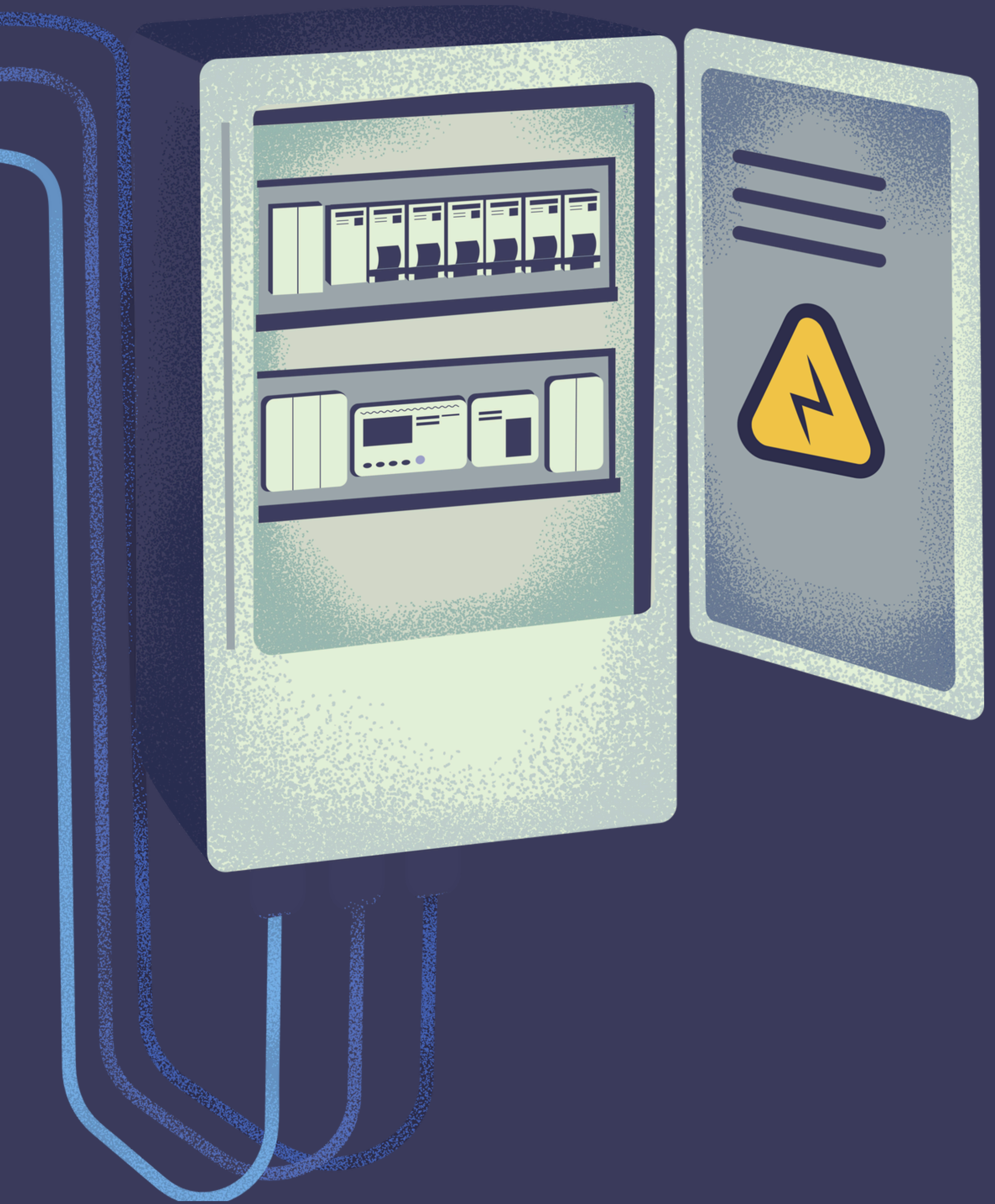
Presented by Rana Alaa

Energy Consumption

Solutions for devising strategies to optimize energy usage.

Introduction
Unveiling Patterns in Energy Consumption: (A Data-Driven Approach)



Energy consumption is a critical factor influencing both operational costs and environmental sustainability.
Understanding consumption patterns allows for data-driven decision-making to enhance efficiency.
This project explores energy usage across various regions, times, and conditions to identify optimization opportunities.
By leveraging data analytics, the study aims to uncover actionable insights for reducing energy costs and improving resource allocation.
The ultimate goal is to promote sustainable practices while ensuring economic benefits for stakeholders.

Dataset Overview:

- Energy Consumption.csv:

This dataset contains energy consumption data collected from network sites, to help us monitor and
track sites' energy consumption, the dataset has 1,380,252 rows and 5 columns:

• DateTime
• Site_id: unique site identifier contains more than one cell.
• Cell_id: unique cell identifier.
• Region: geographical area that identifies site location.
• KWH/hh (per half hour): Energy consumed in half hour intervals.

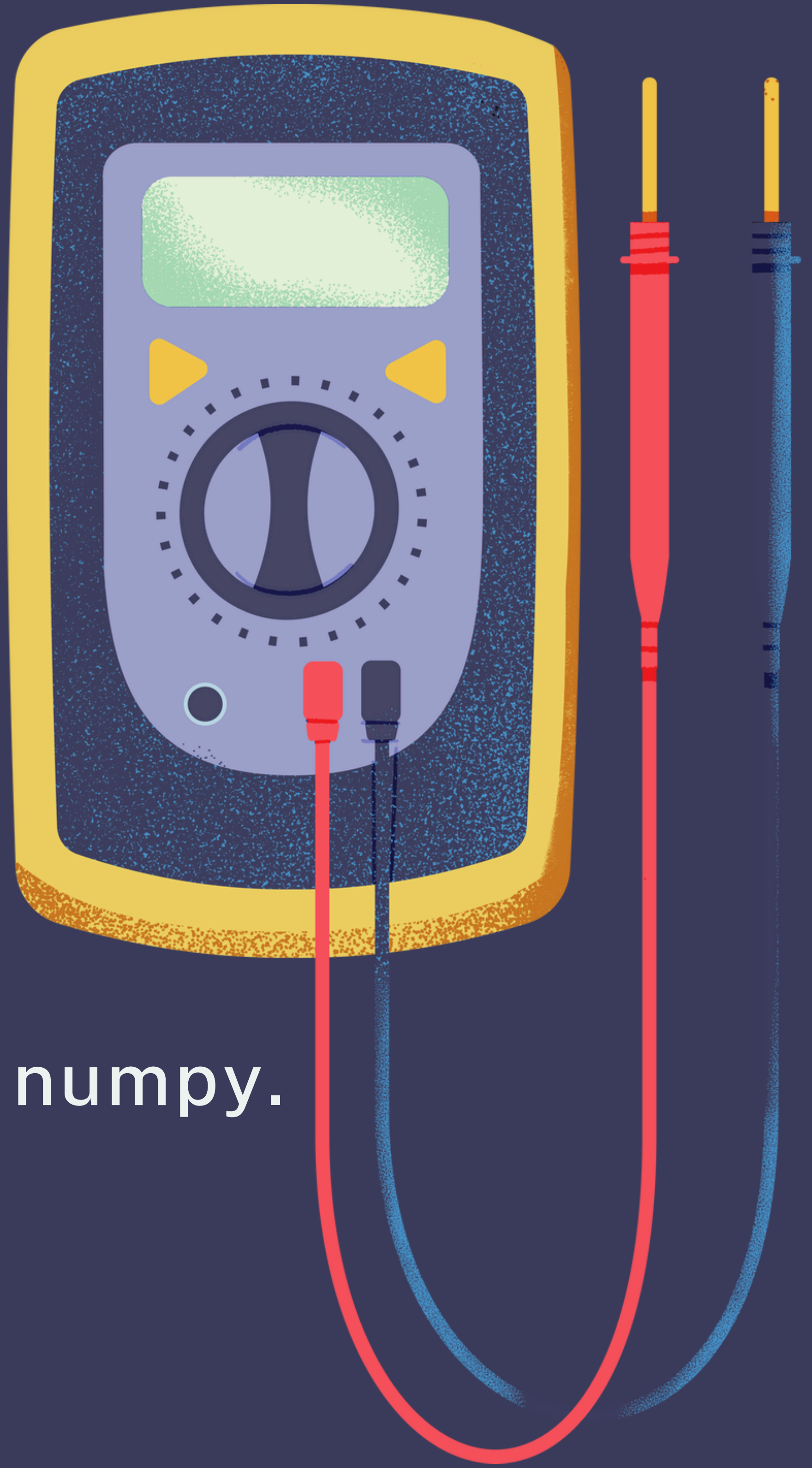
- Power Demand.csv:

This dataset contains energy consumption demand every half hour interval, the dataset has 17,520 rows and 2 columns:

• DemandDateTime
• Demand: energy consumption demand for the time interval

Methodology:

- Steps Followed:
  1. Data Cleaning and Preprocessing.
  2. Exploratory Data Analysis (EDA).
  3. Pattern Identification and Insights Extraction.
  4. Optimization and Recommendations.
- Tools Used:
  Python: Pandas, Matplotlib, Seaborn, os, scipy.stats, f_oneway, numpy.



