A3: Design & Implementation of E2E Pipeline using Snowflake and Tableau

Business Intelligence

Angel Lanto

Master's in Business Analytics

Table of Contents

I. Introduction

a. Business Problem

"How can household-level energy usage patterns be optimized to reduce consumption during peak periods and improve overall energy efficiency?"

Energy efficiency at the household level plays a critical role in broader sustainability and power grid management. Despite the abundance of residential data, many utility providers and policy planners lack granular insights into how and when households consume energy. This gap limits the creation of effective energy-saving strategies, innovative pricing models, and comprehensive renewable integration plans.

The goal of this report is to optimize energy consumption by pinpointing peak load periods, examining appliance-specific usage trends, and identifying anomalous consumption behaviors. By leveraging these insights, stakeholders can implement smarter grid operations, develop targeted energy-saving initiatives, and establish robust policy recommendations that together lower energy consumption during peak periods and improve overall efficiency.

b. Dataset Overview

The analysis is based on the [Individual Household Electric Power Consumption Dataset from the UCI Machine Learning Repository]. The dataset contains over 2 million records from a single household between December 2006 and November 2010, captured at one-minute intervals. It includes the following features:

- Date, Time: Timestamp of observation
- Global_active_power: Total active power consumed (kW)
- Global_reactive_power: Reactive power consumed (kW)
- Voltage: Voltage (V)
- Global_intensity: Current (ampere)
- Sub_metering_1: Kitchen appliances (Wh)
- Sub_metering_2: Laundry appliances (Wh)
- Sub_metering_3: Climate control (AC/heating) (Wh)

II. Step-by-step ETL pipeline explanation

The ETL process ensures the household power consumption dataset is clean, enriched, and structured to support advanced analytics and meaningful visualizations in Tableau.

1. Extract

The .txt file was unzipped and imported into a relational database using standard data ingestion tools.

2. Transform

Transformation was the most intensive step, involving multiple sub-processes:

2.1 Data Cleaning & Type Conversion

a. Combined Date and Time into a unified DATETIME column using: TO_TIMESTAMP_NTZ(CONCAT(DATE, ' ', TIME), 'DD/MM/YYYY HH24:MI:SS')

b. Converted all numeric columns to FLOAT and handled missing values represented as '?' using:
   NVL(NULLIF(column_name, '?')::FLOAT, 0)
   Columns cleaned:
   - GLOBAL_ACTIVE_POWER
   - GLOBAL_REACTIVE_POWER

- VOLTAGE
- GLOBAL_INTENSITY
- SUB_METERING_1, _2, _3

2.2 Feature Engineering
To support time-based and behavioral analysis, the following features were derived:

New Column	Description
YEAR, MONTH, HOUR	Extracted for time trend visualizations
DAY_OF_WEEK	Weekday name (e.g., Mon, Tue, ...)
IS_WEEKEND	TRUE if Saturday or Sunday
IS_PEAK_HOUR	TRUE if time is 6–9 AM or 6–9 PM
PEAK_PERIOD_LABEL	Labeled: 'Morning Peak', 'Evening Peak', 'Off-Peak'
WEEKDAY_CATEGORY	Categorized as 'Weekday' or 'Weekend'

2.3 Aggregation
To improve performance and facilitate dashboard-level analytics, a second table was created with daily granularity that aggregates power usage and retains contextual time columns.
Table: household_power_consumption_aggregated

- DATE, YEAR, MONTH, DAY, DAY_OF_WEEK
- AVG: GLOBAL_ACTIVE_POWER, GLOBAL_REACTIVE_POWER, VOLTAGE, GLOBAL_INTENSITY
- SUM: SUB_METERING_1, SUB_METERING_2, SUB_METERING_3

3. Load
The cleaned and enriched tables “household_power_consumption_cleaned.xlsx” and “household_power_consumption_aggregated” were loaded into a SQL-ready data warehouse and connected directly to Tableau.
• Table 1: household_power_consumption_cleaned – Minute-level data used for hourly and peak/off-peak analyses.
• Table 2: household_power_consumption_aggregated – Daily-level data used for monthly trends, day-based insights, and summary KPIs.
These structured tables enabled flexible drag-and-drop analysis in Tableau, with calculated fields and filters powered by the transformation logic built during the ETL phase.

