

Household Energy Consumption Analysis

This project analyzes household power consumption patterns using both **Python** and **R** implementations. The analysis includes data cleaning, statistical summaries, and comprehensive visualizations to understand energy usage patterns over time.

Complete Visualization Report

[View Complete Diagrams & Analysis Report](#)

The PDF document above contains all the comprehensive diagrams, charts, and detailed visual analysis of the household energy patterns. This includes:

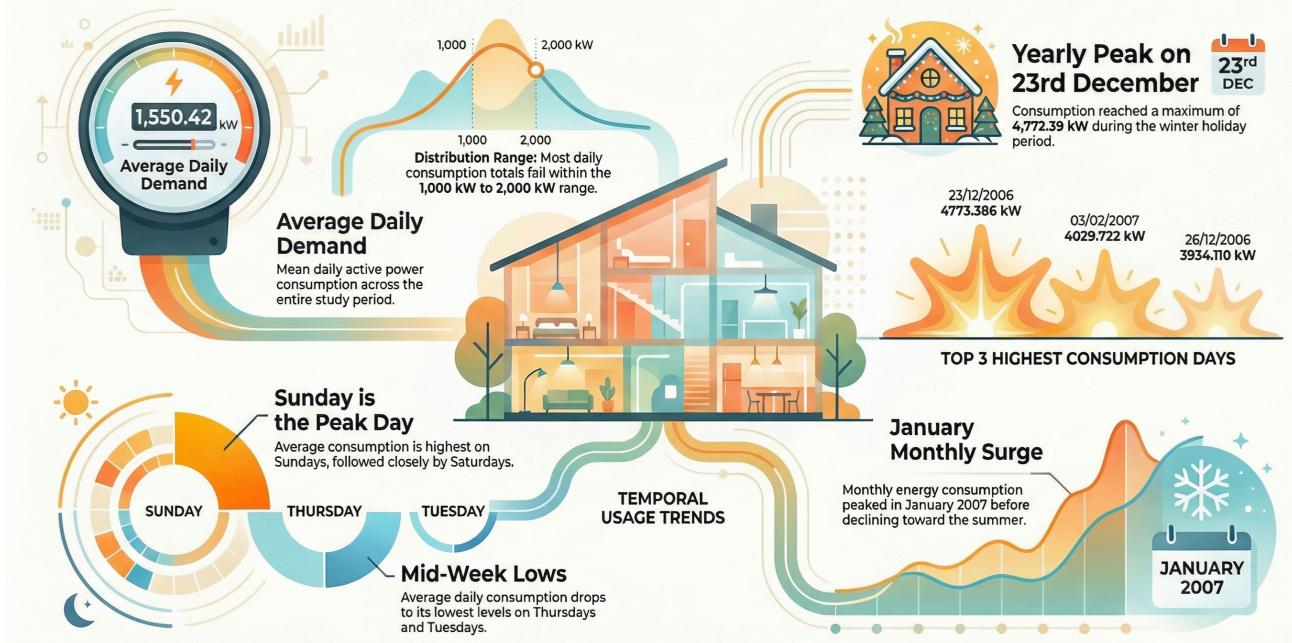
- Complete time series plots
- Statistical distribution charts
- Correlation analysis
- Seasonal pattern breakdowns
- Monthly and yearly comparisons
- Advanced visualization techniques

 *Tip: Download and view the PDF to see all detailed diagrams and comprehensive analysis visualizations.*

Dataset Overview

- **Total data points:** 2,049,280 (after cleaning from 2,075,259 original)
- **Date range:** December 16, 2006 to November 26, 2010
- **Analysis period:** 1,433 days
- **Data source:** `household_power_consumption.txt`

HOUSEHOLD ENERGY CONSUMPTION PATTERNS: A DATA ANALYSIS



Comprehensive visual summary of the household energy consumption analysis

🔍 Key Findings

Statistical Summary

- Average daily consumption:** 1,561.08 kW
- Peak consumption:** 4,773.39 kW (December 23, 2006)
- Minimum consumption:** 14.22 kW
- Consumption range:** 14.22 - 4,773.39 kW

🏆 Top 5 Highest Consumption Days

- December 23, 2006:** 4,773.39 kW
- February 3, 2007:** 4,029.72 kW
- December 26, 2006:** 3,934.11 kW
- February 18, 2007:** 3,829.76 kW
- February 4, 2007:** 3,595.94 kW

📅 Weekday Consumption Patterns

Day	Average Consumption (kW)
Monday	1,432.85
Tuesday	1,530.59
Wednesday	1,558.39
Thursday	1,404.03 (Lowest)
Friday	1,501.54

