UNIVERSITY SCHOOL OF AUTOMATION AND ROBOTICS
GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY
EAST DELHI CAMPUS, SURAJMAL VIHAR, DELHI- 110032

Summer Training Report

On

Demand Energy Consumption Forecasting Model

Submitted in partial fulfillment of the

requirements for the completion of six-week summer internship/training [ART 457]

Name: **B Dhruv**

Enrollment Number **03519011921**

Under the supervision of

Dr. Renu Dalal

DECLARATION

I hereby declare that the Summer Training Report entitled **Demand Energy Consumption Forecasting Model** is an authentic record of work completed as part of the requirements of Summer Training (ART 457) during the period from 24/06/24 to 09/08/24 in **50 Hertz Limited** under the supervision of **Dr. Renu Dalal**.

(Signature of student)

B Dhruv

Enrollment No: 03519011921

Date: ____________________

(Signature of Supervisor)

Dr. Renu Dalal

Date: _____________________

Ref. No - 0721/2024

Date – 09.08.2024

Internship Completion Letter

This is to certify that Mr. B Dhruv S/O Mr. B Balasubramaniam a student of BTech in Artificial Intelligence and Data Science from Guru Gobind Singh Indraprastha University has successfully completed 45 Days (24th June 2024 to 9th August 2024) Internship programme at our organization.

During the period of his assignment, we found him sincere, hardworking and a keen learner.

We wish him all success in his future endeavors.

Sincerely,

50 HERTZ LIMITED

Head – HR & ADM

50 HERTZ LIMITED

Regd. Office : 303, 3rd Floor, D21, Corporate Park, Sector - 21, Dwarka, New Delhi - 110077, Tel : 011 - 4011 0747
Branch Office : 3rd Floor, 'AASTHA' 460 Madurdaha, Off E.M. Bye Pass, Kolkata 700107, Tel : +91 33 4061 0165
CIN - U72900DL2012PLC239290 | E : contactus@50hertz.in | www.50hertz.in

Acknowledgement

I would like to express my sincere gratitude to **Mr. Sanjeev Jain** for his invaluable guidance and support throughout my internship at **50 Hertz Limited**. His mentorship was crucial to the successful completion of the **Demand Energy Consumption Forecasting Model** project. He also provided me with the load data of BRPL Delhi, which was essential for my analysis. His insights greatly enhanced my learning experience.

I am also thankful to the entire team at **50 Hertz Limited** for providing me with the opportunity to work on this project and for their assistance and cooperation. I am particularly grateful to the employees in the Data Science Department, who helped me and guided me whenever I had doubts about the project and explained the theory and workings behind it. The experience gained during this internship has been immensely valuable to my professional growth.

Additionally, I would like to acknowledge my supervisor, **Dr. Renu Dalal** at **University School of Automation and Robotics,** for her support regarding any doubts I had about the internship report and for helping me as needed.

About The Company

50 Hertz Limited is a leading energy data analytics company founded in 2012, renowned for its innovative solutions tailored to the energy sector. Specializing in Solar, Wind, and Demand Energy Forecasting, 50 Hertz Limited provides comprehensive services that drive efficiency and reliability in energy management.

Company Overview

As a multi-domain software and service provider, 50 Hertz Limited offers a range of cutting-edge solutions designed to address the diverse needs of the energy industry. The company is committed to enhancing operational efficiency and decision-making processes for its clients through advanced analytics and technology-driven insights.

Core Domains

Solar: 50 Hertz Limited provides specialized services for solar power generation, including short-term forecasting and scheduling. Their solutions ensure accurate predictions of solar energy production, enabling optimal integration of solar power into the energy grid.

Wind: The company offers expert services for wind energy forecasting, helping clients manage the variability and intermittency of wind power. Their forecasting models support effective scheduling and grid management, enhancing the reliability of wind energy integration.

Demand Energy Forecasting: One of the core areas of expertise at 50 Hertz Limited is demand energy forecasting. The company employs advanced forecasting models to predict electricity demand, enabling utilities and energy providers to optimize grid operations and improve supply reliability.

Solutions and Services

- Enterprise-Level Decision Support System: Designed for power utilities, this system enhances decision-making and operational efficiency by providing actionable insights and comprehensive data analysis.
- Short-Term Forecasting and Scheduling: Specialized services for accurate forecasting and effective scheduling of wind and solar power, ensuring seamless integration into the grid.
- Asset Monitoring & Optimization: Solutions to monitor and optimize the performance of energy assets, maximizing efficiency and reliability.
- Integrated Energy Management: Comprehensive energy management solutions that streamline operations and improve overall energy management strategies.

Clientele

50 Hertz Limited serves a diverse range of clients globally, including state utilities, distribution companies (DISCOMs), independent power producers (IPPs), renewable energy generators, and commercial and industrial customers. Their sector-specific services and solutions are trusted for their accuracy, reliability, and effectiveness in enhancing energy management practices.

Commitment to Innovation

At 50 Hertz Limited, innovation is at the heart of their operations. The company continuously invests in developing and refining their analytical tools and methodologies to stay ahead of industry trends and deliver exceptional value to their clients. Their dedication to driving efficiency, reliability, and sustainability in the energy sector underscores their position as a leader in energy data analytics.

With a focus on delivering cutting-edge software and services, 50 Hertz Limited is committed to supporting the energy sector in navigating the complexities of modern energy management and achieving greater operational excellence.

Table of Content

Sr No.	Topic	Page No.
1	Abstract	1
2	Introduction	2
3	Problem Statement	3
4	Description of various Training Module	4-6
5	Flow Chart	7-8
6	About the Dataset	9-10
7	Data Preprocessing	11-14
8	Forecasting Approaches	15-16
9	Day-Ahead Forecasting	17-18
10	Intraday Forecasting	19-20
11	Hardware Requirements	21
12	Software Requirements	22-23
13	Screenshot of IDE	24-25
14	Result	26-28
15	Screenshot of Data Sheet	29
16	References	30

Abbreviations

1. ARIMA: AutoRegressive Integrated Moving Average
2. SARIMA: Seasonal AutoRegressive Integrated Moving Average
3. SARIMAX: Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors
4. BRPL: BSES Rajdhani Power Limited
5. MAPE: Mean Absolute Percentage Error
6. MAE: Mean Absolute Error
7. TB: Time Block
8. ACF: Autocorrelation Function
9. PACF: Partial Autocorrelation Function
10. RH: Relative Humidity
11. CDS: Copernicus Climate Data Store
12. ERA5: European Centre for Medium-Range Weather Forecasts Reanalysis 5
13. WS: Wind Speed
14. CPU: Central Processing Unit
15. SSD: Solid State Drive
16. IDE: Integrated Development Environment

Abstract

Accurate forecasting of electricity demand is pivotal for optimizing grid operations and ensuring a reliable power supply. This project, titled "Demand Energy Consumption Forecasting Model," is dedicated to enhancing electricity demand predictions for BRPL Delhi using sophisticated time series forecasting techniques. The project employs ARIMA, SARIMA, and SARIMAX models to generate both intraday and day-ahead forecasts by analyzing an extensive dataset spanning 7 years.

The forecasting models are designed to handle the complexities of time series data, including trends, seasonal variations, and the impact of exogenous variables. By incorporating additional data such as temperature, relative humidity, and wind speed, the models aim to capture the multifaceted nature of electricity consumption. The integration of these exogenous variables enhances the models' accuracy and reliability, leading to more precise demand forecasts.

The process involves a comprehensive approach to data preparation, including handling missing values, creating a datetime index, and setting data frequency. SARIMA models are employed with carefully chosen parameters to reflect seasonal and non-seasonal components of the data. Model performance is rigorously evaluated using metrics such as Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE), with the goal of achieving a MAPE of approximately 3%, demonstrating high forecasting accuracy.

The project also focuses on refining model parameters to optimize performance, which includes iterative tuning and retraining to enhance accuracy. By providing detailed forecasts for both intraday and day-ahead periods, the model supports effective grid management and operational decision-making.

The insights derived from this project are intended to advance energy management strategies and improve operational efficiency within the power sector. The implementation of these forecasting models offers valuable contributions towards better energy resource planning and grid stability.

Introduction

Accurate forecasting of electricity demand is essential for optimizing grid operations and ensuring a reliable power supply. The "Demand Energy Consumption Forecasting Model" project is dedicated to predicting future electricity demand for BRPL Delhi using advanced time series forecasting techniques, aiming to enhance both accuracy and operational efficiency.

Effective electricity demand forecasting involves analyzing historical load data to uncover patterns and trends that can inform future consumption predictions. This project utilizes several forecasting models, including ARIMA, SARIMA, and SARIMAX, to provide both intraday and day-ahead forecasts. These models are adept at handling the complexities of time series data, such as seasonal variations and trends, to deliver precise demand forecasts.

The dataset for this project spans 7 years and includes additional exogenous variables such as temperature, relative humidity, and wind speed, which are incorporated into the models to improve forecasting accuracy. The implementation of these models has resulted in a Mean Absolute Percentage Error (MAPE) of approximately 3, demonstrating high accuracy in demand predictions.

The primary goal of this project is to develop a robust forecasting model capable of predicting electricity demand with high precision, thereby supporting effective decision-making and enhancing the reliability of energy supply systems. The insights and findings from this project are intended to contribute to better energy management strategies and improved operational efficiency within the power sector.

Problem Statement

In the energy sector, accurately forecasting electricity demand is crucial for maintaining grid stability, optimizing energy distribution, and ensuring efficient operation of power systems. The challenge lies in predicting future electricity demand with high precision, considering the complexities and variability inherent in historical load data.

