COMPUTER ENGINEERING DEPARTMENT

CSE 464 Case Study 1

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

Mahny Barazandehtar - 20210702004

COMPUTER ENGINEERING DEPARTMENT

I. Executive Summary/Abstract

This study addresses the operational inefficiencies faced by Amendis, the public-private utility provider managing Tetouan City's power distribution network in northern Morocco. Using historical SCADA (Supervisory Control and Data Acquisition) data recorded every 10 minutes throughout 2017, the analysis focuses on understanding energy consumption patterns across three power distribution zones. Descriptive analytics methods were employed to examine the effects of environmental factors such as temperature, humidity, and wind speed, alongside time-of-day and seasonal variations.

Key findings reveal a strong positive correlation between temperature and energy usage, with notable inter-zonal dependencies influencing load balancing. Insights from the analysis guided recommendations for dynamic bandwidth adjustments, renewable energy integration, and demand-side management strategies such as dynamic pricing. By leveraging these data-driven approaches, the study provides actionable solutions to improve operational efficiency, reduce environmental impact, and ensure a reliable energy supply for Tetouan City's growing needs.

II. Background/Introduction

- **About the Client/Company:**
  The case focuses on Tetouan city's power distribution network, managed by Amendis, a public service operator responsible for electricity and water distribution. Established in 2002 under a public-private partnership, Amendis ensures sustainable service delivery in the region. The electricity network serves low and medium voltage consumers across Tetouan through three major substations: Quads, Smir, and Boussafou. However, the company faces challenges in balancing energy supply with fluctuating demand, particularly due to environmental and temporal factors.
- **Industry Context:**
  The energy and utilities sector is dynamic, with growing demands driven by urbanization, economic growth, and evolving consumer behaviors. Forecasting electricity demand is crucial for balancing energy production, minimizing waste, and maintaining grid stability. While advances in machine learning have enabled predictive modeling, understanding historical consumption patterns through descriptive analytics is foundational for strategic decision-making and operational optimization.
- **Goals and Objectives:**
  This study aims to enhance the operational efficiency and sustainability of Tetouan City's power distribution network, managed by Amendis. Using historical SCADA data, the objectives are to uncover consumption trends, evaluate the influence of environmental factors on energy demand, and optimize load balancing. These insights will inform

strategies for renewable energy integration, improve grid stability, and support Amendis in achieving sustainable energy management goals.

III. Challenge/Problem Statement

A. The Problem:

The power distribution network in Tetouan City faces challenges in efficiently balancing energy production with fluctuating consumption patterns across three distinct zones. The company lacks a clear understanding of the effects of environmental factors like temperature, humidity, wind speed, and solar radiation on energy demand, as well as time-of-day and seasonal trends. This has led to inefficiencies in energy supply, suboptimal load balancing, and increased operational costs. Furthermore, the inability to identify inter-zonal dependencies hampers effective resource allocation and strategic planning.

B. Why It Matters:

Efficient energy management is critical for minimizing waste, ensuring uninterrupted supply, and meeting the growing demand of Tetouan City's residents and businesses. If these inefficiencies are not addressed, the company may face several significant challenges. Operational costs are likely to increase due to overproduction during periods of low demand or shortages during peak demand. These inefficiencies could also lead to power outages or surges, which would erode customer trust and satisfaction. Moreover, overproduction contributes to higher carbon emissions, exacerbating environmental concerns and undermining sustainability goals. Lastly, the failure to adopt data-driven energy management strategies could place the company at a competitive disadvantage, especially as other utility providers implement smarter and more efficient solutions.

IV. Solution/Approach

Descriptive analytics was chosen as the primary approach to uncover patterns and trends within the historical SCADA data. This method focuses on analyzing time-of-day consumption, seasonal variations, and the impact of environmental factors on power usage. By leveraging descriptive analytics, the company can gain a comprehensive understanding of its historical data and extract actionable insights to optimize its operations.

One key goal of this analysis is to identify peak and off-peak usage periods. Understanding these fluctuations in demand enables the company to adjust its energy distribution strategies, ensuring adequate supply during high-demand hours and minimizing waste during low-demand periods.

Another critical aspect of the analysis is examining the influence of environmental variables such as temperature, humidity, and wind speed on power consumption. By quantifying these relationships, the company can predict how changes in weather conditions impact energy demand.

Additionally, the analysis aims to examine inter-zonal consumption patterns to improve load balancing. By understanding shared dependencies and unique peaks across the three zones, the company can allocate resources more effectively and address specific regional needs. This comprehensive approach equips the company with the insights needed to enhance operational efficiency and better align with consumer demands.

C. Proposed Solution:

A comprehensive descriptive analysis was conducted on the historical power consumption data from Tetouan City's SCADA system to identify patterns and trends. The first step involved segmenting the data by time, such as hourly, daily, and monthly intervals, to uncover usage patterns and temporal dynamics. This segmentation allowed for a detailed understanding of consumption trends over various timeframes.

The analysis also included a thorough exploration of the relationships between power consumption and environmental factors through correlation analysis. Positive correlations were observed between temperature and power usage, while humidity exhibited a negative correlation with consumption. Wind speed showed a weak positive correlation, suggesting a minimal but noteworthy impact on energy usage.

To effectively communicate these findings, graphs and charts were created to highlight time-of-day consumption trends, seasonal variations, and inter-zonal dependencies. These visualizations provided clear insights into the dynamics of power usage. Additionally, inter-zonal analysis was conducted to identify shared patterns and specific peaks across zones, which were crucial in guiding tailored interventions and optimizing energy distribution strategies.

Based on these insights, the solution involves implementing a system to dynamically adjust the bandwidth of power distribution for each zone according to environmental and temporal factors. For example, during hotter months when temperatures are high and demand increases, bandwidth for Zone 3, which shows the highest correlation with temperature, can be expanded to prevent shortages. Similarly, during periods of lower demand or higher humidity, bandwidth can be reduced to avoid overproduction.

Furthermore, zones with higher exposure to solar radiation or wind speeds, as indicated by diffuse flow and wind speed data, can integrate renewable energy sources like solar panels or wind turbines. By leveraging solar power in areas with strong solar radiation and wind energy in regions with consistent wind patterns, the company can further reduce reliance on traditional energy sources, lowering operational costs and environmental impact. This integrated approach ensures both sustainability and efficiency in meeting Tetouan City's energy demands.

D. Key Features/Technologies:

The project leveraged the following tools and technologies:

Python Libraries:

- **Pandas:** Data preprocessing and manipulation.
- **Numpy:** Numerical and statistical analysis for correlation metrics.
- **Matplotlib:** Visualization of consumption patterns.

