COMPUTER ENGINEERING DEPARTMENT

# CSE 464 Term Project

## INTRODUCTION TO DATA SCIENCE & BIG DATA ANALYTICS (AUTUMN SEMESTER 2024)

**Group 6**
**Mahny Barazandehtar – 20210702004**
**Ali Abbasi Dolatabadi - 20210702109**

# I. Short Story of Business/Organization Challenge

Amendis, a public-private utility provider in Tetouan City, Morocco, manages three distinct power distribution zones. The city's increasing energy demand, coupled with environmental factors and evolving consumption patterns, has led to operational inefficiencies. These inefficiencies include poor load balancing, rising operational costs, and an inability to integrate renewable energy sources effectively. With the global push toward sustainability and smart grid technologies, Amendis faces mounting pressure to optimize its energy management practices.

To address these challenges, Amendis sought to harness historical SCADA (Supervisory Control and Data Acquisition) data and advanced analytical techniques. This project aims to combine descriptive and predictive analytics to uncover actionable insights, optimize resource allocation, and guide the integration of renewable energy.

# II. Problem Summary/Definition

## A. The Problem:

Amendis' power distribution network struggles to efficiently balance energy production with fluctuating consumption patterns across its three zones. The challenges identified include the lack of actionable insights from historical consumption data, a limited understanding of how environmental factors such as temperature and humidity influence energy demand, and inadequate predictive capabilities for planning peak-load management. Furthermore, the company has missed opportunities to integrate renewable energy sources and adopt dynamic pricing strategies to optimize energy consumption.

## B. Why It Matters:

Operational inefficiencies in energy management can lead to increased costs, higher carbon emissions, and the risk of power outages. Such issues undermine customer trust and satisfaction, posing significant challenges to Amendis' reputation and sustainability goals. As energy demands continue to grow, Amendis must adopt data-driven strategies to maintain competitiveness and ensure reliable energy distribution while advancing sustainability objectives.

# III. Solution/Recommendations/Decisions

## C. Proposed Solution:

### a. *Phase 1: Descriptive Analytics*

The first phase focused on exploring historical SCADA data to understand past trends and generate insights using descriptive-analytical methods. The data was aggregated across all three zones to provide a holistic view of total power consumption. A correlation matrix was developed to identify relationships between variables, revealing a strong positive correlation between temperature and power consumption, while humidity showed a weaker negative correlation. Visualizations such as bar charts and histograms were created to illustrate hourly and monthly consumption trends and the distribution of power consumption. For example, the analysis revealed that peak usage occurred during evening hours and summer months, driven largely by air conditioning demands. Two distinct descriptive methods were employed: correlation analysis and temporal trend visualization. Ultimately, temporal trends were selected as the most effective method for generating actionable insights due to their clarity and direct relevance to decision-making.

### b. *Phase 2: Predictive Analytics*

Building upon the insights from Phase 1, the second phase utilized supervised machine learning methods to predict future energy demand. Three models were trained and evaluated: Linear Regression, Gradient Boosting Regressor (GBR), and K-Nearest Neighbors (KNN). The models were assessed based on performance metrics such as $R^2$, RMSE, and MAE. Linear Regression served as a baseline, achieving an $R^2$ score of 0.61, but lacked the complexity needed to accurately model the data. Gradient Boosting Regressor significantly improved accuracy, with an $R^2$ score of 0.91 and a lower RMSE. K-Nearest Neighbors outperformed the other models, achieving an $R^2$ score of 0.98, demonstrating its capability to capture complex relationships in the data. Feature importance analysis revealed that variables such as the hour of the day, temperature, and month were the most critical drivers of power consumption, highlighting the importance of temporal and environmental factors in predictive modeling.

## D. Key Features/Technologies:

The project leveraged Python libraries like Pandas and Numpy for data manipulation, and Matplotlib for visualization. Scikit-learn was utilized to implement machine learning models. The SCADA dataset provided comprehensive records of energy consumption and environmental variables. All analysis was conducted on Google Colab, ensuring reproducibility and streamlined documentation.

The links to the Colab Environment and the SCADA dataset are as follows:

- Colab: https://colab.research.google.com/drive/1jn-WXwXko2dauHk6WSl3CbnQaH1KXW9c?usp=sharing
- SCADA Dataset: https://archive.ics.uci.edu/dataset/849/power+consumption+of+tetouan+city

# IV. Results/Outcomes

The descriptive analysis revealed critical insights. For example, the correlation matrix highlighted a strong positive correlation between temperature and total power consumption, indicating that temperature increases significantly drove energy usage, particularly due to air conditioning. The analysis of power consumption trends showed that peak usage occurred during summer months, particularly between June and August, with an evening peak between 6 PM and 9 PM. Humidity, in contrast, exhibited a negative correlation, suggesting reduced energy usage during more humid conditions.

Visualizations played a significant role in presenting these findings. Bar charts and histograms illustrated distinct patterns, such as hourly peaks and seasonal surges, while scatter plots showed the alignment between predicted and actual values for machine learning models.

The predictive analytics phase further demonstrated the effectiveness of advanced models. Linear Regression provided limited explanatory power with an $R^2$ score of 0.61. Gradient Boosting Regressor improved upon this, achieving an $R^2$ score of 0.91, while K-Nearest Neighbors emerged as the best-performing model with an $R^2$ score of 0.98. The feature importance analysis underscored the significance of variables like hour, temperature, and month in predicting energy demand.

