Second Project

Electricity
Consumption Prediction

ANIKET KUMAR

Problem Statement

- This internship project is focused on leveraging Python for time series prediction of household electricity consumption. The dataset includes essential features such as date, time, global active power, global reactive power, voltage, global intensity, and sub-metering values. The objective is to build robust time series forecasting models that can accurately predict future electricity consumption trends based on historical data. The insights derived from this analysis can empower households to optimize energy usage, plan efficiently, and contribute to sustainable energy practices.

- Dataset Description

- 1. Date: Date of the electricity consumption recording.
- 2. Time: Time of the electricity consumption recording.
- 3. Global_active_power: Total active power consumed by the household.
- 4. Global_reactive_power: Total reactive power consumed by the household.
- 5. Voltage: Voltage level during the electricity consumption period.
- 6. Global_intensity: Total current intensity consumed by the household.
- 7. Sub_metering_1: Electricity consumption in sub-metering 1 (e.g., kitchen).
- 8. Sub_metering_2: Electricity consumption in sub-metering 2 (e.g., laundry).
- 9. Sub_metering_3: Electricity consumption in sub-metering 3 (e.g., water heater).

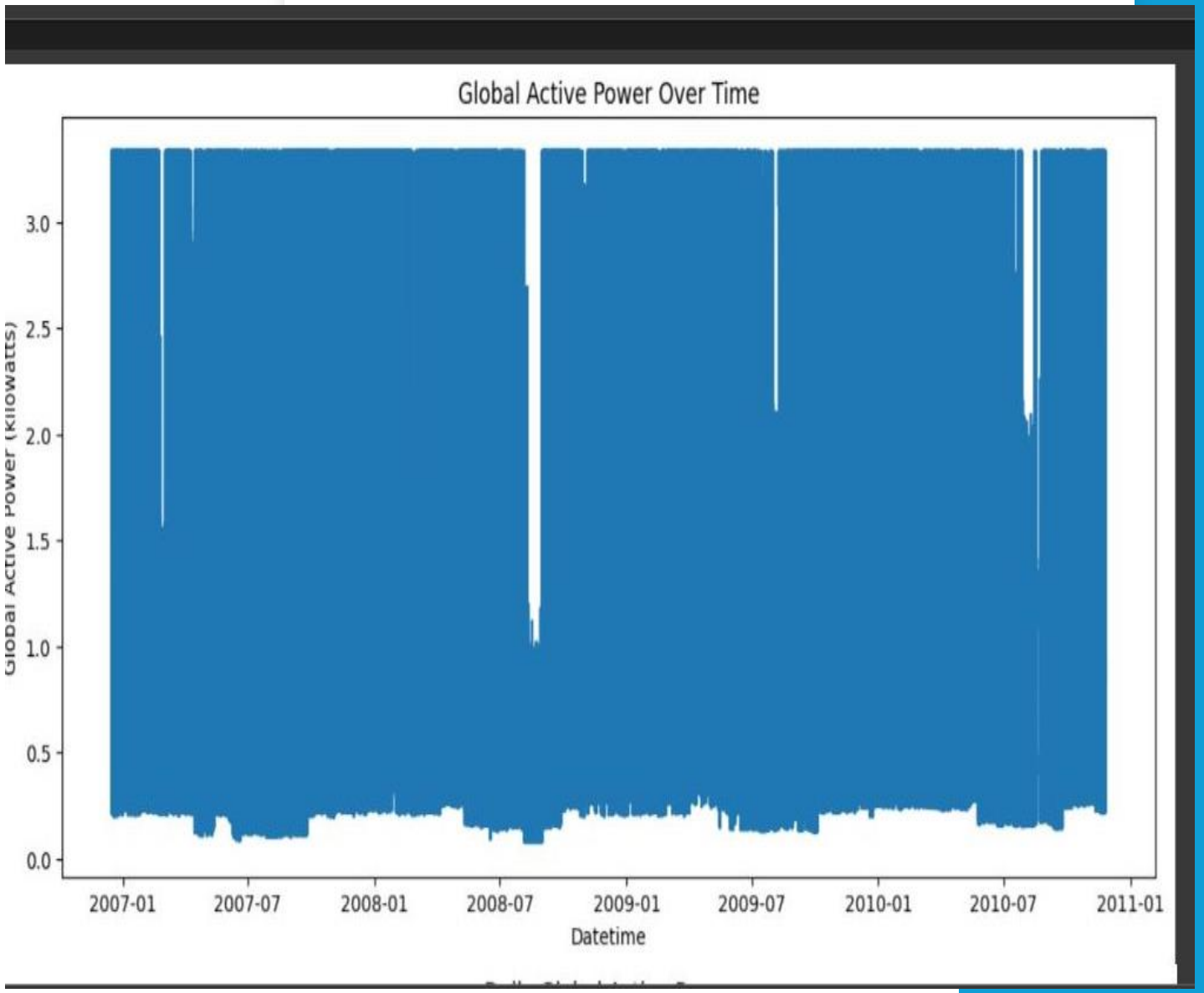
```
df.head()
```

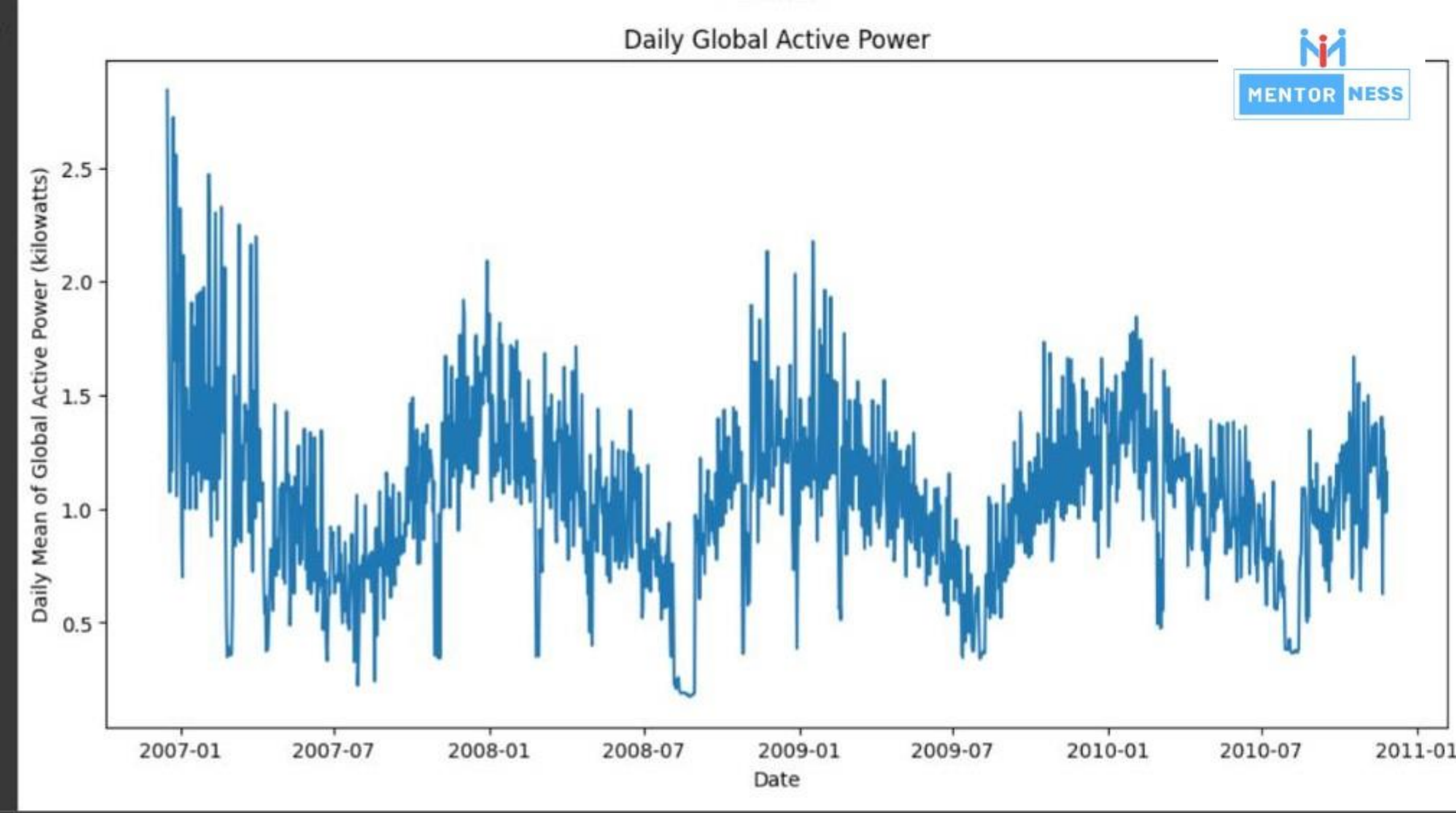
Datetime	Global_active_power	Global_reactive_power	Voltage	Global_intensity	Sub_metering_1	Sub_metering_2	Sub_metering_3
2006-12-16 17:24:00	4.216	0.418	234.84	18.4	0.0	1.0	17.0
2006-12-16 17:25:00	5.360	0.436	233.63	23.0	0.0	1.0	16.0
2006-12-16 17:26:00	5.374	0.498	233.29	23.0	0.0	2.0	17.0
2006-12-16 17:27:00	5.388	0.502	233.74	23.0	0.0	1.0	17.0
2006-12-16 17:28:00	3.666	0.528	235.68	15.8	0.0	1.0	17.0