SCADA Data:
The dataset, titled Power Consumption of Tetouan City, is obtained from UC Irvine's Machine Learning Repository and can be accessed at the following link:
https://archive.ics.uci.edu/dataset/849/power+consumption+of+tetouan+city

Google Colab:
Access the Google Colab environment for this case study using the following link
https://colab.research.google.com/drive/1iLYd7L70yKAR_Ter_leWEzowJ1RkCvi7?usp=sharing

E. Implementation Process:

1. **Data Preparation:**
   - The SCADA data was preprocessed to make it suitable for analysis. Timestamps were converted into meaningful components, such as hours and months, to facilitate the study of temporal trends. Additionally, the DateTime column was set as the index, allowing for efficient time-based operations and grouping. These steps ensured that the data was structured in a way that enabled detailed exploration of trends and patterns.
2. **Analysis:**
   - Descriptive analytics techniques were applied to uncover temporal patterns and inter-zonal dependencies. The analysis segmented the data by hourly, daily, and monthly intervals to reveal consumption trends. Correlation metrics were computed to quantify the influence of environmental factors on energy usage.
3. **Visualization:**
   - Matplotlib was used extensively to create a variety of graphs that effectively communicated the findings. The graphs illustrated relationships between environmental factors and energy usage, and correlation matrices highlighted variable dependencies.
4. **Insights Extraction:**
   - Actionable insights were derived from the analysis to address operational inefficiencies. These insights formed the basis of recommendations for dynamic

bandwidth adjustments, renewable energy integration, and load balancing strategies.

