Energy Consumption And Prediction

Objective of the project:

The objective of this project is to analyze and forecast household energy consumption using the Household Power Consumption dataset to derive actionable insights for efficient energy management. This involves exploring time-series patterns, understanding energy usage trends, and predicting future consumption patterns.

Project vision and mission

To empower households with actionable insights and predictive tools for efficient energy management, promoting sustainability and cost-effectiveness.

01. Analyze Patterns: Leverage time-series data to uncover trends, peak usage times, and anomalies in household power consumption.

02. Build Predictive Models: Develop accurate forecasting systems using ARIMA and Prophet models to predict future energy demands.

03. Promote Energy Efficiency: Provide data-driven recommendations to optimize energy usage, reduce waste, and support susatainable living practices.

Project Overview

01

Data Collection and Preparation

02

Data Analysis and Visualization

03

Data Encoding

04

Model Building

1

DATA PREPARATION

Dataset Description:

This dataset captures electric power consumption in a household with one-minute intervals over nearly four years (December 2006 to November 2010). It provides detailed time-series data on various electrical quantities and energy sub-metering.

Dataset Attributes:

Date, Time, Global_active_power, Global_reactive_power, Voltage, Global_intensity, Sub_metering_1, Sub_metering_2, Sub_metering_3

key steps:

01

- Loading Data:
- Import the dataset into a DataFrame for processing and exploration.
-

02

Basic Data Exploration

03

Datatype conversion

04

Finding Null Values and Unique Values

05

Handling Null Values

06

Data Encoding

2 DATA VISUALIZATION

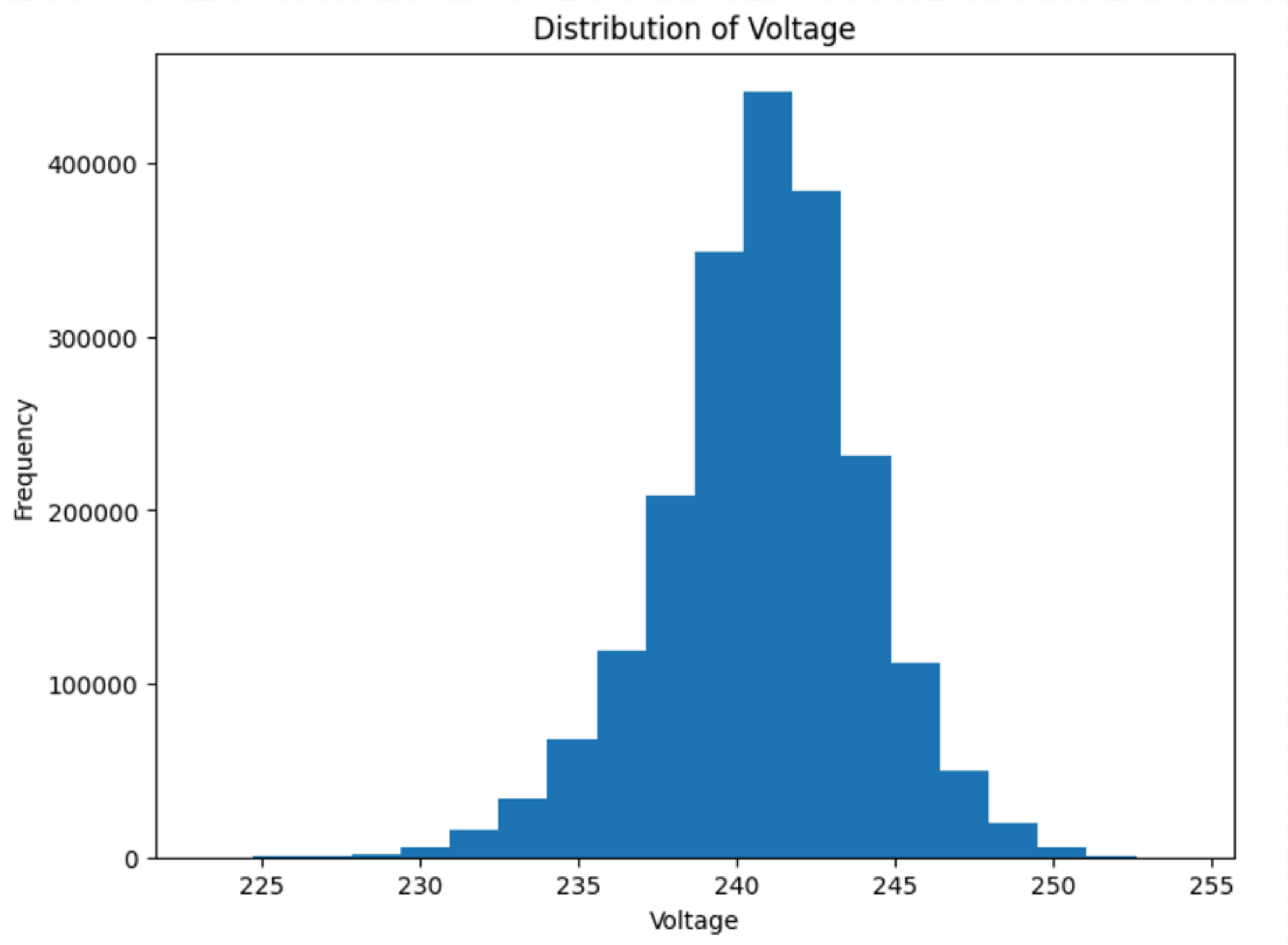
This part focuses on exploring the dataset visually to uncover meaningful patterns and relationships between variables. Through graphs and plots, we analyze energy consumption trends and distributions, helping identify key insights such as peak consumption times and seasonal variations. Additionally, data encoding ensures that time-based attributes (such as Date and Time) and categorical variables are transformed into a format suitable for machine learning models.

HISTOGRAM : Distribution of voltage values

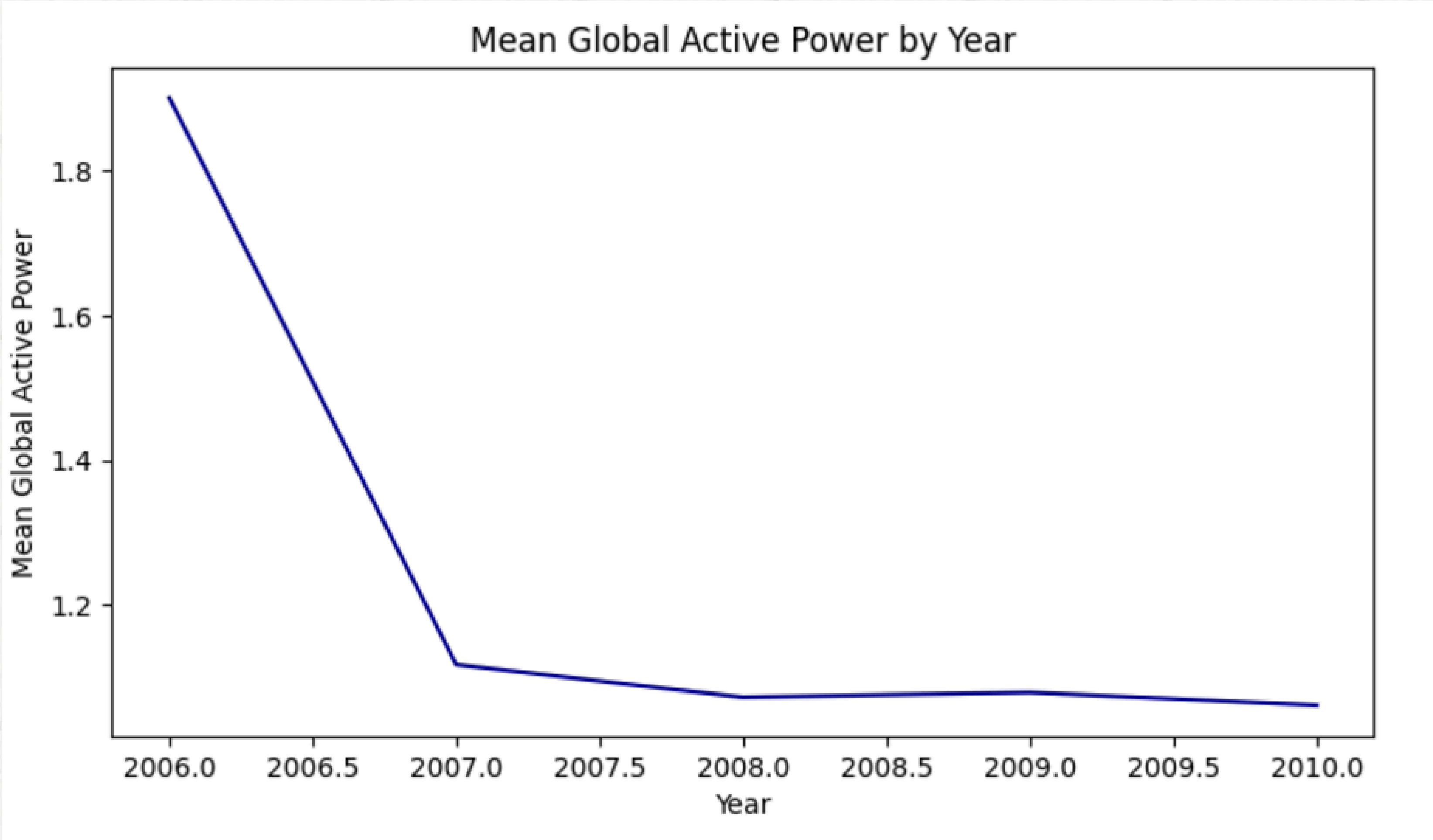
.

Key observations:

- The voltage values seem to be concentrated around a specific range, indicating that most of the readings are within a certain voltage range.
- The distribution is likely uniform with some frequency spikes, suggesting relatively consistent voltage levels in the data.
- Outliers or extreme voltage values are less frequent, as the plot appears to have a narrow spread.
- This suggests that the voltage readings are generally stable with some variation, likely representing typical household or industrial electricity usage.