III. SQL queries & Tableau insights

This section presents the interpretation of SQL queries developed to answer the business question: “How could household-level energy usage patterns be optimized to reduce consumption during peak periods and improve overall energy efficiency?” It also integrates insights from the corresponding Tableau dashboard.

1. Overall Consumption Summary

Query Name	Description	Fields Used
Total Usage by Appliance Category	Calculates total energy usage by appliance zones: kitchen, laundry, heating/cooling.	['SUB_METERING_1', 'SUB_METERING_2', 'SUB_METERING_3']
Monthly Energy Consumption (kWh)	Shows average daily power usage per month to identify seasonal trends.	['YEAR', 'MONTH', 'SUM_GLOBAL_ACTIVE_POWER']

Daily Energy Consumption (kWh)	Identifies the lowest and highest total daily power consumption.	['DATE', 'SUM_GLOBAL_ACTIVE_POWER']

These queries provided a high-level overview of household energy consumption patterns. Total energy usage was calculated across sub-metered zones such as the kitchen, laundry, and heating/cooling. Total monthly energy usage trends revealed seasonal variations, while identifying the highest and lowest consumption days offered insight into behavior-driven spikes or anomalies. In the Tableau dashboard, KPI cards displayed total consumption by appliance category. Charts visualized total monthly and daily energy usage trends, highlighting winter peaks and mid-year lows.

2. Time-Based Patterns

Query Name	Description	Fields Used
Average Hourly Consumption	Analyzes energy usage across each hour of the day to highlight daily peaks.	['HOUR', 'AVG_GLOBAL_ACTIVE_POWER']
Peak Period Consumption	Compares average usage across peak and off-peak time windows.	['PEAK_PERIOD_LABEL', 'GLOBAL_ACTIVE_POWER']
Load Distribution Across the Week	Shows total weekly load per weekday based on daily averages.	['DAY_OF_WEEK', 'AVG_GLOBAL_ACTIVE_POWER']
Energy Consumption by Day	Shows average power usage per weekday to compare daily patterns.	['DAY_OF_WEEK', 'AVG_GLOBAL_ACTIVE_POWER']
Weekend vs Weekday Consumption	Compares energy use on weekends versus weekdays.	['WEEKDAY_CATEGORY', 'AVG_GLOBAL_ACTIVE_POWER']
Hourly Usage Heatmap	Combines day of week and hour to visualize energy intensity across the week.	['DAY_OF_WEEK', 'HOUR', 'GLOBAL_ACTIVE_POWER']

These queries provided a high-level overview of household energy consumption patterns. Total energy usage was calculated across sub-metered zones such as the kitchen, laundry, and heating/cooling. Monthly average energy usage trends revealed seasonal variations, while identifying the highest and lowest consumption days offered insight into behavior-driven spikes or anomalies. In the Tableau dashboard, KPI cards displayed total consumption by appliance category. Line charts visualized monthly and daily energy usage trends, highlighting winter peaks and mid-year lows.

3. Variability and Anomaly Detection

Query Name	Description	Fields Used
Daily Variability in Power Usage	Shows most inconsistent days by standard deviation of daily power usage.	['DATETIME', 'GLOBAL_ACTIVE_POWER']
Days with Anomalously High Consumption	Flags days exceeding average power usage by more than 2 standard deviations.	['DATE', 'AVG_GLOBAL_ACTIVE_POWER']
Correlation between Active and Reactive Power	Computes Pearson correlation coefficient between active and reactive power.	['GLOBAL_ACTIVE_POWER', 'GLOBAL_REACTIVE_POWER']

Queries in this category were used to detect irregular consumption and inefficient usage behaviors. Days with the highest standard deviation were flagged as most variable, while days exceeding two standard deviations above the average were identified as anomalies. Correlation analysis between active and reactive power was also conducted to explore potential inefficiencies. In Tableau, a variability line chart and a bar chart of anomalous days provided clear visual evidence of outliers, enhancing data-driven decision-making on energy interventions.

4. Tableau Dashboard Integration

Figure 1. Tableau Dashboard

For an interactive experience that offers granular insights and drill-down options, please hover over the visual elements on the Tableau Public dashboard.

The interactive Tableau dashboard consolidated the SQL-based insights into a visual narrative. Filters for year, month, and day of the week enhanced usability, allowing stakeholders to explore trends across different time frames. Visualizations aligned directly with the three main insight categories: overall consumption, time-based behavior, and variability detection. Together, the SQL queries and Tableau visuals provided actionable intelligence to optimize energy usage and support sustainability goals at the household level.