V. Results/Outcomes

F. Quantifiable Results:

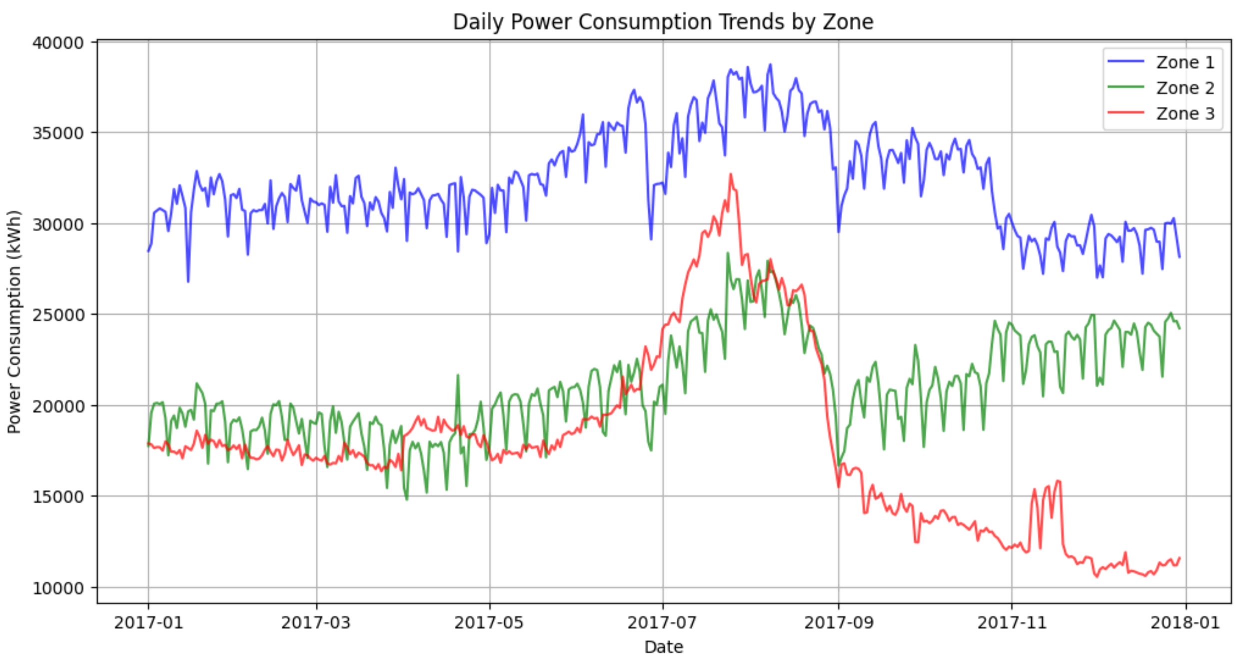

Figure 1: Daily Power Consumption Trends

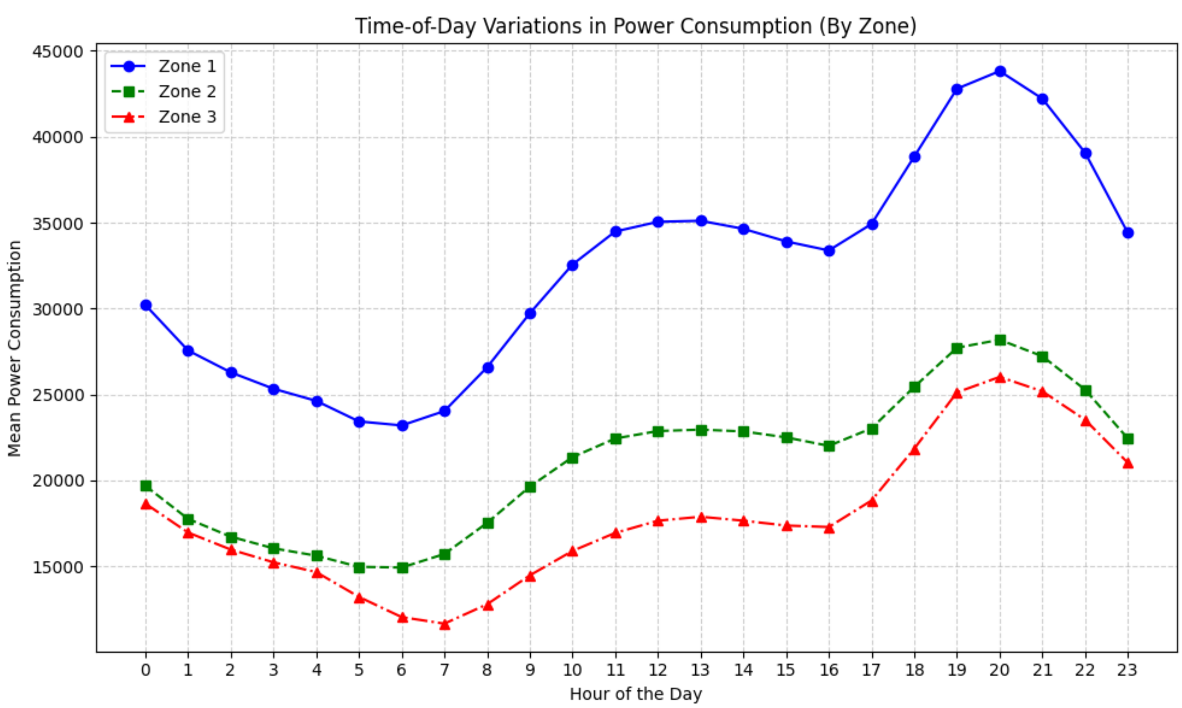

Figure 2: Time-of-Day Power Consumption Trends

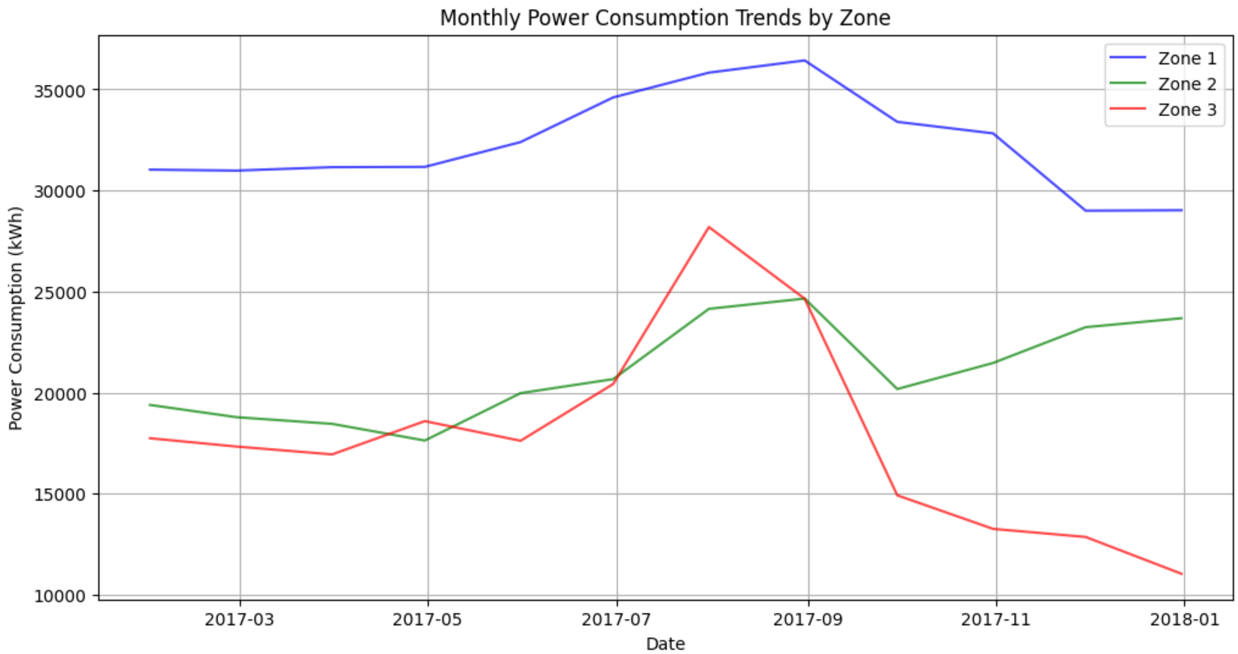

Figure 3: Monthly Power Consumption Trends

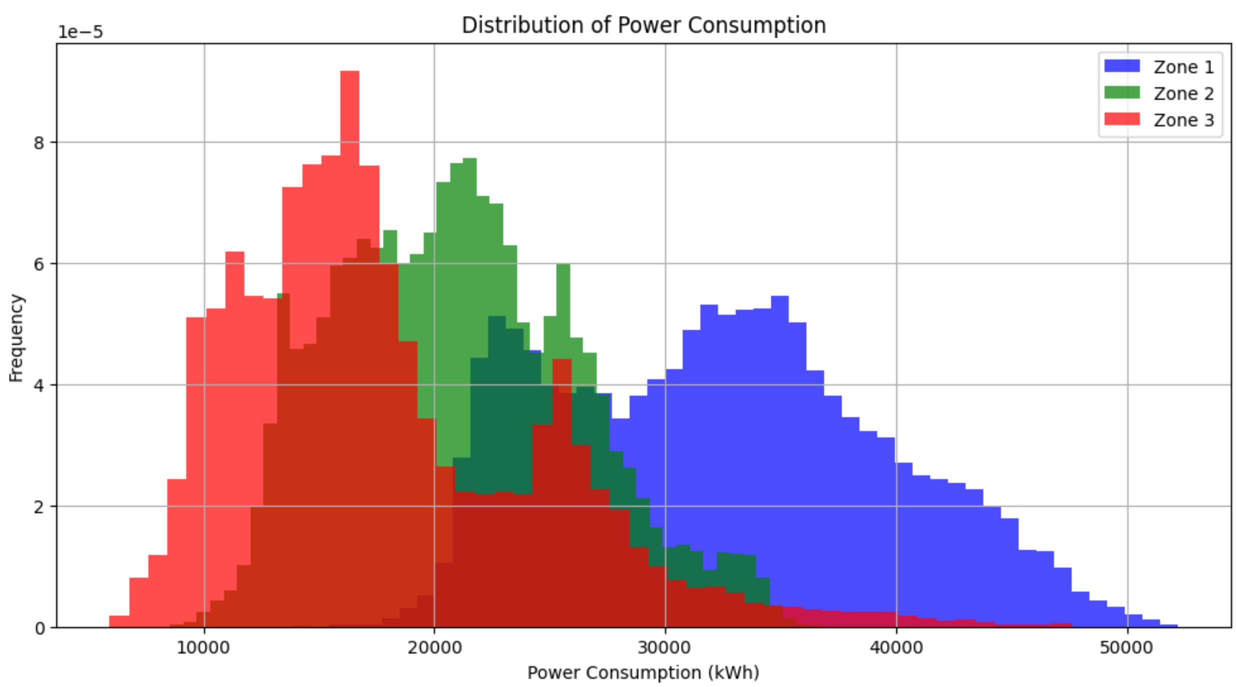