Time-Series Plots for Global Active Power

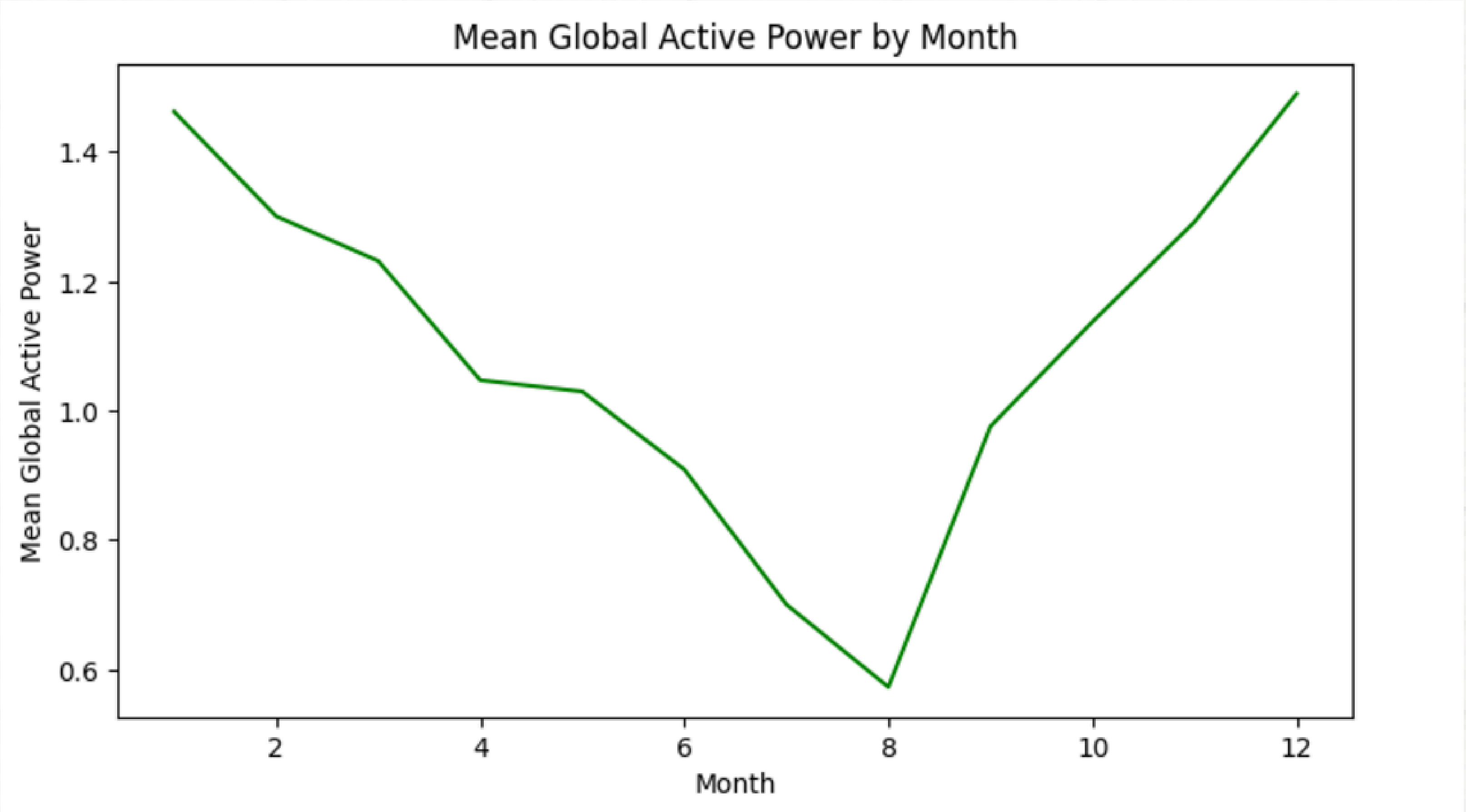


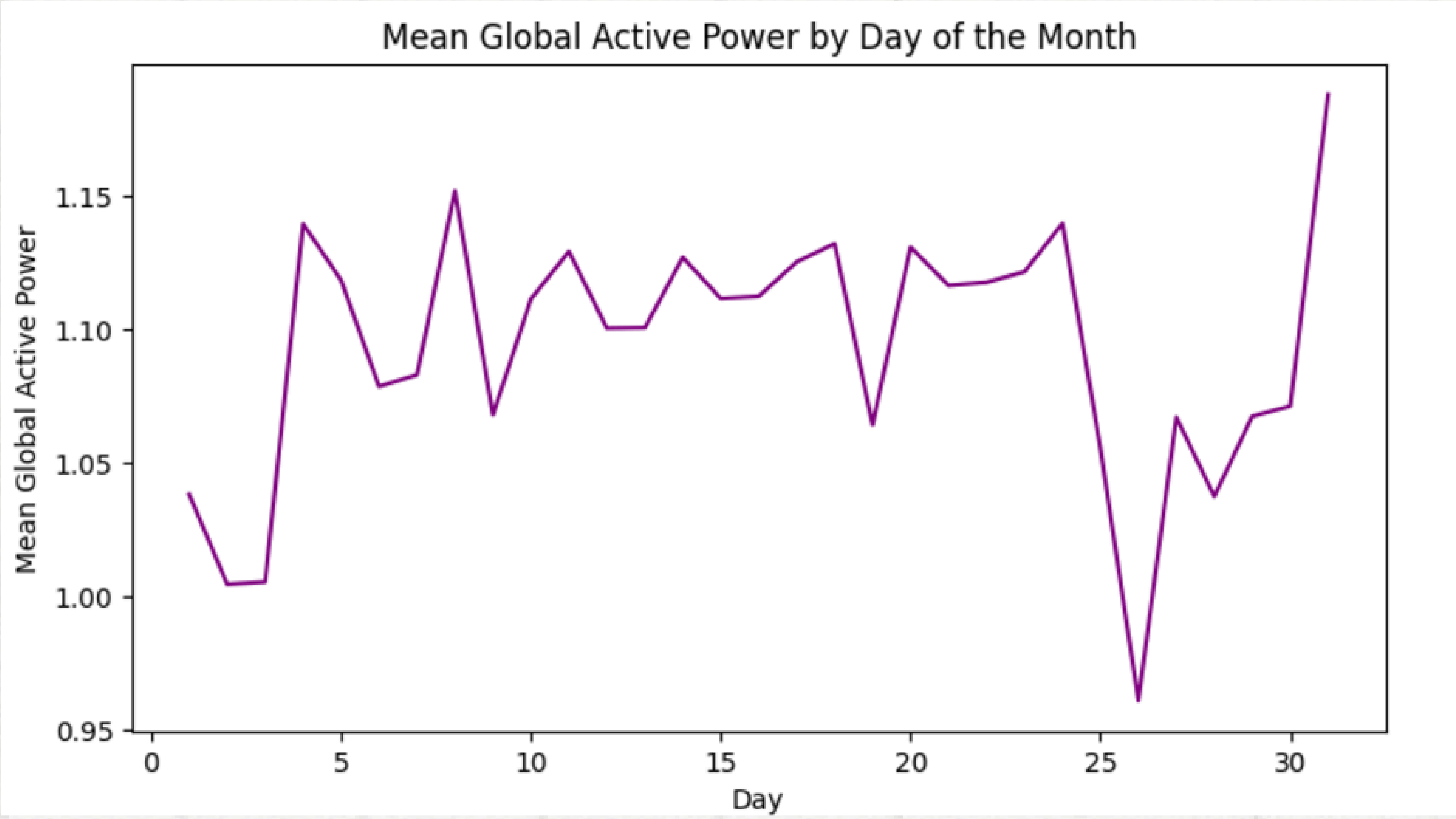
Key observations:

- **Fluctuations:** Global active power shows yearly peaks and valleys.
- **Trends:** Consumption varies, possibly due to seasonal or external factors.
- **Yearly Impact:** Certain years may experience higher demand, influenced by factors like weather or trends.

Key observations:

- **Monthly fluctuations:** Higher power usage in winter (heating) and summer (cooling).
- **Lower usage:** Spring and fall months show reduced power consumption.
- **Seasonal influence:** Energy consumption is heavily influenced by weather and temperature control needs.



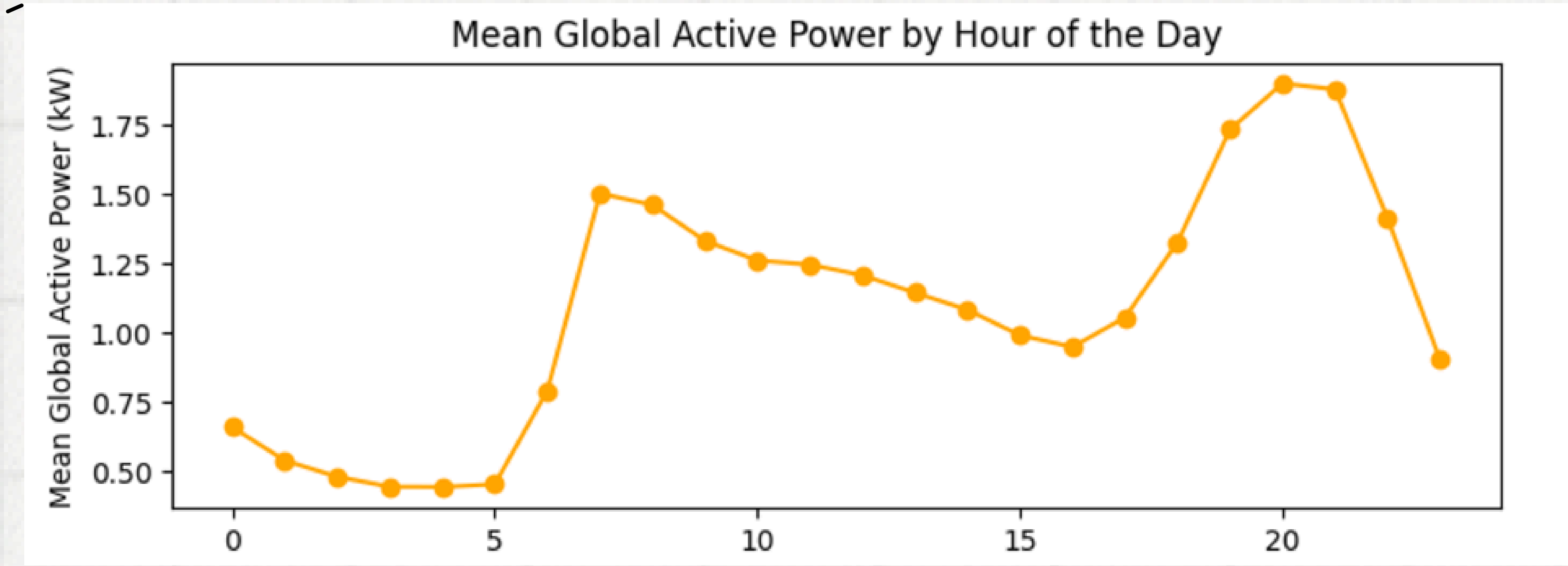


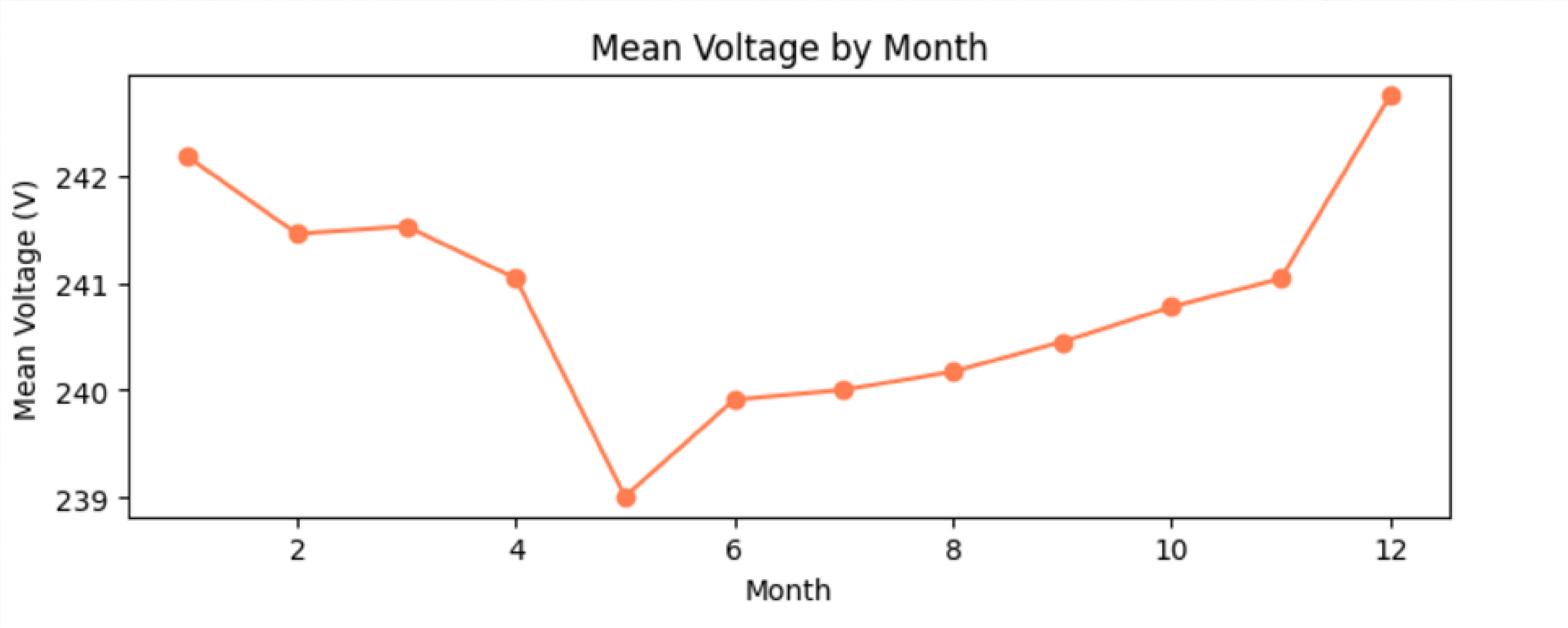
Key observations:

- Power usage fluctuates daily, with higher consumption on weekends and holidays.
- Weekdays show consistent usage, influenced by work and routine activities.
- Irregular spikes may indicate special events or anomalies in consumption.