Data Cleaning and Preprocessing:

1- Energy DataFrame:

- The dataset consists of 1,380,252 rows and 5 columns, covering energy consumption data.
- No null values were present in the dataset after initial inspection.
- Columns include cell_id, DateTime, KWH/hh (per half hour), site_id, and region.
- The DateTime column initially required conversion to a proper datetime format for analysis.
- Identified and removed 942 duplicate rows to ensure data integrity.
- Converted the DateTime column into a datetime format for consistency and ease of time-series analysis.
- Invalid datetime values were safely handled and set to NaT (Not a Time).

	cell_id	DateTime	KWH/hh (per half hour)	site_id	region
0	MAC000002	2013-01-01 00:00:00	0.219	A	A
1	MAC000002	2013-01-01 00:30:00	0.241	A	A
2	MAC000002	2013-01-01 01:00:00	0.191	A	A
3	MAC000002	2013-01-01 01:30:00	0.235	A	A
4	MAC000002	2013-01-01 02:00:00	0.182	A	A
5	MAC000002	2013-01-01 02:30:00	0.229	A	A
6	MAC000002	2013-01-01 03:00:00	0.194	A	A
7	MAC000002	2013-01-01 03:30:00	0.201	A	A

2- Demand DataFrame:

- The Demand.xlsx dataset contains 17,520 rows and 2 columns.
- No null values or duplicates were found in the dataset.
- Initial inspection revealed inconsistent values in the Demand column:
  - Variations included 'Normal', 'Low', 'Normall', 'High', 'normal', 'high'.
- Unique values were standardized using:
  - Capitalization for consistency.
  - Typo correction: 'Normall' was replaced with 'Normal'.
- Standardization
  - The Demand column now contains consistent values: ['Normal', 'Low', 'High'].
- This ensures reliable categorization for subsequent analysis.

	DemandDateTime	Demand
0	2013-01-01 00:00:00	Normal
1	2013-01-01 00:30:00	Normal
2	2013-01-01 01:00:00	Normal
3	2013-01-01 01:30:00	Normal
4	2013-01-01 02:00:00	Normal
5	2013-01-01 02:30:00	Normal

3- Merged DataFrame:

- Merged the energy usage data (energy_df) with demand data (demand) using an inner join on datetime columns:
  DateTime from energy_df.
  DemandDateTime from demand.
- Ensured rows are synchronized based on timestamps, resulting in a comprehensive dataset for analysis.
- Dropped unnecessary columns after merging to simplify the dataset structure.
- Cost Calculation:
  Defined a pricing dictionary for demand categories:
  High: $77.21, Normal: $19.46, Low: $6.89.
  Computed electricity cost using:
  Cost=Demand Price×KWH/hh (per half hour)
- Extract day of the week from DateTime to identify weekday patterns (0 = Monday, 6 = Sunday)
- Extract hour of the day to study hourly variations in energy usage.
- Created a binary feature (is_weekend) to distinguish between weekdays (0) and weekends (1).
- Time of day : Categorized hours into time periods:
  Morning: 0–11, Afternoon: 12–17, Evening: 18–23.

	cell_id	DateTime	KWH/hh (per half hour)	site_id	region	Demand	Cost	day_of_week	hour	is_weekend	time_of_day
0	MAC000002	2013-01-01 00:00:00	0.219	A	A	Normal	4.26174	1	0	1	Morning
1	MAC000002	2013-01-01 00:30:00	0.241	A	A	Normal	4.68986	1	0	1	Morning
2	MAC000002	2013-01-01 01:00:00	0.191	A	A	Normal	3.71686	1	1	1	Morning
3	MAC000002	2013-01-01 01:30:00	0.235	A	A	Normal	4.57310	1	1	1	Morning
4	MAC000002	2013-01-01 02:00:00	0.182	A	A	Normal	3.54172	1	2	1	Morning

Data Insights:

1- Energy Consumption and Cost Analysis per Site

- **Total Energy Consumption per Site:**

Calculated by summing the KWH/hh (per half hour) for each site.

Findings:

Site E has the highest energy consumption (33,243.78 kWh).

Site U has the lowest energy consumption (2,706.34 kWh).

This shows a significant variation in energy usage between sites, potentially reflecting differences in size, activity, or operational hours.

- **Total Cost per Site:**

Determined by summing the Cost column for each site.

Findings:

Site E incurs the highest cost (677,459.84 currency units).

Site U has the lowest cost (56,098.60 currency units).

The cost aligns proportionally with energy consumption, as expected, due to direct correlation.

- **Sites with High Energy Usage:**

Sites like E, H, D, and Q consume significantly more energy and have correspondingly higher costs. These sites may benefit from targeted energy efficiency strategies to reduce costs.

- **Sites with Low Energy Usage:**

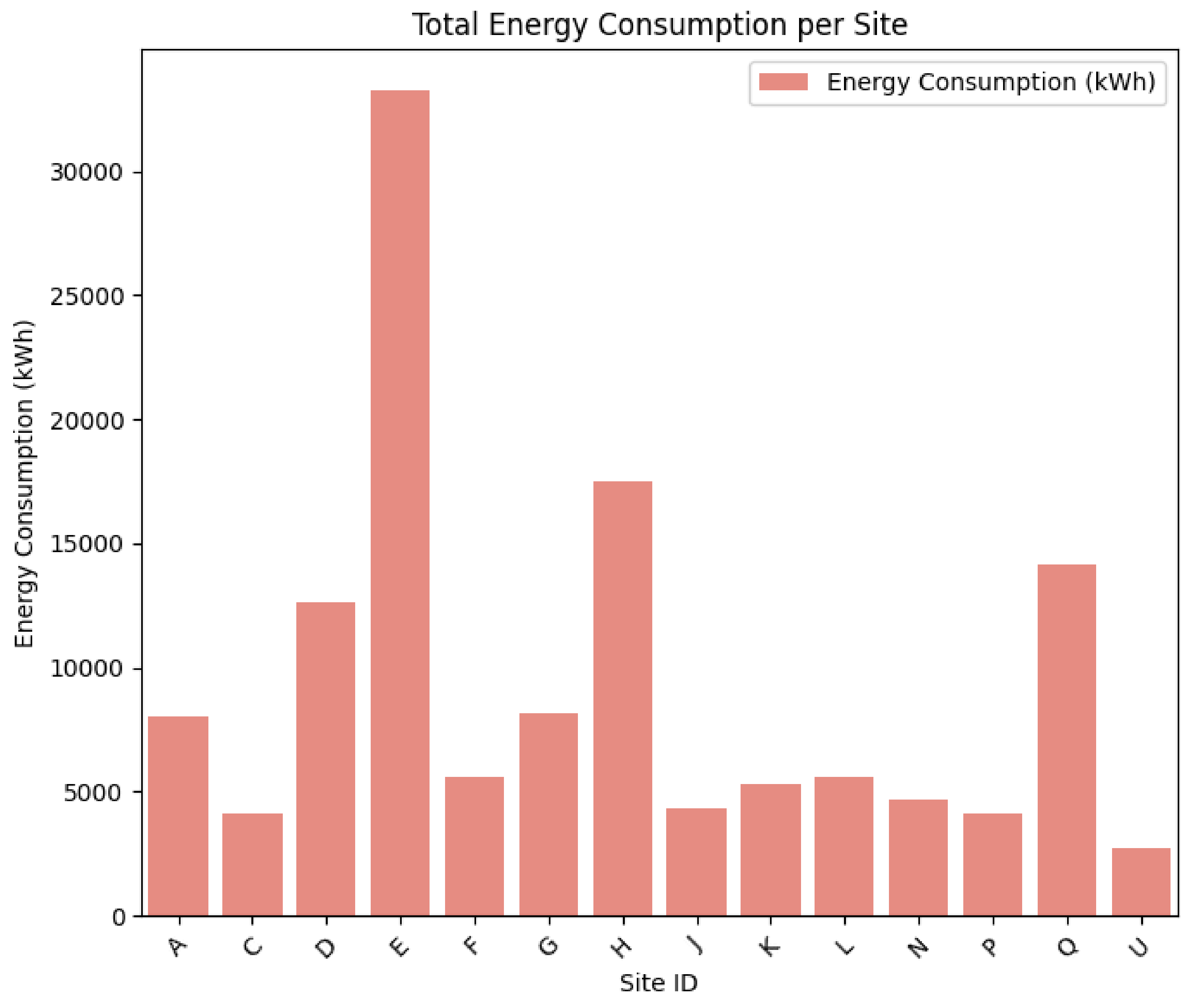
Sites like U, P, and N have minimal energy usage and costs, indicating efficient operations or lower activity levels.