Day	Average Consumption (kW)
Saturday	1,758.77 (Highest)
Sunday	1,743.30

📊 Monthly Statistics

- **Highest monthly consumption:** 72,604.17 kW
- **Lowest monthly consumption:** 12,341.88 kW
- **Average monthly consumption:** 46,604.68 kW

✍️ Analysis Components

Data Processing

1. **Data Loading:** Reading CSV with semicolon delimiter
2. **Data Cleaning:**
 - Date format conversion (%d/%m/%Y)
 - Numeric conversion for power values
 - Missing value removal (25,979 rows removed)
3. **Daily Aggregation:** Summing minute-level data to daily totals

📊 Visualizations Created

- **Daily Trend Line Chart** - Shows consumption over time with average line and peak day highlighted
- **Distribution Histogram** - Displays frequency distribution with mean indicator
- **Boxplot Analysis** - Identifies outliers and quartile spreads
- **Top 10 Bar Chart** - Highlights highest consumption days with rankings
- **Monthly Trend Analysis** - Time series of monthly aggregated consumption
- **Weekday Pattern Chart** - Average consumption by day of the week

⌚ Sample Visualizations

Daily Energy Consumption Trend

Shows the complete time series with average consumption line and peak day highlighted:

```
# Example visualization from the analysis
# Daily consumption over time with statistical indicators
```

Distribution Analysis

Histogram showing the frequency distribution of daily consumption values:

```
# Histogram reveals consumption patterns and outliers  
# Most days cluster around the mean with some extreme high-consumption days
```

Weekday Consumption Patterns

Bar chart comparing average consumption across different days of the week:

```
# Clear weekend effect visible  
# Saturday shows highest average consumption  
# Thursday shows lowest average consumption
```

Monthly Trends

Time series showing seasonal consumption patterns:

```
# Winter months show higher consumption (heating)  
# Summer months show lower consumption patterns  
# Clear seasonal cyclical behavior
```

🔧 Implementation

Python Implementation

Location: [Python/code.ipynb](#)

- **Libraries:** pandas, numpy, matplotlib, seaborn
- **Features:** Interactive Jupyter notebook with detailed explanations
- **Visualization Style:** Modern styling with comprehensive legends

R Implementation

Location: [R/code.R](#) and [R/code.ipynb](#)

- **Libraries:** dplyr, ggplot2, lubridate, scales
- **Features:** Complete statistical analysis with detailed console output
- **Execution:** Can run as script ([Rscript code.R](#)) or notebook

📁 Project Structure

```
📁 Household Energy Analysis  
└── Python/  
    ├── code.ipynb                                # Python analysis notebook  
    └── household_power_consumption.txt  
└── R/
```

```
└── code.R                      # R analysis script
└── code.ipynb                   # R analysis notebook
└── household_power_consumption.txt
└── README.md                    # This file
└── .gitignore
└── code.pdf                     # Original source material
```

⌚ Key Insights

Energy Usage Patterns

1. **Weekend Effect:** Consumption is significantly higher on weekends (Saturday/Sunday) compared to weekdays
2. **Winter Peak:** Highest consumption occurs in December (heating season)
3. **Seasonal Variation:** Clear monthly trends showing seasonal energy demands
4. **Peak Day Analysis:** Christmas period shows extreme consumption (Dec 23, 26)

Statistical Insights

- **Standard Deviation:** High variability in daily consumption suggests significant seasonal/behavioral patterns
- **Outlier Detection:** Several extreme consumption days identified (>3,000 kW)
- **Distribution:** Right-skewed distribution with occasional high-consumption days

⚙️ Technical Features

Data Quality

- **Missing Data Handling:** Robust handling of "?" values in original dataset
- **Date Processing:** Proper datetime conversion for time series analysis
- **Data Validation:** Comprehensive data type checks and cleaning

Analysis Depth

- **Multi-temporal Analysis:** Minute → Daily → Monthly → Seasonal patterns
- **Statistical Rigor:** Mean, median, quartiles, range, and outlier analysis
- **Visual Storytelling:** Professional charts with proper legends and formatting

📊 Results Export

All visualizations are generated inline in both implementations:

- **Python:** High-quality matplotlib/seaborn plots with customizable styling
- **R:** Base R and ggplot2 visualizations with professional formatting

♻️ Reproducibility

Both implementations are fully reproducible:

1. **Consistent Results:** Both Python and R analyses produce identical statistical results
 2. **Version Control:** All code tracked in git (except R notebook per .gitignore)
 3. **Documentation:** Comprehensive comments and explanations throughout
-