Key observations:

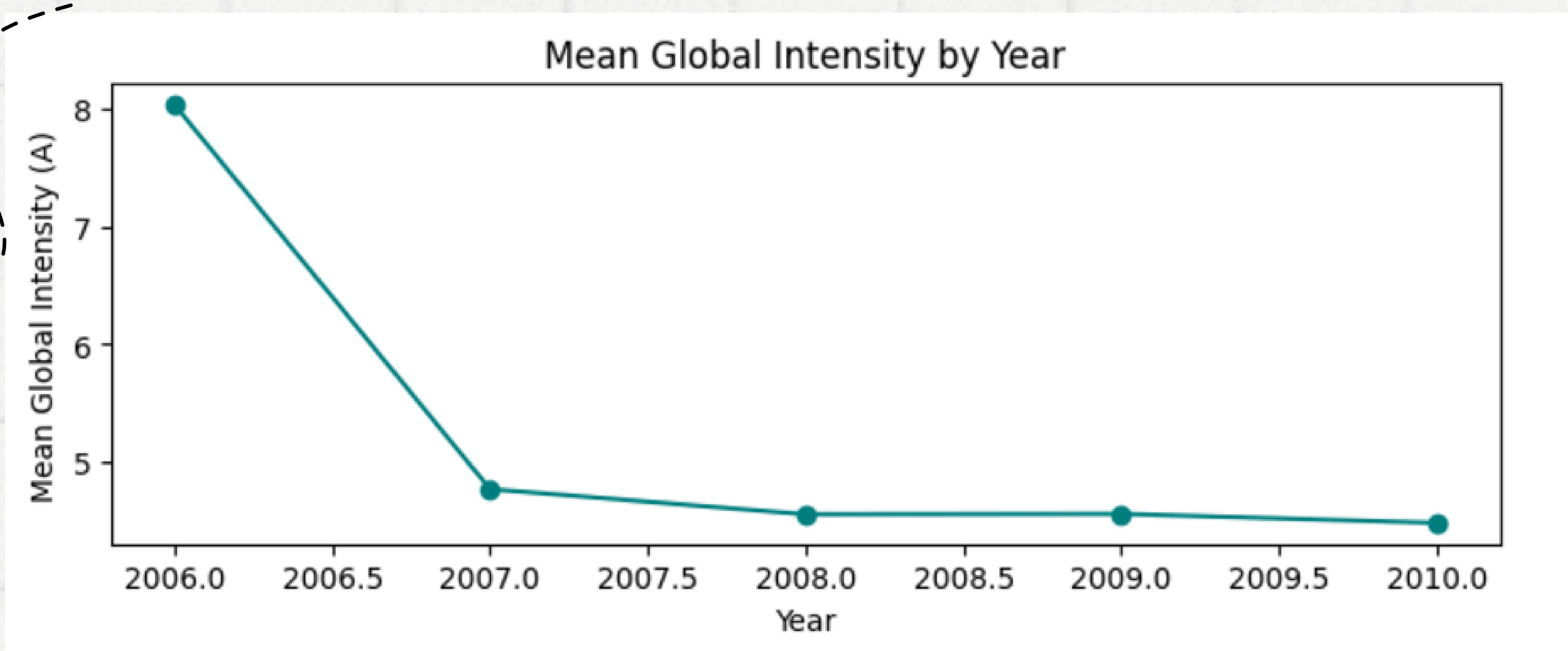
- Peak usage occurs in the early morning and evening due to daily activities.
- Lower usage happens late at night or early morning when activities are minimal.
- This analysis helps identify peak hours for better energy management and load balancing.

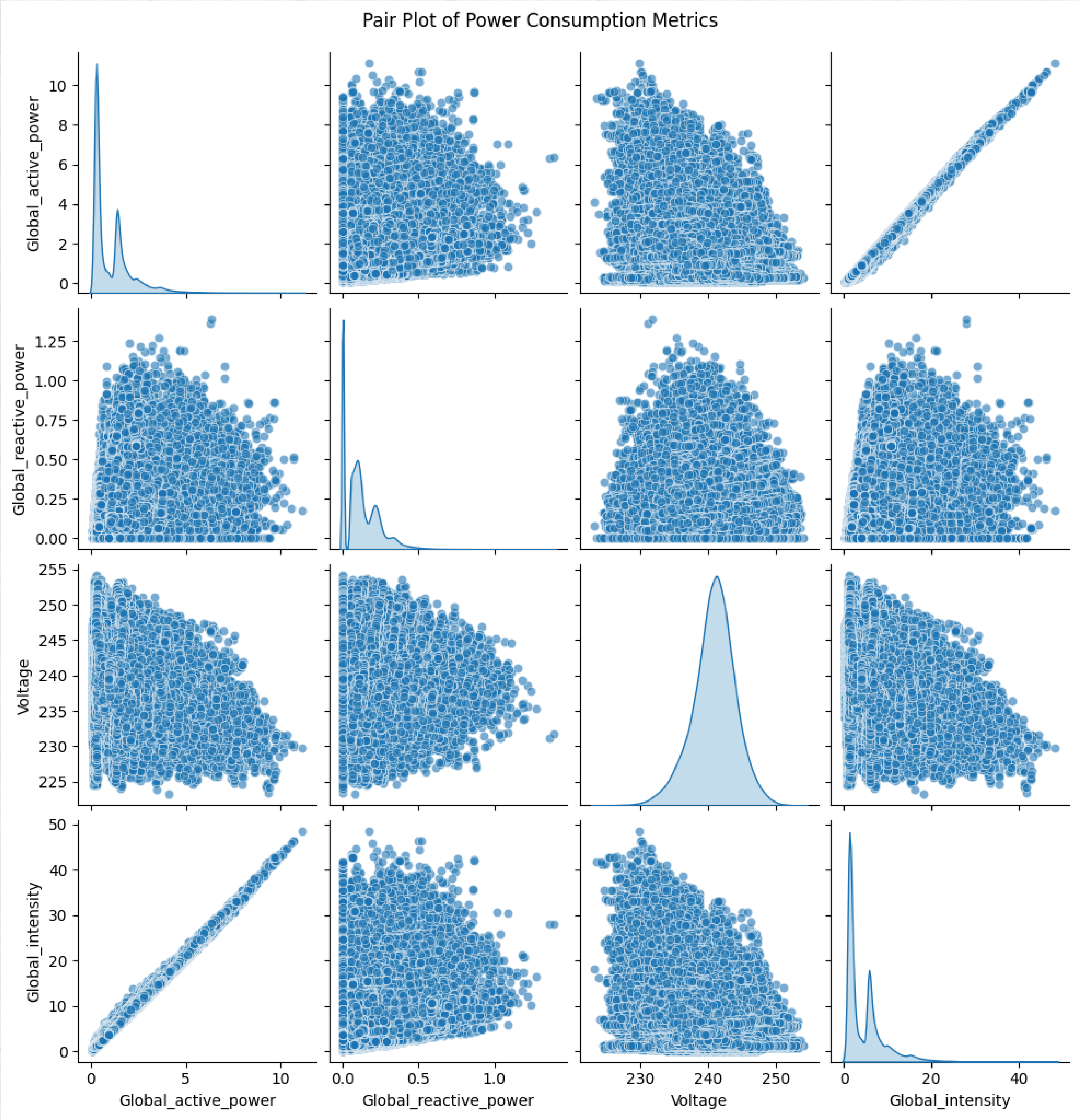




Key observations:
- Global intensity shows fluctuations over the years, indicating changes in power consumption patterns.
- Variations in intensity suggest shifts in electricity demand and grid performance.
- This analysis helps understand demand peaks and the power distribution system's health.

Key observations:
- Global intensity trends vary yearly, reflecting shifts in power consumption or grid performance.
- Markers highlight yearly fluctuations, aiding in understanding demand patterns and grid health.





Pair Plot for Global Active Power and Voltage

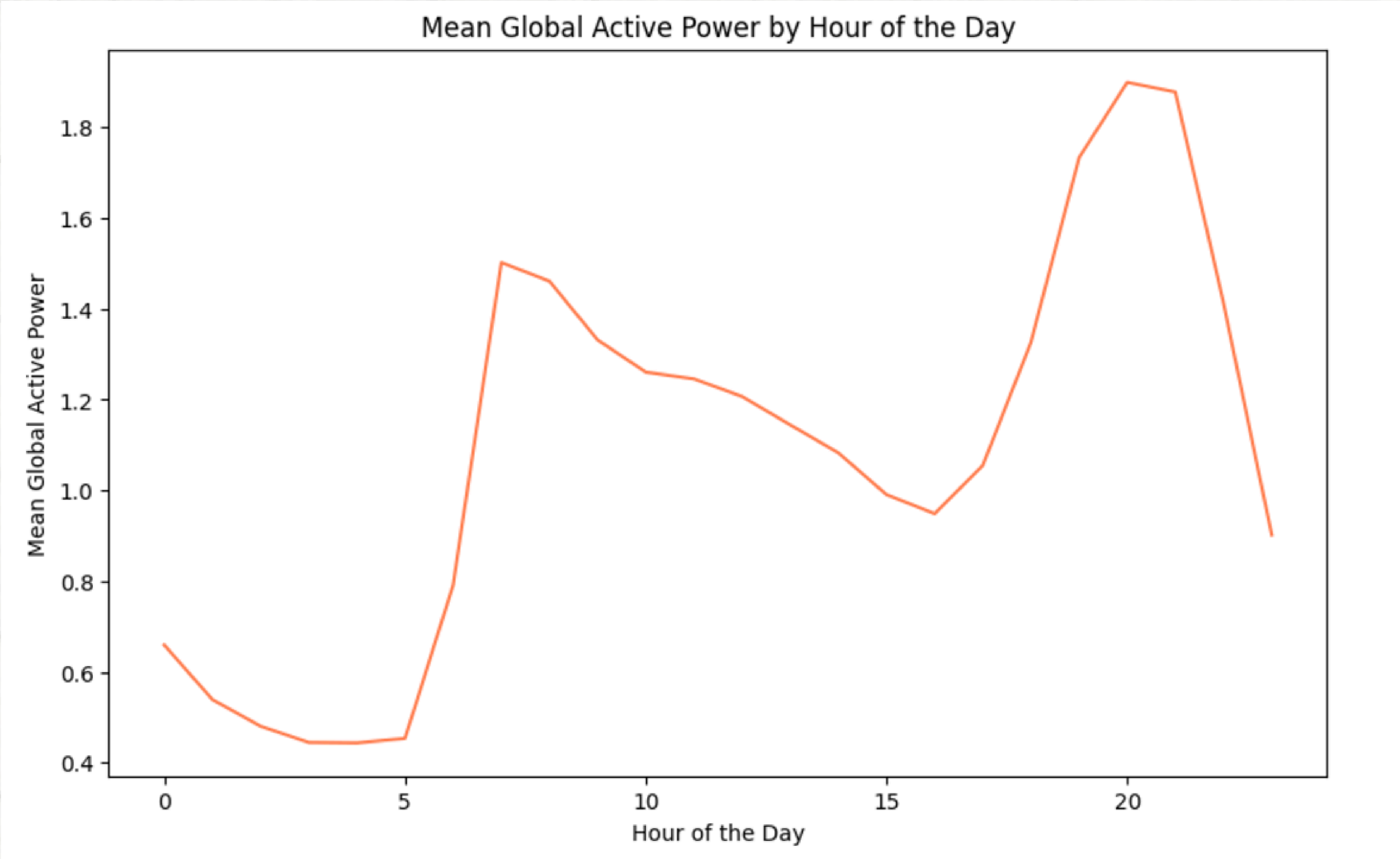
Key observations:

- KDE plots show data distribution for each feature.
- Scatter plots reveal correlations and outliers.
- Global Intensity correlates with Global Active Power.
- Weak correlation between Active and Reactive Power.
- Voltage remains stable, while power metrics vary more.

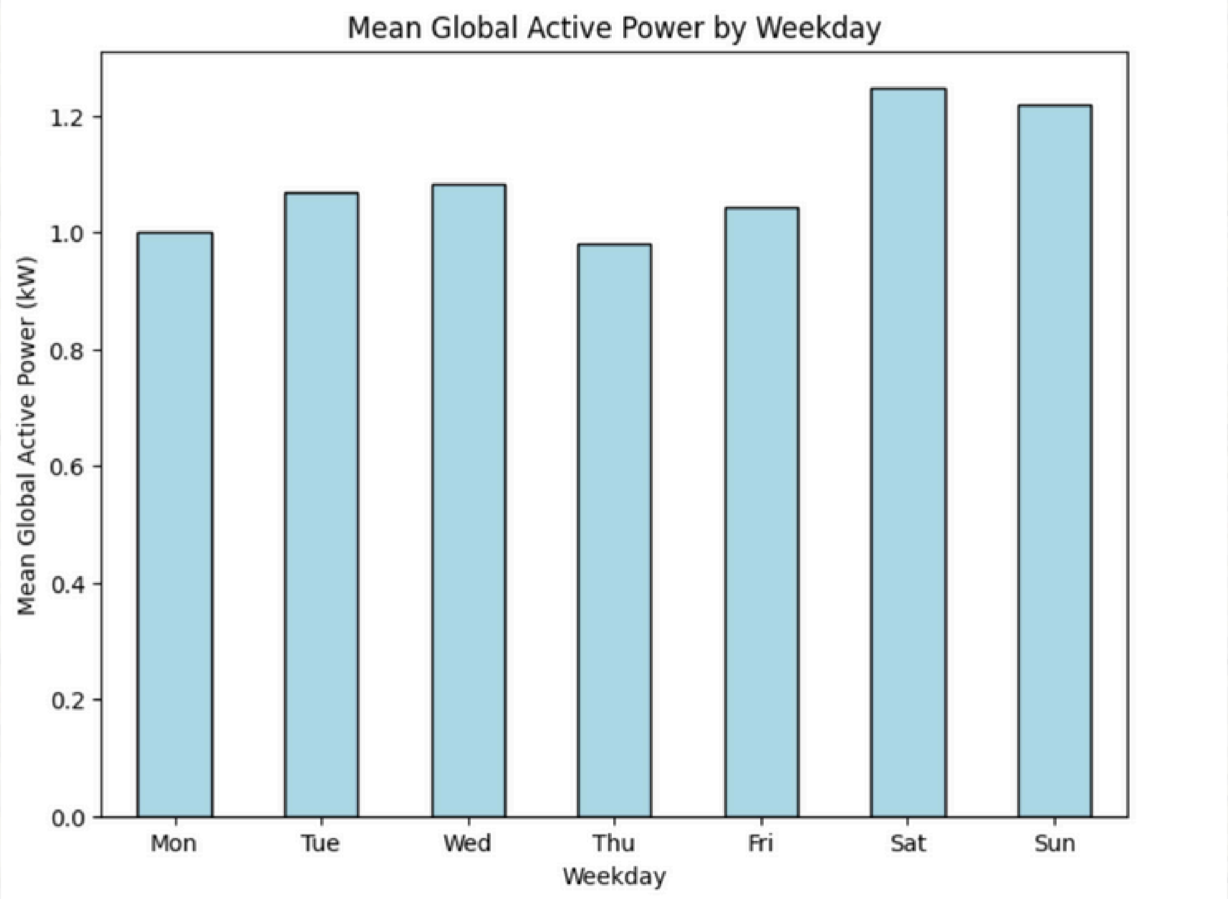
Line Plot of Mean Global Active Power by Hour

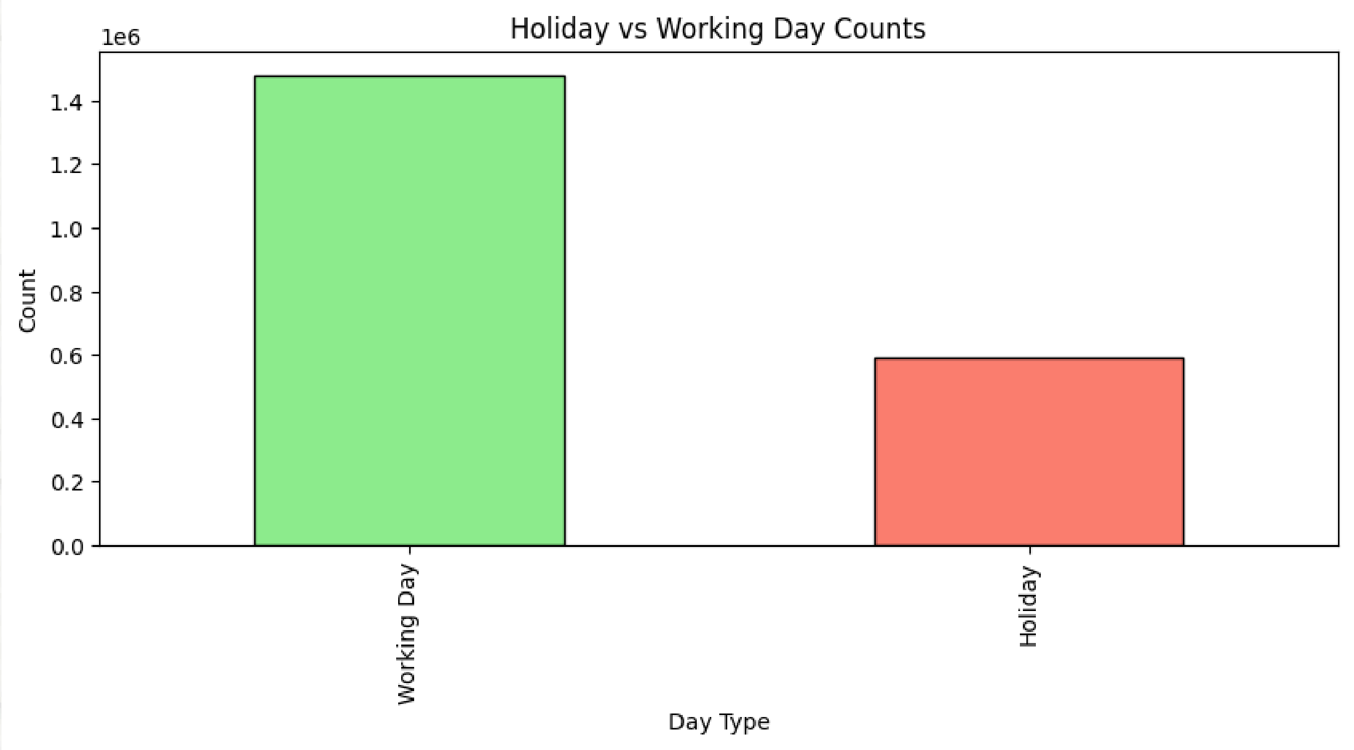
The plot shows hourly power trends:

- Low consumption at night (midnight to early morning).
- Peak usage during daytime (morning to evening).
- Drop in late evening as activities wind down.
- Useful for identifying peak demand times.