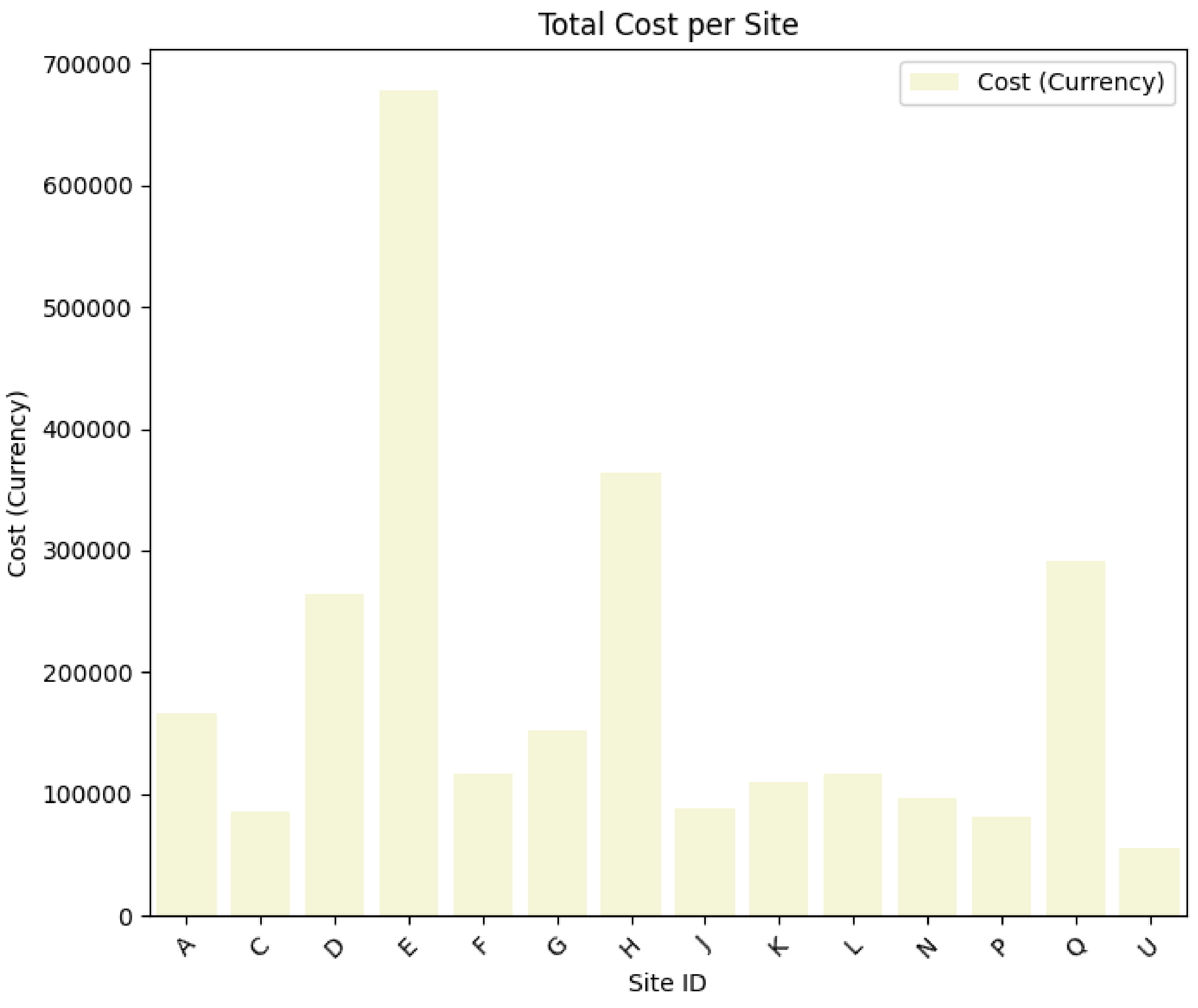
- **Actionable Insight:**

Identify factors driving high consumption in specific sites.
Explore optimization opportunities to reduce costs, especially in high-consuming sites like E and H

- **Conclusion:**

Sites with high energy consumption contribute the most to overall costs, presenting clear opportunities for energy-saving initiatives.





2- Energy Consumption and Cost Analysis by Region:

- **High-Consumption Regions:**

Regions A, B, and C show significantly higher energy usage, leading to increased costs.
Efforts to reduce energy waste or optimize usage in these regions could result in substantial cost savings.

- **Low-Consumption Region:**

Region D stands out with minimal consumption and cost, potentially indicating smaller-scale activities or efficient energy practices.

- **Correlations:**

The cost aligns proportionally with energy consumption across all regions, suggesting uniform pricing structures.

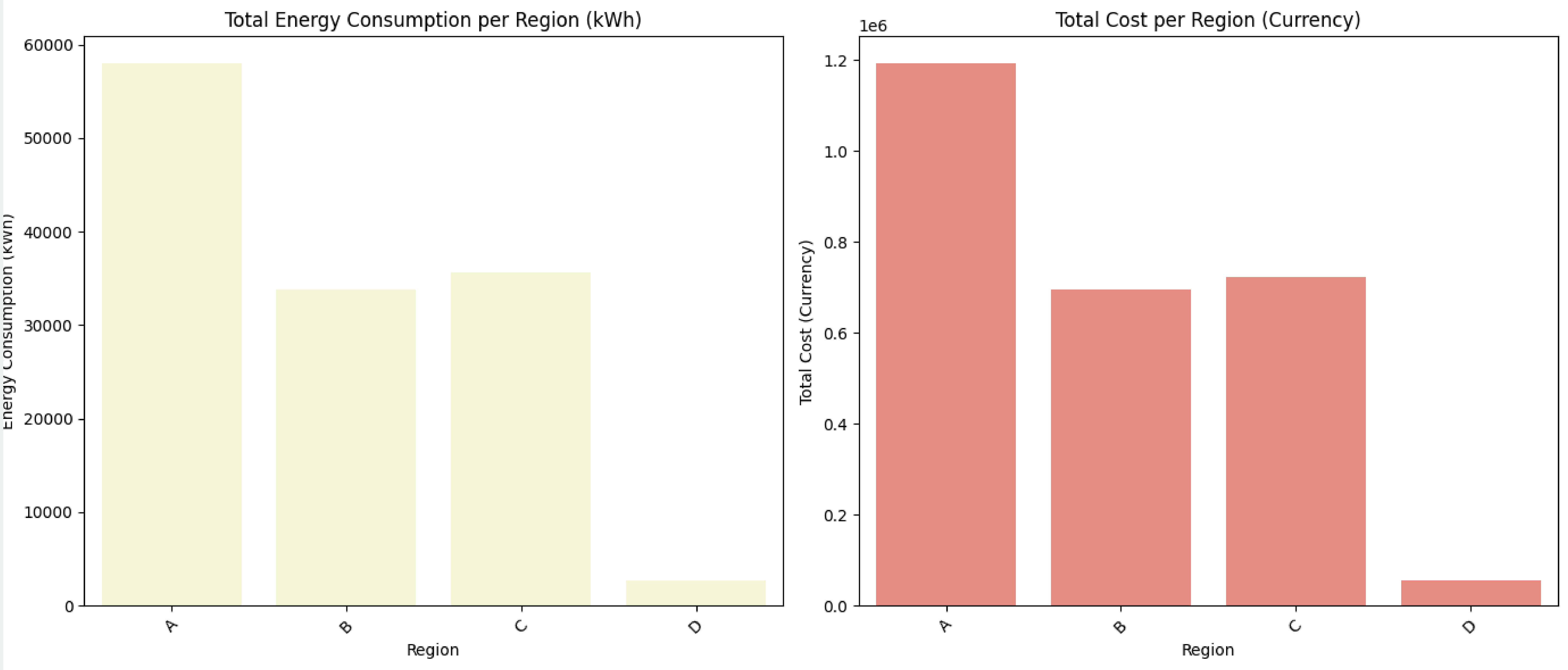
- **Suggested Actions:**
- **For Region A:**

Conduct energy audits to identify inefficiencies.
Introduce energy-saving measures for large-scale operations.

- **For Regions B and C:**

Assess industrial or residential consumption patterns.
Implement region-specific energy efficiency programs.

- **Conclusion:**
  The analysis provides a clear overview of energy consumption and costs at the regional level, enabling targeted interventions to enhance efficiency and reduce expenditures.



3- Average Energy Consumption by Time of Day:

- **Key Insights:**
- **Energy Consumption Patterns:**

The pie chart shows the average energy usage across three key time periods: Morning, Afternoon, and Evening.