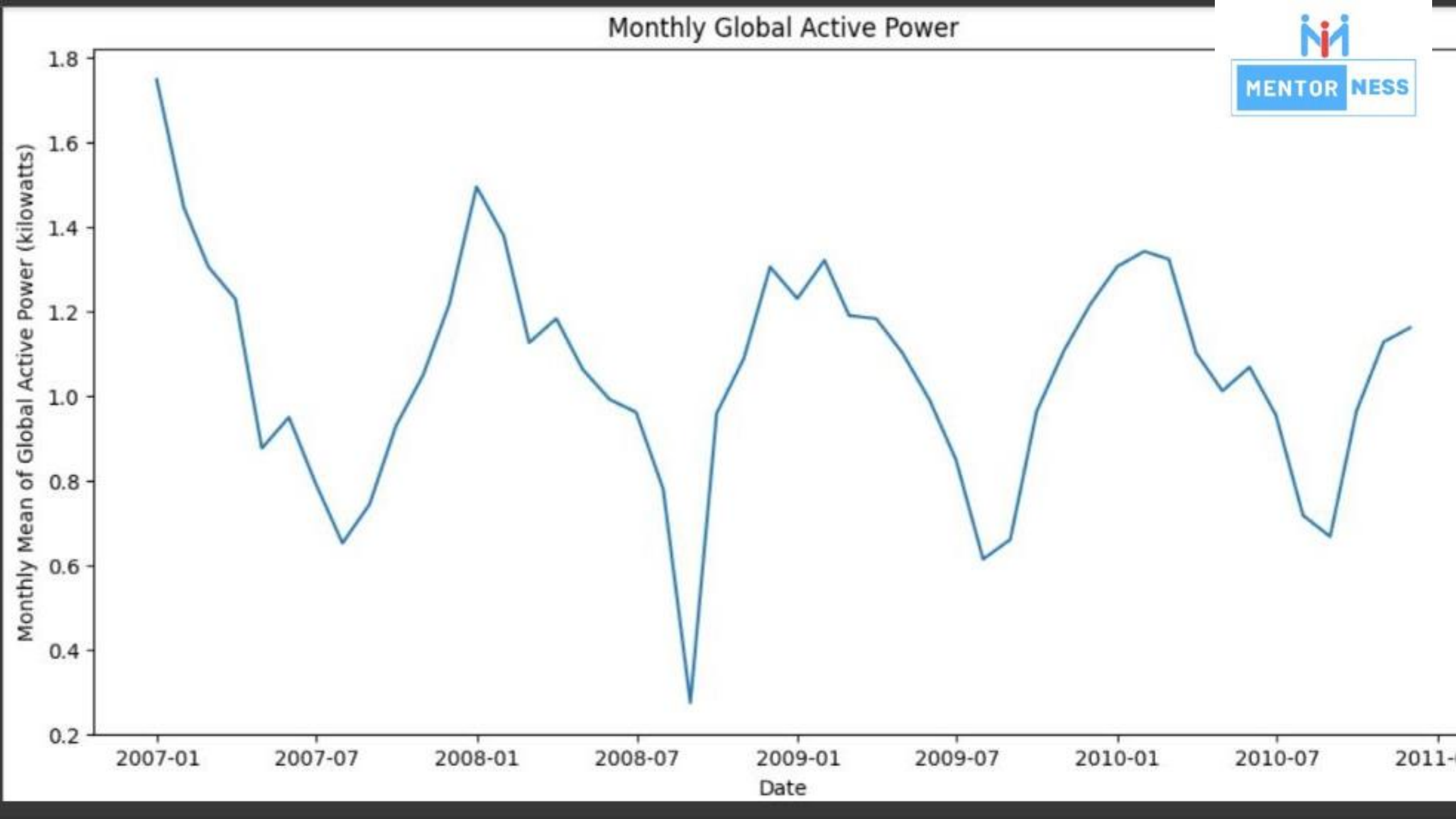
Data Preprocessing

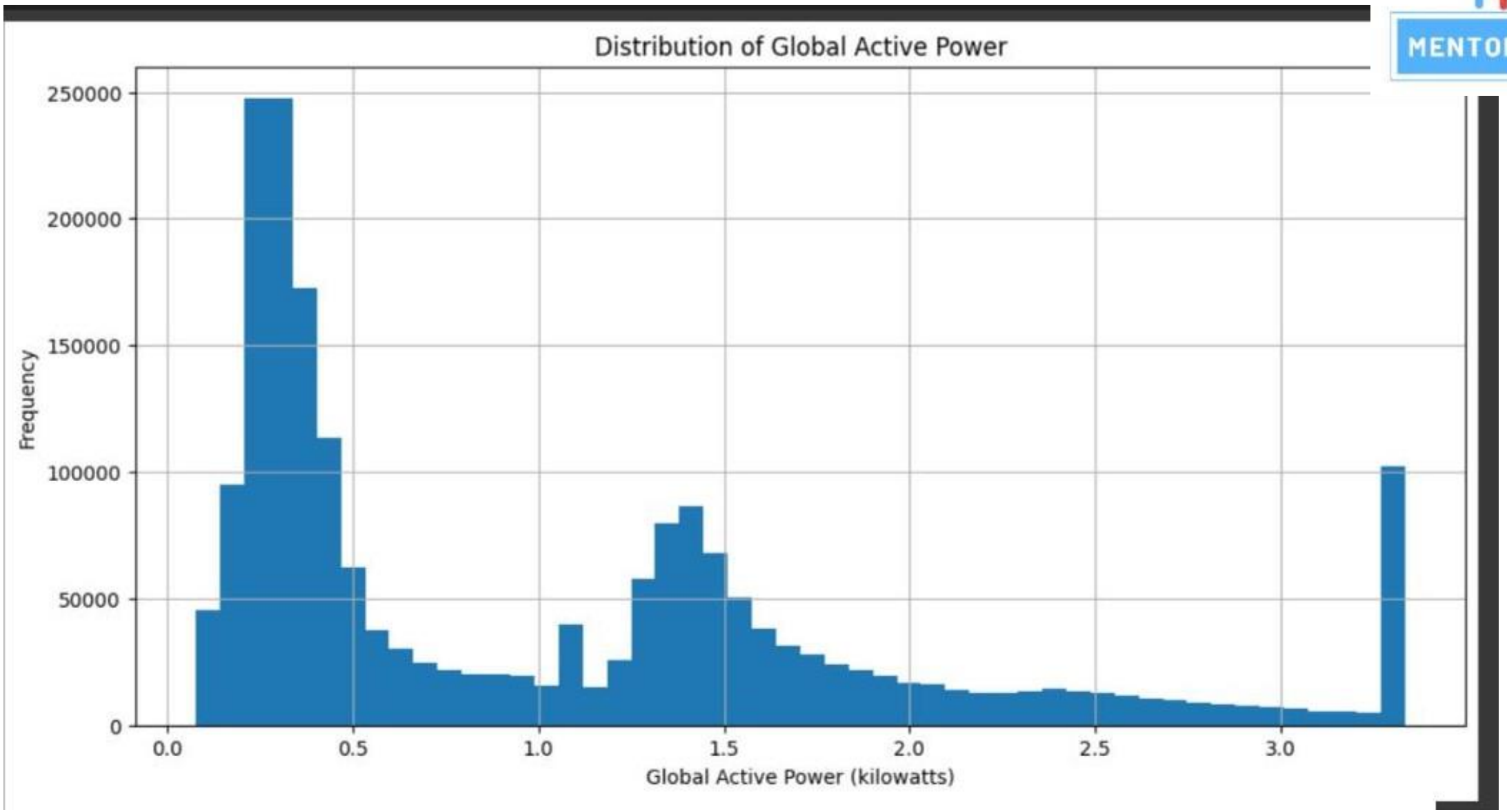