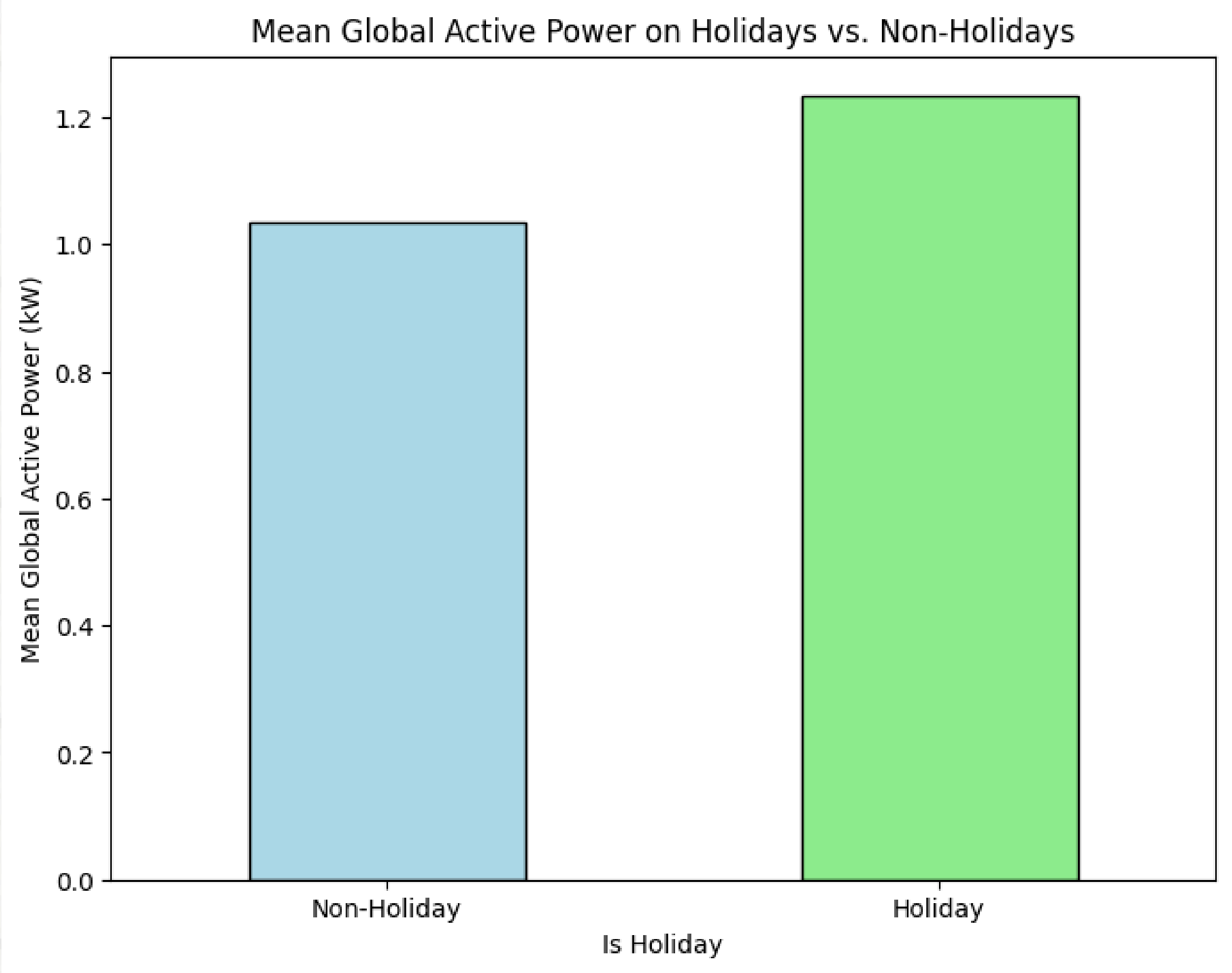


Mean Global Active Power

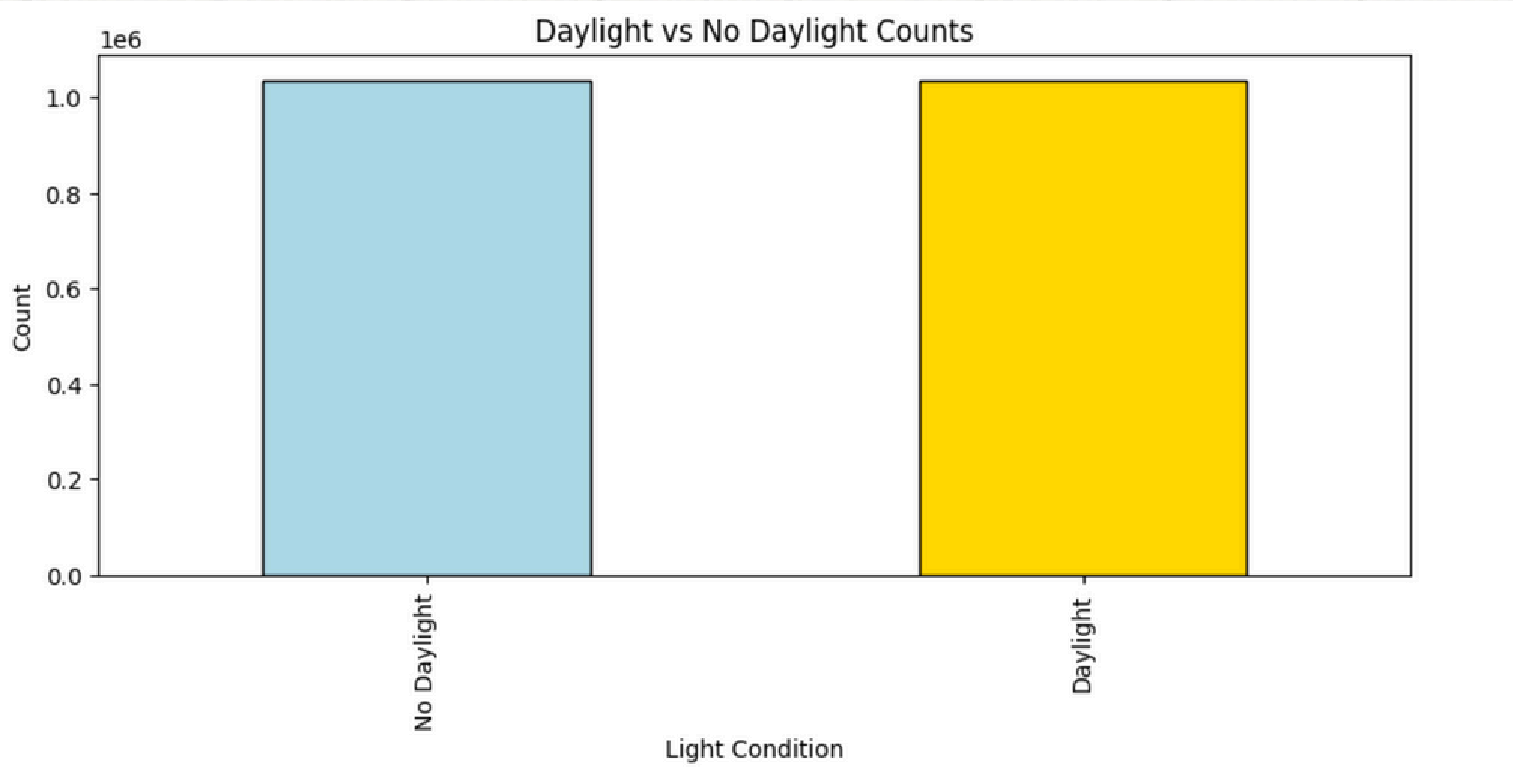




- Holidays vs. Non-Holidays: Power consumption may differ, with holidays showing higher or lower usage depending on activities.
- Weekday Trends: Weekdays often reflect consistent patterns due to routine activities, while weekends might show variations.
- Energy Insights: The analysis reveals how day types influence energy use, aiding in better demand planning.



Checking sunlight

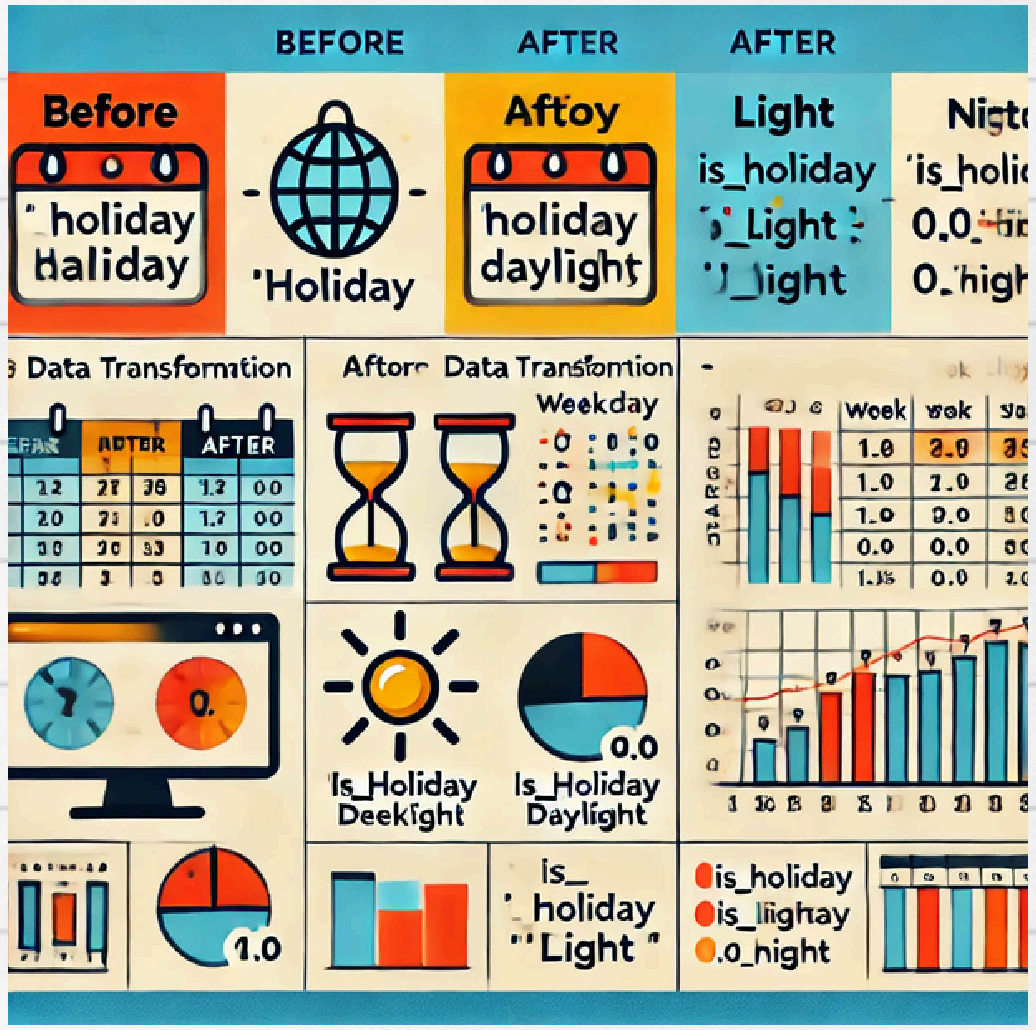


- The plot compares daylight and no daylight occurrences.
- Light Blue bar represents daylight, Gold bar represents no daylight.
- Highlights the frequency of daylight versus nighttime instances.