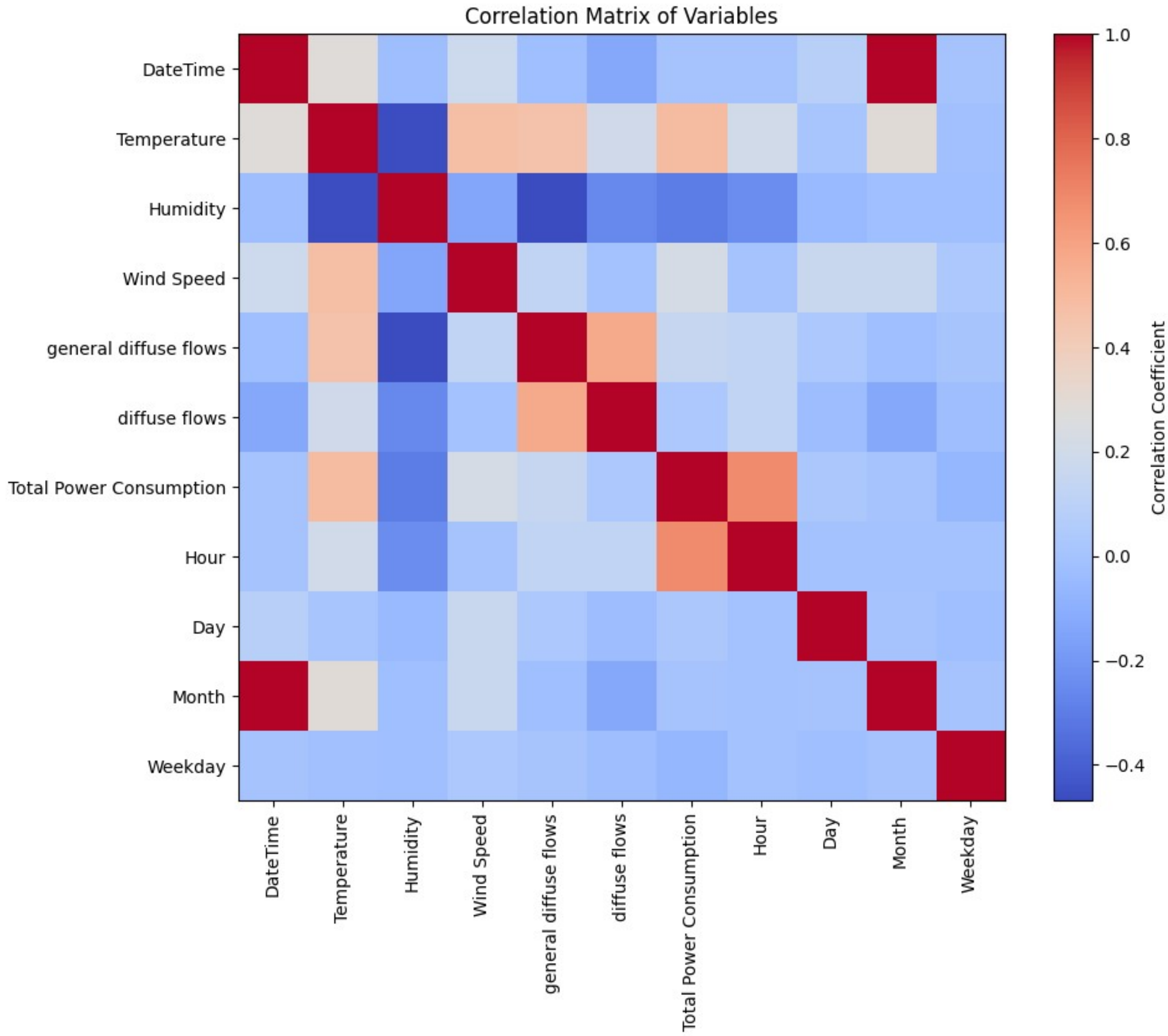


*Figure 1: Correlation Matrix of Variables*

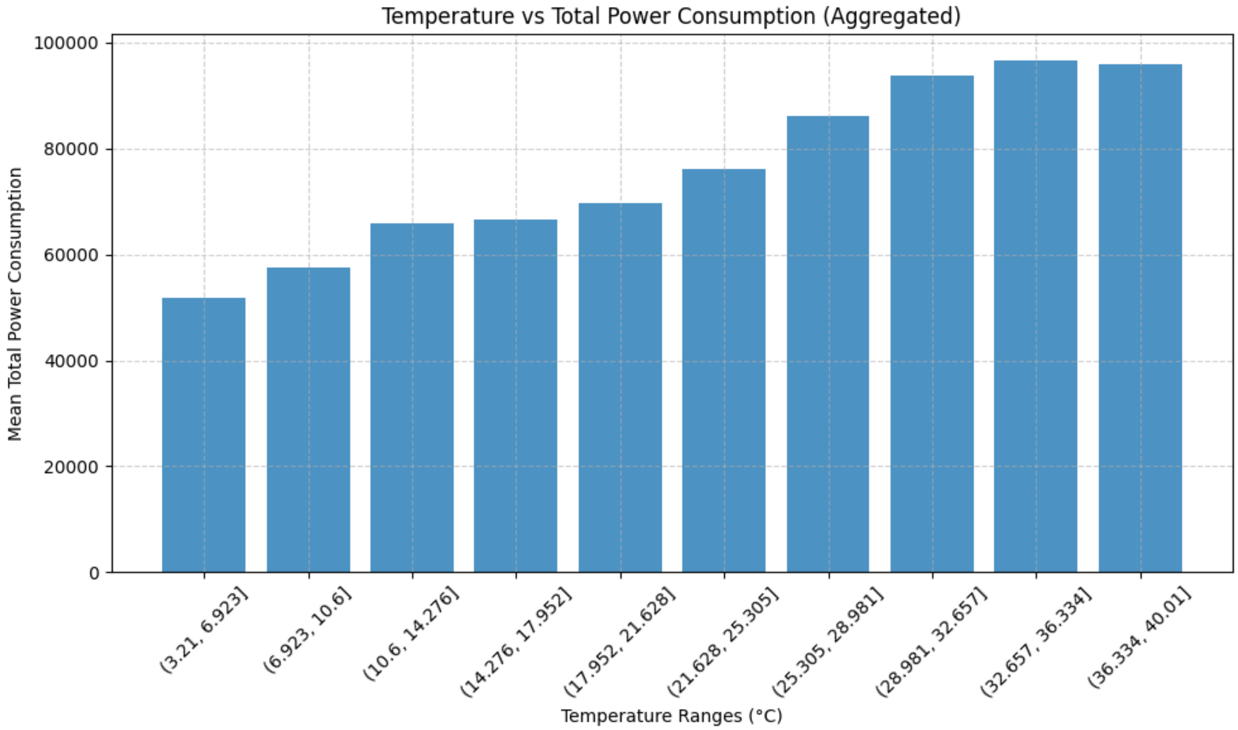


*Figure 2: Temperature vs Total Power Consumption (Aggregated)*

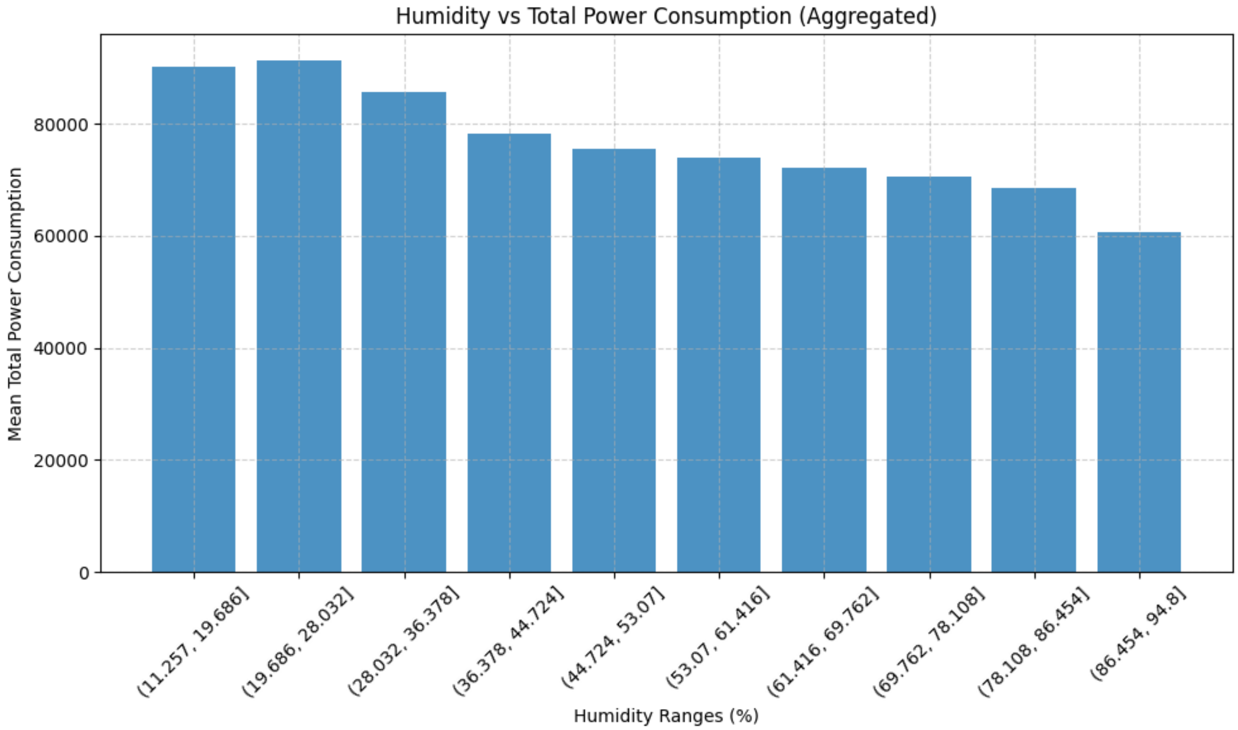


*Figure 3: Humidity vs Total Power Consumption (Aggregated)*

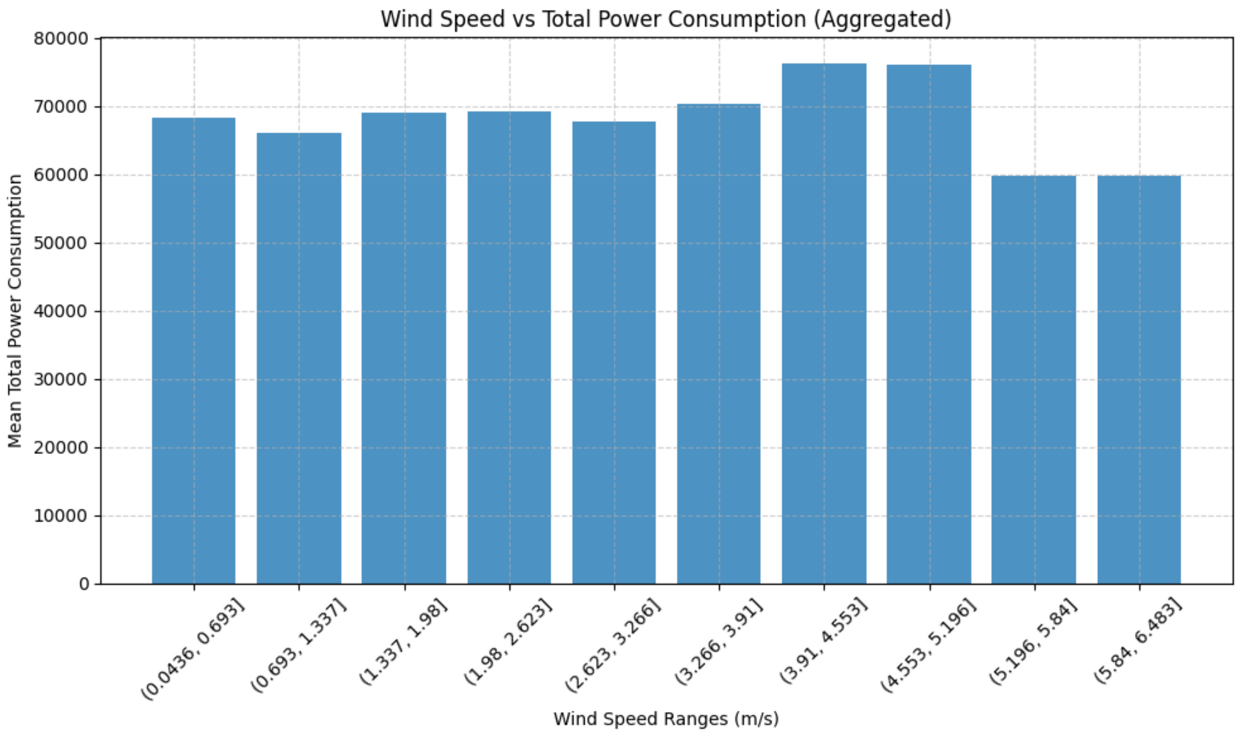


*Figure 4: Wind Speed vs Total Power Consumption (Aggregated)*

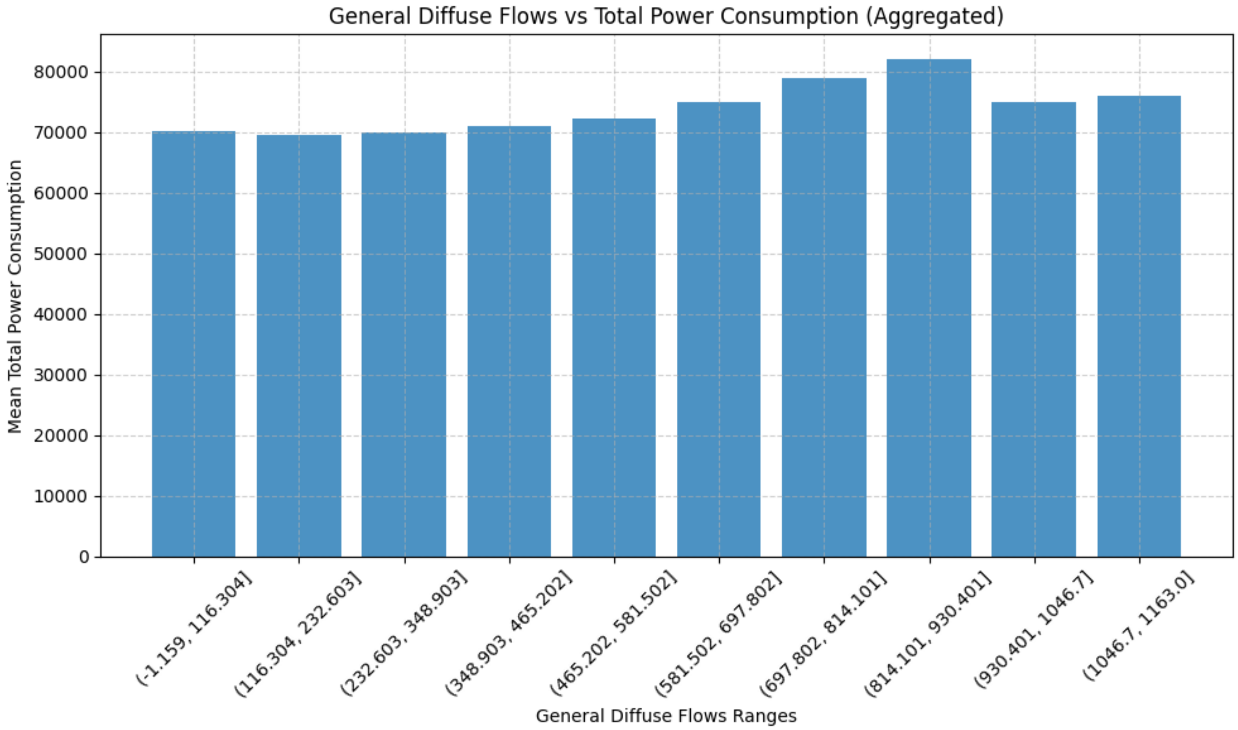


*Figure 5: General Diffuse Flows vs Total Power Consumption (Aggregated)*

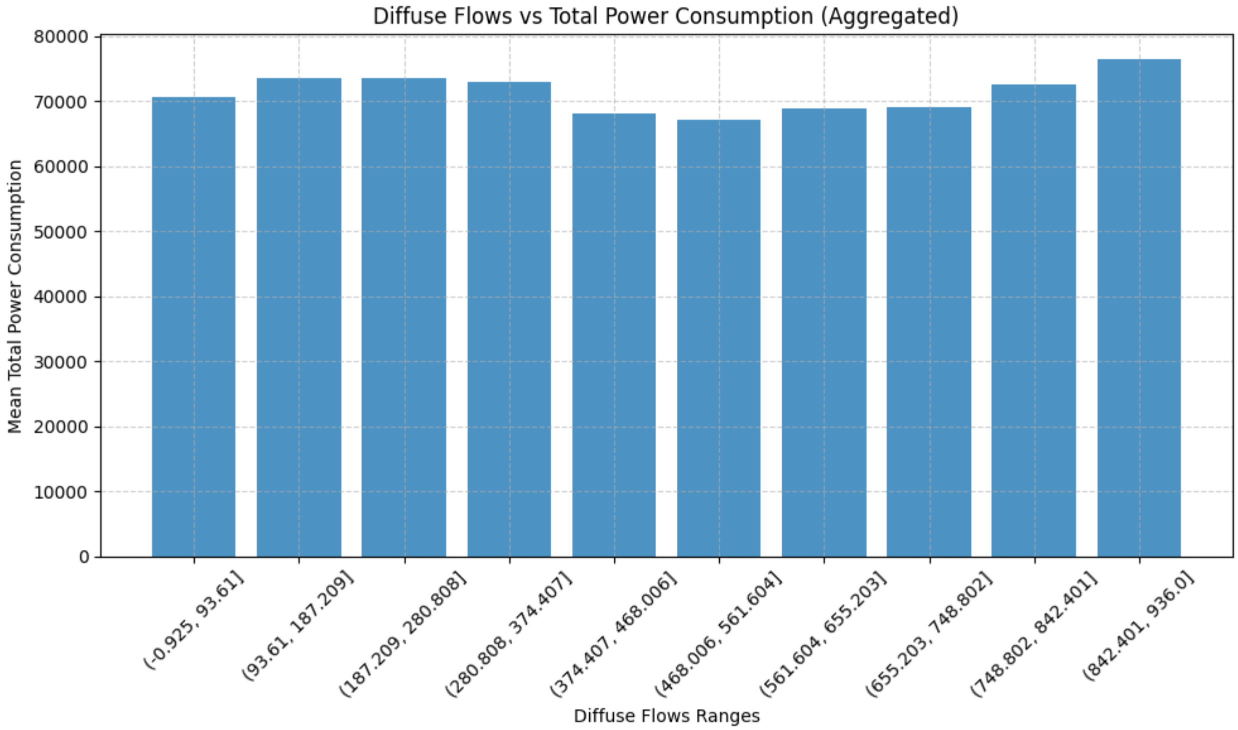