MENTOR NESS

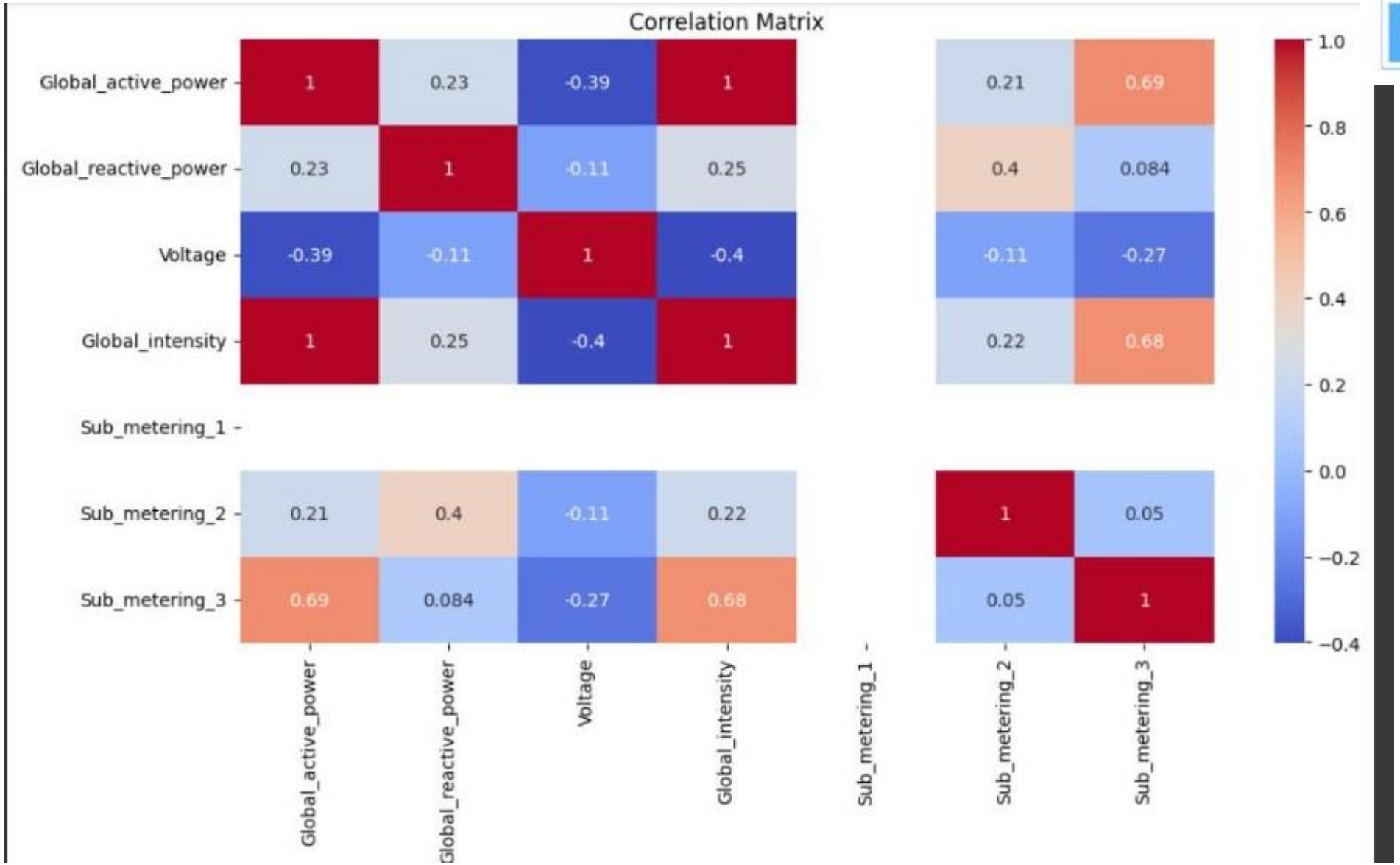
Exploratory Da Analysis (EDA)