3 DATA ENCODING

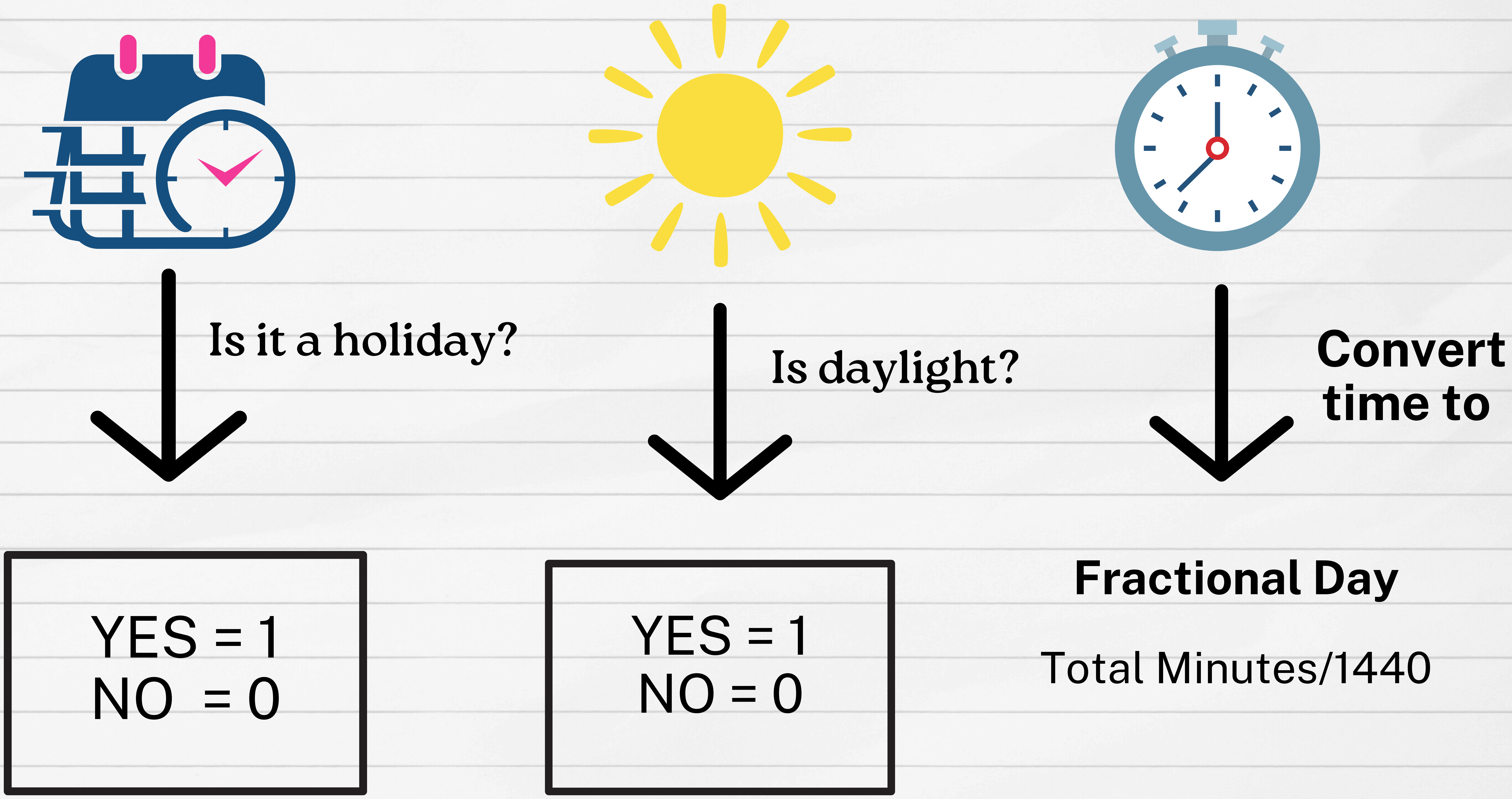
Objective of Data Encoding

- Simplifies categorical and time-related data for machine learning and analysis.
- Prepares features for identifying patterns like holidays, daylight hours, and time trends.



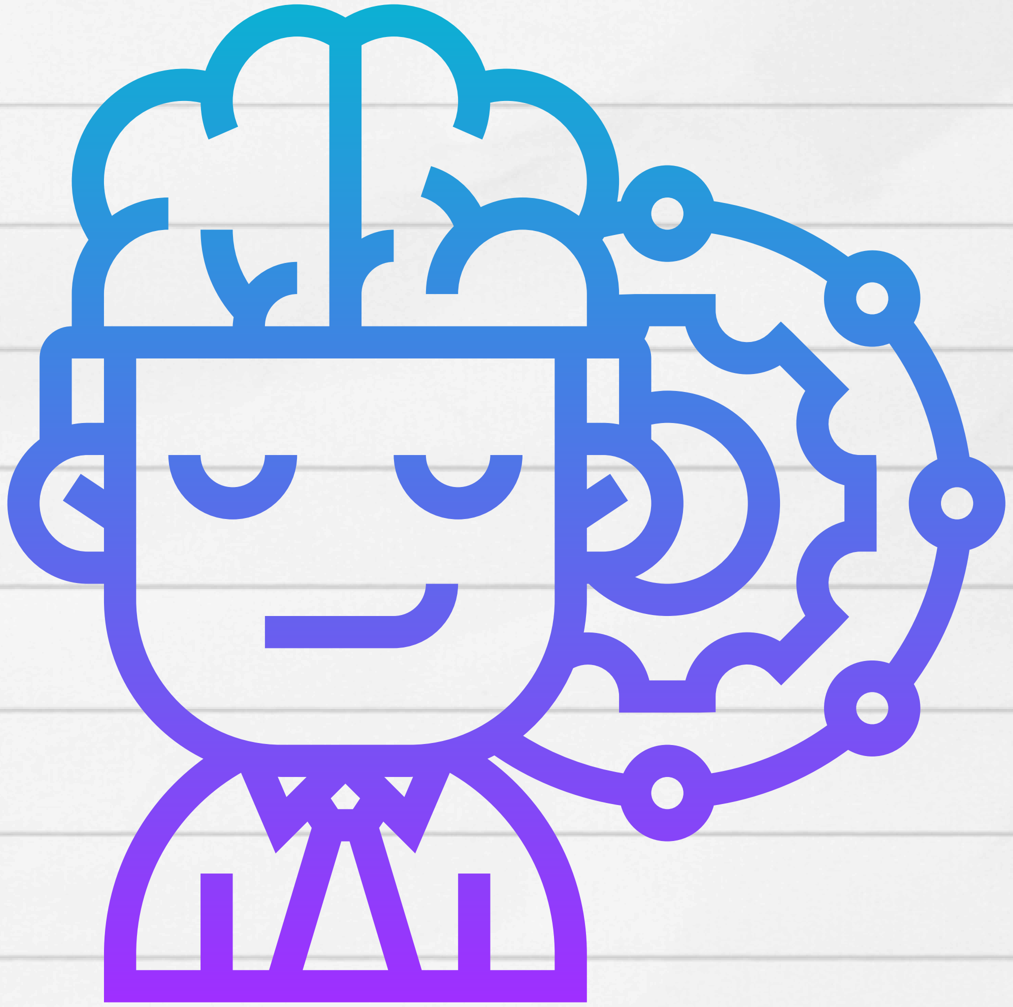
Overview of Encoded Features

- Is_holiday: Marks weekends (Saturday, Sunday) as holidays (1.0) and weekdays as non-holidays (0.0).
- Light: Labels daylight hours (06:00 to 17:59) as 1, night hours as 0.
- Time: Converts time into a fractional representation of the 24-hour day.


Is it a holiday?
YES = 1
NO = 0
Is daylight?
YES = 1
NO = 0
Convert time to
Fractional Day
Total Minutes/1440

4 MACHINE LEARNING MODEL BUILDING

Machine Learning (ML) models are algorithms that enable computers to learn patterns from data and make predictions or decisions without being explicitly programmed.



Common models

- Linear Regression
- lasso Regression
- Rigid Regression

MODEL TRAINING

MODEL EVALUATION

Model Performance Summary: RMSE and Accuracy

Model	RMSE	Accuracy (R^2)
Linear Regression	0.0402547211826	0.998532718724319
Lasso Regression	0.2330151550891	0.950835903130767
Ridge Regression	0.0402547212425	0.998532718719626

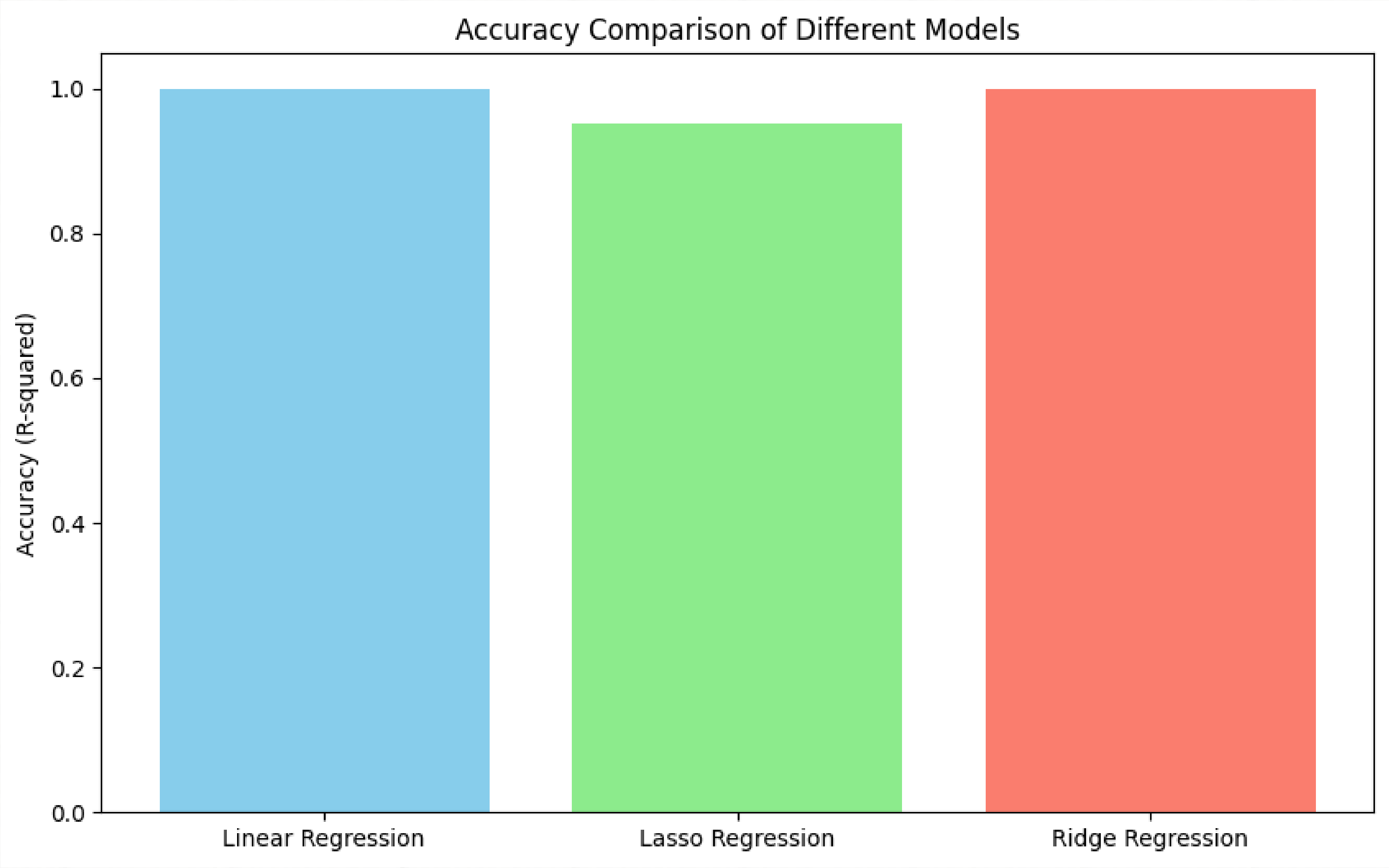
1. RMSE (Root Mean Squared Error):