*Figure 6: Diffuse Flows vs Total Power Consumption (Aggregated)*

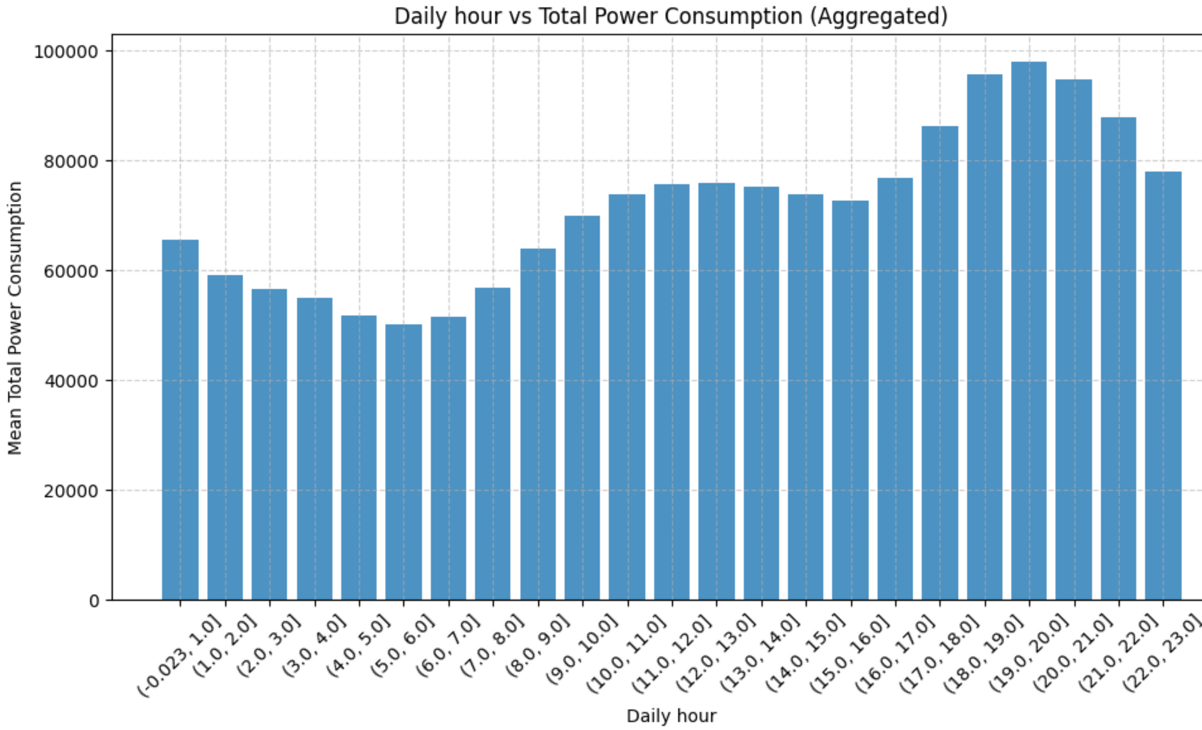


*Figure 7: Daily Hour vs Total Power Consumption (Aggregated)*

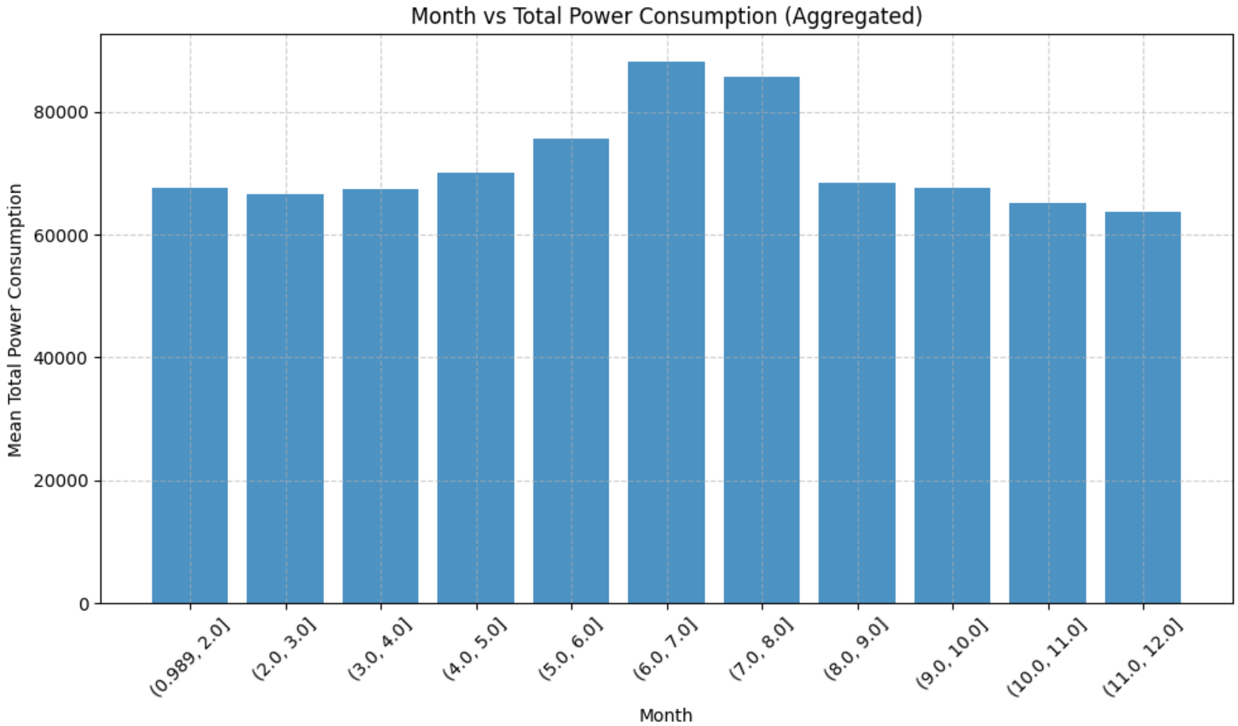


*Figure 8: Month vs Total Power Consumption (Aggregated)*

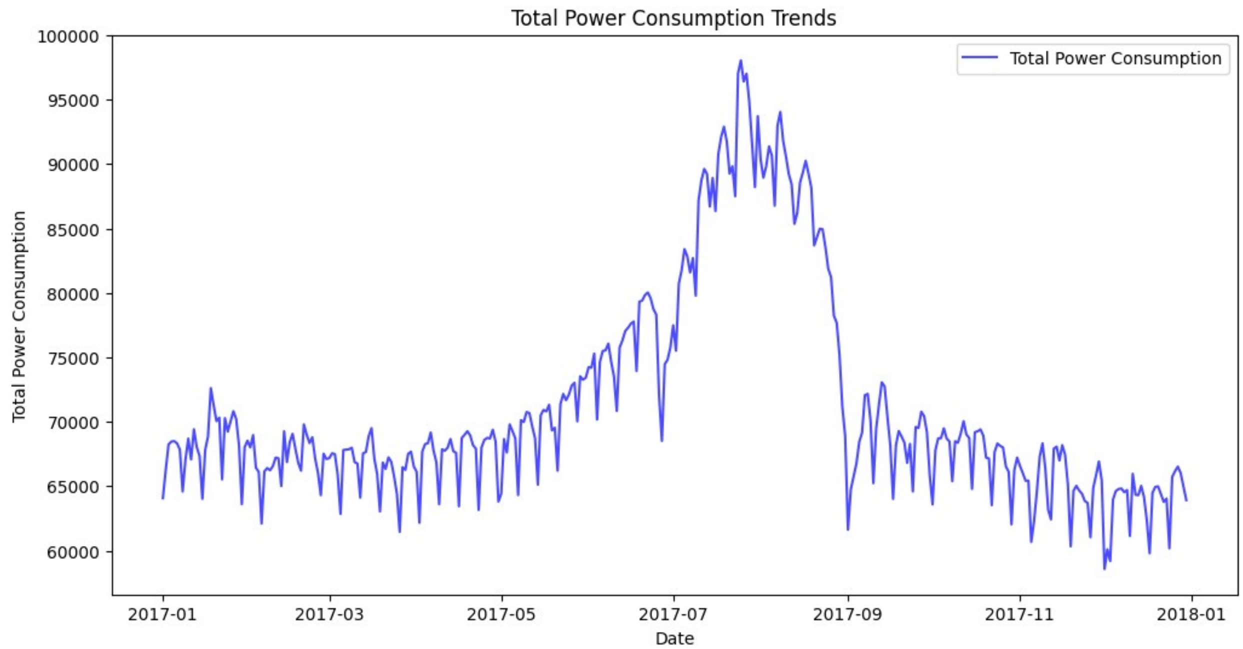


*Figure 9: Total Power Consumption Trends*

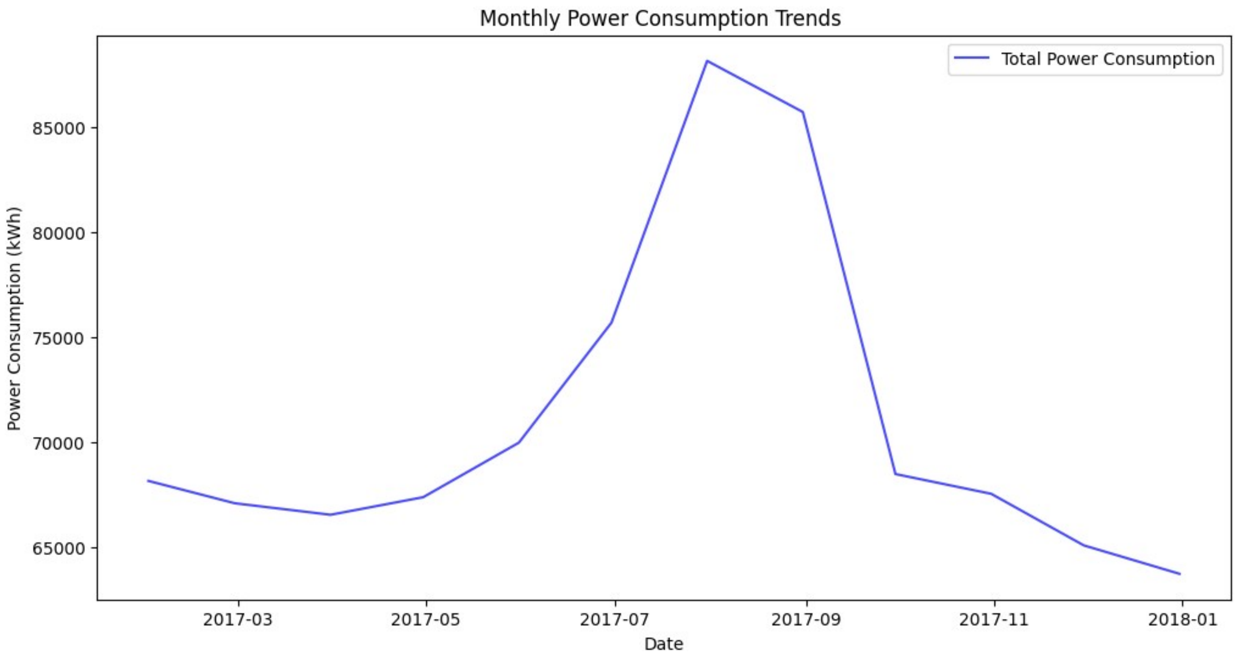


*Figure 10: Monthly Power Consumption Trends*

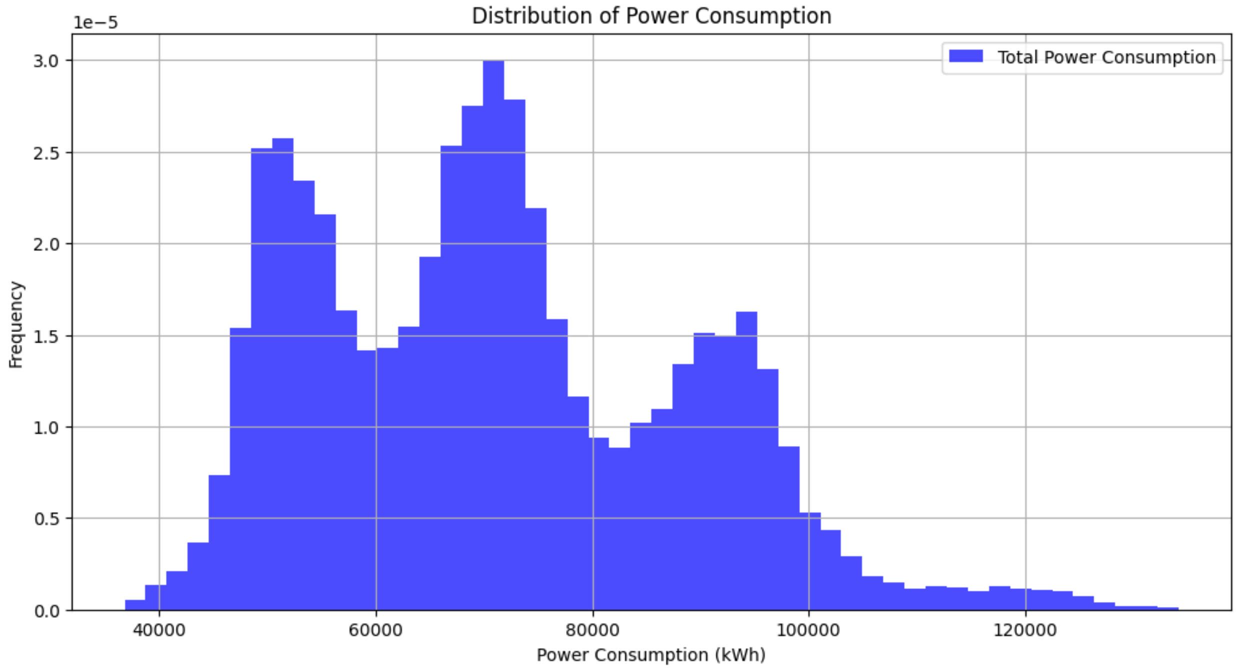


*Figure 11: Distribution of Power Consumption*

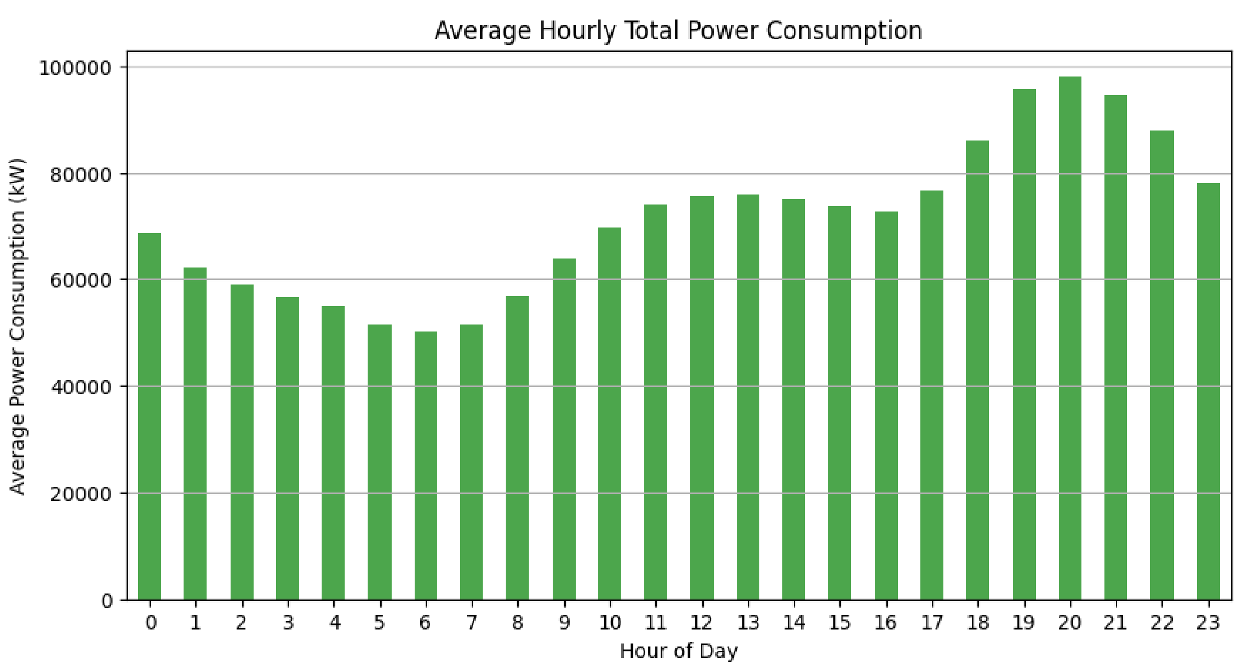


*Figure 12: Average Hourly Total Power Consumption*