Figure 4: Distribution of Power Consumption

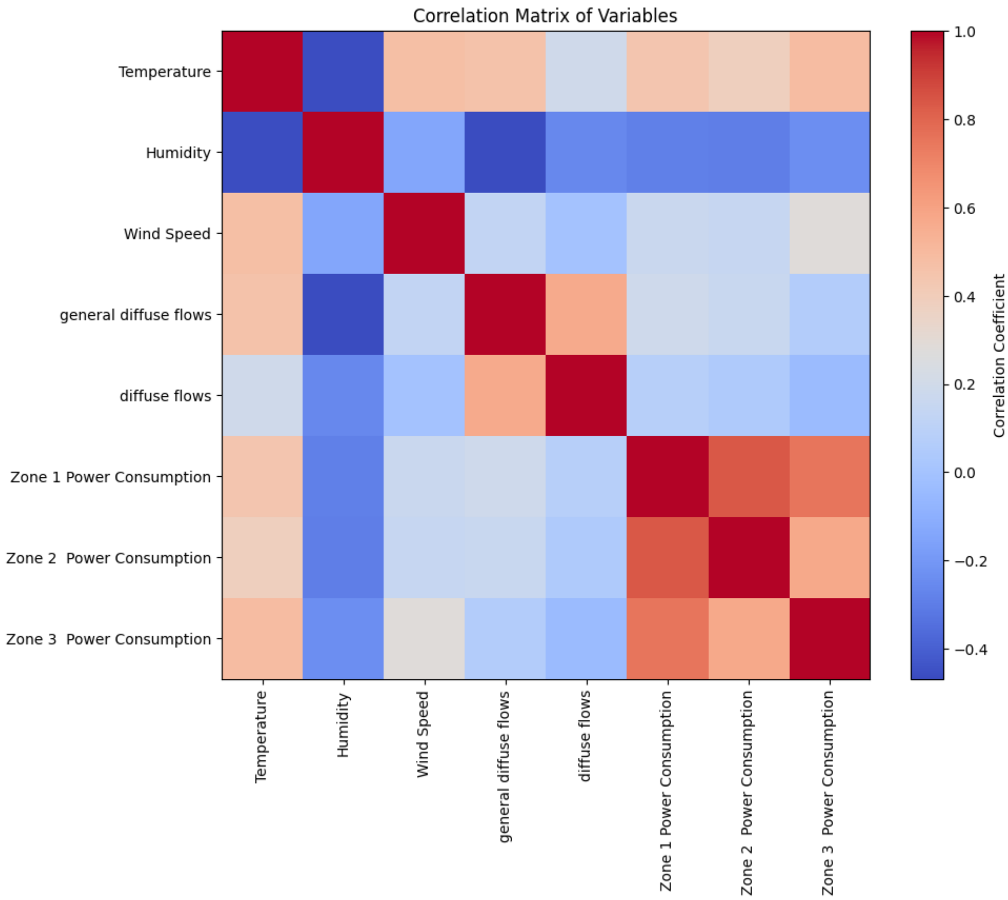

Figure 5: Correlation Matrix of Variables

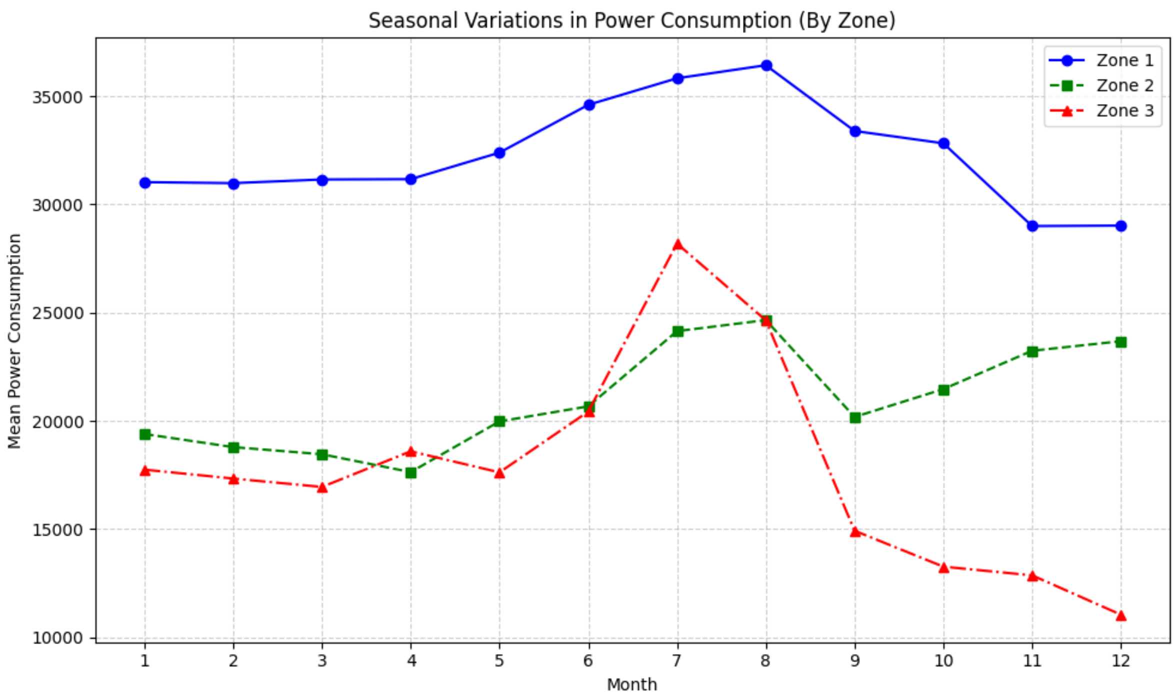

Figure 6: Seasonal Variations in Power Consumption (By Zone)

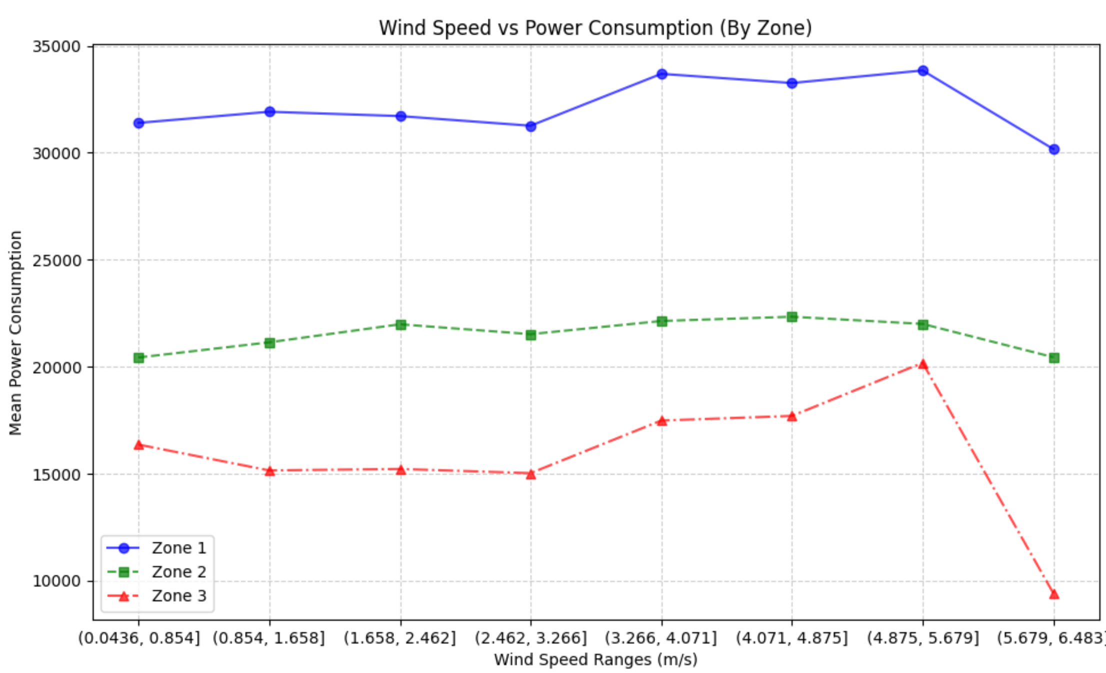

Figure 7: Wind Speed vs Power Consumption (By Zone)