Evening accounts for the largest slice of energy consumption, indicating peak energy usage during this time.

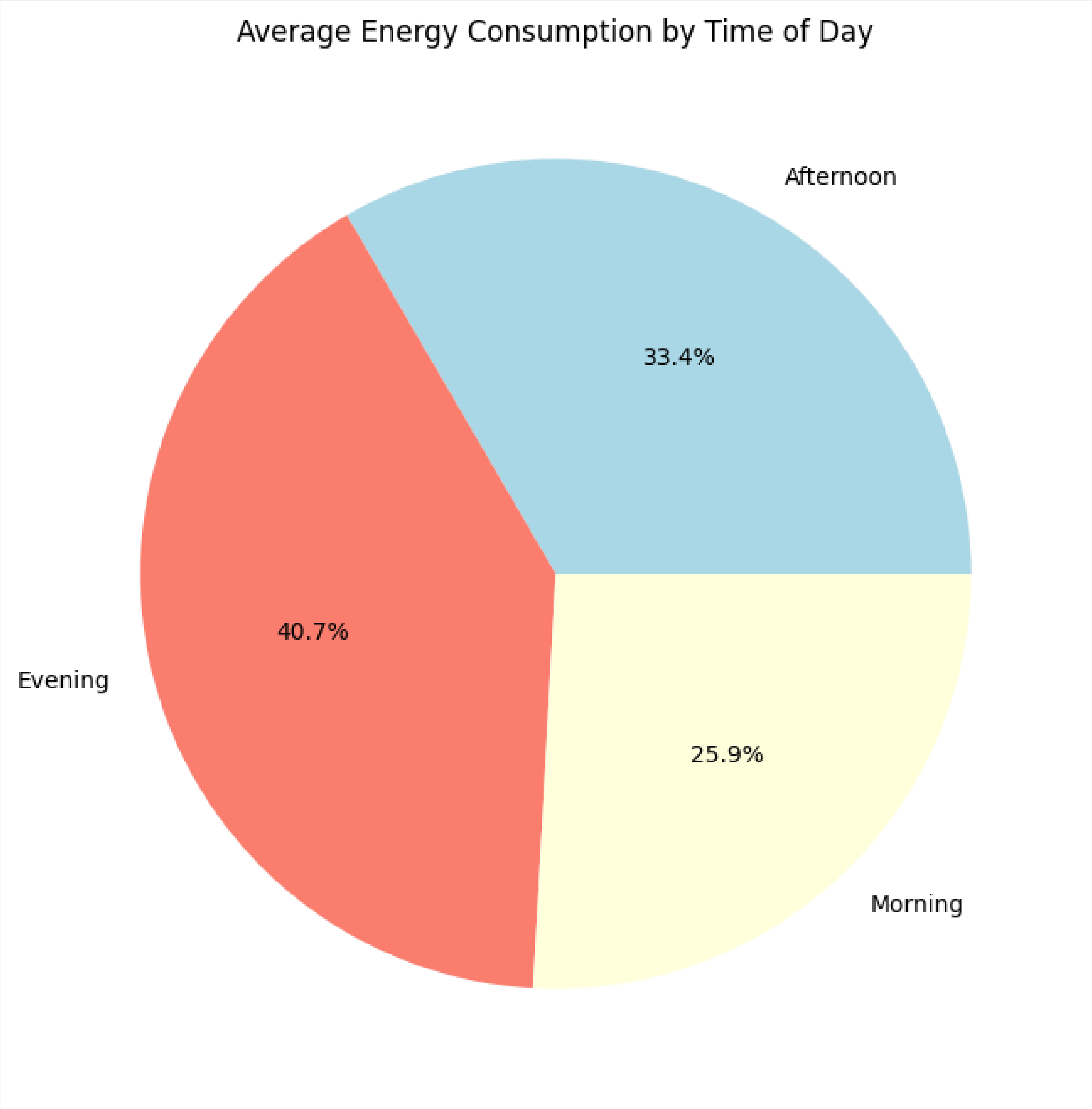
- **Significance of Evening Peak:**

High evening energy usage suggests increased activity during this period, such as residential lighting, appliance use, and commercial operations transitioning into off-peak hours.

- **Opportunities for Optimization:**

The evening peak provides a strong case for implementing demand-side energy management strategies.

- **Conclusion:**
  By targeting the Evening peak with tailored interventions, we can achieve:
  Cost savings for users.
  Reduced stress on the power grid.
  Enhanced sustainability in energy usage.
  These strategies can significantly contribute to a greener and more resilient energy system.



4- Energy Consumption: Weekdays vs. Weekends:

- **Key Insights:**
- **Energy Usage Distribution:**

The pie chart demonstrates a significant difference in energy consumption between weekdays and weekends.

Weekends account for the majority of energy consumption, with 85.1%, while weekdays contribute only 14.9%.

- **Behavioral Implications:**

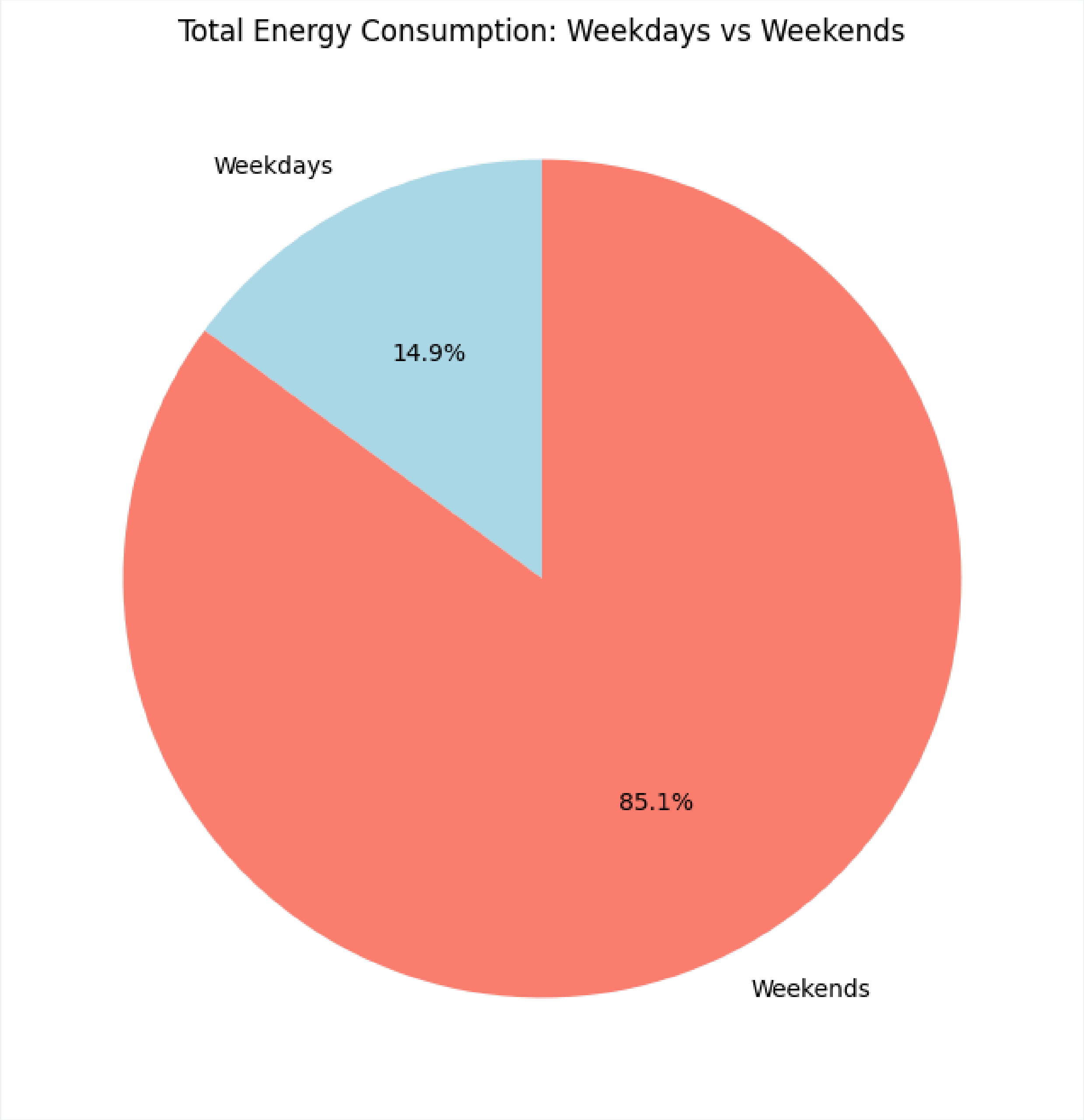
Higher energy usage during weekends indicates increased household activities, such as cooking, laundry, entertainment, and other home-centric tasks.