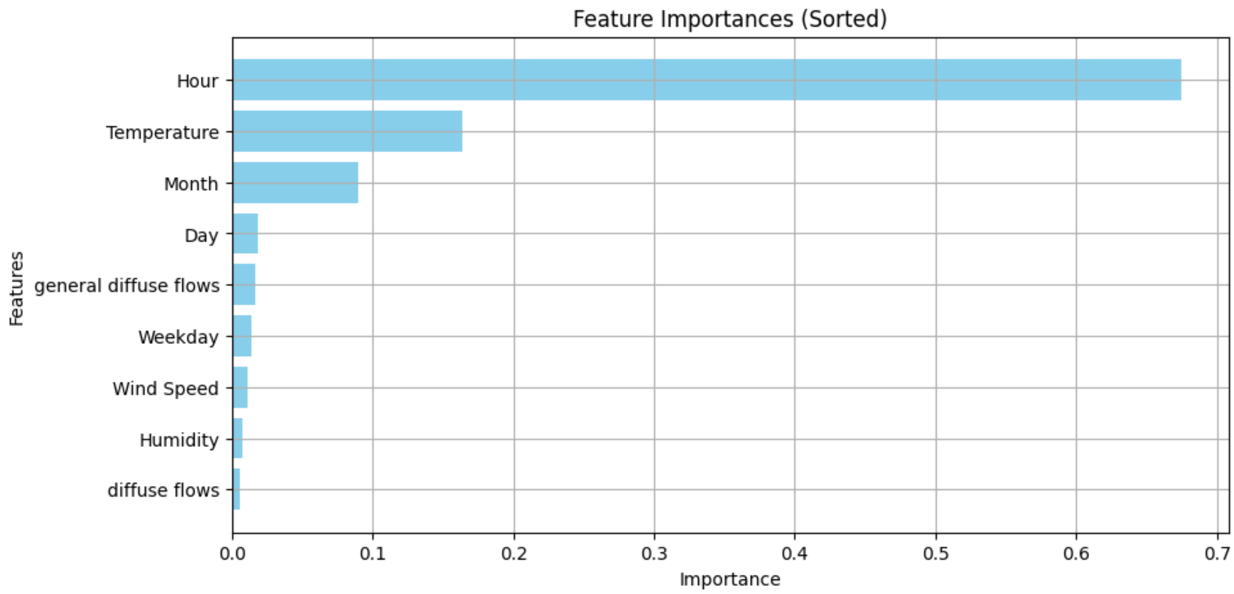


*Figure 13: Feaure Importance (Sorted)*

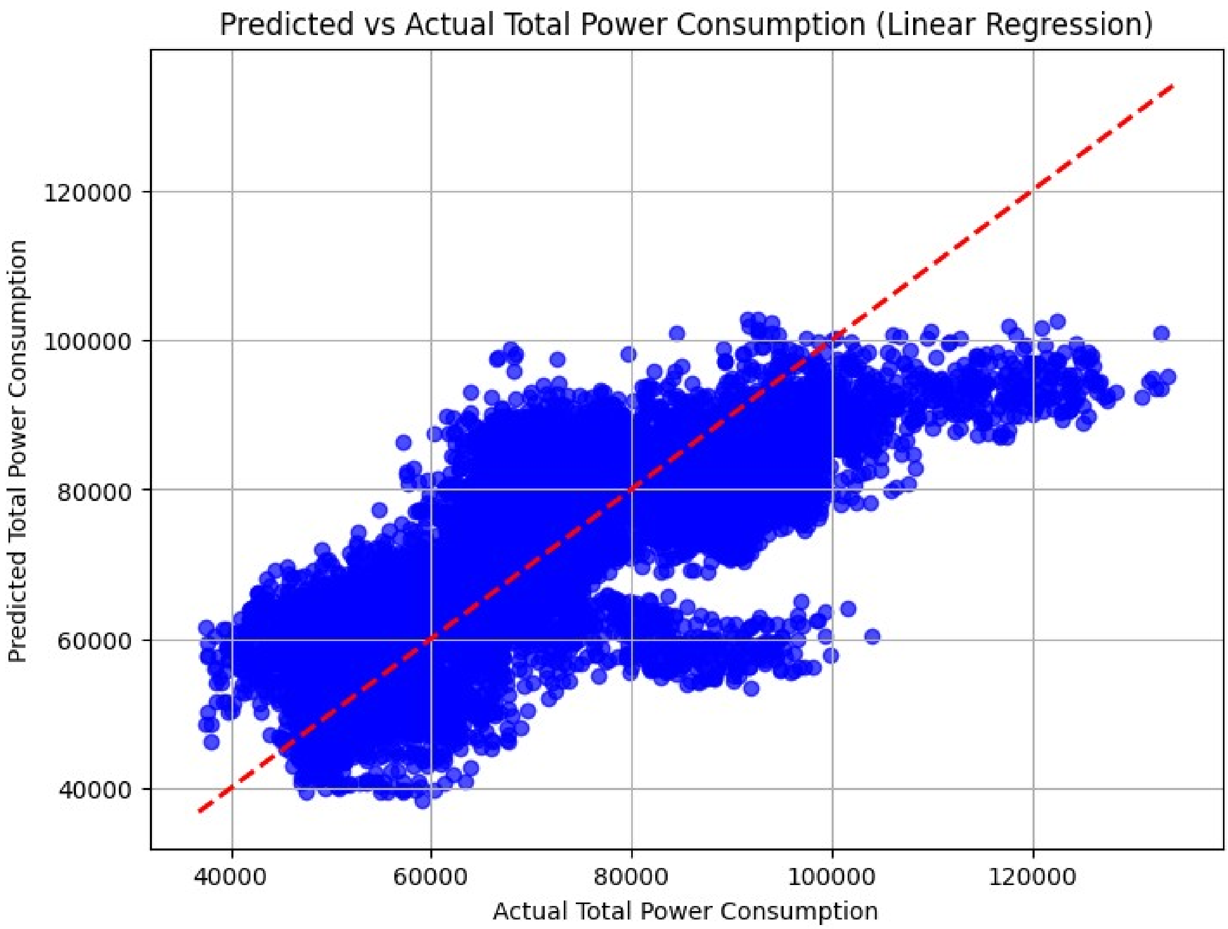


*Figure 14: Predicted vs Actual Total Power Consumption (Linear Regression)*

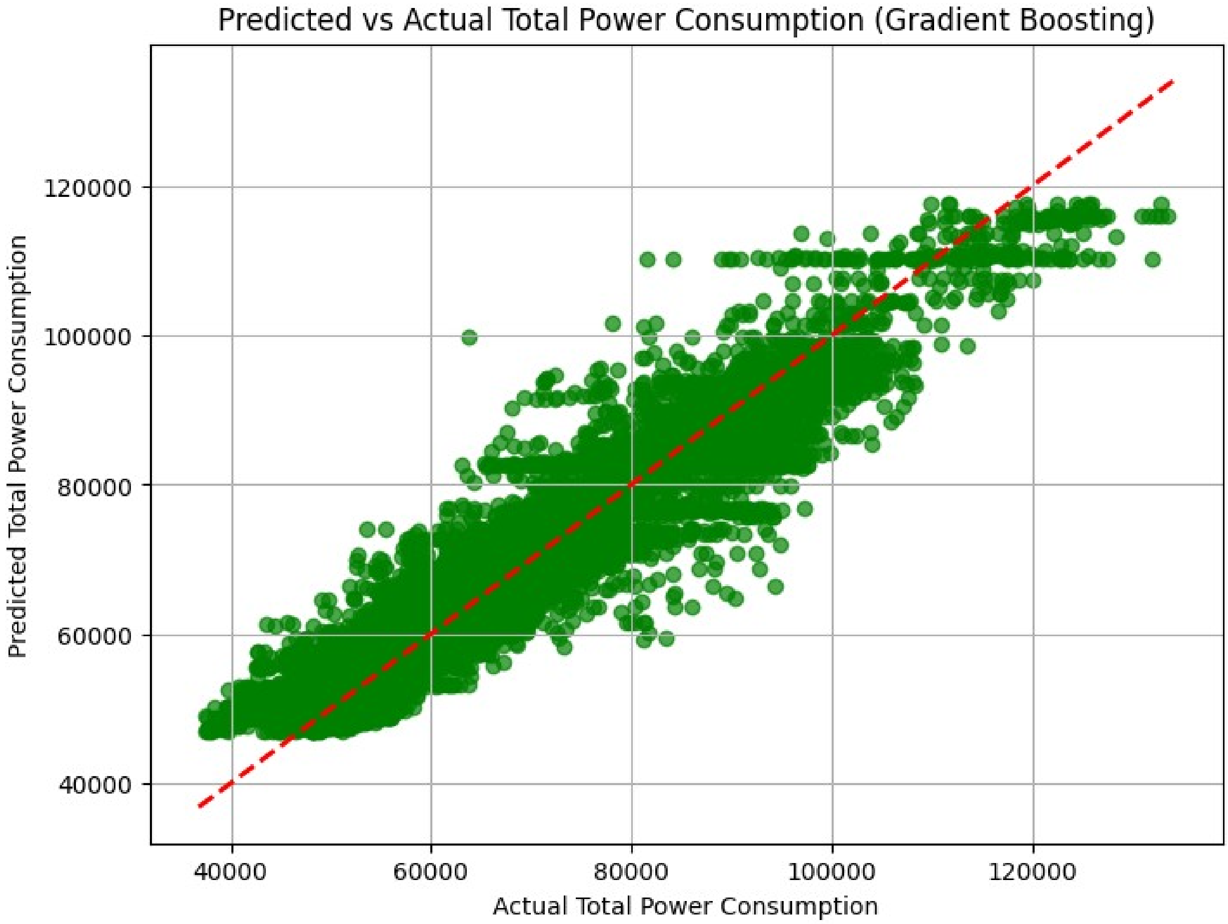


*Figure 15: Predicted vs Actual Total Power Consumption (Gradient Boosting)*

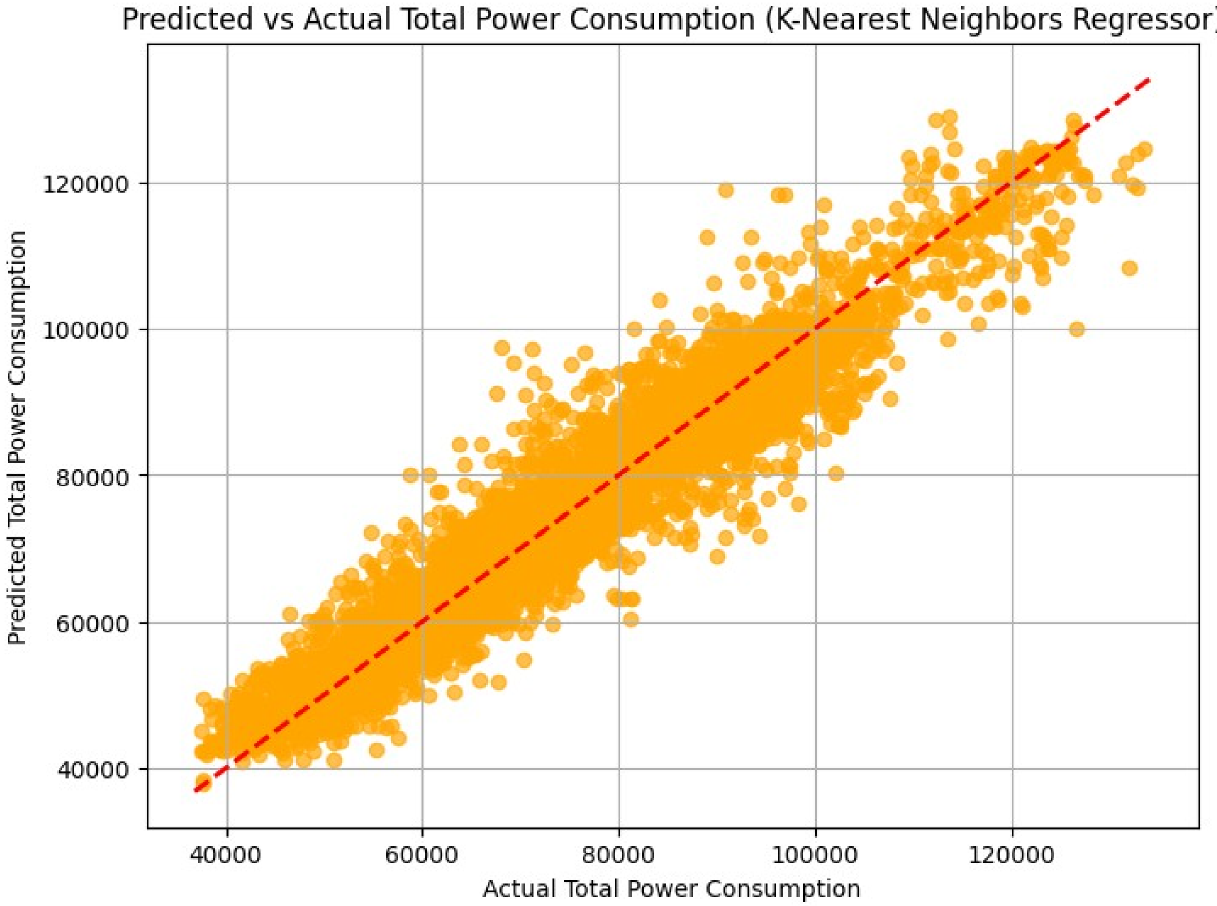


*Figure 16: Predicted vs Actual Total Power Consumption (K-Nearest Neighbors Regressor)*

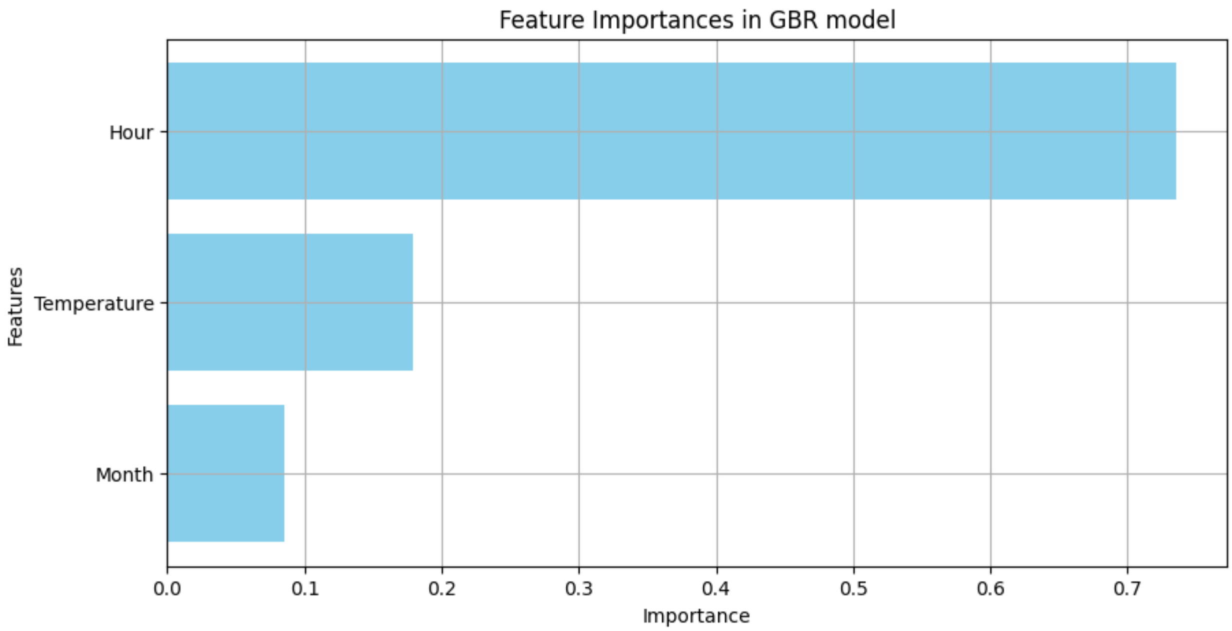


*Figure 17: Feature Importances in GBR Model*