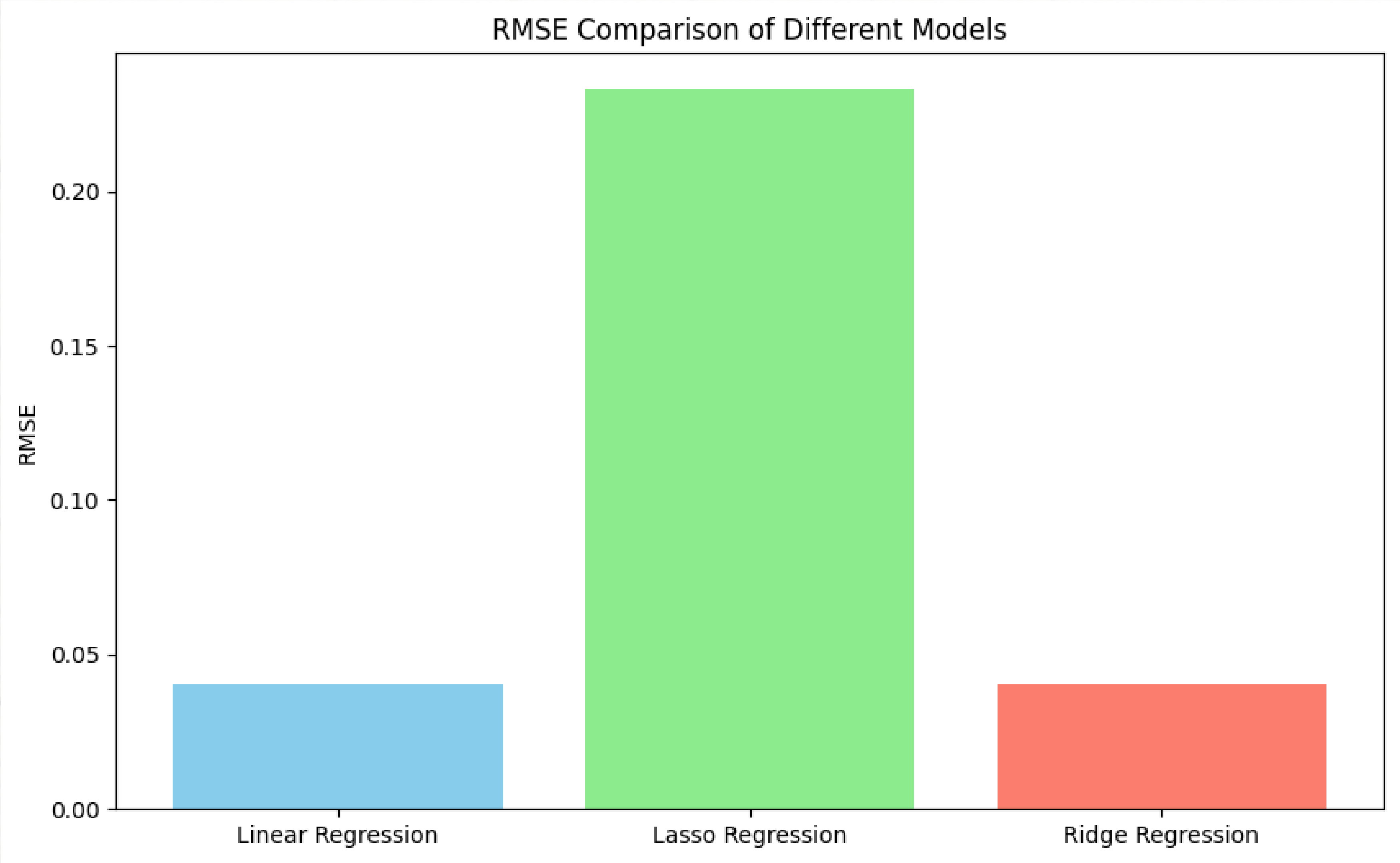
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

2. R^2 (R-squared):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

PLOT VISUALISATION





```
Linear Regression - RMSE: 0.04025472118256601 Accuracy (R^2): 0.9985327187243319
Lasso Regression - RMSE: 0.23301515508914317 Accuracy (R^2): 0.9508359031307767
Ridge Regression - RMSE: 0.04025472124250169 Accuracy (R^2): 0.9985327187199626
```

Observations:

- **Linear & Ridge Regression:** Both exhibit high accuracy ($R^2 \approx 0.9985$) and low RMSE (~0.040), indicating similar and strong performance.
- **Lasso Regression:** Lower accuracy ($R^2 \approx 0.9508$) and higher RMSE (~0.233), likely due to regularization penalizing coefficients.

Conclusion: Linear and Ridge Regression are optimal for this dataset, offering the best balance of accuracy and error.

ARIMA MODEL

ARIMA is a statistical model used for time series forecasting.

The **ARIMA (AutoRegressive Integrated Moving Average)** model is built on three components:

1. **AR (AutoRegressive)**: Models the relationship between a value and its past values (lags).

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

2. **I (Integrated)**: Makes the series stationary by differencing it d-times (subtracting current values from previous ones).

$$y'_t = y_t - y_{t-1}$$

3. **MA (Moving Average)**: Models the relationship between a value and past forecast errors (residuals).

$$y_t = \mu + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Combined, the ARIMA model is represented as $\text{ARIMA}(p, d, q)$, where:

- p: Number of AR terms.
- d: Degree of differencing for stationarity.
- q: Number of MA terms.

PROPHET MODEL

- The Prophet model, developed by Facebook, is an additive time series forecasting tool.
- Handles datasets with strong seasonality, long-term trends, and missing data.

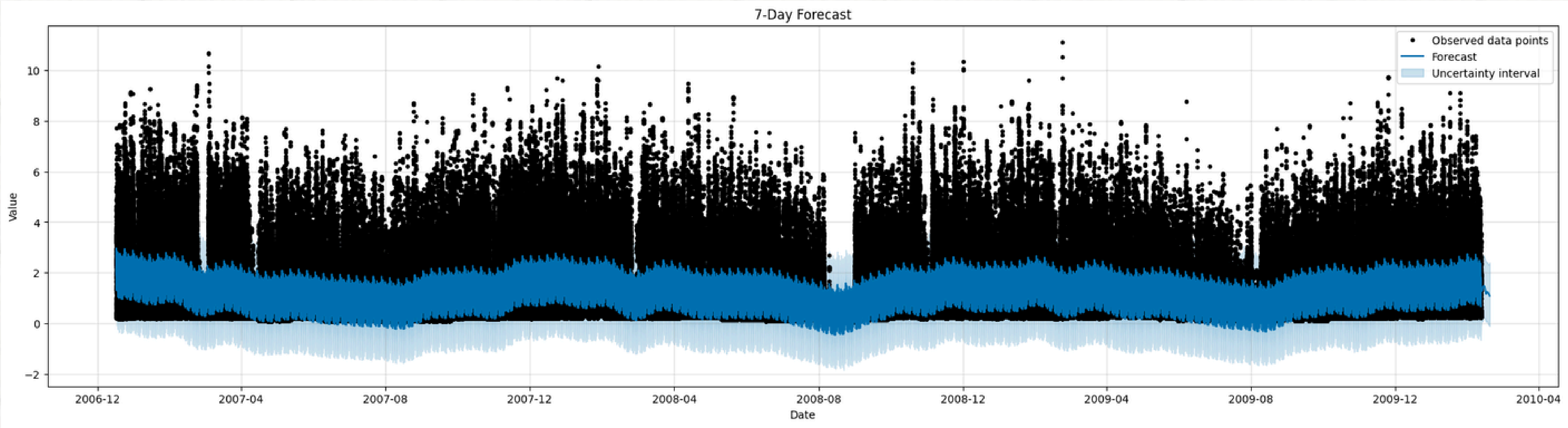
Prophet Equation:

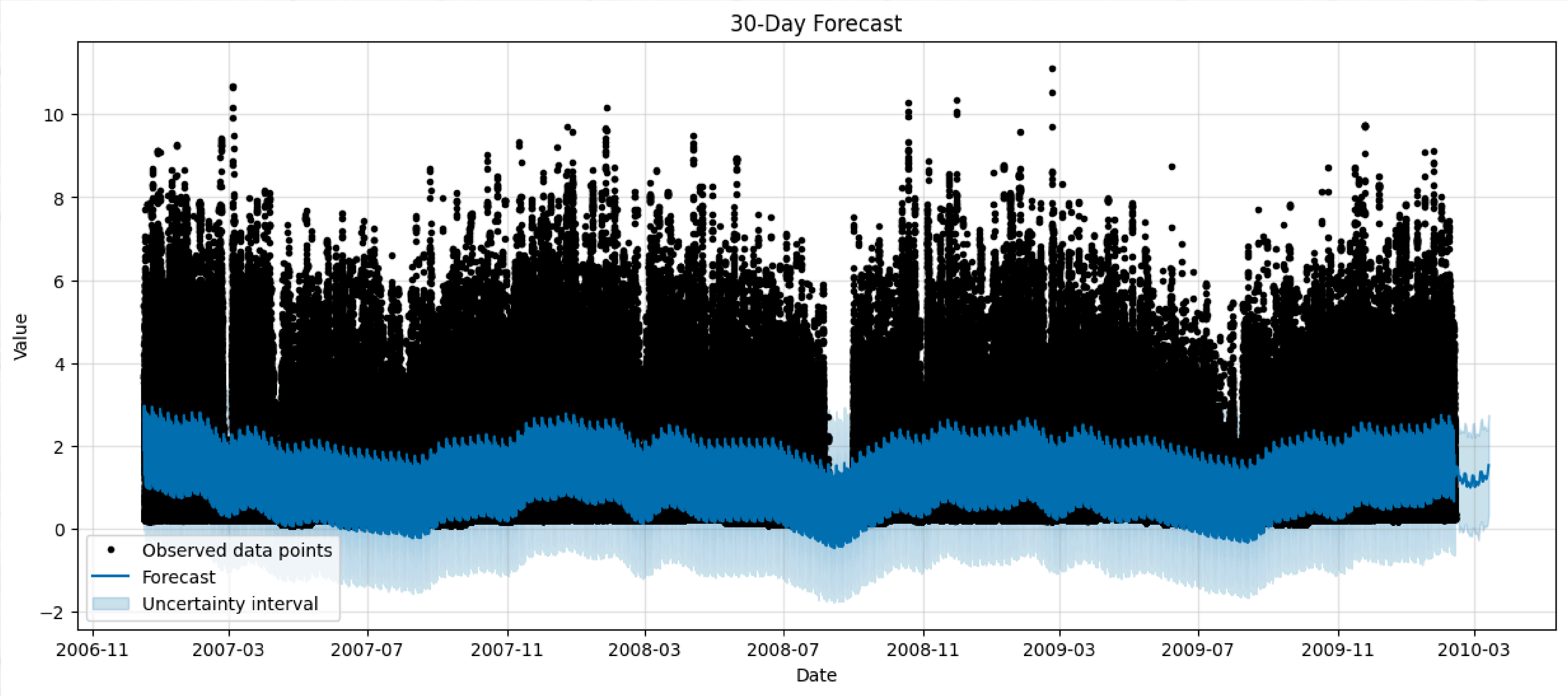
$$y(t) = g(t) + s(t) + h(t) + \epsilon$$

Where:

1. **g(t)** - **Trend Component**
   - Models the overall **growth or decline** over time.
   - Supports **linear or logistic growth** based on data characteristics.
2. **s(t)** - **Seasonality Component**
   - Captures recurring patterns (e.g., **daily**, **weekly**, **yearly**).
   - Auto-detects seasonality, with options for custom periodicities.
3. **h(t)** - **Holiday Effect**
   - Quantifies the **impact of holidays or special events** on the series.
4. **ϵ** - **Noise**
   - Represents **random or unexplained variations** in the data.

Prophet Forecast Graphs





Weekly Forecast (7 Days):

- Predicts trends and patterns for the next week
- Includes trend lines, seasonality, and confidence intervals.

Monthly Forecast (30 Days):

- Projects long-term trends for the next month.
- Highlights seasonal variations with uncertainty bounds.

SUMMARY OF ANALYSIS

Daily Pattern:

- Low consumption early morning (12 AM to 6 AM).
- Increase starts at 7 AM, peaking around 8 PM, then drops late at night.

Yearly Pattern:

- Higher usage in winter (Jan-Mar) and during holidays (Sep-Dec).
- Lower usage in summer months (May-Jul).