In contrast, reduced energy consumption on weekdays reflects time spent outside homes due to work or school.

- **Conclusion:**

The stark contrast in energy consumption between weekdays and weekends underscores the need for customized energy management strategies.

Focusing on weekend consumption offers a valuable opportunity to reduce energy demand and encourage sustainable usage patterns.



5- Energy Consumption by Day of the Week:

- **Consumption Peaks:**

The energy consumption is highest on Day 1 (assumed to be Monday), peaking above 20,500 kWh. Day 0 (Sunday) also shows relatively high consumption close to 19,400 kWh.

- **Consumption Drop:**

A sharp decline is visible on Day 2 (Tuesday), with energy consumption dropping to its lowest level around 17,300 kWh.

- **Mid-Week Stability:**

Between Days 3 to 5 (Wednesday to Friday), the energy consumption gradually increases but remains stable, fluctuating between 17,500 kWh and 18,100 kWh.

- **Recovery by Day 6:**

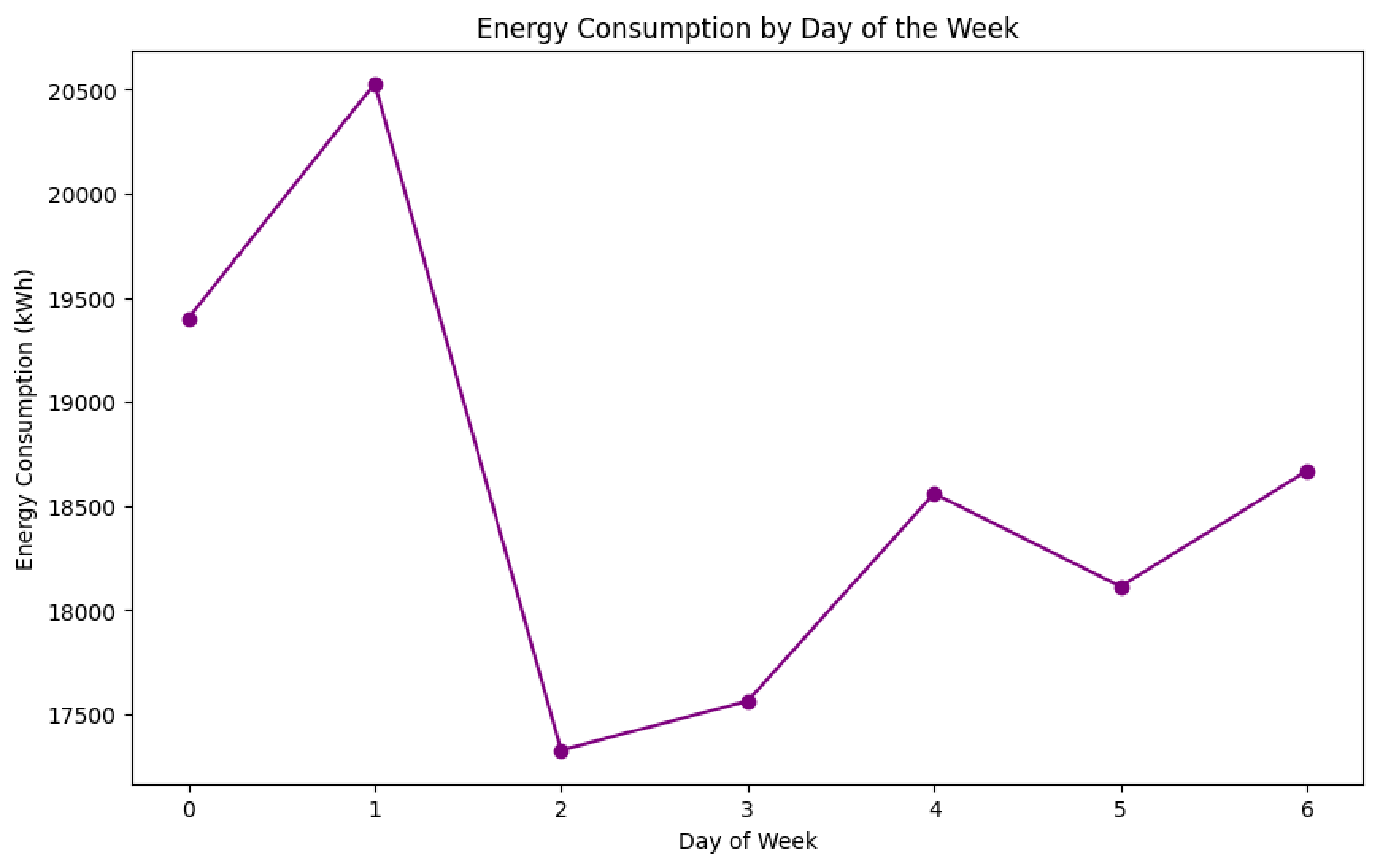
On Day 6 (Saturday), energy consumption rebounds slightly to around 18,600 kWh, indicating moderate usage.

- **Analysis:**
- **High Consumption (Sunday & Monday):**

Increased household activities at the start and end of the week might explain the peaks. Higher energy usage could result from tasks postponed for weekends or heavy weekday preparations.

- **Lowest on Tuesday:**

Possibly due to reduced household activities or work/school routines stabilizing mid-week.



5- Energy Consumption: Weekdays vs Weekends by Time of Day:

- **Key Trends:**

Evening Consumption: Energy consumption in the evening is consistently higher than in the morning and afternoon for both weekdays and weekends.

Afternoon Consumption: Energy usage in the afternoon shows a moderate level compared to the evening.

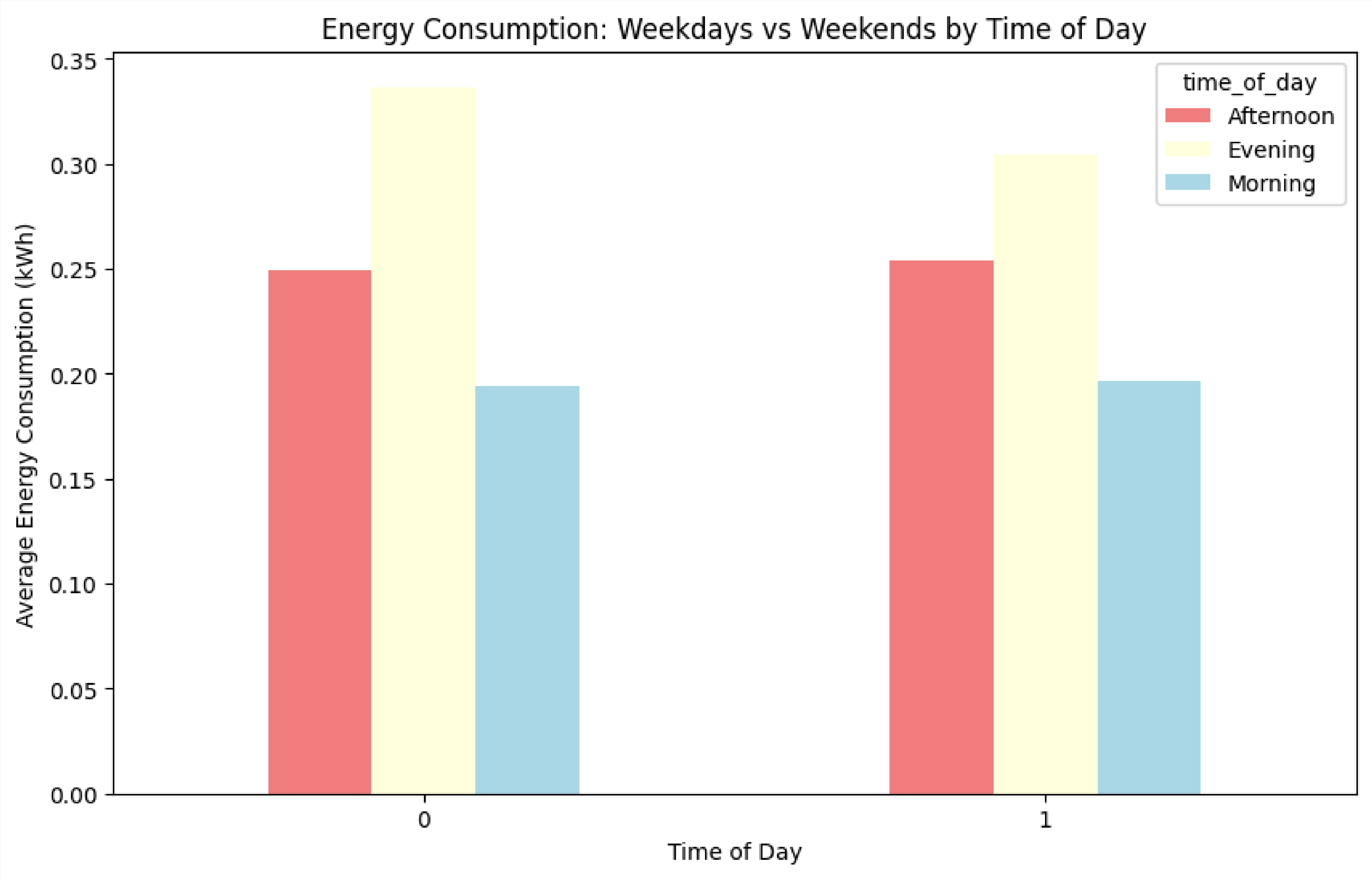
Morning Consumption: Morning hours consistently have the lowest energy consumption across both weekdays and weekends.