For the BRPL Delhi region, there is a need to develop a forecasting model that can effectively handle the intricacies of time series data, including trends, seasonality, and the impact of exogenous variables. Traditional forecasting methods may fall short in capturing the dynamic patterns of electricity consumption, leading to less accurate predictions and suboptimal grid management.

The primary problem addressed by the "Demand Energy Consumption Forecasting Model" project is the development of a reliable and accurate forecasting model for electricity demand that can provide both intraday and day-ahead forecasts. This involves:

1. **Handling Large and Complex Datasets:** Efficiently managing and preprocessing 7 years of historical load data, incorporating additional exogenous variables like temperature, relative humidity, and wind speed.
2. **Developing Accurate Forecasting Models:** Implementing advanced time series forecasting techniques, including ARIMA, SARIMA, and SARIMAX, to capture the underlying patterns and trends in the data.
3. **Achieving High Forecasting Accuracy:** Ensuring that the models achieve a low Mean Absolute Percentage Error (MAPE), ideally around 3, to provide reliable predictions that support effective grid management and operational decision-making.

<u>Description of Various Training Modules</u>

During my internship at 50 Hertz Limited, I acquired significant knowledge in time series forecasting, data analysis, and modeling. Below are the key concepts and techniques I learned, along with their practical applications in the field of energy demand forecasting:

1. Decomposing Time Series Data

One of the first techniques I learned was decomposing time series data into its essential components (trend, seasonality, and residual) using the seasonal_decompose function from the statsmodels library. This technique enabled me to analyze the different patterns in the data and understand their impact on the overall demand forecast.

2. Stationarity Testing and Data Validation

To ensure the data was suitable for time series modeling, I applied the Augmented Dickey-Fuller (ADF) test using the adfuller function from statsmodels. This test helped me validate whether the series was stationary, which is a crucial step in preparing the data for ARIMA and SARIMA models.

3. Analyzing Autocorrelation

Autocorrelation is a fundamental concept in time series analysis, and I learned how to visualize and interpret autocorrelation functions (ACF) and partial autocorrelation functions (PACF) using plot_acf and plot_pacf. These functions helped me understand how past values influence future values and provided insights into the selection of model parameters.

4. SARIMAX Model Development

The SARIMAX model from statsmodels was a key tool in my training. I used it to build seasonal models that incorporated exogenous variables, such as temperature, wind speed, and relative humidity. This allowed me to enhance the forecasting accuracy for both intraday and day-ahead electricity demand predictions.

5. Identifying Trends and Seasonality in Data

I learned to recognize and analyze the presence of long-term trends and seasonal patterns in the load data. Understanding these patterns was essential for building models that could accurately predict variations in electricity demand over time.

6. Modeling Dependencies Using ACF and PACF

The concepts of autocorrelation and partial autocorrelation are vital in time series modeling. I used ACF and PACF plots to determine the appropriate lag values for autoregressive (AR) and moving average (MA) processes, which helped me define the parameters for ARMA and ARIMA models.

7. Exploring ARMA Models

I gained an in-depth understanding of ARMA (AutoRegressive Moving Average) models, which combine the strengths of both AR and MA processes. This was an important step toward understanding more complex models like ARIMA and SARIMA.

8. Introduction to Time Series and Data Preparation

Before diving into advanced models, I learned the basics of time series data, including how to structure it and handle common issues such as missing values, outliers, and irregular intervals. This foundational knowledge was key to ensuring accurate modeling.

9. Dealing with Non-Stationary Data Using ARIMA

One of the critical challenges in time series analysis is dealing with non-stationary data. I learned to apply ARIMA models to handle non-stationarity by differencing the data and capturing both trend and seasonality.

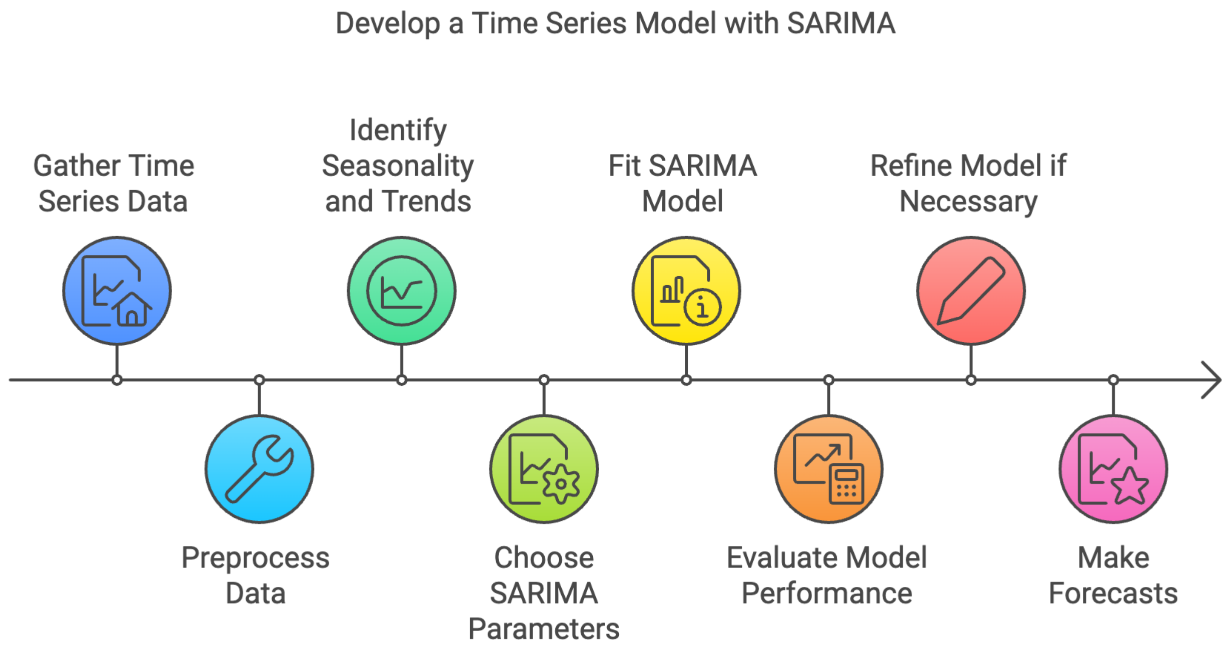
10. Conducting Dickey-Fuller Tests

The Dickey-Fuller and Augmented Dickey-Fuller tests were essential in my training for detecting unit roots and confirming whether differencing was necessary to stabilize the time series. These tests guided the pre-processing of the data before applying ARIMA-based models.

11. Building ARIMA, SARIMA, and SARIMAX Models

I gained expertise in building and fine-tuning ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), and SARIMAX (Seasonal ARIMA with exogenous variables) models. These models formed the backbone of the demand forecasting system I developed during the internship, particularly for accurately predicting electricity demand in the BRPL Delhi region.

FlowChart



The flowchart visually represents the sequential steps involved in developing a Seasonal AutoRegressive Integrated Moving Average (SARIMA) time series model. Each step is crucial for ensuring that the model is accurate and reliable. Here’s a breakdown of the process:

1. Gather Time Series Data: The BRPL data is a time series dataset where the LOAD value depends on various factors such as time of day, day of the week, and seasonal patterns. The objective is to model and forecast the LOAD values accurately using SARIMA.

2. Identify Seasonality and Trends: Analyze the data to detect any seasonal patterns (e.g., monthly, quarterly) and trends (upward or downward). This analysis is crucial for understanding the underlying behavior of the data.

3. Preprocess Data: Prepare the data by handling missing dates and values, creating a datetime index for the dataframe, and setting the frequency of the dataframe. This ensures that the data is clean and ready for modeling.

4. Choose SARIMA Parameters: Select appropriate values for the SARIMA model's parameters. Specifically, determine the values for the autoregressive order (p), integrated order (d), and moving average order (q), as well as the seasonal components: seasonal autoregressive order (P), seasonal integrated order (D), and seasonal moving average order (Q).

5. Fit SARIMA Model: Train the SARIMA model using the preprocessed data and the chosen parameters. This involves estimating the model parameters to best fit the data.

6. Evaluate Model Performance: Assess the model's performance using metrics such as Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). These metrics are used to quantify the accuracy of the forecasts.

7. Refine Model if Necessary: If the model's performance is not satisfactory, adjust the order and seasonal order parameters to hypertune the model. Retrain the model with these adjusted parameters to improve accuracy.

8. Make Forecasts: Use the fitted SARIMA model to generate predictions for future time periods. This step provides the forecasted LOAD values based on the trained model.

About the Dataset

The dataset used for this project was provided by my mentor, Sanjeev Jain, at 50 Hertz Limited. It comprises detailed electricity load data from the BRPL (BSES Rajdhani Power Limited) Delhi region. This dataset spans from April 1, 2017, to June 26, 2024, and includes a comprehensive range of information crucial for forecasting and analysis.

Load Data

The primary dataset features electricity load data recorded at 15-minute intervals, resulting in 96 data points per day. Each data entry includes the following columns:

Load: The electricity demand recorded in megawatts (MW) for each 15-minute interval.

Block: An identifier for each 15-minute interval within a day, ranging from 1 to 96.

Date: The date corresponding to each data entry.

Time: The specific time of day for each 15-minute interval.

This detailed data format provides a granular view of electricity consumption, essential for accurate forecasting and trend analysis.

Climate Data

In addition to the load data, climate variables were also integrated into the analysis. The climate data was sourced from the Copernicus Climate Data Store (CDS), specifically the ERA5 dataset, which provides hourly data from 1940 to the present. For this project, climate variables for the Delhi region were extracted for the same period as the load data.