IV. References

ChatGPT. (2025). Chatgpt.com. https://chatgpt.com/c/67f3c43f-385c-8002-bbba-ba3db8f9a025

ChatGPT. (2025). Chatgpt.com. https://chatgpt.com/c/67ee7204-adc8-8002-b4f2-2455c4ac9ae3

V. Appendix

Appendix A. Overall Consumption Summary: Results and Discussion

1. Application Usage
   The top-left indicators show that the household has an average daily consumption of approximately 35 kWh, culminating in around 1,050 kWh per month. Over the period analyzed, there is a subtle upward trend in overall consumption by about 3–5% year-over-year. Peak periods occasionally approach 50 kWh in a single hour. These numbers highlight the significant energy footprint of the household. The gradual upward trend underlines the need for interventions that curb energy usage growth. By implementing smart metering and real-time consumption feedback, households can better understand and ultimately reduce their daily and monthly usage. Focusing on specific strategies during peak hours can help shift some of this load to off-peak times, reducing the stress on the grid and contributing to overall energy efficiency.

2. Monthly Energy Consumption
   Analysis on a monthly scale shows that consumption peaks in the winter months such as December to January, whereas the summer months drop by about 20% relative to winter. This reflects the additional demand from heating systems during colder periods. The seasonal variation underscores the impact of climate control on energy use. Investing in better insulation, energy-efficient heating systems, or programmable thermostats can reduce the winter peak load. From a broader perspective, if households can moderate these seasonal peaks through improved technologies and practices, overall energy efficiency can be enhanced while also easing grid pressure during the cold season.

3. Daily Energy Consumption (kWh)
   The Daily Energy Consumption (kWh) figure shows relatively consistent usage across days, but with slightly higher consumption on weekends. This implies that households tend to be more active during the weekend, which increases energy demand. To improve efficiency, households could be encouraged to stagger high-energy activities throughout the week, especially if Time-of-Use (TOU) pricing is in effect, to reduce concentrated weekend loads.

Appendix B. Time-Based Patterns: Results and Discussion

1. Average Hourly Consumption
   The average hourly consumption profile reveals low usage during the early morning hours at around 5AM, rising steadily to a pronounced peak at 8-9 PM. There is a secondary, less prominent peak around 7AM, which might be tied to start of the day activities. This hourly breakdown pinpoints the critical windows where energy demands are highest. With the evening peak being nearly four times higher than early morning levels, households can significantly benefit by rescheduling high-energy tasks like laundry or dishwashing to earlier or later hours. Time-of-use pricing models or smart appliance automation could be key in shifting energy-intensive activities to off-peak hours, thereby optimizing energy efficiency.

2. Peak Period Consumption
   This focused chart isolates peak period consumption, revealing that during the 7–9 PM window, consumption can spike to values higher than the average evening usage. This figure directly addresses the business question, highlighting the critical impact of peak period usage on the household's overall energy profile. By employing smart scheduling strategies such as delaying the operation of washing machines or electric vehicle charging, users can attenuate these peaks. Demand-side management programs and incentives can further encourage households to shift energy loads, thereby reducing the stress on the energy grid during peak times and improving efficiency.

3. Load Distribution Across the Week
   The bar chart detailing daily load distribution illustrates that Mondays and Sundays present the highest loads, with average consumption on these days about 10–15% above the mid-week average. Tuesdays through Thursdays are generally lower, indicating a steadier daily routine during the workweek. Recognizing these specific high-load days provides an opportunity for targeted interventions. For instance, households can plan high-energy activities on Tuesday or Wednesday instead of Monday or Sunday. Additionally, the data can inform utility companies in designing day-specific demand-response or time-of-use rate programs, further incentivizing consumers to adjust their energy use behavior.

4. Energy Consumption (kWh) by Day
   The day-to-day chart indicates that weekend usage exceeds weekday usage. Specific days, such as Sundays and Saturdays, show spikes above the daily average. Conversely, mid-weekdays maintain consumption closer to the baseline. This differentiation suggests that behavioral patterns related to weekday versus weekend routines contribute significantly to overall consumption. Households can work on moderating weekend usage by planning energy-heavy activities during less congested times, or by leveraging scheduling tools on appliances. The findings also provide clear direction for utilities to design demand-response incentives that encourage load shifting on higher consumption days.