# V. Recommendations

The analysis led to several actionable recommendations. First, Amendis should implement dynamic bandwidth adjustments based on temperature and time-of-day patterns to optimize supply during peak demand periods. By leveraging these insights, the company can allocate resources more effectively and reduce operational inefficiencies. Second, the integration of renewable energy sources should be prioritized. Solar panels can be deployed in areas with high solar radiation, while wind energy projects should be explored in zones with consistent wind speeds. Finally, Amendis should introduce dynamic pricing strategies to incentivize consumers to shift their energy usage to off-peak hours. This strategy can help reduce grid strain during peak periods, ultimately leading to more stable and efficient energy distribution.

# VI. Follow Up and Evaluation Plan

## i. Ongoing Data Collection and Analysis

To ensure continuous improvement, Amendis should implement real-time monitoring of energy consumption and environmental variables. By updating machine learning models periodically with new data, the company can refine its predictions and maintain high accuracy in forecasting energy demand.

## ii. Stakeholder Engagement

Engagement with stakeholders, including local businesses and households, is crucial for successful implementation. Awareness campaigns should be conducted to highlight the benefits of dynamic pricing and renewable energy integration. By aligning strategies with consumer needs, Amendis can build stronger community support for its initiatives.

## iii. Performance Metrics

The success of the proposed strategies should be measured using both quantifiable and qualitative metrics. Quantitatively, a reduction in peak demand by at least 10% through dynamic pricing and an increase in renewable energy contribution by 15% within three years would signify significant progress. Qualitatively, enhanced customer satisfaction due to reduced outages and improved public perception of Amendis as a sustainable energy provider would be key indicators of success.

## iv. Sustainability Goals

By adopting the proposed strategies, Amendis can develop a scalable smart grid framework to support long-term energy efficiency. These efforts align with Morocco's national goals to reduce carbon emissions and promote renewable energy adoption, positioning Amendis as a leader in sustainable energy management.