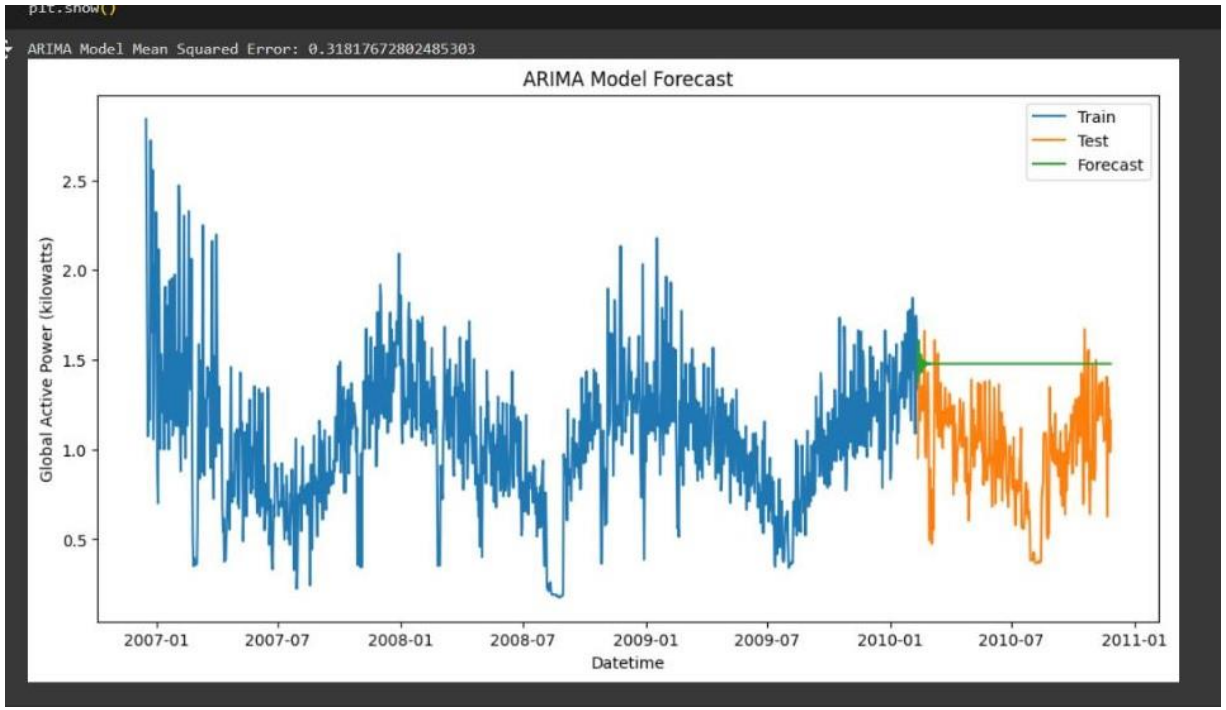

Daily Global Active Power
MENTOR NESS
Daily Mean of Global Active Power (kilowatts)
2.5
2.0
1.5
1.0
0.5
2007-01
2007-07
2008-01
2008-07
2009-01
2009-07
2010-01
2010-07
2011-01
Date


Monthly Global Active Power
Monthly Mean of Global Active Power (kilowatts)
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
2007-01
2007-07
2008-01
2008-07
2009-01
2009-07
2010-01
2010-07
Date
MENTOR NESS

