Power Pulse: Household Energy Usage Forecast – Project Report

**Approach:**

The approach follows a structured workflow:

1. Data Understanding and Exploration

* Load dataset and examine structure
* Identify missing values, duplicates, and data types.

1. Data Preprocessing

* Handle missing values using appropriate strategies.
* Handle placeholder values
* Convert data types where necessary.
* Detect Outliers
* Handle Outliers

1. Feature Engineering

* Create new features such as Hour, DayOf Week, Month, WeekOfYeat, IsWeekend, IsPeakHour, Daily\_Consumption, Unmetered\_Energy, Short\_Term\_Avg\_Power, Hourly\_Avg\_Power, Daily\_Avg\_Power, Power\_Deviation\_10min, Power\_Anomaly\_Flag, Season, TimeOfDay.

1. Exploratory Data Analysis

* Analyse trends, outliers, skewness
* Visualize correlations among variables

1. Model Selection and Training

* Train Multiple regression models

1. Evaluation

Compare models using MAE, MSE, R2 Score, RMSE.

**Data Analysis:**

The dataset consists of energy consumption measurements taken from households over a 4-year period. The preprocessing steps performed include

1. **Data Cleaning & Preprocessing**

✔ **Handled missing values** by dropping the null values rows.  
✔ **Converted date and time columns** into a single datetime format.  
✔ **Dropped duplicate rows** to ensure data integrity.  
✔ **Identified outliers** using IQR-based filtering, z-score and visualize outlier using boxplot.  
✔ **Performed skewness & kurtosis analysis** to detect data distribution anomalies.

1. **Feature Engineering**

✔ **Created new features** such as:

* **Hour, DayOf Week, Month, IsWeekend, IsPeakHour**
* **Daily energy consumption, unmetered energy,**  
  ✔ **Extracted seasonal trends** to detect cyclic energy usage patterns.  
  ✔ **Applied correlation** to identify significant relationships between features.

**Model Selection & Evaluation**

Multiple regression models were trained and evaluated using **2 lakh samples** (extracted from the full dataset of 20+ lakh records).

**1) Models Used**

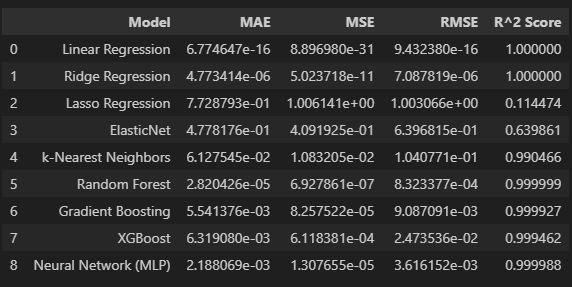
* **Linear Regression**
* **Ridge Regression**
* **Lasso Regression**
* **ElasticNet**
* **K-Nearest Neighbors (KNN)**
* **Random Forest**
* **Gradient Boosting**
* **XGBoost**
* **Neural Network (MLP Regressor)**

**2) Evaluation Metrics**

The models were evaluated based on:

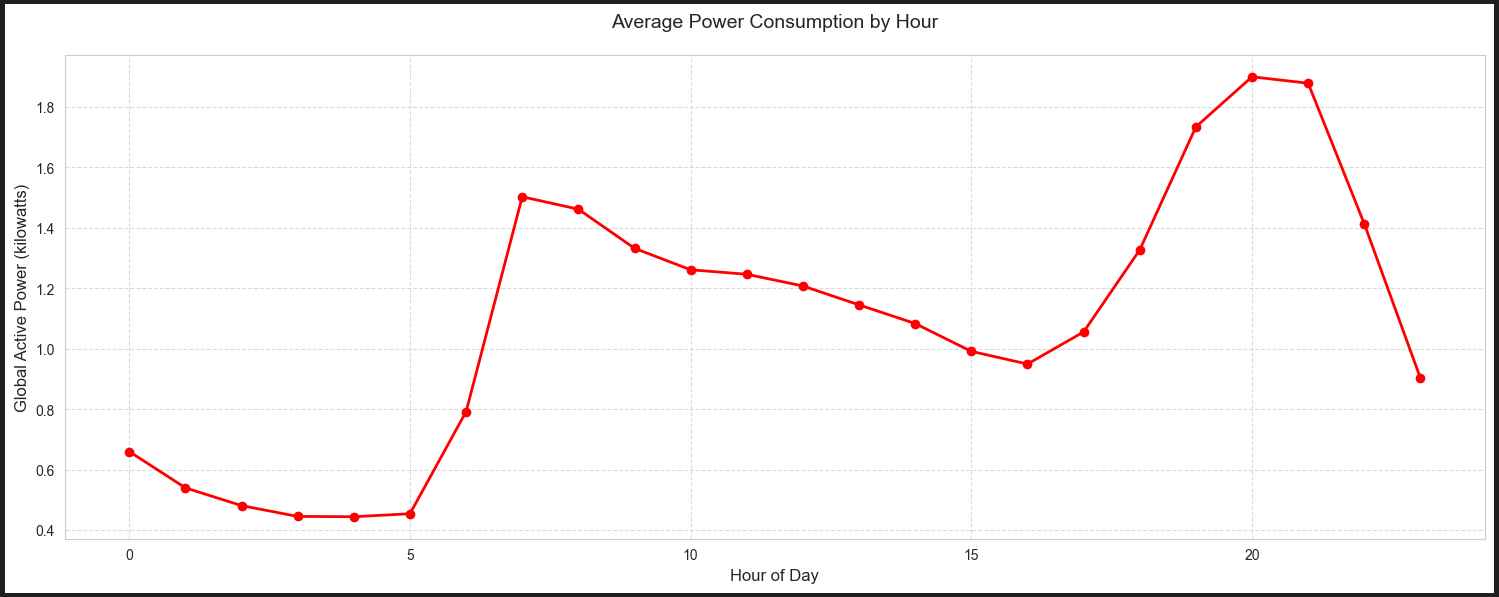
* **Mean Absolute Error (MAE)**
* **Mean Squared Error (MSE)**
* **Root Mean Squared Error (RMSE)**
* **R² Score**