The extracted climate variables include:

Wind Speed: The hourly average wind speed recorded in meters per second (m/s).

Relative Humidity (RH): The hourly relative humidity percentage, which impacts electricity demand patterns.

Temperature: The hourly temperature readings in degrees Celsius (°C).

Combining these climate variables with the load data allows for a more comprehensive analysis, enabling the incorporation of external factors that influence electricity demand. This integration is crucial for improving the accuracy of forecasting models and understanding the relationship between climate conditions and energy consumption.

Data Preprocessing

Data preprocessing is a crucial step in preparing raw data for analysis and modeling. For this project, several preprocessing tasks were carried out to ensure the data's quality and usability. Here is a detailed overview of the preprocessing steps undertaken:

1. Loading the Data

The initial step involved loading the dataset into a DataFrame for analysis. The dataset, provided by Sanjeev Jain, included detailed load data recorded at 15-minute intervals, along with columns for date, time, and block. This data was loaded into a DataFrame using Python's pandas library, which facilitated subsequent data manipulation and analysis.

2. Handling Missing Dates

During the preprocessing, it was discovered that there was a missing date, specifically June 16, 2023. To address this gap, a new entry for June 16, 2023, was added to the DataFrame with None values for all load data. This step ensured that the dataset maintained continuous time series coverage, which is essential for accurate forecasting and modeling.

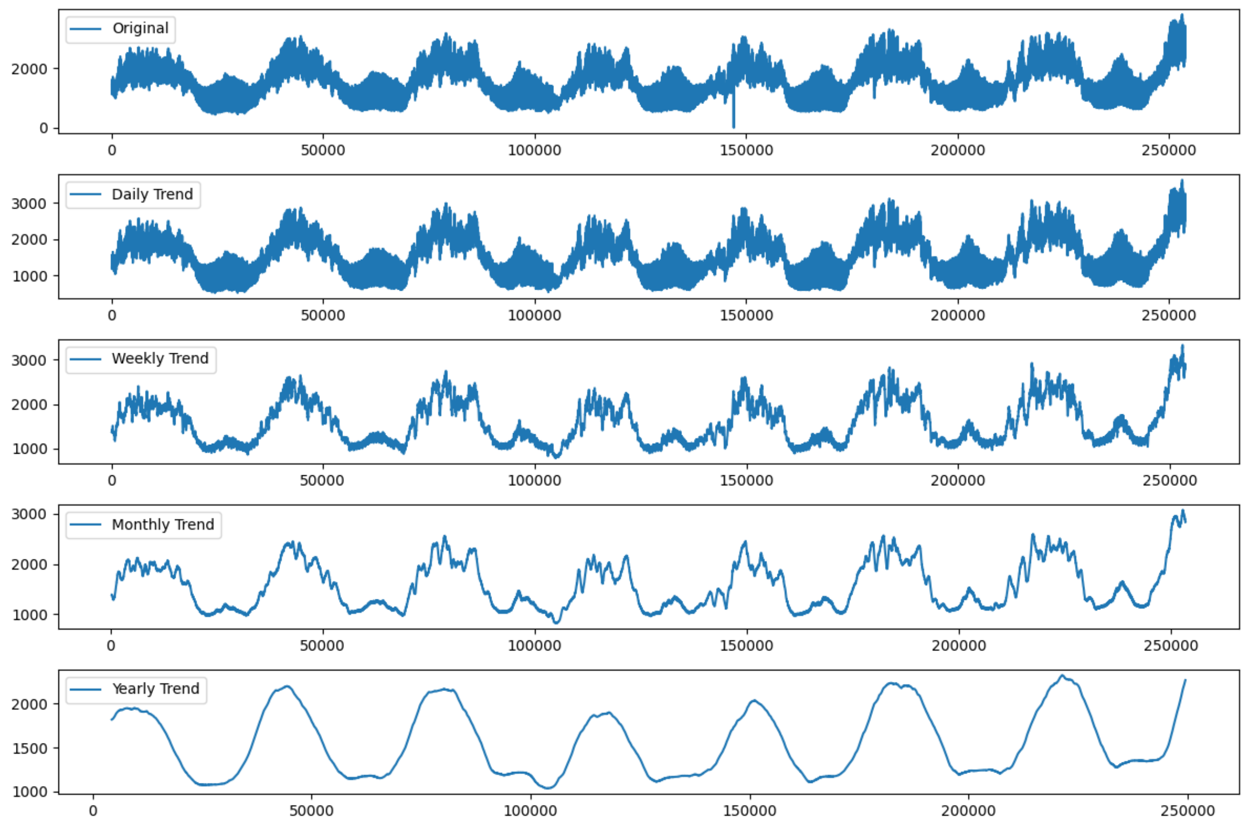
3. Filling Missing Load Values

Several missing load values needed to be addressed to complete the dataset:

- June 14, 2023
- June 16, 2023 (the previously added date)
- June 29, 2023
- September 9, 2023
- January 10, 2024

To fill these missing values, linear interpolation was used. This method estimates the missing values based on the values before and after the missing data points, providing a reasonable estimate for continuity in the time series.

4. Analyzing Trends in Data



Analyzing the data for patterns was essential to understand the underlying behavior and trends. The following observations were made:

Daily Patterns: Load data exhibited a clear daily cycle. The load decreases from 00:00 to 09:00, then increases until 15:00, decreases again until 19:00, and finally increases once more. This pattern reflects typical daily consumption trends, likely influenced by the varying demand throughout the day.

Monthly Patterns: On a monthly basis, the data showed an upward trend. This general increase in load over time could be attributed to growing demand or increasing population and industrial activity.

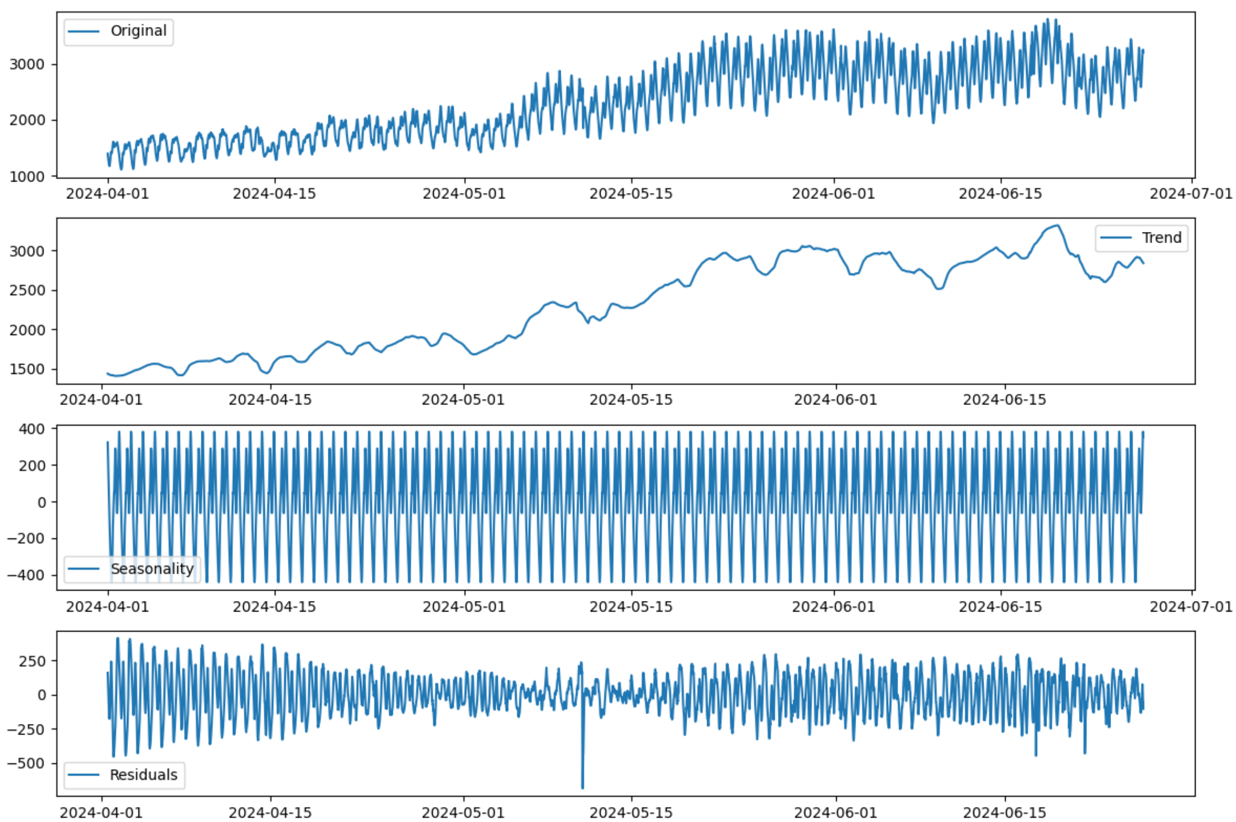
Yearly Patterns: Seasonal variations were evident in the yearly data. During the summer months (April to September), the load values were notably higher, while in the winter months (October to

March), the load values were lower. This trend aligns with higher electricity consumption during the hotter months due to increased air conditioning use and lower consumption in cooler months.

Impact of COVID-19: The dataset revealed a noticeable decrease in load values during the COVID-19 pandemic years of 2020 and 2021. This decrease likely reflects reduced industrial activity and lower electricity consumption due to lockdowns and work-from-home arrangements.

