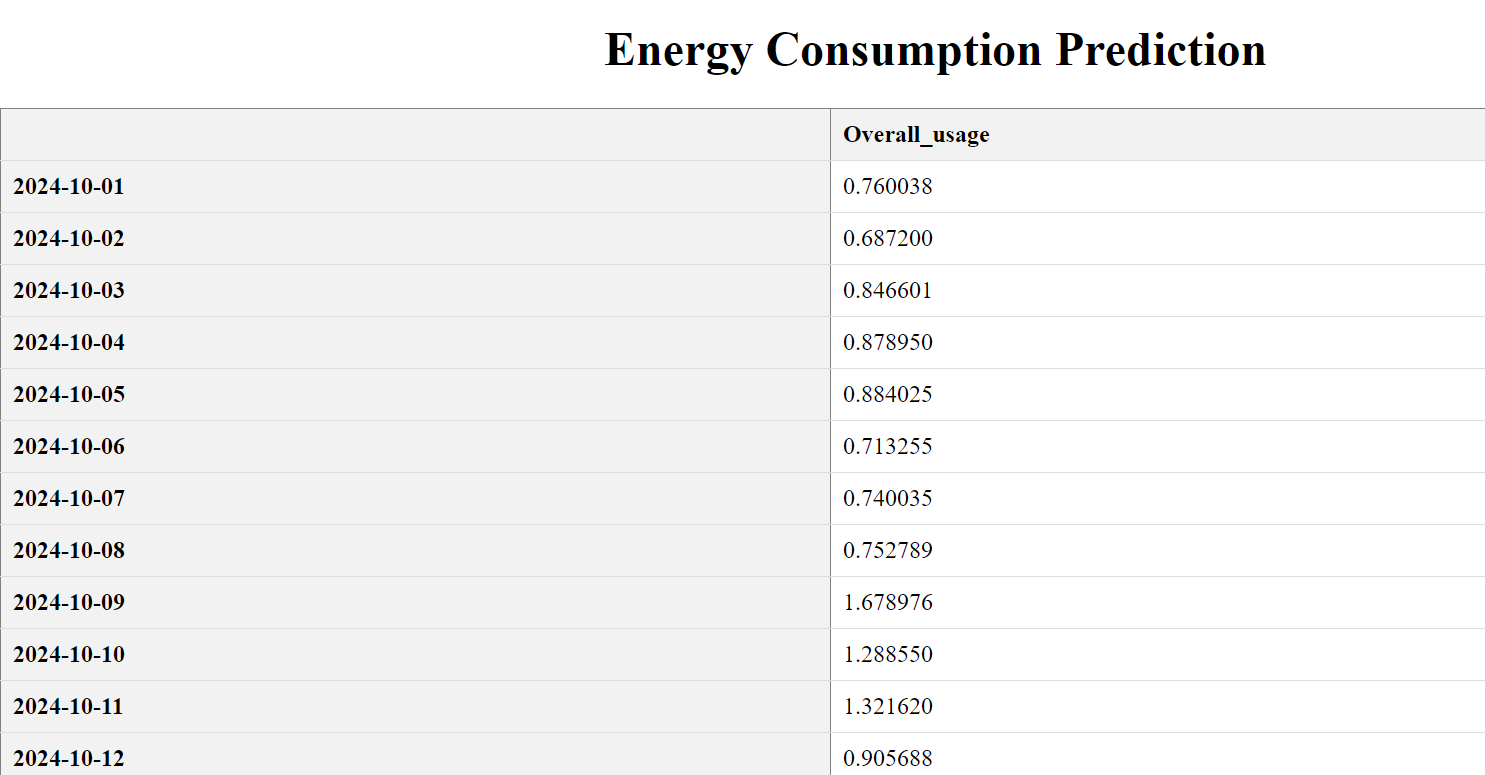


From the table above it is clear that the ARIMA rolling model outperforms other models . so we have decided to use the ARIMA model for predicting future energy consumption values.