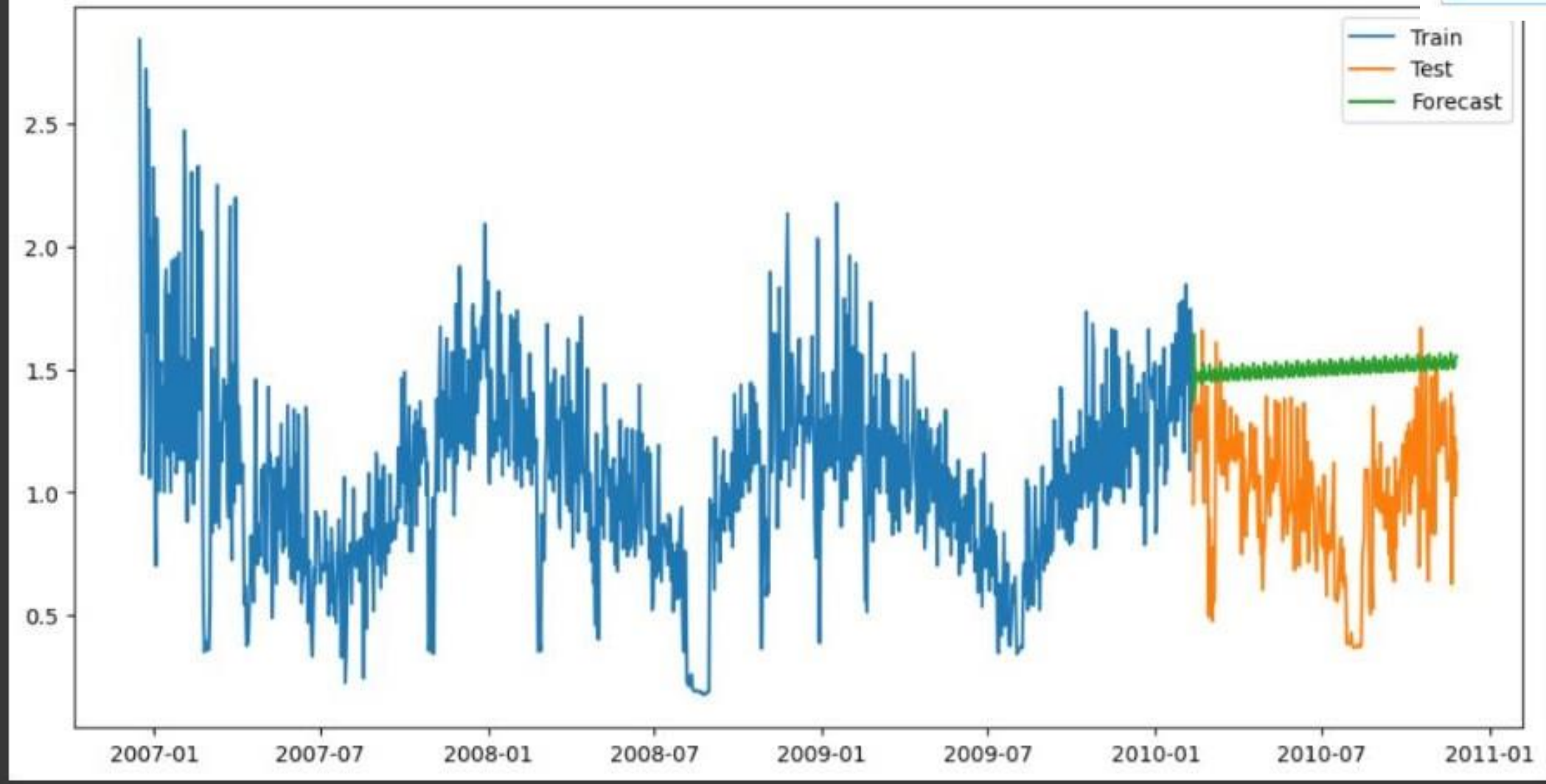

Distribution of Global Active Power
Frequency
250000
200000
150000
100000
50000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Global Active Power (kilowatts)


Correlation Matrix
Global_active_power
Global_reactive_power
Voltage
Global_intensity
Sub_metering_1
Sub_metering_2
Sub_metering_3
1
0.23
-0.39
1
0.21
0.69
0.23
1
-0.11
0.25
0.4
0.084
-0.39
-0.11
1
-0.4
-0.11
-0.27
1
0.25
-0.4
1
0.22
0.68
0.21
0.4
-0.11
0.22
1
0.05
0.69
0.084
-0.27
0.68
0.05
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