5. Weekday vs. Weekend Usage
   The pie chart reveals that weekdays account for roughly 60% of overall energy consumption while weekends are responsible for about 40%. However, on a per-day basis, weekends show higher individual consumption rates. This dual perspective implies that while weekdays contribute more overall due to the higher number of days, the intensity of usage during weekends is comparatively higher. Optimizing usage may involve targeted strategies such as promoting off-peak incentives during weekend

mornings or early afternoons. Moreover, education campaigns and real-time monitoring can empower households to distribute their energy usage more evenly across the week.

6. Hourly Usage Heatmap
   The heatmap visualization shows pronounced “hot spots” in the evening around 8 PM. There are also moderate usage intensities during midday on weekends. The darkest cells represent consumption values higher than off-peak hours. By visualizing energy consumption intensity over both time and days, this heatmap enables households to pinpoint exact periods of excessive usage. Such granular data supports the case for automated energy management systems that can delay the operation of non-critical appliances until lower-demand periods. This targeted shifting not only reduces peak period loads but also contributes to smoother overall energy consumption curves, enhancing efficiency.

Appendix C. Variability and Anomalies: Results and Discussion

1. Daily Variability in Power Usage

   The daily variability chart displays a standard deviation of approximately 1.2 kWh in daily consumption, with some days deviating significantly from the norm, particularly during periods of increased activity or adverse weather conditions. This variability indicates that while there is a consistent consumption pattern, unexpected surges can lead to inefficiencies during peak periods. Targeted interventions such as smart thermostats that adjust heating/cooling dynamically or scheduling notifications when usage is unusually high can reduce this variability. By smoothing out these peaks, the household can maintain a more balanced energy usage profile, contributing to overall efficiency gains.

2. Days with Anomalously High Consumption (Bottom-Right List or Chart)
   This chart identifies specific outlier days where consumption spikes by more than 50% compared to the baseline. Analysis shows that power usage is normal for this household. These anomalous days often coincide with extreme weather events or unusual household activities, such as holiday gatherings or major appliance malfunctions. The identification of these anomalous days is crucial for troubleshooting and targeted behavioral interventions. When households recognize that high consumption correlates with events like severe weather or equipment issues, they can take proactive measures ranging from scheduled maintenance to temporary adjustments in usage to mitigate these spikes. From a business intelligence perspective, such insights can lead to more accurate forecasting models and adaptive strategies to maintain grid stability during unforeseen demand surges.

Appendix D. Conclusion

In conclusion, the proponent's analysis demonstrates that leveraging detailed household-level energy data can greatly enhance energy efficiency by tailoring interventions to real-time usage patterns. The end-to-end pipeline from data extraction, thorough transformation, and feature engineering to comprehensive visualization in Tableau enabled the proponent to pinpoint critical trends such as high appliance-specific energy consumption and time-based usage patterns. Our SQL queries revealed that heating/cooling systems dominate energy use, while kitchen and laundry appliances also contribute significantly, underscoring the potential benefits of shifting high-demand activities away from peak periods. Monthly and hourly analyses highlighted clear seasonal patterns and pronounced evening peaks. These insights support the deployment of smart scheduling, such as delaying the use of energy-intensive appliances, and adopting dynamic pricing models to incentivize off-peak consumption.