- **General Pattern:**

Evening hours dominate energy usage, which aligns with typical household activities like lighting, cooking, and entertainment.

Weekends show marginally higher energy consumption during evening hours, reflecting increased household presence and activities.

- **Key Observations:**
Evening Peaks: Both weekdays and weekends experience peak energy consumption during the evening.
Minimal Morning Use: Mornings exhibit the least energy demand, possibly because people are away or activities are minimal.
Weekend Increase: The slight increase in evening consumption on weekends indicates extended leisure or household activities.



6- Average Energy Cost by Hour of Day:

- **Peak Hours for Cost:**

The highest average energy cost occurs between 18:00 (6 PM) and 20:00 (8 PM), peaking at 8.2. This suggests a significant increase in energy demand during early evening hours, possibly due to household activities like cooking, lighting, and appliance use after work.

- **Lowest Cost Periods:**

The cost is at its lowest between 3:00 AM to 5:00 AM, hovering around 2.7 to 2.8. This is likely due to reduced energy usage during late-night hours when most households are inactive.

- **Rising Trend in Early Morning:**

Starting from 6:00 AM, there is a gradual increase in cost, indicating the start of daily activities. Costs spike noticeably around 8:00 AM, coinciding with morning routines.

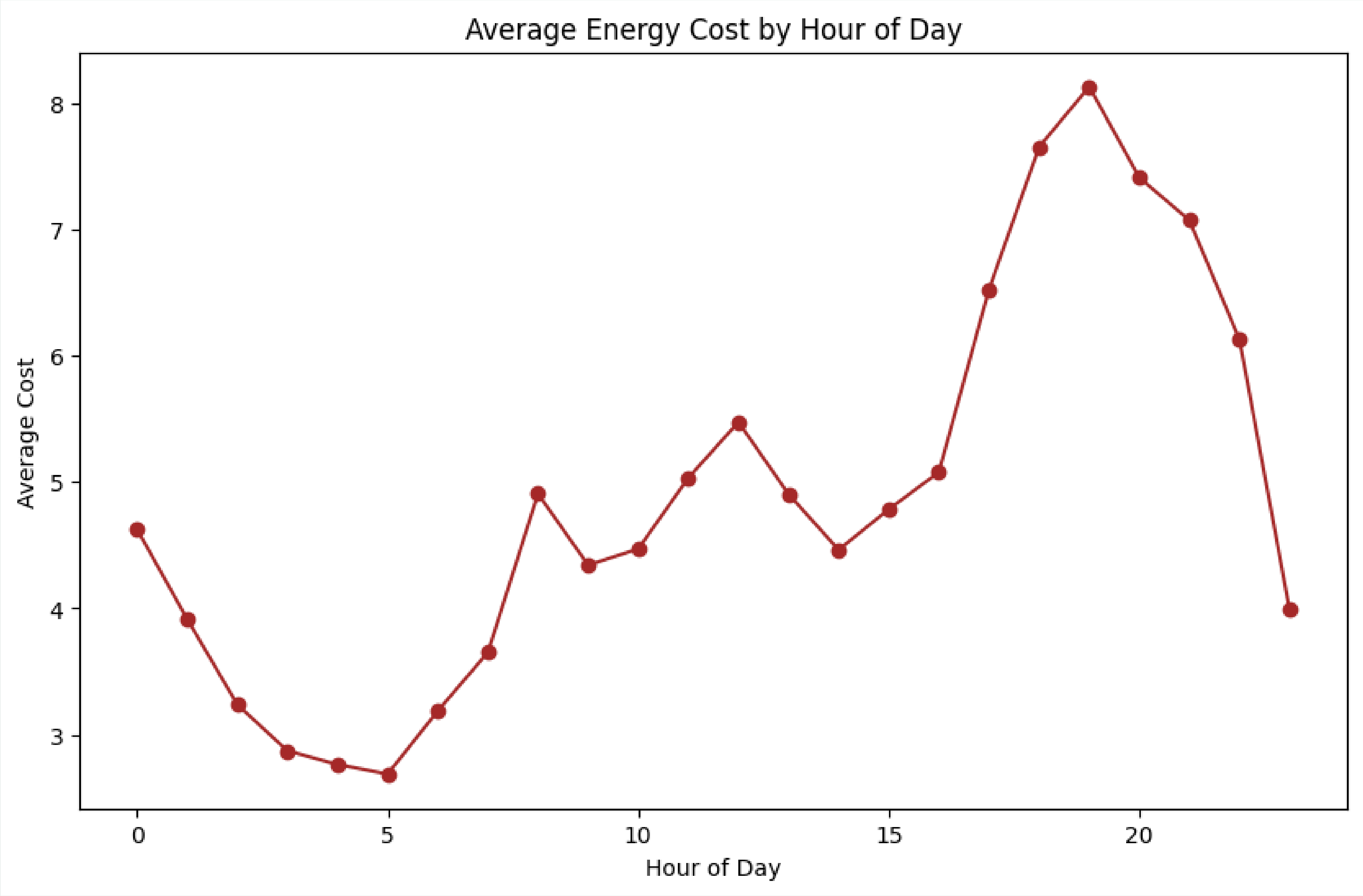
- **Midday Stability:**

Between 10:00 AM and 3:00 PM, the energy cost remains relatively stable, fluctuating between 4.3 and 5.5.

- **Key Observations:**

Evening Peaks: Higher energy demand during evening hours causes a sharp cost increase. Targeting these hours for energy-saving measures is crucial.

Off-Peak Opportunities: Late-night hours (3 AM–5 AM) offer lower energy costs, ideal for activities like running energy-intensive appliances (e.g., washing machines).



Impact of Region on Energy Consumption: ANOVA Analysis:

1. **Does the region significantly affect energy consumption?**

- **Hypotheses Formulation:**

H0 (Null Hypothesis): The region does not significantly affect energy consumption (means are equal).

Ha (Alternative Hypothesis): The region significantly affects energy consumption (means differ).

- **Test Selection and Justification:**

The test compares the means of energy consumption across multiple groups (regions). Since we have categorical data (regions) and continuous data (energy consumption), ANOVA is appropriate.