5. Analyzing Daily Seasonal Decomposition

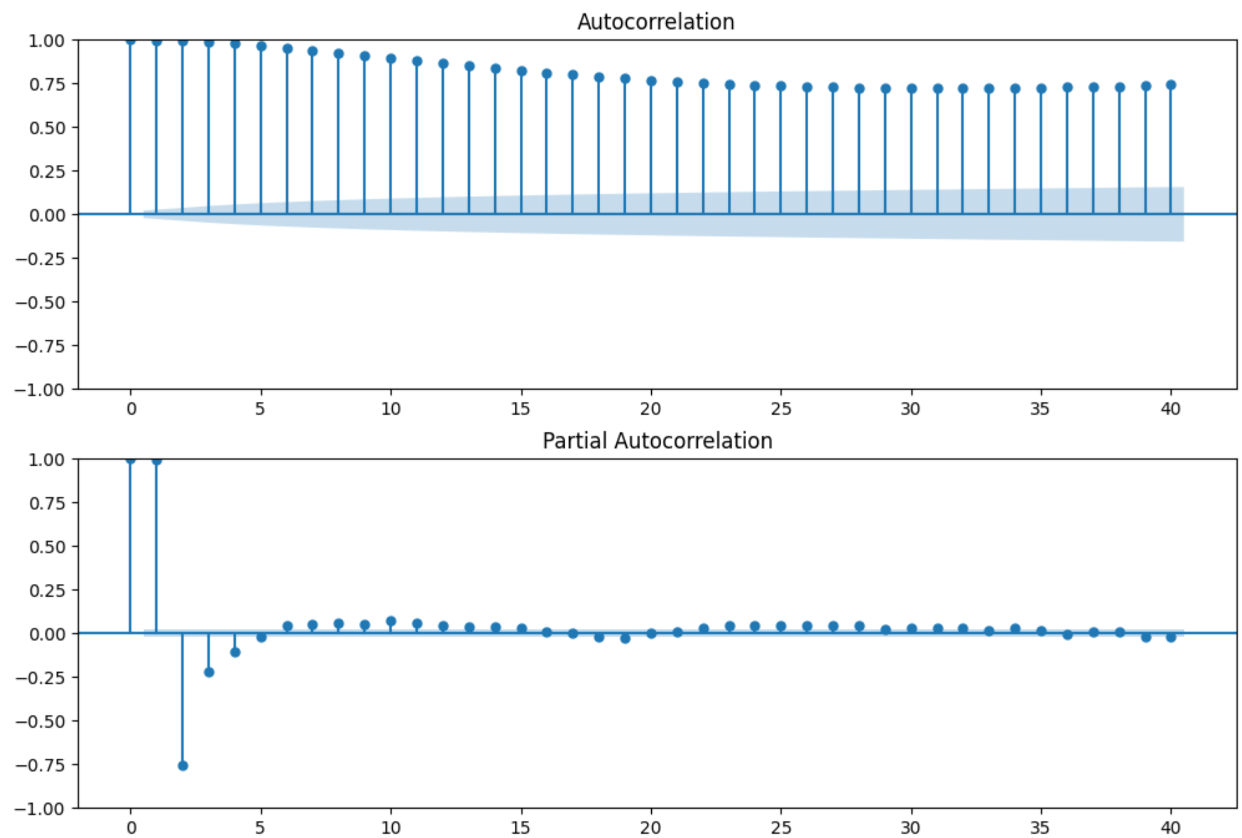
Daily seasonal decomposition is an essential technique for analyzing time series data, allowing us to break down the load data into its fundamental components. This process separates the data into three key elements: the trend, which reveals long-term changes; the seasonal component, which highlights repeating daily patterns; and the residual, which captures any remaining noise or irregularities.



By isolating these components, we gain a clearer understanding of the underlying patterns and fluctuations in electricity demand, which enhances the accuracy of our forecasting models and helps in refining predictions.

6. Identifying the Order and Seasonal Order for the SARIMAX Model

Determining the order and seasonal order for the SARIMAX model involves selecting the appropriate non-seasonal and seasonal parameters. The non-seasonal orders—autoregressive (p), integrated (d), and moving average (q)—are identified using ACF and PACF plots. For the seasonal components, the seasonal autoregressive (P), integrated (D), and moving average (Q) orders are chosen based on observed seasonal patterns and periodicity. This process ensures the model captures both the temporal and seasonal dynamics of the data effectively.



This preprocessing and analysis provided a solid foundation for building accurate forecasting models, allowing for a deeper understanding of the factors influencing electricity demand.

Forecasting Approaches for Demand Energy Consumption Forecasting

There are two main types of forecasting approaches used for Demand Energy Consumption Forecasting: Day-Ahead Forecasting and Intraday Forecasting. Both methods help anticipate electricity demand to ensure efficient grid operations.

1) Day-Ahead Forecasting

In Day-Ahead Forecasting, the model is trained using historical data up until 8:00 AM on the day of the forecast. At this time, the model has actual load data available only up to 7:15 AM, maintaining a 3-time block (TB) lag. Based on this data, it predicts the demand for the entire next day, covering all 96-time blocks (15-minute intervals) for the upcoming 24 hours. In this project, two different training periods were used: 30 days and 60 days of historical data.

Model Run Time: 8:00 AM.

Actual Load Data Available: Up to 7:15 AM.

Training Period: 30 days and 60 days.

Forecast Window: The entire next day (96-time blocks).

This forecasting helps plan energy generation and schedule resources efficiently for the upcoming day.

2) Intraday Forecasting

Intraday Forecasting is used for real-time load predictions within the same day. The model also uses a 3 TB lag of actual load data. For example, if the model is run at 10:30 PM, it only has actual load data available until 9:15 PM. It then forecasts the demand for the next 6 TB (1.5 hours) ahead of the model run time, predicting the load for the period from 12:00 AM to 1:30 AM.

This process is repeated throughout the day, with the model run 16 times a day, updating its training data and forecasting the next 6 TB at each run.

Model Run Frequency: 16 times a day.

Lag: 3 TB (45 minutes of actual load data).

Forecast Window: Next 6 TB (1.5 hours) after the model run time.

This dynamic forecasting approach allows grid operators to adjust their predictions frequently and maintain real-time accuracy throughout the day.

Comparison Between Day-Ahead and Intraday Forecasting

- At the end of both forecasting approaches, the target was to achieve accurate predictions with acceptable error margins. In Day-Ahead Forecasting, we aimed for forecasted values with a Mean Absolute Percentage Error (MAPE) close to 6%, while in Intraday Forecasting, we aimed for MAPE values near 3%.

- From running the models in both cases, it was observed that the model run time for Day-Ahead Forecasting was approximately 10 times less than that of Intraday Forecasting. However, the MAPE for Day-Ahead was around 6% due to irregular load patterns. For example, some days had an earlier peak load at 2 PM instead of 3 PM, which increased the forecasting error.

- On the other hand, with Intraday Forecasting, only 2-3 forecasted periods out of the 16 daily runs for 6 TB showed a high MAPE, and the rest of the forecasts remained accurate even with irregular load data. The drawback was that since the model needed to be run 16 times per day, the model run time was 10 times more than that of Day-Ahead Forecasting to complete the forecasts for the entire day.

Day-Ahead Forecasting

Day-Ahead Forecasting is a vital component of Demand Energy Consumption forecasting that aims to predict the load for the following day based on historical data. The process involves the following steps:

Forecasting Process

For Day-Ahead Forecasting, two SARIMA (Seasonal AutoRegressive Integrated Moving Average) models were used:

- Model with 1 Month Training Data: This model utilizes historical load data from the past month to generate predictions for the next day.
- Model with 2 Months Training Data: This model incorporates a longer history of data from the past two months, aiming to enhance the accuracy of the forecasts.

Each day, the forecasting process begins at 8:00 AM, using load data available up to 7:15 AM to predict the demand for the entire next day. This setup maintains a 3-time block (TB) lag, meaning that the forecasts are made based on data that is 45 minutes old. The forecast covers all 96 time blocks (15-minute intervals) for the next 24 hours.

The forecasting was carried out over a period of 26 days, from June 1 to June 26, 2024. During this period, forecasts were generated daily, and the performance of the models was evaluated by comparing the forecasted values with the actual load data.

Performance Evaluation

The accuracy of the Day-Ahead Forecasting models was assessed using the Mean Absolute Percentage Error (MAPE). For June 2024, the average MAPE was 6.234%, indicating the overall accuracy of the forecasts. This metric reflects the average deviation between the predicted and actual load values.

To better understand the performance of the models:

Comparison: The forecasted values were compared to the actual load data for each day.

Visualization: Graphs were created to display the training load values, test load values, and forecasted load values. These visualizations help in analyzing how well the SARIMA models predicted the next day’s demand.

This approach provided a clear view of the SARIMA models' effectiveness in predicting demand based on historical data, offering valuable insights for energy planning and resource scheduling for the next day.

Intraday Forecasting

Intraday Forecasting is designed to predict energy demand within the same day, enabling real-time adjustments and accurate load predictions. This approach involves frequent updates and forecasts based on the latest available data. Here’s a detailed description of the process and models used:

Model Development and Approach

1. SARIMA Models for Monthly Forecasting

- March 2024 Forecasting
- April 2024 Forecasting
- May 2024 Forecasting
- June 2024 Forecasting

2. SARIMAX Models for June 2024 with Exogenous Variables

- SARIMAX with Temperature as the Exogenous Variable: This model incorporated temperature data to enhance the accuracy of the forecasts, recognizing the impact of temperature on energy demand.
- SARIMAX with Relative Humidity (RH) as the Exogenous Variable: Relative Humidity was included in this model to account for its effect on load demand.
- SARIMAX with Wind Speed (WS) as the Exogenous Variable: Wind Speed was used to adjust forecasts based on its influence on energy consumption patterns.
- SARIMAX with All Three Variables (Temperature, RH, and WS) Together: A comprehensive model combining temperature, relative humidity, and wind speed to provide a more robust forecast by integrating multiple climatic factors.