Furthermore, the detection of anomalous consumption days and high daily variability provides a framework for real-time alerts, enabling prompt corrective actions like maintenance or behavior adjustments. The strong correlation between active and reactive power suggests further opportunities to refine energy management systems for improved load balancing. Overall, by aligning strategy with the granular insights derived from this data, stakeholders from households to utility companies can design more effective demand-response programs. These initiatives not only reduce peak demand but also lower overall energy consumption, contributing to grid stability and environmental sustainability. The interactive Tableau dashboard serves as a vital tool, offering additional context and drill-down capabilities for continuous monitoring and adaptive decision-making in an ever-evolving energy landscape.

```
-- ================================================
-- HOUSEHOLD POWER CONSUMPTION DATA PIPELINE
-- ================================================

-- STEP 1: DATA CLEANING & TYPE CONVERSION
-- Converts date/time columns into a single timestamp and numeric columns
into floats, replacing missing values with 0
CREATE OR REPLACE TABLE household_power_consumption_cleaned AS
SELECT
    TO_TIMESTAMP_NTZ(CONCAT(DATE, ' ', TIME), 'DD/MM/YYYY HH24:MI:SS') AS
DATETIME,
    NVL(NULLIF(GLOBAL_ACTIVE_POWER, '?')::FLOAT, 0) AS GLOBAL_ACTIVE_POWER,
    NVL(NULLIF(GLOBAL_REACTIVE_POWER, '?')::FLOAT, 0) AS
GLOBAL_REACTIVE_POWER,
    NVL(NULLIF(VOLTAGE, '?')::FLOAT, 0) AS VOLTAGE,
    NVL(NULLIF(GLOBAL_INTENSITY, '?')::FLOAT, 0) AS GLOBAL_INTENSITY,
    NVL(NULLIF(SUB_METERING_1, '?')::FLOAT, 0) AS SUB_METERING_1,
    NVL(NULLIF(SUB_METERING_2, '?')::FLOAT, 0) AS SUB_METERING_2,
    NVL(NULLIF(SUB_METERING_3, '?')::FLOAT, 0) AS SUB_METERING_3
FROM
    household_power_consumption;


-- STEP 2: DATA QUALITY CHECK
-- Describe table structure and check for any remaining NULL values
DESCRIBE TABLE household_power_consumption_cleaned;

SELECT
    COUNT(*) AS TOTAL_ROWS,
    COUNT_IF(DATETIME IS NULL) AS NULL_DATETIME,
    COUNT_IF(GLOBAL_ACTIVE_POWER IS NULL) AS NULL_GLOBAL_ACTIVE_POWER,
    COUNT_IF(GLOBAL_REACTIVE_POWER IS NULL) AS NULL_GLOBAL_REACTIVE_POWER,
    COUNT_IF(VOLTAGE IS NULL) AS NULL_VOLTAGE,
    COUNT_IF(GLOBAL_INTENSITY IS NULL) AS NULL_GLOBAL_INTENSITY,
    COUNT_IF(SUB_METERING_1 IS NULL) AS NULL_SUB_METERING_1,
    COUNT_IF(SUB_METERING_2 IS NULL) AS NULL_SUB_METERING_2,
    COUNT_IF(SUB_METERING_3 IS NULL) AS NULL_SUB_METERING_3
FROM household_power_consumption_cleaned;


-- STEP 3: DESCRIPTIVE STATISTICS
```

```
-- Calculate statistical metrics for numeric columns to understand data
distribution
SELECT
    COLUMN_NAME,
    MIN(VALUE) AS MIN_VALUE,
    MAX(VALUE) AS MAX_VALUE,
    AVG(VALUE) AS AVG_VALUE,
    STDDEV(VALUE) AS STDDEV_VALUE,
    MEDIAN(VALUE) AS MEDIAN_VALUE
FROM (
    SELECT 'GLOBAL_ACTIVE_POWER' COLUMN_NAME, GLOBAL_ACTIVE_POWER AS VALUE
FROM household_power_consumption_cleaned UNION ALL
    SELECT 'GLOBAL_REACTIVE_POWER', GLOBAL_REACTIVE_POWER FROM
household_power_consumption_cleaned UNION ALL
    SELECT 'VOLTAGE', VOLTAGE FROM household_power_consumption_cleaned UNION
ALL
    SELECT 'GLOBAL_INTENSITY', GLOBAL_INTENSITY FROM
household_power_consumption_cleaned UNION ALL
    SELECT 'SUB_METERING_1', SUB_METERING_1 FROM