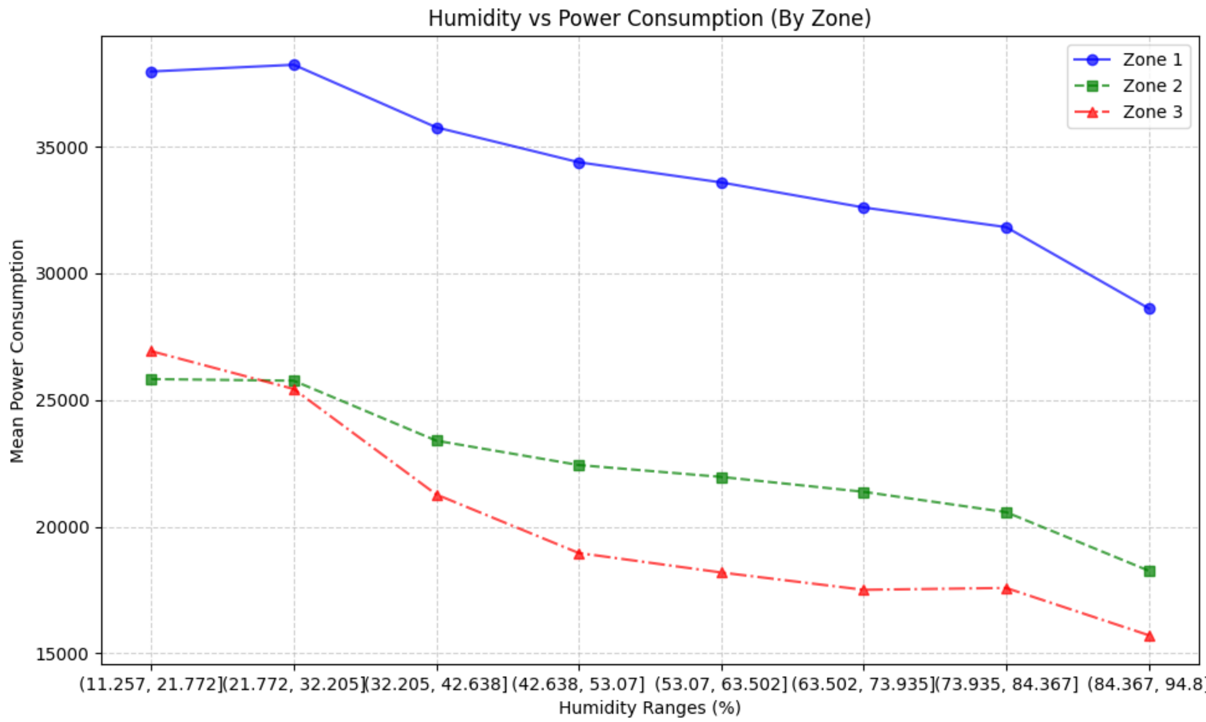

Figure 8: Humidity vs Power Consumption (By Zone)

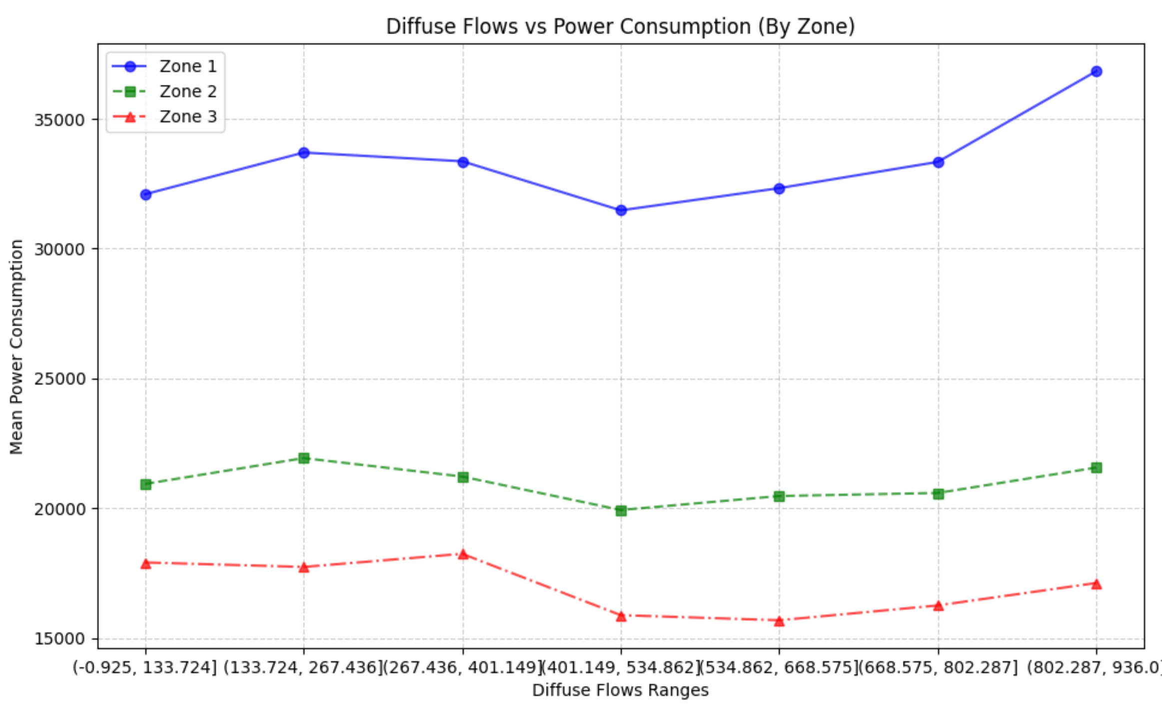

Figure 9: Diffuse Flows vs Power Consumption (By Zone)