Forecasting Process

The forecasting process operates on a dynamic, real-time basis:

Model Run Time: If the model is run at 10:30 PM, it has actual load data available only up to 9:15 PM.

Lag: 3 time blocks (TB) of actual load data are used, which corresponds to a lag of 45 minutes.

Forecast Window: The model predicts load for the next 6 TB (1.5 hours){After Skipping 6TB}, covering the period from 12:00 AM to 1:30 AM.

This process is repeated 16 times a day, with each run updating the training data and forecasting the subsequent 6 TB of load. For example, if the forecast for 12:00 AM to 1:30 AM is completed, the model is then trained to forecast the next 6 TB period starting from 1:30 AM to 3:00 AM, and so on.

Combining Forecasts and Evaluation

After generating forecasts for each of the 16 periods per day, the individual forecasts are combined to create a full-day forecast. This comprehensive forecast is then compared to actual load data to evaluate accuracy.

Mean Absolute Percentage Error (MAPE): Calculated for each 6 TB forecast and then averaged to determine the overall daily MAPE.

Mean Absolute Error (MAE): Similarly calculated and used to assess the accuracy of forecasts.

Reporting

The results of the Intraday Forecasting are compiled into an Excel report using pandas and xlsxwriter. The report includes:

1. June Forecasted Hourly Data: Hourly forecast values for June.
2. June Forecasted Daily Data: Daily aggregated forecast values.
3. June Test Data: Actual load data for comparison.
4. MAPE: Mean Absolute Percentage Error for each forecasted period.
5. MAE: Mean Absolute Error for each forecasted period.
6. June Temperature Data: Only included in the SARIMAX model.
7. June RH Data: Only included in the SARIMAX model.
8. June Wind Speed Data: Only included in the SARIMAX model.

Hardware Requirements

For successfully running SARIMA and SARIMAX models and handling large datasets over extended periods in demand energy forecasting, the following hardware specifications are recommended:

1. Processor (CPU)

- Type: Intel Core i7/i9 or AMD Ryzen 7/9
- Cores: 4-8 cores minimum
- Clock Speed: 3.0 GHz or higher
- Purpose: A multi-core processor is crucial for performing complex time series forecasting computations efficiently, especially when running multiple iterations of SARIMA and SARIMAX models.

2. RAM

- Capacity: 16 GB or higher
- Type: DDR4 or DDR5
- Purpose: Handling large time series datasets requires significant memory, particularly when performing data transformations, running models, and processing multiple months of data.

3. Storage

- Type: Solid State Drive (SSD)
- Capacity: 256 GB or higher
- Purpose: High-speed storage is essential for quick access to large datasets and fast read/write operations and exporting forecast results to Excel files using pandas.

Software Requirements

For running SARIMA and SARIMAX models efficiently in the Demand Energy Forecasting project and handling large time series datasets, the following software specifications are recommended:

1. Operating System

- Type: 64-bit Windows 10/11, macOS, or Linux
- Purpose: A 64-bit operating system ensures compatibility with Python libraries and allows efficient processing of large datasets.

2. Python

- Version: Python 3.8 or higher
- Purpose: Python is used for implementing SARIMA/SARIMAX models, data manipulation, and running forecasting algorithms. Compatibility with relevant libraries is essential.

3. Python Libraries

- pandas: For data manipulation, handling large datasets, and exporting forecast results to Excel files.
- numpy: For numerical computations and efficient array handling.
- statsmodels: To build SARIMA and SARIMAX models for time series forecasting.
- scikit-learn: For calculating performance metrics like Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE).
- matplotlib and seaborn: For visualizing trends and forecast results through graphs.
- xlsxwriter: To export forecast data and performance metrics into structured Excel reports.

4. IDE/Code Editor

- Recommendation: Visual Studio Code or Kaggle or Jupyter Notebook

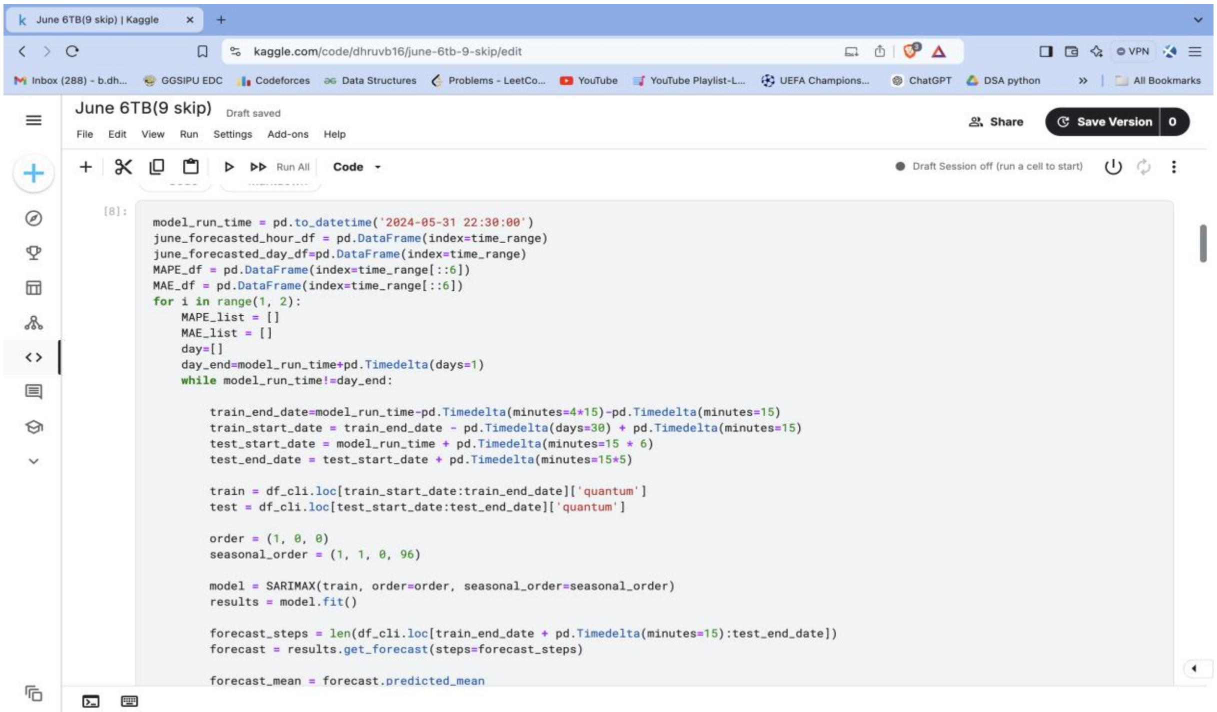
- Purpose: These environments offer tools for debugging, code completion, and easy library management, essential for smooth execution of forecasting models.

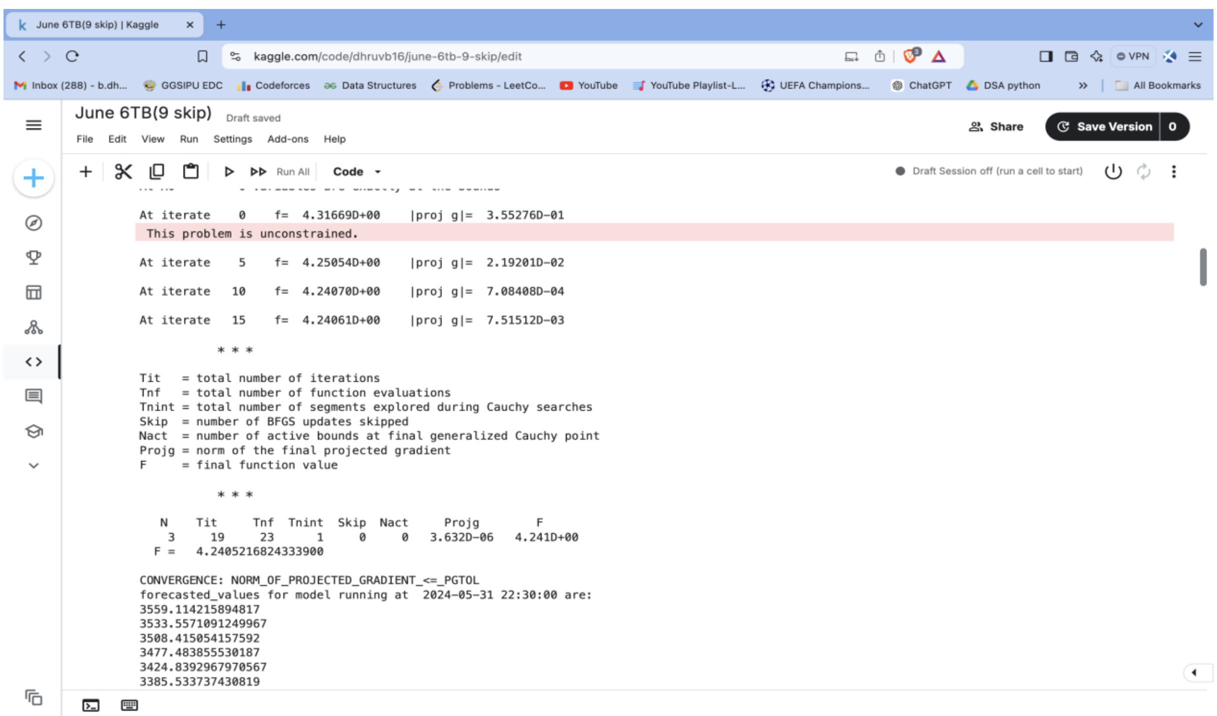
5. Excel Software

- Type: Microsoft Excel
- Purpose: To view and analyze the Excel reports generated from the forecasting process. These reports include forecasted values, actual load data, and error metrics like MAPE and MAE.