household_power_consumption_cleaned UNION ALL
    SELECT 'SUB_METERING_2', SUB_METERING_2 FROM
household_power_consumption_cleaned UNION ALL
    SELECT 'SUB_METERING_3', SUB_METERING_3 FROM
household_power_consumption_cleaned
)
GROUP BY COLUMN_NAME;


-- STEP 4: DATA ENRICHMENT (FINAL VERSION WITH HOUR EXTRACTION)
-- Add columns for time-based analysis including peak periods,
weekday/weekend labels, and hour

ALTER TABLE household_power_consumption_cleaned
ADD YEAR INT,
    MONTH INT,
    DAY_OF_WEEK VARCHAR,
    IS_WEEKEND BOOLEAN,
    IS_PEAK_HOUR BOOLEAN,
    PEAK_PERIOD_LABEL VARCHAR,
    WEEKDAY_CATEGORY VARCHAR,
    HOUR INT;

UPDATE household_power_consumption_cleaned
```

```sql
SET
    YEAR = EXTRACT(YEAR FROM DATETIME),
    MONTH = EXTRACT(MONTH FROM DATETIME),
    DAY_OF_WEEK = DAYNAME(DATETIME),
    IS_WEEKEND = CASE
                    WHEN DAYNAME(DATETIME) IN ('Sat', 'Sun') THEN TRUE
                    ELSE FALSE
                 END,
    IS_PEAK_HOUR = CASE
                    WHEN EXTRACT(HOUR FROM DATETIME) BETWEEN 6 AND 9
                      OR EXTRACT(HOUR FROM DATETIME) BETWEEN 18 AND 21
                    THEN TRUE
                    ELSE FALSE
                   END,
    PEAK_PERIOD_LABEL = CASE
                            WHEN EXTRACT(HOUR FROM DATETIME) BETWEEN 6 AND 9
THEN 'Morning Peak'
                            WHEN EXTRACT(HOUR FROM DATETIME) BETWEEN 18 AND
21 THEN 'Evening Peak'
                            ELSE 'Off-Peak'
                        END,
    WEEKDAY_CATEGORY = CASE
                            WHEN EXTRACT(DOW FROM DATETIME) IN (0, 6) THEN
'Weekend' -- Sunday = 0, Saturday = 6
                            ELSE 'Weekday'
                       END,
    HOUR = EXTRACT(HOUR FROM DATETIME);

-- STEP 5: DAILY AGGREGATION
-- Aggregate data daily to simplify analyses
CREATE OR REPLACE TABLE household_power_consumption_aggregated AS
SELECT
    DATE_TRUNC('day', DATETIME) AS DATE,
    EXTRACT(YEAR FROM DATE_TRUNC('day', DATETIME)) AS YEAR,
    EXTRACT(MONTH FROM DATE_TRUNC('day', DATETIME)) AS MONTH,
    EXTRACT(DAY FROM DATE_TRUNC('day', DATETIME)) AS DAY,
    DAYNAME(DATE_TRUNC('day', DATETIME)) AS DAY_OF_WEEK,

    AVG(GLOBAL_ACTIVE_POWER) AS AVG_GLOBAL_ACTIVE_POWER,
    AVG(GLOBAL_REACTIVE_POWER) AS AVG_GLOBAL_REACTIVE_POWER,
    AVG(VOLTAGE) AS AVG_VOLTAGE,
    AVG(GLOBAL_INTENSITY) AS AVG_GLOBAL_INTENSITY,
```

```sql
    SUM(SUB_METERING_1) AS TOTAL_SUB_METERING_1,
    SUM(SUB_METERING_2) AS TOTAL_SUB_METERING_2,
    SUM(SUB_METERING_3) AS TOTAL_SUB_METERING_3
FROM household_power_consumption_cleaned
GROUP BY DATE
ORDER BY DATE;

-- ==================================================
-- STEP 6: QUERYING FOR BUSINESS INSIGHTS
-- Organized by Insight Category
-- ==================================================

-- =======================================================
-- 1. OVERALL CONSUMPTION SUMMARY
-- Gives a bird’s-eye view of total and average usage
-- =======================================================

-- 6.1: Total Usage by Appliance Category
SELECT
    SUM(SUB_METERING_1) AS TOTAL_KITCHEN_USAGE,
    SUM(SUB_METERING_2) AS TOTAL_LAUNDRY_USAGE,
    SUM(SUB_METERING_3) AS TOTAL_HEATING_COOLING_USAGE
FROM household_power_consumption_cleaned;

-- 6.2: Total Monthly Active Power
SELECT YEAR, MONTH, SUM(GLOBAL_ACTIVE_POWER) AS TOTAL_MONTHLY_ACTIVE_POWER