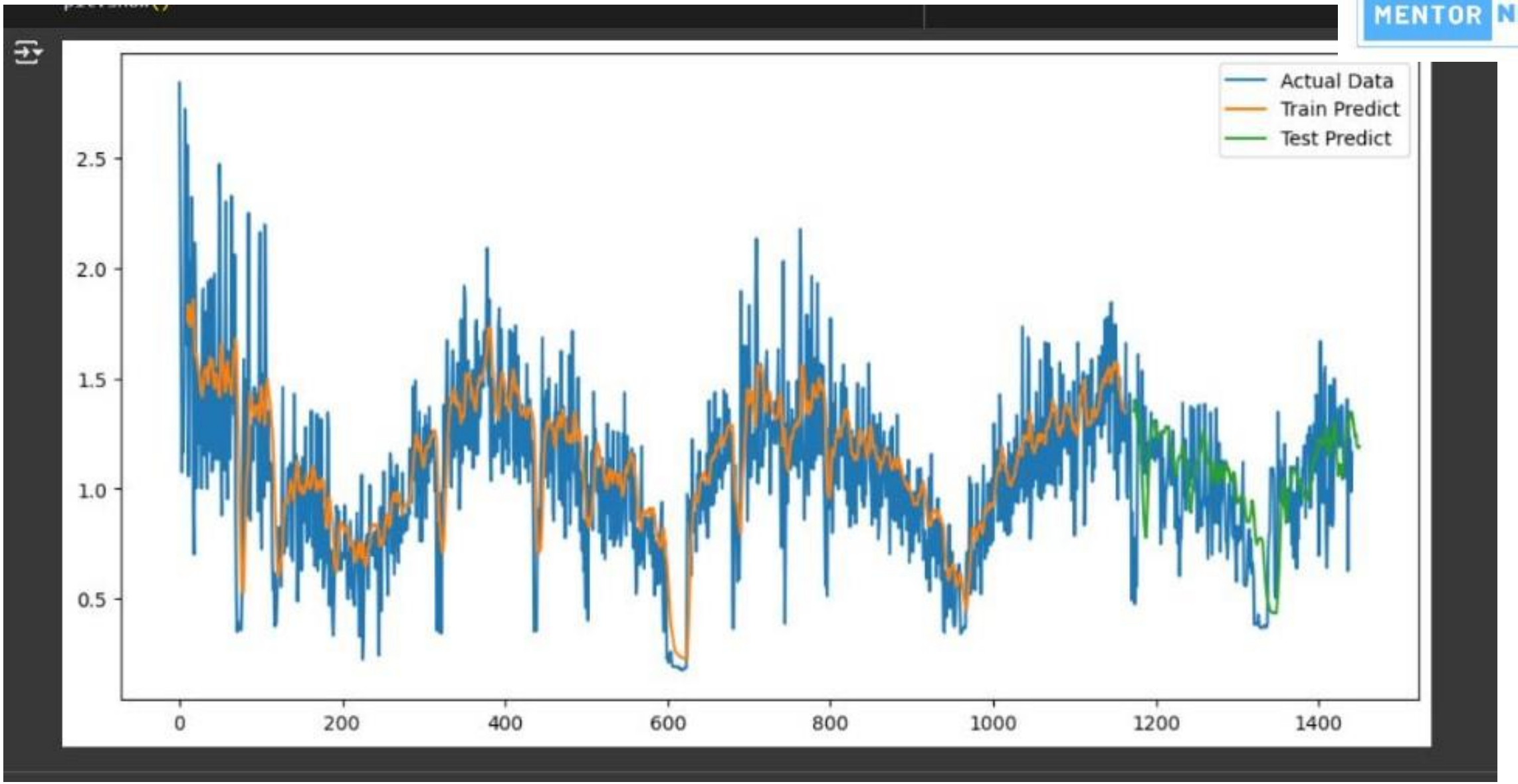
Time Series Forecasting Models



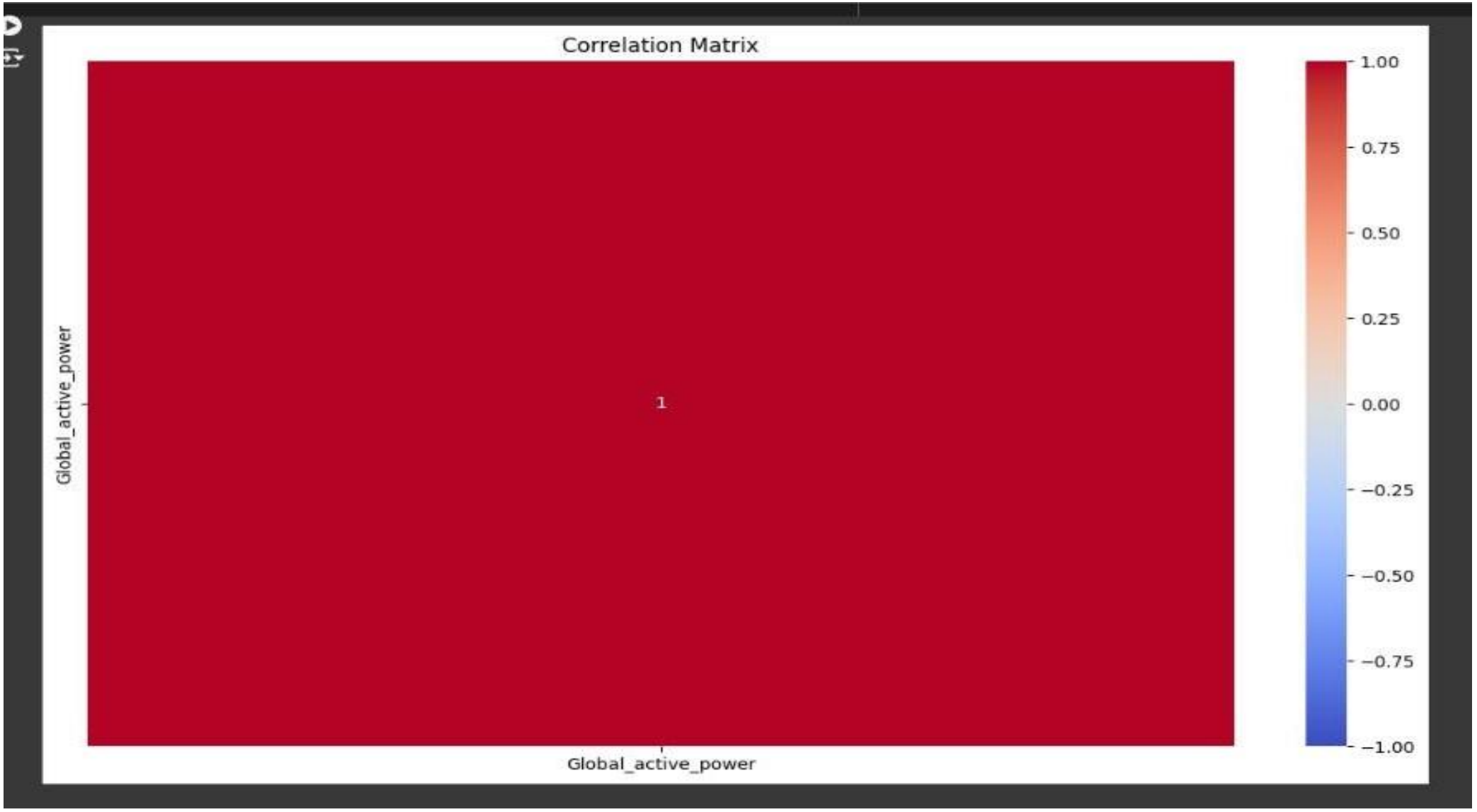
```
SARIMA Model Mean Squared Error: 0.34519292613182917
```