We have created a date series starting from 2024-10-01 we are going to predict the energy usage for the next 30 days

The result we get from the model is as follows:





**CHAPTER 8**

**CONCLUSION**

In conclusion, we evaluated the performance of seven machine learning models, including Moving average, ARIMA rolling, ARIMA single, SARIMAX, LSTM, LightGBM and XGBoost. The models were trained on a given dataset, and their performance was evaluated using MAE, MSE and RMSE. Our results indicate that the ARIMA rolling model performed the best and LightGBM model performed the least on the given dataset.

We did univariate and multivariate time series algorithms on the given dataset. As per our analysis, univariate models were better than multivariate models. Firstly, we tried the Moving average model with window size as 5,20 and 30 on a daily sampled data , since it is better at smoothing out fluctuations. On doing the ADF test, we found the data to be stationary and decided to use the ARIMA algorithm. As our data showed additive seasonality, the SARIMAX algorithm was also tried. The LSTM model underperformed than expected due to insufficient data. We also performed hyperparameter tuning on the multivariate models , XGBoost and LightGBM models using GridSearchCV.

On analyzing ACF and PACF, and trying different p, d, and q, we fixed them as 2,0,5 respectively.ARIMA rolling algorithm was trained with these parameters and the resultant error metrics were MAE=O.175, MSE=0.069 and RMSE=0.263.

Summarizing, as per our analysis ARIMA rolling outperformed the rest of the models with this data, but we cannot purely say that it's the best, since data insufficiency concerned the multivariate models to a certain extent. Still we can say that using the historical weather data and time information, we can make good predictions of the future energy consumption. The findings and conclusions presented here have contributed to the existing body of knowledge in this field and can serve as a foundation for future research and practical applications.

REFERENCES

1.Taranveer Singh Anttal(2019) Smart Home Dataset with weather information’

-<https://www.kaggle.com/datasets/taranvee/smart-home-dataset-with-weather-information/discussion>

2. Dr.Avishek Pal,Dr. PKS Prakash (2017) ‘Practical Time series

Analysis’ 238(1)

3.Energy consumption prediction using machine learning; a review by Amir Mosavi and Abdullah Bahmani, 2019

4.R. A. Rashid, L. Chin, M. A. Sarijari, R. Sudirman and T. Ide, "Machine Learning for Smart Energy Monitoring of Home Appliances Using IoT", 2019

5. Energy consumption prediction by using machine learning for smart building: Case study in Malaysia”, by Mel Keytingan M. Shapi, Nor Azuana Ramli, Lilik J. Awalin, 2021

6.A machine-learning ensemble model for predicting energy consumption in smart homes”, by Ishaani Priyadarshini, Sandipan Sahu, Raghvendra Kumar, David Taniar,2022

7.Mohammad Malekzadeh(2019), Data understanding & EDA & time-series modeling

<https://www.kaggle.com/malekzadeh/smart-home-data-processing-weather-vs-energy>

8.Smith, J. (2021). "Energy Consumption Prediction Using Machine Learning Techniques." Journal of Energy Engineering, 45(2), 78-94

9.Johnson, R., & Smith, M. (2019). "Time Series Analysis: Forecasting and Control." Wiley.

10.Zhang, Y., et al. (2017). "Long Short-Term Memory Networks for Time Series Forecasting." Neural Computing and Applications, 31(10), 7973-7987.

11.J Parikh, S., & Yousuf, T. (2019). Energy consumption prediction using machine learning algorithms: A review. 2019 11th International Conference on Computational Intelligence and Communication Networks (CICN), 10-15.

12.Li, X., & Hong, T. (2020). A review of machine learning techniques for building energy consumption prediction: From single-building to campus-scale applications. Applied Energy, 261, 114452. ohnson, R., & Smith, M. (2019). "Time Series Analysis: Forecasting and Control." Wiley.