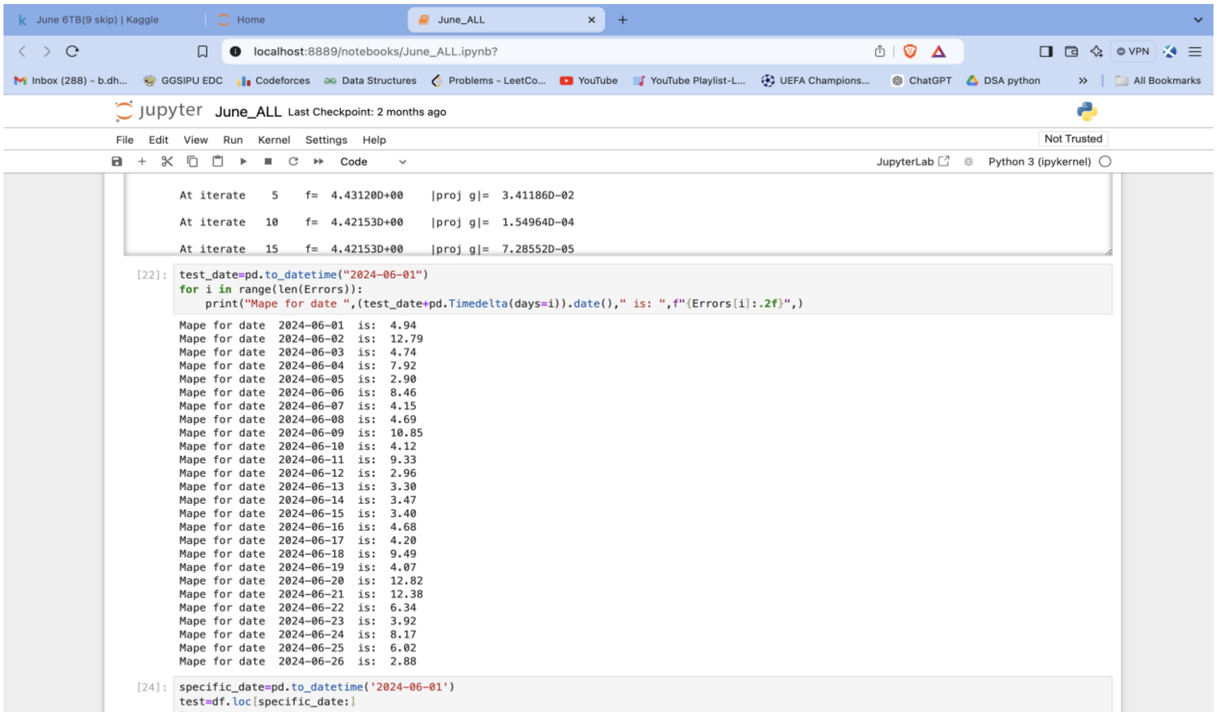
By ensuring these software requirements are met, smooth execution of SARIMA and SARIMAX models for energy forecasting can be guaranteed.

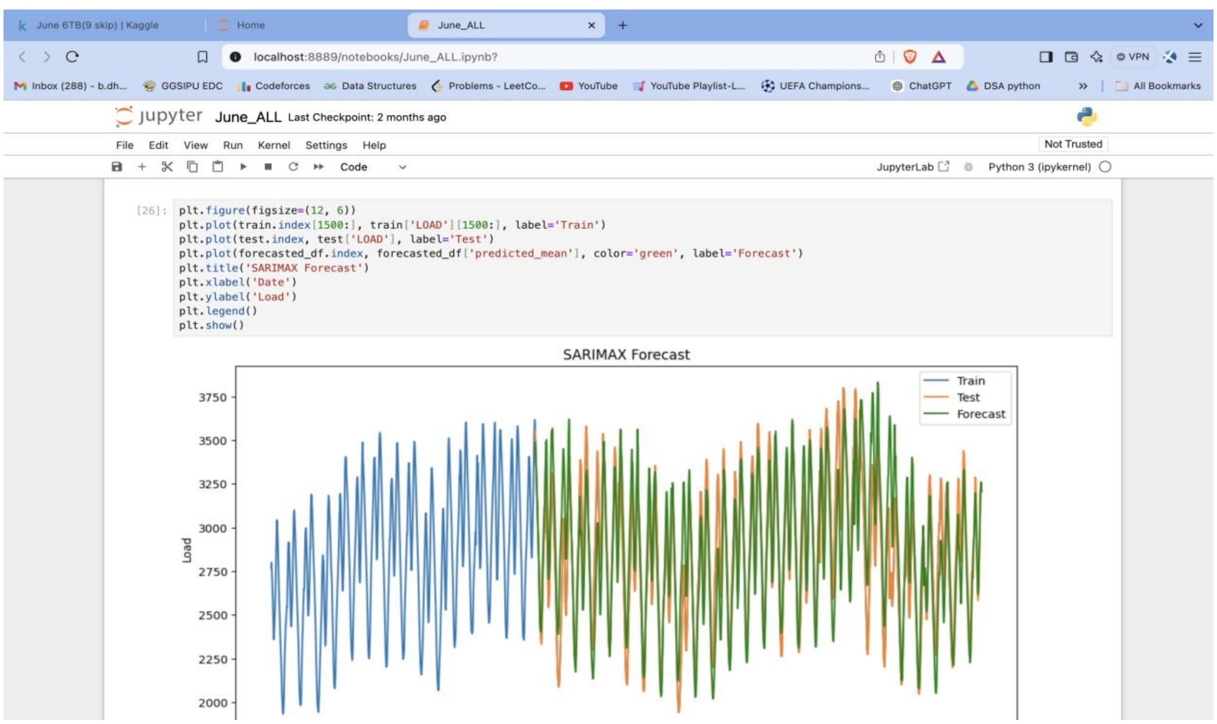
Snapshot of IDE (Kaggle)





Snapshot of IDE (Jupyter Notebook)





Result Obtained

Below are the results obtained from the Day-Ahead and Intraday Forecasting models. These results indicate the Mean Absolute Percentage Error (MAPE) values for each day of June 2024.

1. Day-Ahead Forecasting Results:

- MAPE for date 2024-06-01 is: 4.94
- MAPE for date 2024-06-03 is: 4.74
- MAPE for date 2024-06-05 is: 2.90
- MAPE for date 2024-06-07 is: 4.15
- MAPE for date 2024-06-09 is: 10.85
- MAPE for date 2024-06-11 is: 9.33
- MAPE for date 2024-06-13 is: 3.30
- MAPE for date 2024-06-15 is: 3.40
- MAPE for date 2024-06-17 is: 4.20
- MAPE for date 2024-06-19 is: 4.07
- MAPE for date 2024-06-21 is: 12.38
- MAPE for date 2024-06-23 is: 3.92
- MAPE for date 2024-06-25 is: 6.02

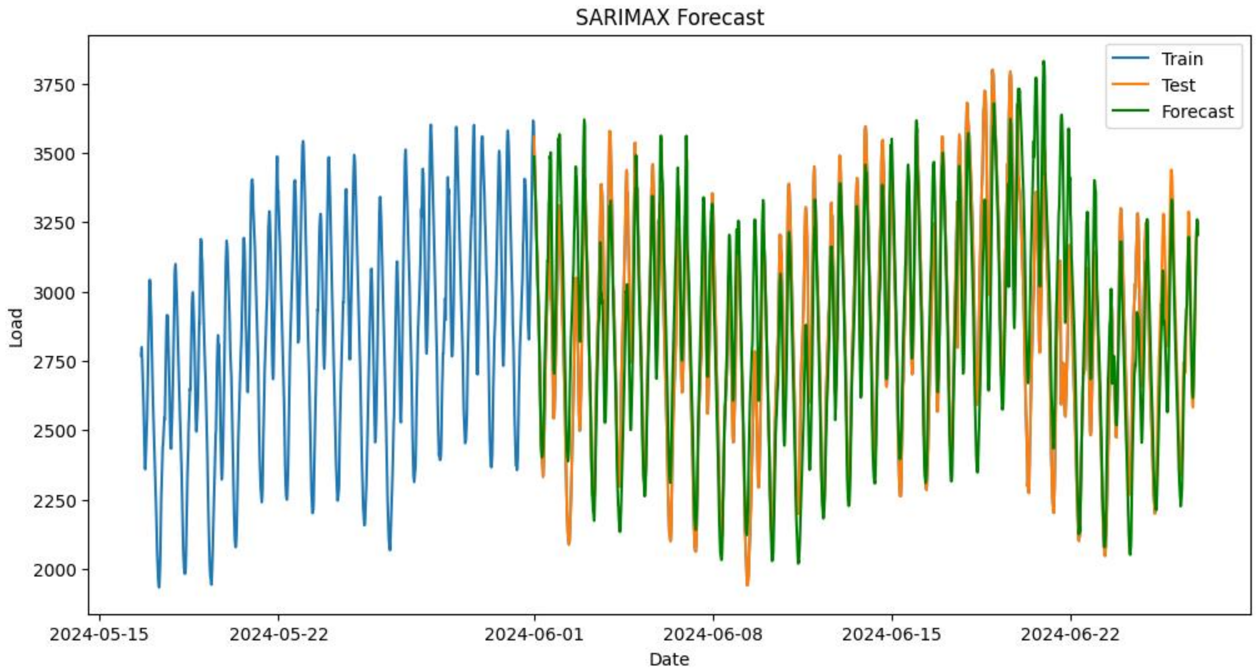
- MAPE for date 2024-06-02 is: 12.79
- MAPE for date 2024-06-04 is: 7.92
- MAPE for date 2024-06-06 is: 8.46
- MAPE for date 2024-06-08 is: 4.69
- MAPE for date 2024-06-10 is: 4.12
- MAPE for date 2024-06-12 is: 2.96
- MAPE for date 2024-06-14 is: 3.47
- MAPE for date 2024-06-16 is: 4.68
- MAPE for date 2024-06-18 is: 9.49
- MAPE for date 2024-06-20 is: 12.82
- MAPE for date 2024-06-22 is: 6.34
- MAPE for date 2024-06-24 is: 8.17
- MAPE for date 2024-06-26 is: 2.88