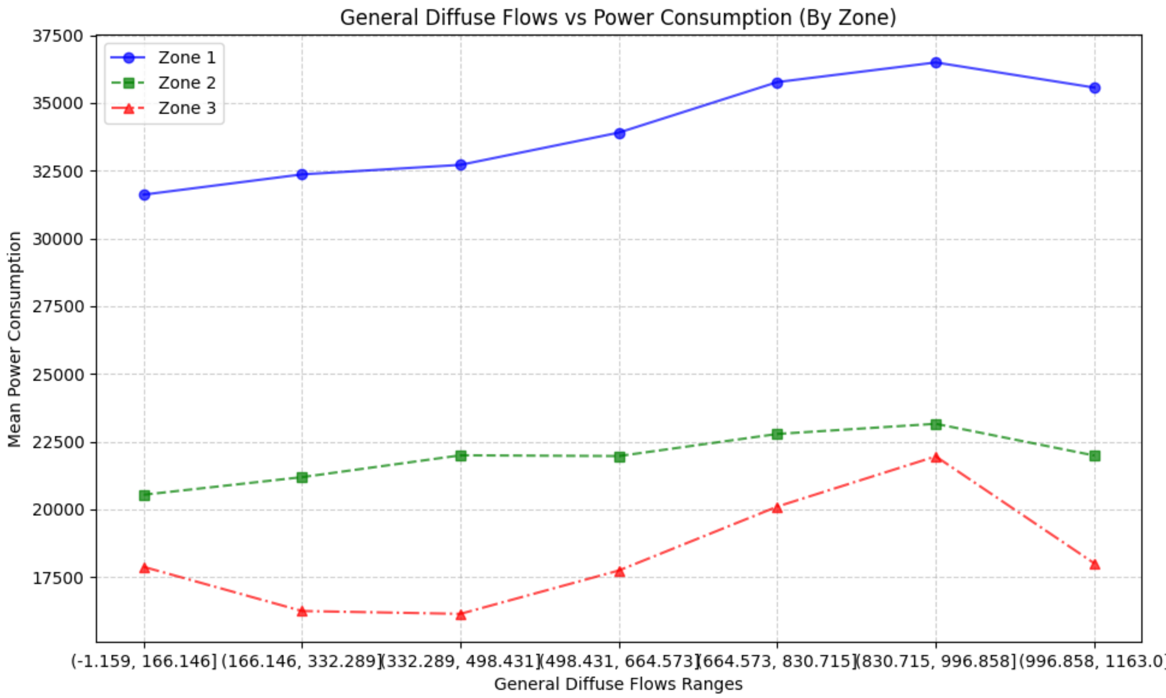

Figure 10: General Diffuse Flows vs Power Consumption (By Zone)

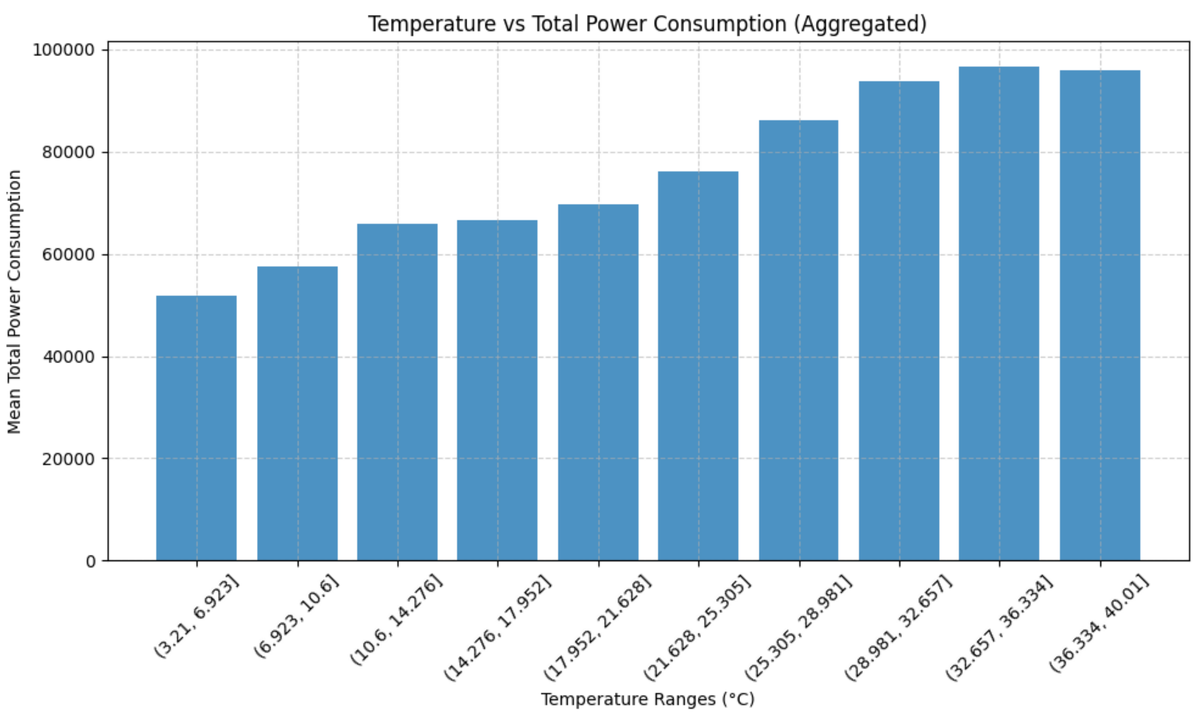

Figure 11: Temperature vs Total Power Consumption (Aggregated)

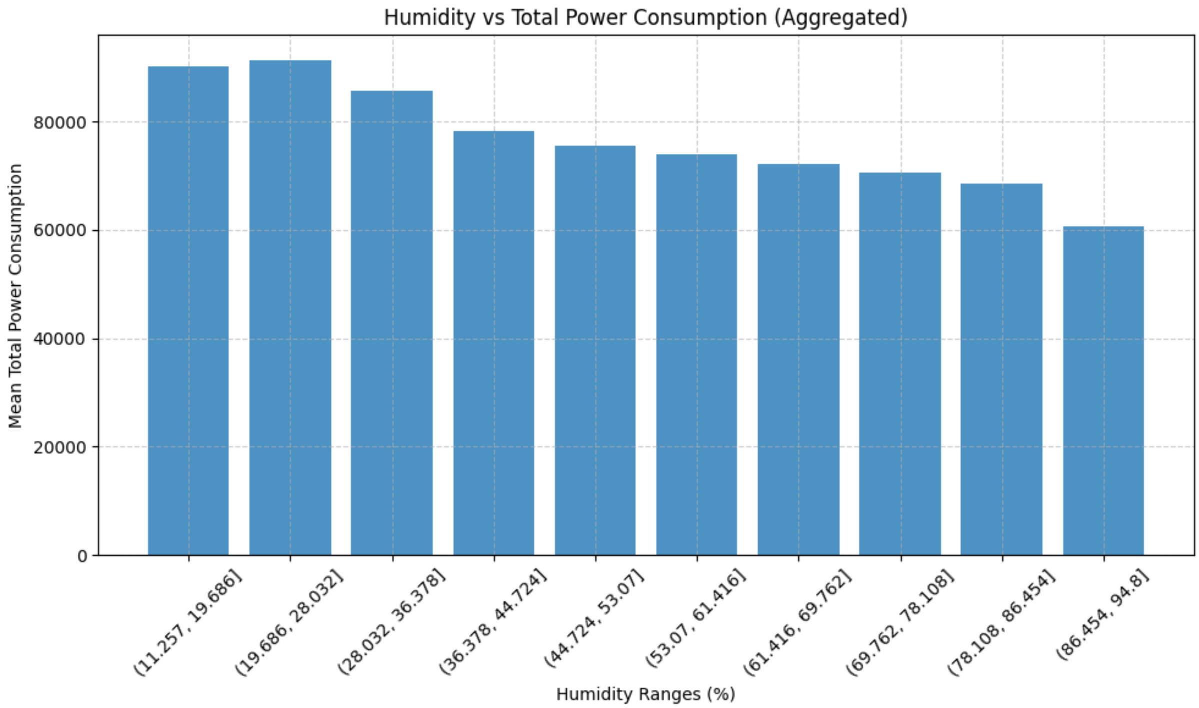

Figure 12: Humidity vs Total Power Consumption (Aggregated)