- **Steps Performed:**

1. Group energy consumption data by region.

2. Apply the One-Way ANOVA test using f_oneway to compare the means.

3. Interpret the results based on the F-statistic and p-value.

- **Results of ANOVA Test:**

1. F-statistic = 38.83:

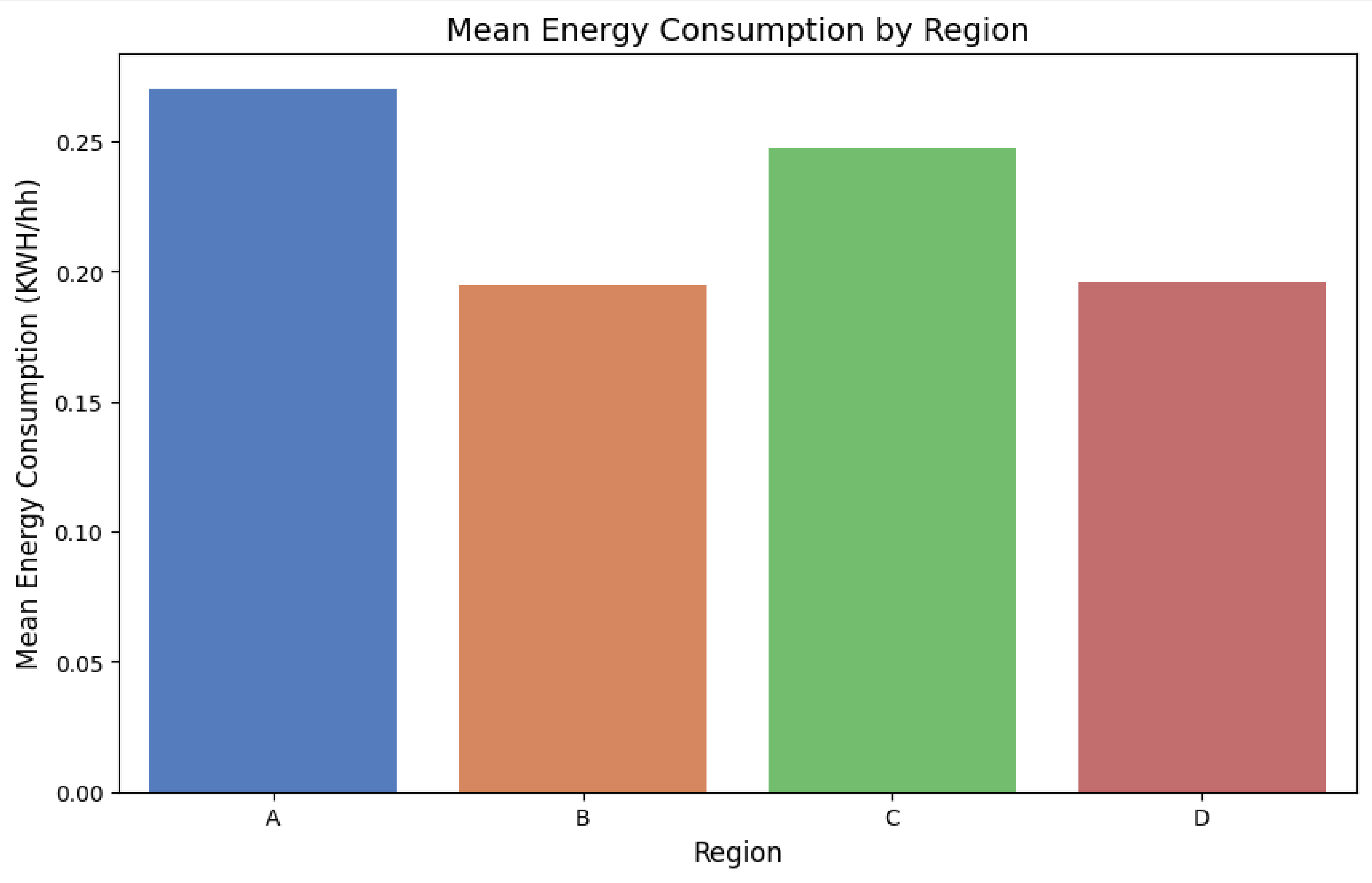
This value reflects the ratio of variance between the groups (regions) to the variance within the groups.

2. P-value = 4.39e-25

This value is much smaller than 0.05.

It indicates strong evidence to reject the null hypothesis (H0).

- **Conclusion:**
  Reject the null hypothesis (H0): There is a statistically significant difference in energy consumption between regions.
  The alternative hypothesis (Ha): confirms that the region has a significant effect on energy consumption.
- **What does this mean practically?**
  Different regions exhibit different energy consumption patterns.
  The reason for this difference could be geographic, economic, or demographic factors.



Impact of Region on Energy Consumption: ANOVA Analysis:

2. **Is there a relationship between the time of day and energy consumption?**

- **Hypotheses Formulation:**

Null Hypothesis (H0): There is no relationship between the time of day and energy consumption.
Alternative Hypothesis (Ha): There is a statistically significant relationship between the time of day and energy consumption.

- **Test Selection and Justification:**

The One-Way ANOVA test is appropriate when comparing the means of three or more independent groups (in this case, the different times of day: Morning, Afternoon, Evening) to see if there is a significant difference in energy consumption.

- **Steps Performed:**

1. Group energy consumption data by time of day.
2. Apply the One-Way ANOVA test using f_oneway to compare the means.
3. Interpret the results based on the F-statistic and p-value.

- **Results of ANOVA Test:**

1. F-statistic = 119.14:
indicates a significant difference in energy consumption between the times of day.
2. P-value = 1.85e-52
This value is much smaller than 0.05.
This provides strong evidence to reject the null hypothesis, meaning there is a significant difference in energy consumption based on the time of day.

- **Conclusion:**

Reject the null hypothesis (H0): There is a statistically significant relationship between the time of day and energy consumption.

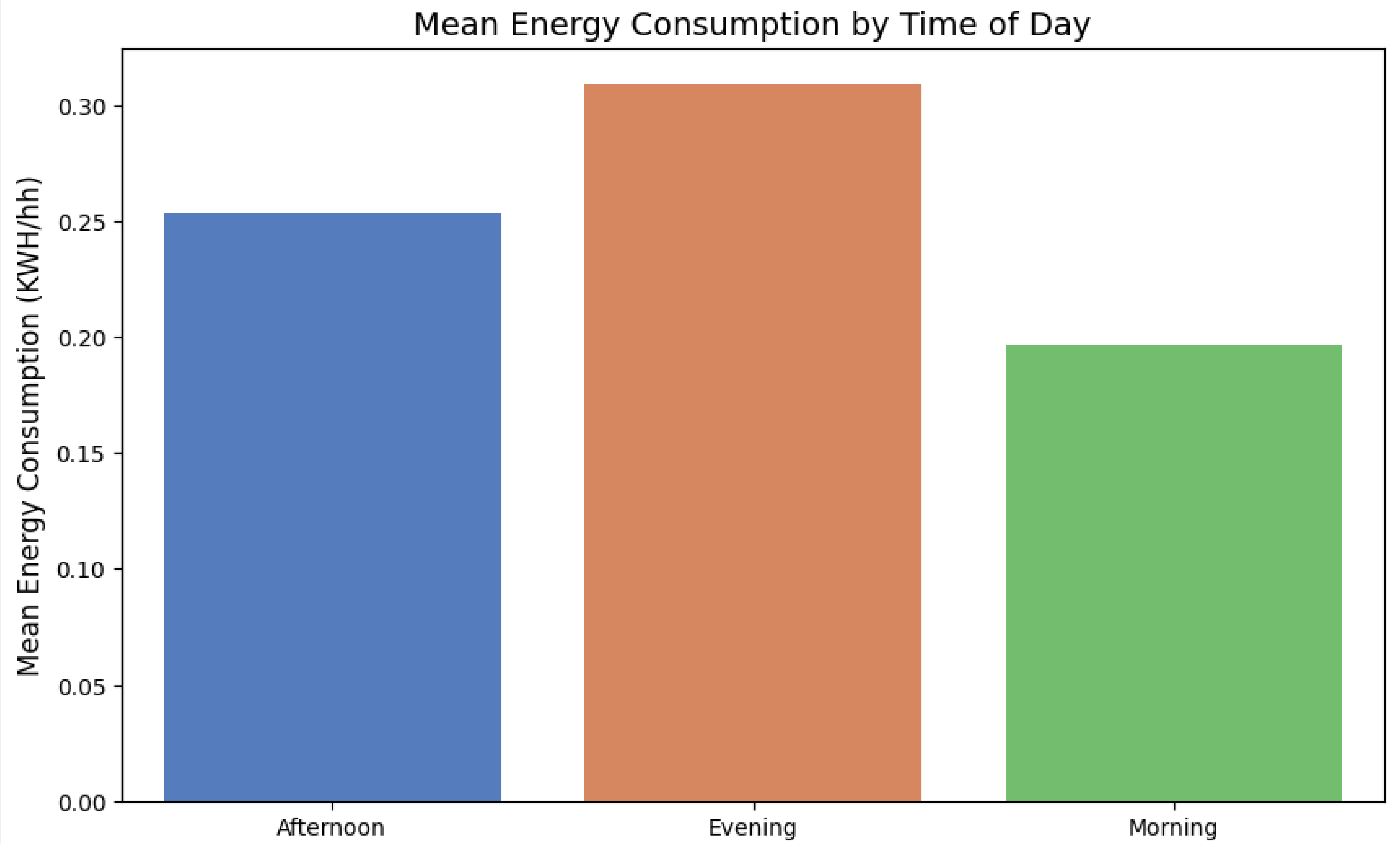
The alternative hypothesis (Ha): confirms that the time of day significantly affects energy consumption.

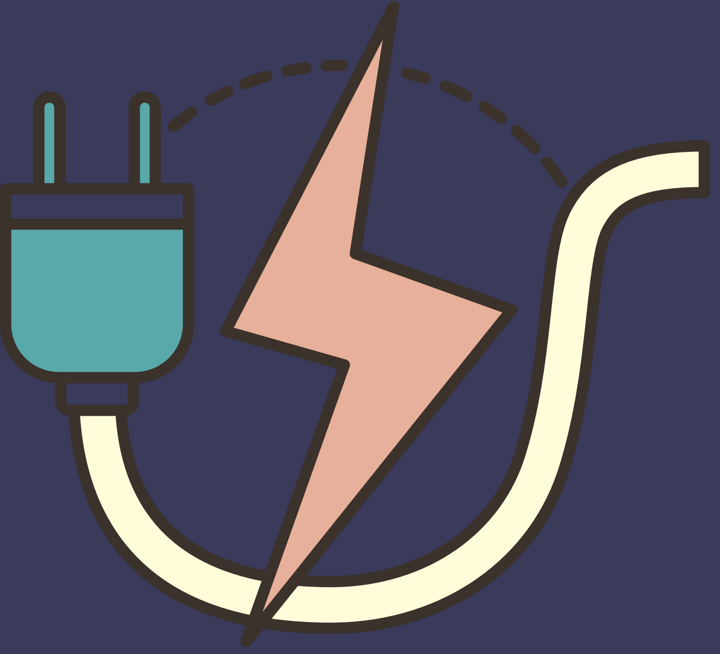
- **What does this mean practically?**

Energy consumption varies significantly across different times of day.