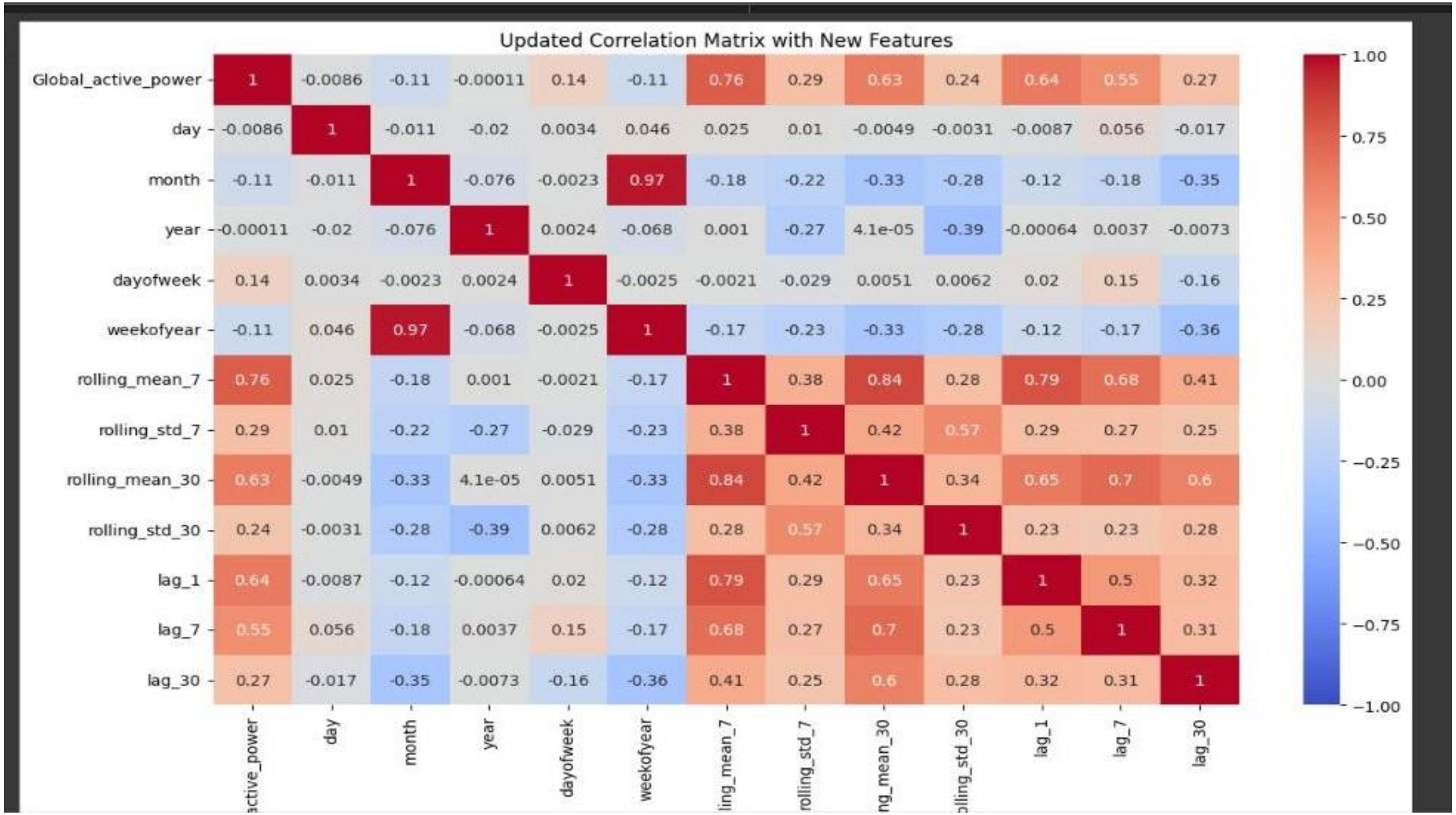


LSTM


Actual Data
Train Predict
Test Predict
2.5
2.0
1.5
1.0
0.5
0
200
400
600
800
1000
1200
1400