**3) Model Performance Summary**

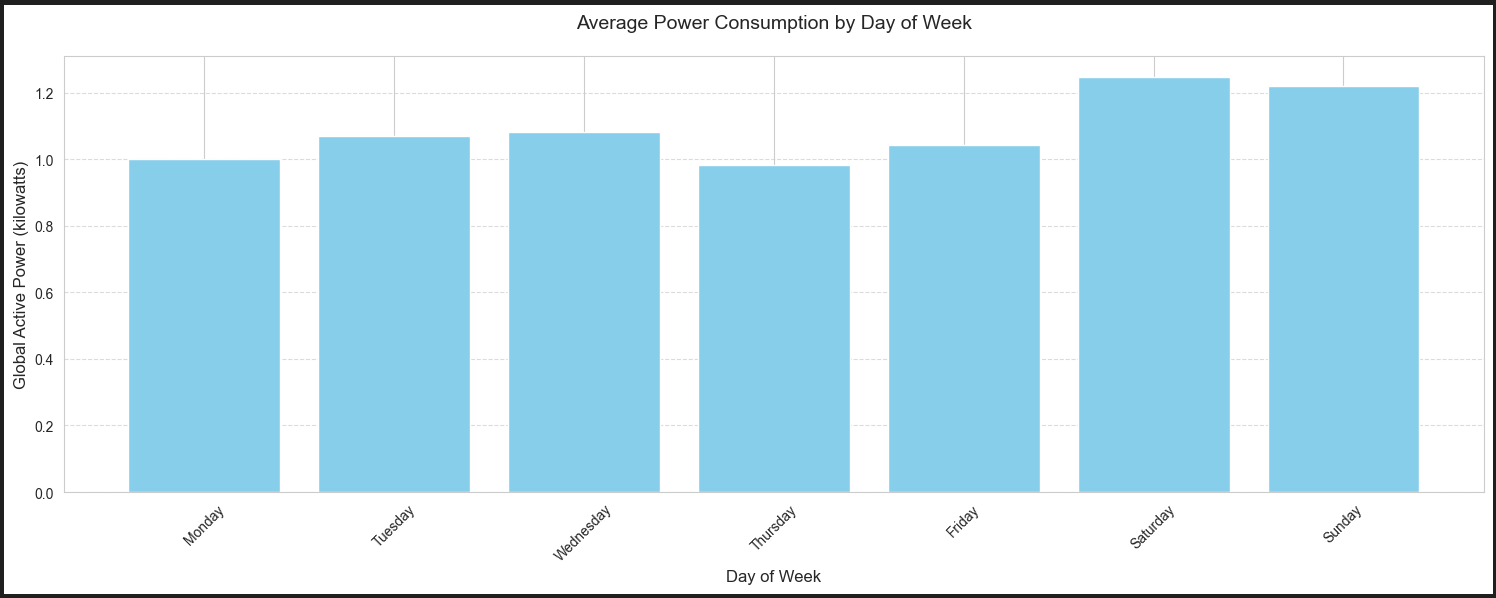


**Insights**

1. Daily Energy Consumption pattern.



1. Weekly Consumption pattern



1. Time of day consumption pattern