This means that certain times of the day, such as peak hours, might exhibit higher consumption compared to off-peak hours.

Companies or energy providers can use this information to develop strategies for optimizing energy usage, such as offering lower prices during off-peak hours when energy consumption is lower.





Energy Consumption Optimization: Threshold Determination and Cost Savings Analysis

- **Problem Statement:**

The objective is to optimize energy consumption by identifying cells with low energy consumption that can be shut down to reduce energy costs. A threshold must be defined to determine which cells should be shut down based on their energy consumption.

- **Goal:**

The goal of this analysis is to find an optimal energy consumption threshold that minimizes energy usage while ensuring no significant operational disruption.

- **Threshold Determination:**

1. **Definition of Threshold:**

The threshold is the energy consumption level below which cells are considered for shutdown. We define the threshold using the 25th percentile of energy consumption across all cells, meaning we target the bottom 25% of cells that consume the least energy.

2. **Rationale:**

By setting the threshold at the 25th percentile, we focus on the lowest-consuming cells, which are more likely to offer substantial savings without affecting core operations.

- **Identifying Low Consumption Cells:**

Identification Criteria:

Cells with average energy consumption below the calculated threshold were identified as "low consumption" cells. These cells are candidates for shutdown to optimize energy use.

- **Cost and Energy Savings Calculation:**

Total Energy Savings (KWH):

The total energy savings are the sum of the energy consumption for all cells that fall below the threshold. These cells are shut down to reduce overall energy consumption.

Cost Savings (EGP):

The savings in Egyptian Pounds (EGP) are calculated by multiplying the total energy savings by the average cost per KWH, based on historical consumption and cost data.

```python
# Calculate the total energy savings in KWH by summing the energy consumption of the selected cells
total_savings_kwh = low_consumption_df['KWH/hh (per half hour) '].sum()

# Calculate the average cost per KWH (cost per unit of energy consumed)
avg_cost_per_kwh = merged_df['Cost'].sum() / merged_df['KWH/hh (per half hour) '].sum()

# Calculate the total savings in EGP by multiplying the KWH saved by the average cost per KWH
total_savings_egp = total_savings_kwh * avg_cost_per_kwh
```

- The calculated cost savings were **38,044.72 EGP** for a total of **1857.65 KWH** saved by shutting down the identified low consumption cells.

Alternate Threshold Analysis

1. **First Threshold (0.0650 KWH):**
    - Potential Savings:
    - Energy: 1857.65 KWH
    - Cost: 38,044.72 EGP
    - Number of Cells Eligible for Shutdown: 7 cells
    - Effectiveness: Significant energy and cost savings, but it requires ensuring the shutdown does not impact operational processes.
2. **Second Threshold (0.1102 KWH):**
    - Potential Savings:
    - Energy: 1.56 KWH
    - Cost: 1.56 EGP
    - Number of Cells Eligible for Shutdown: Very few cells.
    - Effectiveness: Ineffective for optimizing energy consumption or cost reduction.

- **Analysis:**

The first threshold offers significant energy and cost savings, while the second threshold does not lead to meaningful savings, making it ineffective for this optimization.

Recommendations

1. **Primary Threshold to Adopt:**

The first threshold of 0.0650 KWH is recommended, as it results in a substantial reduction in energy consumption and cost. This provides a balance between energy savings and operational disruption.

1. **Shutdown Implementation:**
   - Operational Considerations: Ensure that critical operations are not impacted by shutting down low consumption cells.
   - Redistribution of Load: If applicable, consider redistributing loads among cells rather than shutting down completely.
2. **Time of Shutdown:**

It is advisable to shut down cells during periods of low demand (e.g., off-peak hours, holidays) to minimize any potential negative impact on operations.

1. **Monitoring and Adjustments:**

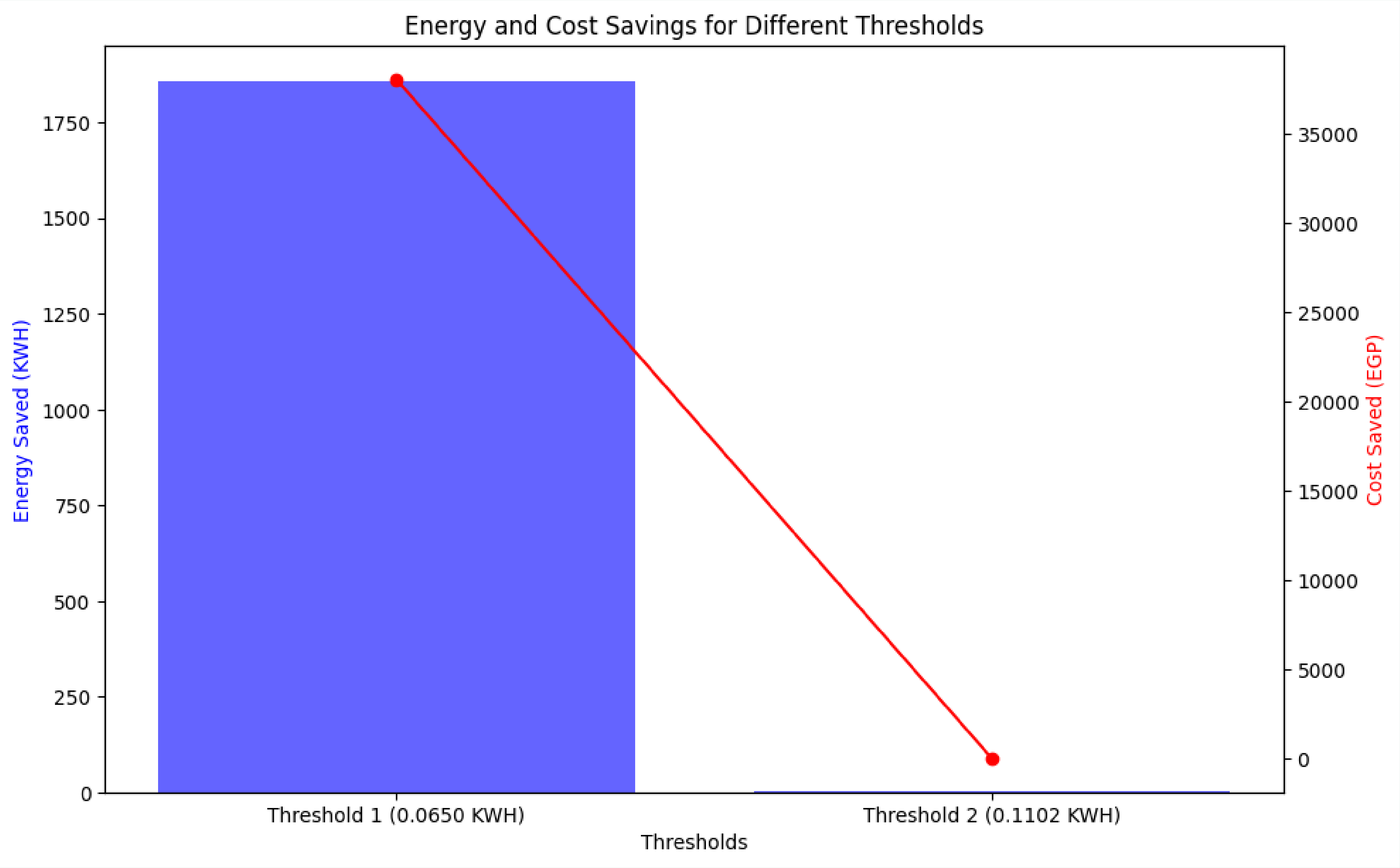
Regularly track energy consumption and the operational impact of cell shutdowns. If needed, adjust the threshold or implement further optimizations.

- **Conclusion**

  By adopting the first threshold of 0.0650 KWH, significant energy and cost savings can be achieved. This strategy offers a balanced approach to optimizing energy consumption without compromising operational efficiency.

- **Final Recommendation:**

  It is crucial to monitor the implementation and impact of this plan, making adjustments as necessary to ensure continued efficiency and cost reduction.



Presented by Rana Alaa

Thank You