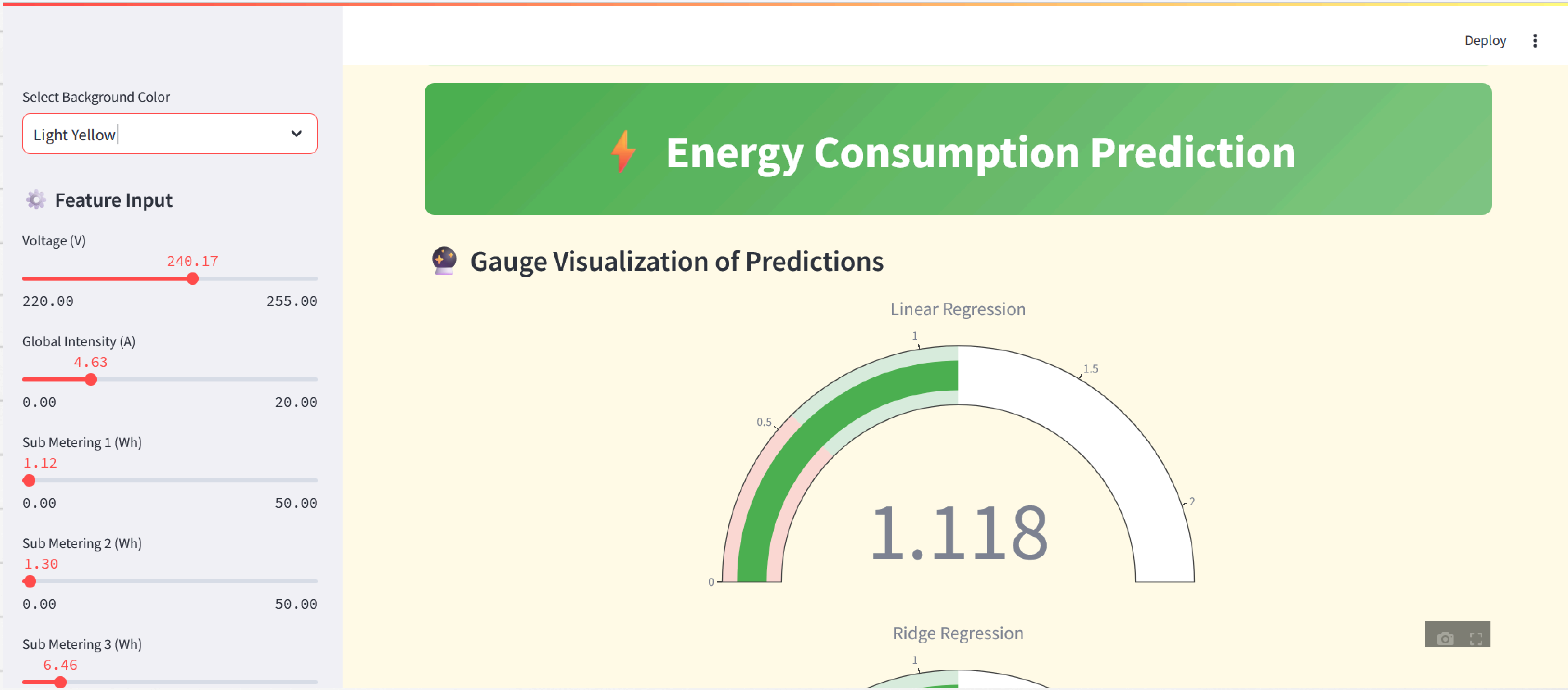
Weekly Pattern:

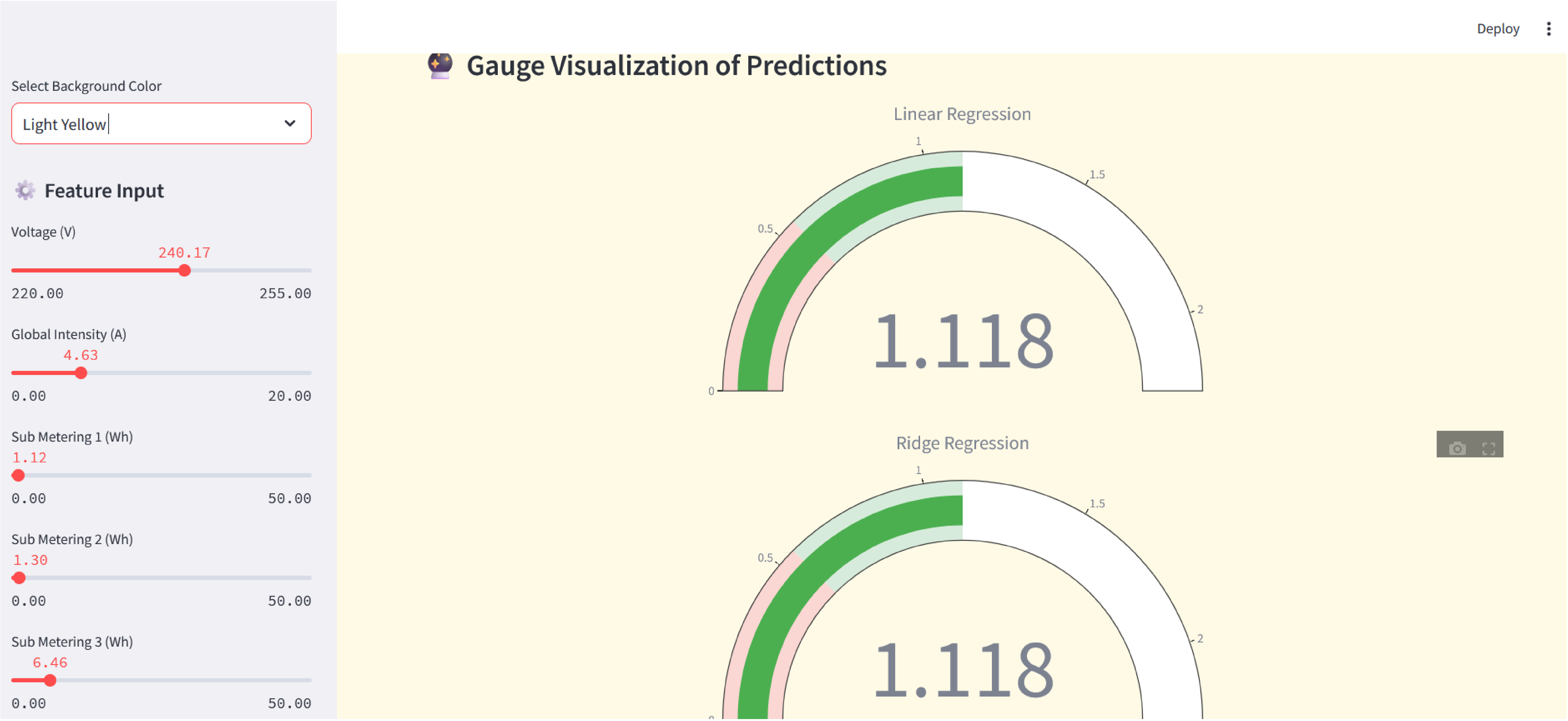
- Lowest usage on Sundays, highest on Fridays and Saturdays, with moderate consumption on weekdays.

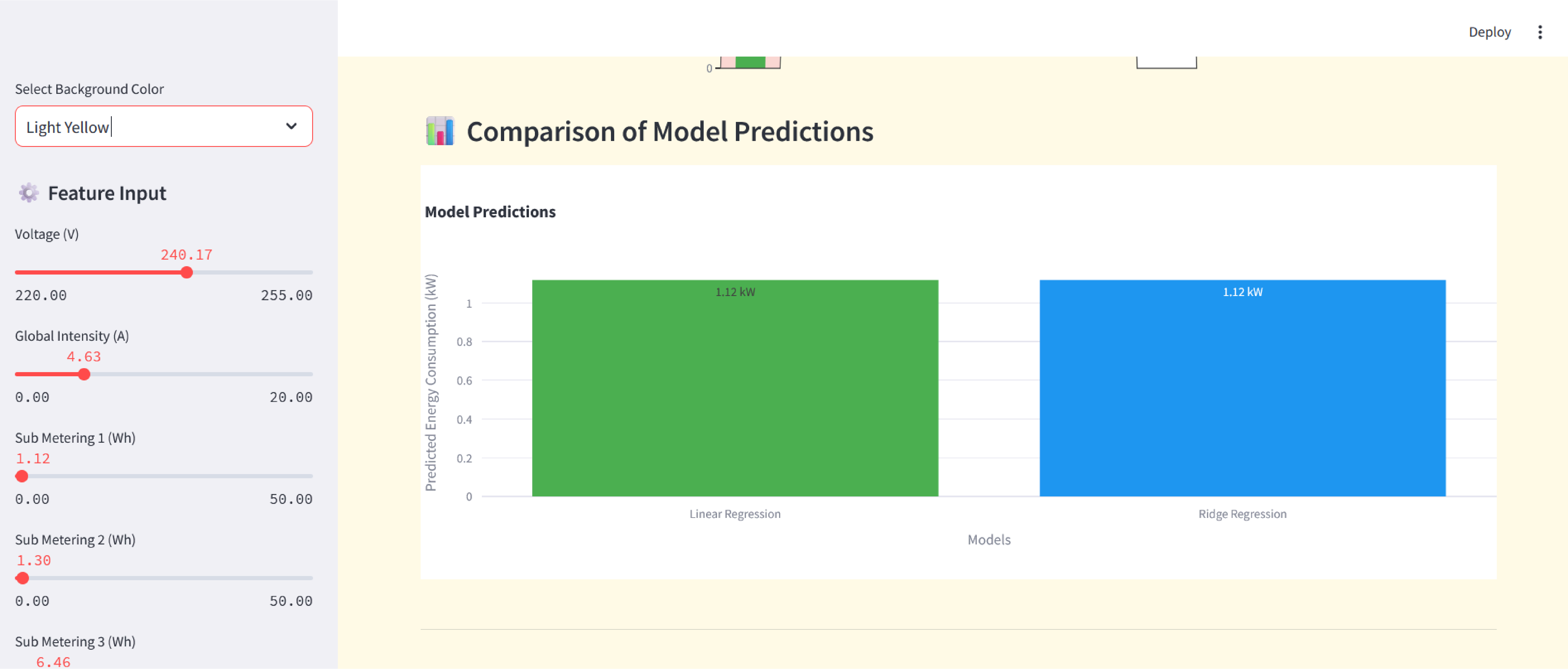
Overall Trend:

- Decline in energy use from 2007 to 2009, with a slight recovery from 2010 onwards.

STREAMLIT APPLICATION




Deploy
Gauge Visualization of Predictions
Select Background Color
Light Yellow
Feature Input
Voltage (V)
240.17
220.00
255.00
Global Intensity (A)
4.63
0.00
20.00
Sub Metering 1 (Wh)
1.12
0.00
50.00
Sub Metering 2 (Wh)
1.30
0.00
50.00
Sub Metering 3 (Wh)
6.46
0.00
50.00
Linear Regression
0
0.5
1
1.5
2
1.118
Ridge Regression
0.5
1
1.5
2
1.118


Deploy
Select Background Color
Light Yellow
Feature Input
Voltage (V)
240.17
220.00
255.00
Global Intensity (A)
4.63
0.00
20.00
Sub Metering 1 (Wh)
1.12
0.00
50.00
Sub Metering 2 (Wh)
1.30
0.00
50.00
Sub Metering 3 (Wh)
6.46
Comparison of Model Predictions
Model Predictions
Predicted Energy Consumption (kW)
1.12 kW
1.12 kW
Linear Regression
Ridge Regression
Models

Thank you very much!