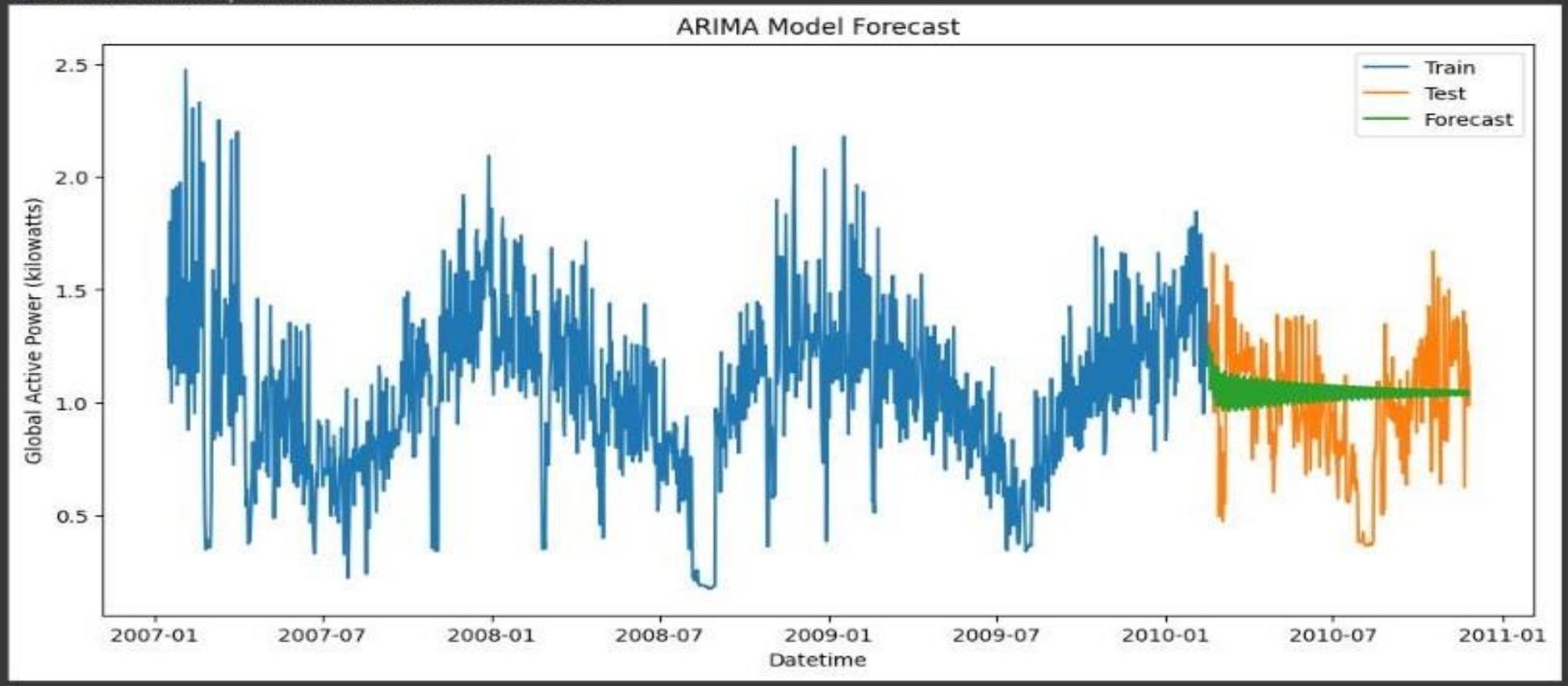
Feature Engineering



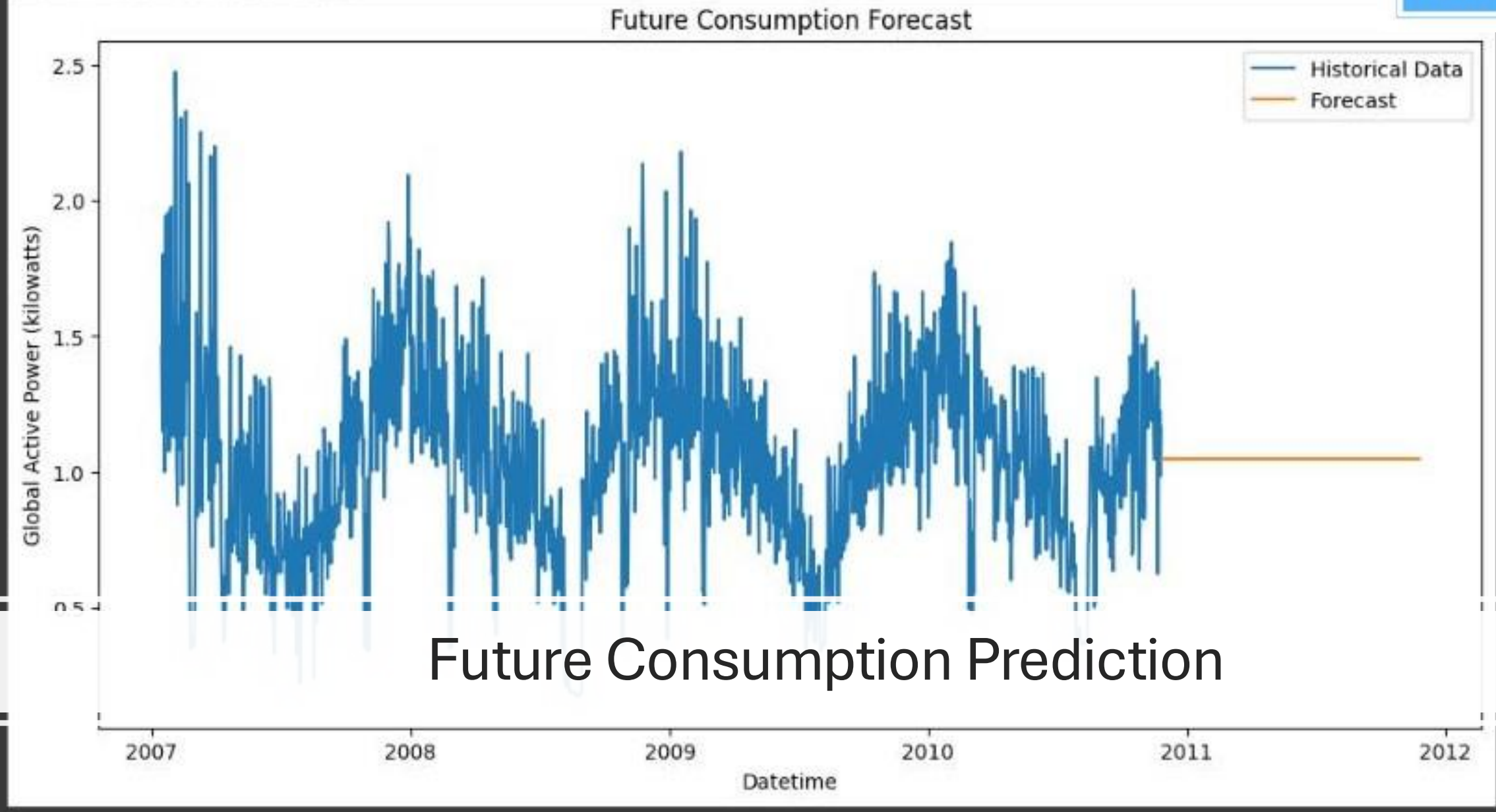

Updated Correlation Matrix with New Features
Global_active_power 1 -0.0086 -0.11 -0.00011 0.14 -0.11 0.76 0.29 0.63 0.24 0.64 0.55 0.27
day -0.0086 1 -0.011 -0.02 0.0034 0.046 0.025 0.01 -0.0049 -0.0031 -0.0087 0.056 -0.017
month -0.11 -0.011 1 -0.076 -0.0023 0.97 -0.18 -0.22 -0.33 -0.28 -0.12 -0.18 -0.35
year -0.00011 -0.02 -0.076 1 0.0024 -0.068 0.001 -0.27 4.1e-05 -0.39 -0.00064 0.0037 -0.0073
dayofweek 0.14 0.0034 -0.0023 0.0024 1 -0.0025 -0.0021 -0.029 0.0051 0.0062 0.02 0.15 -0.16
weekofyear -0.11 0.046 0.97 -0.068 -0.0025 1 -0.17 -0.23 -0.33 -0.28 -0.12 -0.17 -0.36
rolling_mean_7 0.76 0.025 -0.18 0.001 -0.0021 -0.17 1 0.38 0.84 0.28 0.79 0.68 0.41
rolling_std_7 0.29 0.01 -0.22 -0.27 -0.029 -0.23 0.38 1 0.42 0.57 0.29 0.27 0.25
rolling_mean_30 0.63 -0.0049 -0.33 4.1e-05 0.0051 -0.33 0.84 0.42 1 0.34 0.65 0.7 0.6
rolling_std_30 0.24 -0.0031 -0.28 -0.39 0.0062 -0.28 0.28 0.57 0.34 1 0.23 0.23 0.28
lag_1 0.64 -0.0087 -0.12 -0.00064 0.02 -0.12 0.79 0.29 0.65 0.23 1 0.5 0.32
lag_7 0.55 0.056 -0.18 0.0037 0.15 -0.17 0.68 0.27 0.7 0.23 0.5 1 0.31
lag_30 0.27 -0.017 -0.35 -0.0073 -0.16 -0.36 0.41 0.25 0.6 0.28 0.32 0.31 1
day
month
year
dayofweek
weekofyear
lag_1
lag_7
lag_30
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Model Evaluation and Tuning

```
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/statespace/sarimax.py:966: UserWarning: Non-stationary starting autoregressive parameters f
  warn('Non-stationary starting autoregressive parameters'
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/statespace/sarimax.py:978: UserWarning: Non-invertible starting MA parameters found. Using
  warn('Non-invertible starting MA parameters found.'
/usr/local/lib/python3.10/dist-packages/statsmodels/base/model.py:607: ConvergenceWarning: Maximum Likelihood optimization failed to converge. Chec
  warnings.warn("Maximum Likelihood optimization failed to "
ARIMA Model Mean Squared Error: 0.08013433088179023
```



```
  warnings.warn( 'Maximum Likelihood optimization failed to
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/base/tsa_model.py:836: ValueWarning: No supported index is avail
  return get_prediction_index(
/usr/local/lib/python3.10/dist-packages/statsmodels/tsa/base/tsa_model.py:836: FutureWarning: No supported index is avai
  return get_prediction_index(
```



Future Consumption Prediction

Conclusion

This project has equipped us with hands-on experience in time series analysis, forecasting, and feature engineering. It contributes to the broader goal of promoting energy-efficient practices in households, empowering them to make informed decisions about their electricity usage.

THANK YOU