Average MAPE for Day-Ahead Forecasting: 6.05%

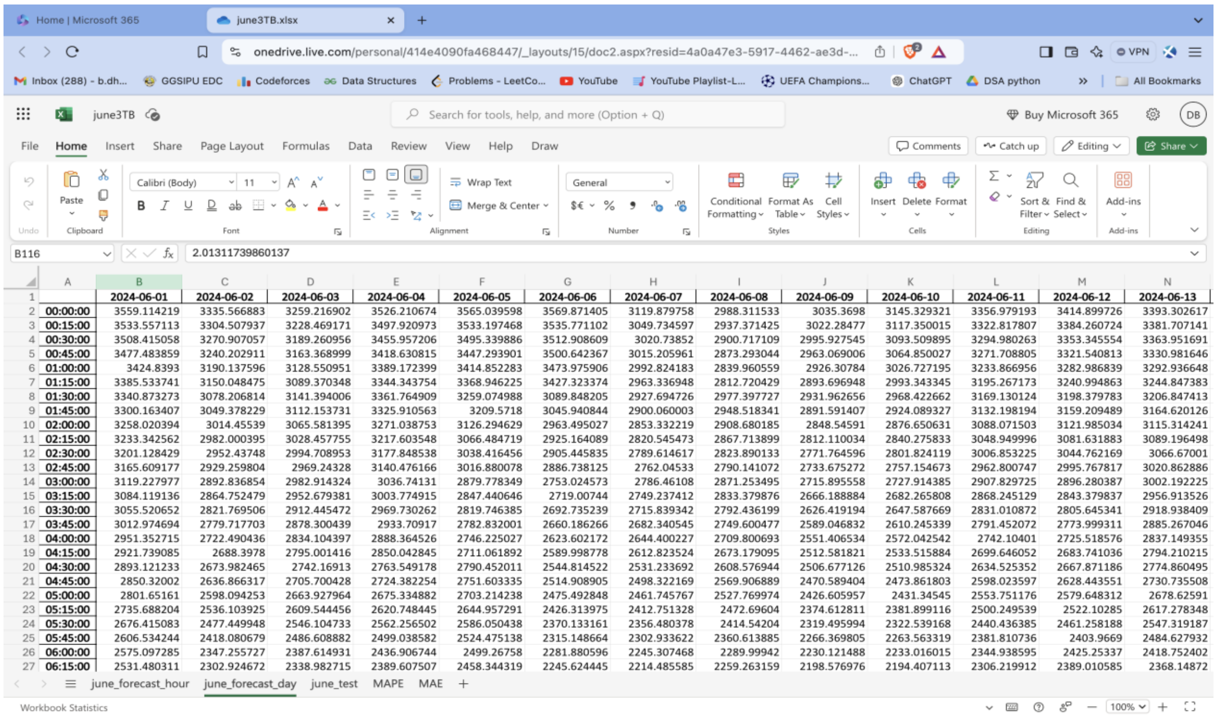
2. Intraday Forecasting Results:

- MAPE for date 2024-06-01 is: 3.25
- MAPE for date 2024-06-02 is: 3.81
- MAPE for date 2024-06-03 is: 3.57
- MAPE for date 2024-06-04 is: 2.63
- MAPE for date 2024-06-05 is: 2.47
- MAPE for date 2024-06-06 is: 2.88
- MAPE for date 2024-06-07 is: 2.30
- MAPE for date 2024-06-08 is: 2.79
- MAPE for date 2024-06-09 is: 4.27
- MAPE for date 2024-06-10 is: 2.75
- MAPE for date 2024-06-11 is: 1.89
- MAPE for date 2024-06-12 is: 1.19
- MAPE for date 2024-06-13 is: 0.90
- MAPE for date 2024-06-14 is: 2.28
- MAPE for date 2024-06-15 is: 1.82
- MAPE for date 2024-06-16 is: 3.12
- MAPE for date 2024-06-17 is: 3.24
- MAPE for date 2024-06-18 is: 2.79
- MAPE for date 2024-06-19 is: 2.12
- MAPE for date 2024-06-20 is: 3.72
- MAPE for date 2024-06-21 is: 7.49
- MAPE for date 2024-06-22 is: 3.44
- MAPE for date 2024-06-23 is: 3.36
- MAPE for date 2024-06-24 is: 2.81
- MAPE for date 2024-06-25 is: 3.06
- MAPE for date 2024-06-26 is: 2.27
- MAPE for date 2024-06-27 is: 4.41
- MAPE for date 2024-06-28 is: 7.24
- MAPE for date 2024-06-29 is: 5.05
- MAPE for date 2024-06-30 is: 4.09

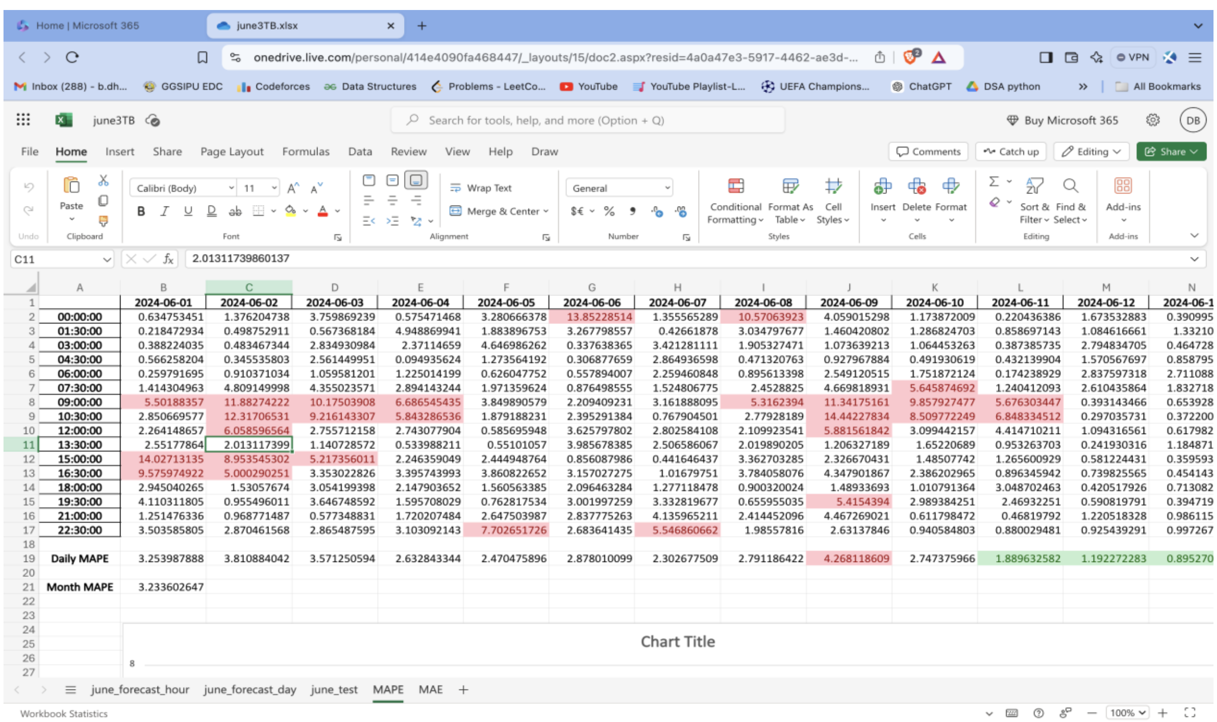
Average MAPE for Intraday Forecasting: 3.30%


SARIMAX Forecast
Train
Test
Forecast
Load
Date
3750
3500
3250
3000
2750
2500
2250
2000
2024-05-15
2024-05-22
2024-06-01
2024-06-08
2024-06-15
2024-06-22