FROM household_power_consumption_cleaned
GROUP BY YEAR, MONTH
ORDER BY YEAR, MONTH;

-- 6.3: Days with the Highest Energy Consumption
SELECT
    DATE_TRUNC('day', DATETIME) AS DAY,
    SUM(GLOBAL_ACTIVE_POWER) AS TOTAL_DAILY_ACTIVE_POWER
FROM household_power_consumption_cleaned
GROUP BY DATE_TRUNC('day', DATETIME)
ORDER BY TOTAL_DAILY_ACTIVE_POWER DESC
LIMIT 10;

-- 6.4: Days with the Lowest Energy Consumption
SELECT
    DATE_TRUNC('day', DATETIME) AS DAY,
    SUM(GLOBAL_ACTIVE_POWER) AS TOTAL_DAILY_ACTIVE_POWER
```

```sql
FROM household_power_consumption_cleaned
GROUP BY DATE_TRUNC('day', DATETIME)
ORDER BY TOTAL_DAILY_ACTIVE_POWER ASC
LIMIT 10;

-- ============================================
-- 2. TIME-BASED PATTERNS
-- Helps identify consumption behavior over time
-- ============================================

-- 6.5: Average Hourly Consumption Across Dataset
SELECT
    HOUR,
    AVG(GLOBAL_ACTIVE_POWER) AS AVG_ACTIVE_POWER
FROM household_power_consumption_cleaned
GROUP BY HOUR
ORDER BY HOUR;

-- 6.6: Peak Period Consumption
SELECT
    PEAK_PERIOD_LABEL,
    AVG(GLOBAL_ACTIVE_POWER) AS AVG_ACTIVE_POWER
FROM household_power_consumption_cleaned
GROUP BY PEAK_PERIOD_LABEL
ORDER BY AVG_ACTIVE_POWER DESC;

-- 6.7: Load Distribution Across the Week
SELECT
    DAY_OF_WEEK,
    SUM(AVG_GLOBAL_ACTIVE_POWER) AS TOTAL_WEEKDAY_LOAD
FROM household_power_consumption_aggregated
GROUP BY DAY_OF_WEEK
ORDER BY TOTAL_WEEKDAY_LOAD DESC;

-- 6.8: Average Active Power by Day of the Week
SELECT
    DAY_OF_WEEK,
    AVG(AVG_GLOBAL_ACTIVE_POWER) AS AVG_ACTIVE_POWER
FROM household_power_consumption_aggregated
GROUP BY DAY_OF_WEEK
ORDER BY AVG_ACTIVE_POWER DESC;

-- 6.9: Weekend vs Weekday Consumption
```

```sql
SELECT
    WEEKDAY_CATEGORY,
    AVG(GLOBAL_ACTIVE_POWER) AS AVG_ACTIVE_POWER
FROM household_power_consumption_cleaned
GROUP BY WEEKDAY_CATEGORY;

-- 6.10: Hourly Usage Heatmap
-- Visualize average energy usage by hour of day and day of week
SELECT
    DAY_OF_WEEK,
    HOUR,
    SUM(GLOBAL_ACTIVE_POWER) AS TOTAL_HOURLY_ACTIVE_POWER
FROM household_power_consumption_cleaned
GROUP BY DAY_OF_WEEK, HOUR
ORDER BY DAY_OF_WEEK, HOUR;

-- =========================================================
-- 3. VARIABILITY & ANOMALY DETECTION
-- Spot irregular usage patterns or energy inefficiencies
-- =========================================================

-- 6.11: Daily Variability in Power Usage (Top 10 by Std Dev)
SELECT
    DATE_TRUNC('day', DATETIME) AS DATE,
    STDDEV(GLOBAL_ACTIVE_POWER) AS STDDEV_ACTIVE_POWER
FROM household_power_consumption_cleaned
GROUP BY DATE_TRUNC('day', DATETIME)
ORDER BY STDDEV_ACTIVE_POWER DESC
LIMIT 10;

-- 6.12: Days with Anomalously High Consumption
SELECT DATE, AVG_GLOBAL_ACTIVE_POWER
FROM household_power_consumption_aggregated
WHERE AVG_GLOBAL_ACTIVE_POWER > (
    SELECT AVG(AVG_GLOBAL_ACTIVE_POWER) + 2 * STDDEV(AVG_GLOBAL_ACTIVE_POWER)
    FROM household_power_consumption_aggregated
)
ORDER BY AVG_GLOBAL_ACTIVE_POWER DESC;

-- 6.13: Correlation between Active and Reactive Power
SELECT
    CORR(GLOBAL_ACTIVE_POWER, GLOBAL_REACTIVE_POWER) AS
CORRELATION_ACTIVE_REACTIVE
```

```
FROM household_power_consumption_cleaned;

-- STEP 7: QUICK CHECK FINAL DATA
SELECT * FROM household_power_consumption_cleaned LIMIT 10;
SELECT * FROM household_power_consumption_aggregated LIMIT 10;
```