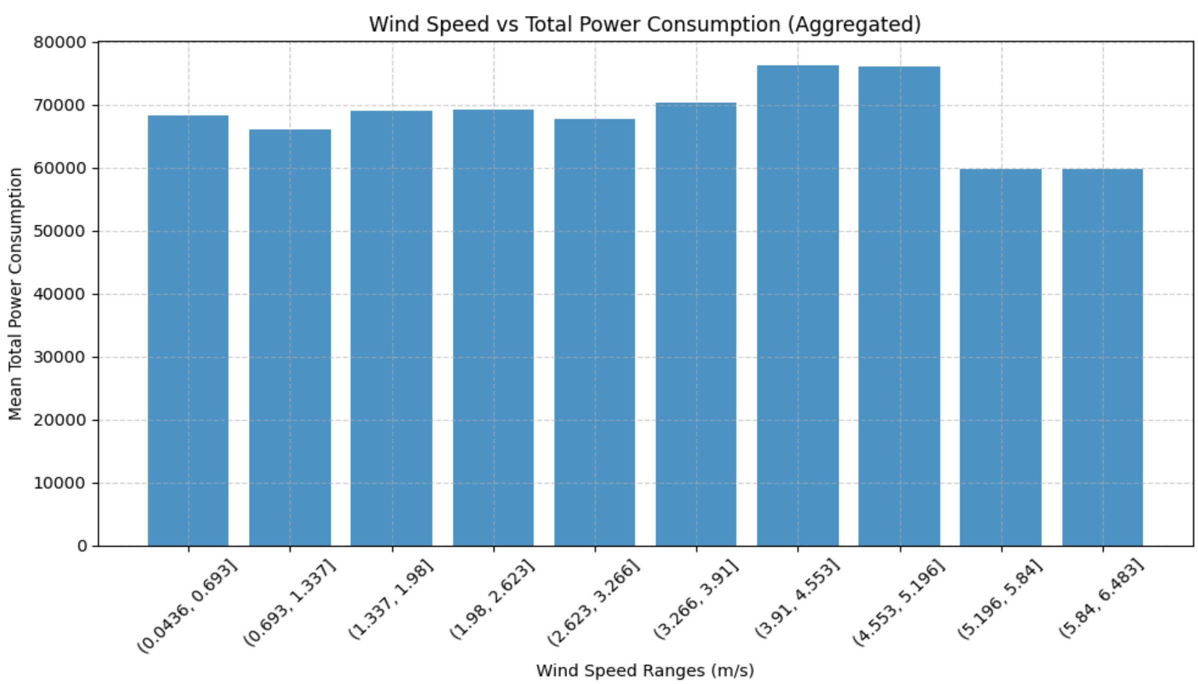

Figure 13: Wind Speed vs Total Power Consumption (Aggregated)

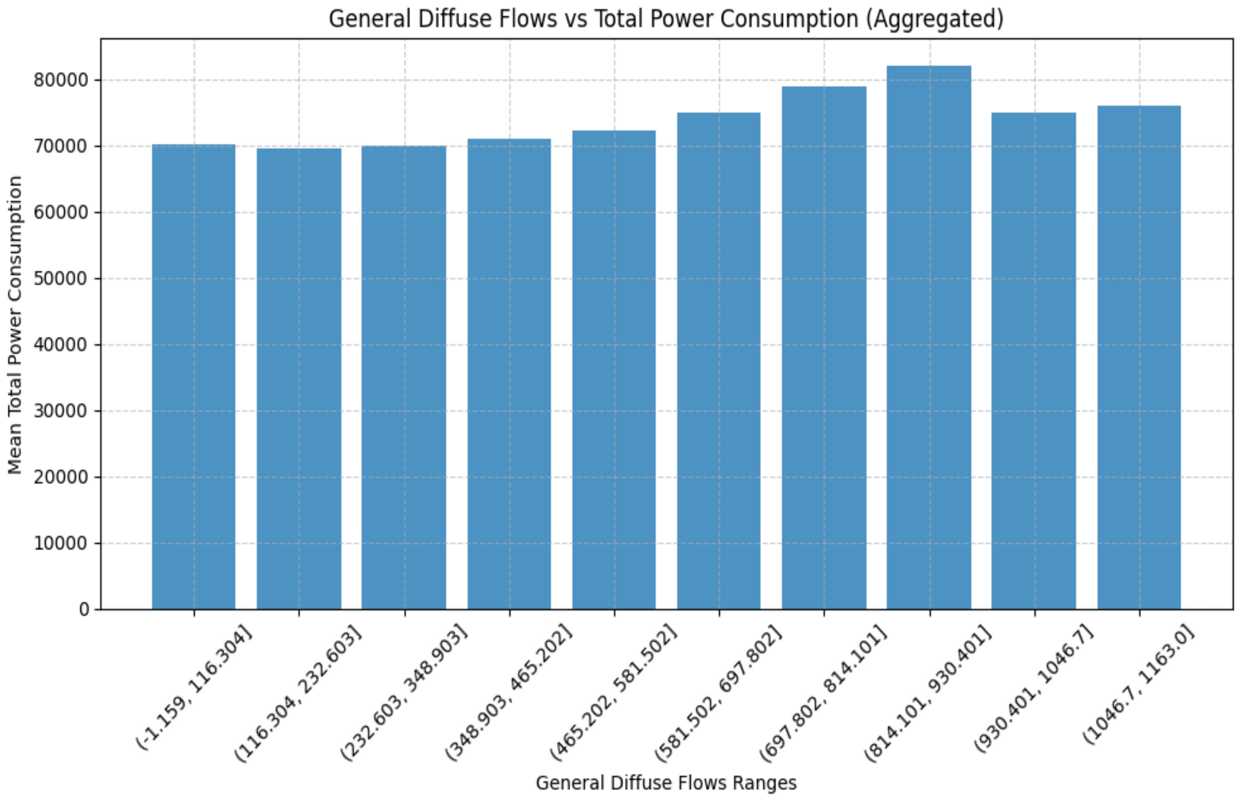

Figure 14: General Diffuse vs Total Power Consumption (Aggregated)