Snapshot of Data Sheet



	2024-06-01	2024-06-02	2024-06-03	2024-06-04	2024-06-05	2024-06-06	2024-06-07	2024-06-08	2024-06-09	2024-06-10	2024-06-11	2024-06-12	2024-06-13
00:00:00	3559.114219	3335.566883	3259.216902	3526.210674	3565.039598	3569.871405	3119.879758	2988.311533	3035.3698	3145.329321	3356.979193	3414.899726	3393.302617
00:15:00	3533.557113	3304.507937	3228.469171	3497.920973	3533.197468	3535.771102	3049.734597	2937.371425	3022.28477	3117.350015	3322.817807	3384.260724	3381.707141
00:30:00	3508.415058	3270.907057	3189.260956	3455.957206	3495.339886	3512.908609	3020.73852	2900.717109	2995.927545	3093.509895	3294.980263	3353.345554	3363.951691
00:45:00	3477.483859	3240.202911	3163.368999	3418.630815	3447.293901	3500.642367	3015.205961	2873.293044	2963.069006	3064.850027	3271.708805	3321.540813	3330.981646
01:00:00	3424.8393	3190.137596	3128.550951	3389.172399	3414.852283	3473.975906	2992.824183	2839.960559	2926.30784	3026.727195	3233.866956	3282.986839	3292.936648
01:15:00	3385.533741	3150.048475	3089.370348	3344.343754	3368.946225	3427.323374	2963.336948	2812.720429	2893.696948	2993.343345	3195.267173	3240.994863	3244.847383
01:30:00	3340.873273	3078.206814	3141.394006	3361.764909	3259.074988	3089.848205	2927.694726	2977.397727	2931.962656	2968.422662	3169.130124	3198.379783	3206.847413
01:45:00	3300.163407	3049.378229	3112.153731	3325.910563	3209.5718	3045.940844	2900.060003	2948.518341	2891.591407	2924.089327	3132.198194	3159.209489	3164.620126
02:00:00	3258.020394	3014.45539	3065.581395	3271.038753	3126.294629	2963.495027	2853.332219	2908.680185	2848.54591	2876.650631	3088.071503	3121.985034	3115.314241
02:15:00	3233.342562	2982.000395	3028.457755	3217.603548	3066.484719	2925.164089	2820.545473	2867.713899	2812.110034	2840.275833	3048.949996	3081.631883	3089.196498
02:30:00	3201.128429	2952.43748	2994.708953	3177.848538	3038.416456	2905.445835	2789.614617	2823.890133	2771.764596	2801.824119	3006.853225	3044.762169	3066.67001
02:45:00	3165.609177	2929.259804	2969.24328	3140.476166	3016.880078	2886.738125	2762.04533	2790.141072	2733.675272	2757.154673	2962.800747	2995.767817	3020.862886
03:00:00	3119.227977	2892.836854	2982.914324	3036.74131	2879.778349	2753.024573	2786.46108	2871.253495	2715.895558	2727.914385	2907.829725	2896.280387	3002.192225
03:15:00	3084.119136	2864.752479	2952.679381	3003.774915	2847.440646	2719.00744	2749.237412	2833.379876	2666.188884	2682.265808	2868.245129	2843.379837	2956.913526
03:30:00	3055.520652	2821.769506	2912.445472	2969.730262	2819.746385	2692.735239	2715.839342	2792.436199	2626.419194	2647.587669	2831.010872	2805.645341	2918.938409
03:45:00	3012.974694	2779.717703	2878.300439	2933.70917	2782.832001	2660.186266	2682.340545	2749.600477	2589.046832	2610.245339	2791.452072	2773.999311	2885.267046
04:00:00	2951.352715	2722.490436	2834.104397	2888.364526	2746.225027	2623.602172	2644.400227	2709.800693	2551.406534	2572.042542	2742.10401	2725.518576	2837.149355
04:15:00	2921.739085	2688.3978	2795.001416	2850.042845	2711.061892	2589.998778	2612.823524	2673.179095	2512.581821	2533.515884	2699.646052	2683.741036	2794.210215
04:30:00	2893.121233	2673.982465	2742.16913	2763.549178	2790.452011	2544.814522	2531.233692	2608.576944	2506.677126	2510.985324	2634.525352	2667.871186	2774.860495
04:45:00	2850.32002	2636.866317	2705.700428	2724.382254	2751.603335	2514.908905	2498.322169	2569.906889	2470.589404	2473.861803	2598.023597	2628.443551	2730.735508
05:00:00	2801.65161	2598.094253	2663.927964	2675.334882	2703.214238	2475.492848	2461.745767	2527.769974	2426.605957	2431.34545	2553.751176	2579.648312	2678.62591
05:15:00	2735.688204	2536.103925	2609.544456	2620.748445	2644.957291	2426.313975	2412.751328	2472.69604	2374.612811	2381.899116	2500.249539	2522.10285	2617.278348
05:30:00	2676.415083	2477.449948	2546.104733	2562.256502	2586.050438	2370.133161	2356.480378	2414.54204	2319.495994	2322.539168	2440.436385	2461.258188	2547.319187
05:45:00	2606.534244	2418.080679	2486.608882	2499.038582	2524.475138	2315.148664	2302.933622	2360.613885	2266.369805	2263.563319	2381.810736	2403.9669	2484.627932
06:00:00	2575.097285	2347.255727	2387.614931	2436.906744	2499.26758	2281.880596	2245.307468	2289.99942	2230.121488	2233.016015	2344.938595	2425.25337	2418.752402
06:15:00	2531.480311	2302.924672	2338.982715	2389.607507	2458.344319	2245.624445	2214.485585	2259.263159	2198.576976	2194.407113	2306.219912	2389.010585	2368.14872



	2024-06-01	2024-06-02	2024-06-03	2024-06-04	2024-06-05	2024-06-06	2024-06-07	2024-06-08	2024-06-09	2024-06-10	2024-06-11	2024-06-12	2024-06-1
00:00:00	0.634753451	1.376204738	3.759869239	0.575471468	3.280666378	13.85228514	1.355565289	10.57063923	4.059015298	1.173872009	0.220436386	1.673532883	0.390995
01:30:00	0.218472934	0.498752911	0.567368184	4.948869941	1.883896753	3.267798557	0.42661878	3.034797677	1.460420802	1.286824703	0.858697143	1.084616661	1.33210
03:00:00	0.388224035	0.483467344	2.834930984	2.37114659	4.646986262	0.337638365	3.421281111	1.905327471	1.073639213	1.064453263	0.387385735	2.794834705	0.464728
04:30:00	0.566258204	0.345535803	2.561449951	0.094935624	1.273564192	0.306877659	2.864936598	0.471320763	0.927967884	0.491930619	0.432139904	1.570567697	0.858795
06:00:00	0.259791695	0.910371034	1.059581201	1.225014199	0.626047752	0.557894007	2.259460848	0.895613398	2.549120515	1.751872124	0.174238929	2.837597318	2.711088
07:30:00	1.414304963	4.809149998	4.355023571	2.894143244	1.971359624	0.876498555	1.524806775	2.4528825	4.669818931	5.645874692	1.240412093	2.610435864	1.832718
09:00:00	5.50188357	11.88274222	10.17503908	6.686545435	3.849890579	2.209409231	3.161888095	5.3162394	11.34175161	9.857927477	5.676303447	0.393143466	0.653928
10:30:00	2.850669577	12.31706531	9.216143307	5.843286536	1.879188231	2.395291384	0.767904501	2.77928189	14.44227834	8.509772249	6.848334512	0.297035731	0.372200
12:00:00	2.264148657	6.058596564	2.755712158	2.743077904	0.585695948	3.625797802	2.802584108	2.109923541	5.881561842	3.099442157	4.414710211	1.094316561	0.617982
13:30:00	2.55177864	2.013117399	1.140728572	0.533988211	0.55101057	3.985678385	2.506586067	2.019890205	1.206327189	1.65220689	0.953263703	0.241930316	1.184871
15:00:00	14.02713135	8.953545302	5.217356011	2.246359049	2.444948764	0.856087986	0.441646437	3.362703285	2.326670431	1.48507742	1.265600929	0.581224431	0.359593
16:30:00	9.575974922	5.000290251	3.353022826	3.395743993	3.860822652	3.157027275	1.01679751	3.784058076	4.347901867	2.386202965	0.896345942	0.739825565	0.454143
18:00:00	2.945040265	1.53057674	3.054199398	2.147903652	1.560563385	2.096463284	1.277118478	0.900320024	1.48933693	1.010791364	3.048702463	0.420517926	0.713082
19:30:00	4.110311805	0.955496011	3.646748592	1.595708029	0.762817534	3.001997259	3.332819677	0.655955035	5.4154394	2.989384251	2.46932251	0.590819791	0.394719
21:00:00	1.251476336	0.968771487	0.577348831	1.720207484	2.647503987	2.837775263	4.135965211	2.414452096	4.467269021	0.611798472	0.46819792	1.220518328	0.986115
22:30:00	3.503585805	2.870461568	2.865487595	3.103092143	7.702651726	2.683641435	5.546860662	1.98557816	2.63137846	0.940584803	0.880029481	0.925439291	0.997267
Daily MAPE	3.253987888	3.810884042	3.571250594	2.632843344	2.470475896	2.878010099	2.302677509	2.791186422	4.268118609	2.747375966	1.889632582	1.192272283	0.895270
Month MAPE	3.233602647												

References

1. https://pandas.pydata.org/pandas-docs/stable/
2. https://numpy.org/doc/stable/
3. https://www.statsmodels.org/stable/index.html
4. https://www.datacamp.com/community/tutorials/time-series-analysis-python
5. https://machinelearningmastery.com/sarima-for-time-series-forecasting-with-python/
6. https://en.wikipedia.org/wiki/Mean_absolute_percentage_error
7. https://en.wikipedia.org/wiki/Mean_absolute_error
8. https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.ExcelWriter.html
9. https://towardsdatascience.com/time-series-forecasting-with-sarimax-in-python-6e3f1c0c2f0f
10. https://realpython.com/pandas-dataframe/