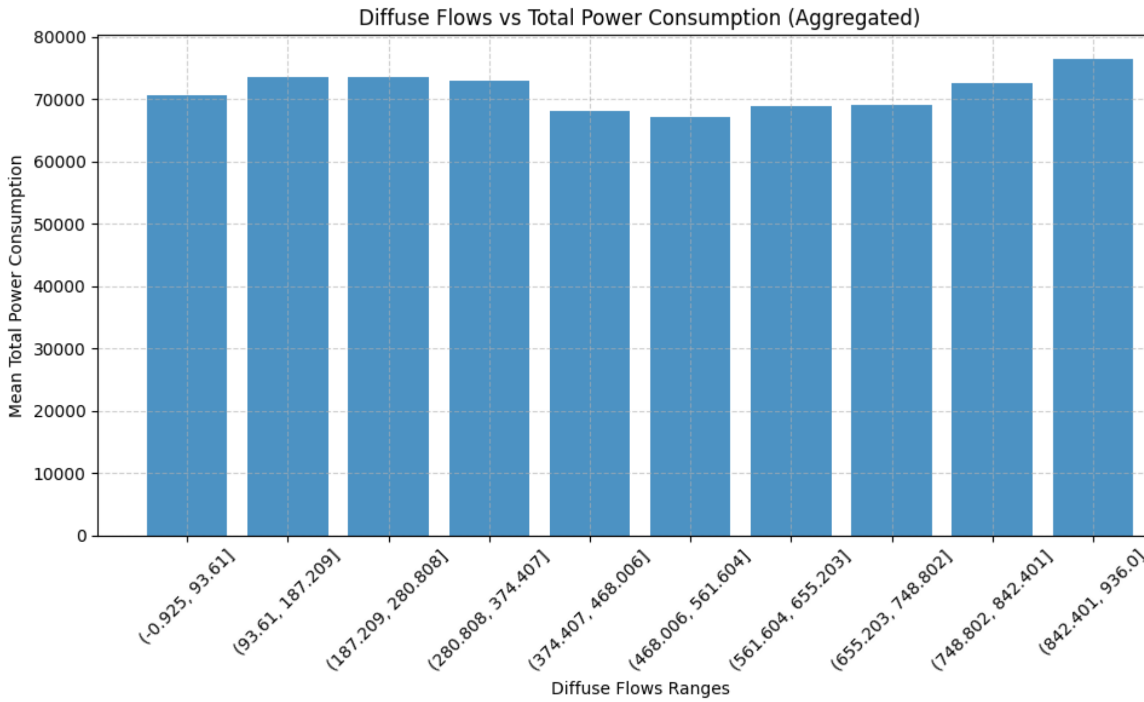

Figure 15: Diffuse vs Total Power Consumption (Aggregated)

G. Qualitative Impact:

Across all zones, clear seasonal trends emerge, with power consumption peaking during the summer months (June to August), potentially driven by cooling demands. Zone 1 consistently exhibits the highest energy usage, followed by Zone 2, with Zone 3 showing the lowest. This trend reflects the influence of industrial, commercial, and residential activities in each zone. Time-of-day variations reveal a common pattern: peak consumption occurs in the evening (6 PM to 9 PM) across all zones, while off-peak hours are observed during the early morning (2 AM to 6 AM). These temporal trends suggest the need for load balancing and dynamic pricing strategies to manage demand more effectively and reduce strain on the power grid during peak hours.

Temperature demonstrates a strong positive correlation with energy consumption across all zones, with Zone 3 showing the steepest increase. This indicates that rising temperatures significantly drive power usage, likely due to cooling systems in residential and commercial sectors. Conversely, humidity negatively correlates with power consumption, suggesting reduced energy demand in humid conditions, possibly linked to cooler weather or decreased outdoor activities. Wind speed exhibits a weak positive correlation with consumption, but its impact remains minimal, indicating limited direct influence on energy usage.

Diffuse and general diffuse solar radiation show a weak but consistent positive correlation with energy consumption, particularly in Zone 1. This relationship highlights an opportunity to harness solar energy during high-radiation periods to supplement traditional power sources. While Zones 2 and 3 show smaller variations, they could still benefit from targeted solar energy integration, especially during peak summer months.

Strong correlations between zones suggest shared energy demand patterns, with inter-zonal dependencies evident during both seasonal and time-of-day variations. For instance, Zones 1 and 2 display higher synchronization in energy consumption compared to Zone 3, emphasizing the need for region-specific load balancing strategies.

VI. Recommendations

H. Key Takeaways:

Based on these findings, the company should prioritize demand forecasting and resource allocation to address seasonal and temporal energy needs. Renewable energy integration, particularly solar power, can significantly reduce reliance on traditional sources during highdemand periods in summer. Wind energy projects should be explored in areas with higher wind speeds, such as parts of Zone 3. Additionally, dynamic pricing strategies can incentivize consumers to shift their energy usage to off-peak hours, reducing peak loads and enhancing grid stability.

The unique relationships between environmental variables and power consumption across zones underscore the necessity of tailored solutions. For instance, Zone 1's high and consistent demand calls for robust energy-saving programs, while Zone 3's variability highlights the need for adaptive strategies to address its fluctuating consumption. By leveraging these insights, the company can optimize its energy distribution, reduce operational costs, and enhance customer satisfaction, ensuring a sustainable and efficient power supply system.

I. Challenges Overcome:

This study faced several challenges that required careful consideration and problem-solving to ensure reliable and actionable results:

1. **Limited Data Duration:**
   The dataset provided only one year of SCADA data (2017), which may not fully capture long-term trends, such as multi-year seasonal variations or the impacts of major infrastructure changes. To address this limitation, the analysis focused on identifying patterns within the available data while acknowledging that longer-term studies could refine the findings.
2. **Data Preparation and Processing:**
   Preparing the dataset for analysis was a critical step. The raw SCADA data required transformation to make it suitable for descriptive analytics. Specifically:

- The timestamps were converted into meaningful components, such as hour of the day and month, to allow for detailed temporal trend analysis.
- The dataset was restructured to use the DateTime column as the index, enabling more efficient manipulation and time-based grouping.

3. **Data Granularity and Consistency:**
   The SCADA data, recorded every 10 minutes, provided a high level of granularity. While this richness allowed for detailed analysis, it also increased computational demands and complexity. Aggregating data into appropriate time intervals for specific analyses (e.g., hourly or monthly) was a challenge that required balancing detail with interpretability.
4. **Handling Environmental Variables:**
   The environmental data included in the analysis, such as temperature, humidity, and wind speed, exhibited varying levels of correlation with energy consumption. In the absence of domain expertise, statistical techniques and exploratory data analysis were employed to identify meaningful relationships among these variables and to address potential noise in the dataset.
5. **Visualization Clarity:**
   Presenting the results in a clear and actionable manner posed another challenge. Creating visualizations that effectively communicated trends while avoiding misinterpretation required careful design and iterative refinement.

By overcoming these challenges, the study successfully delivered insights that addressed Amendis's operational inefficiencies and supported actionable recommendations for optimizing Tetouan City's power distribution network.