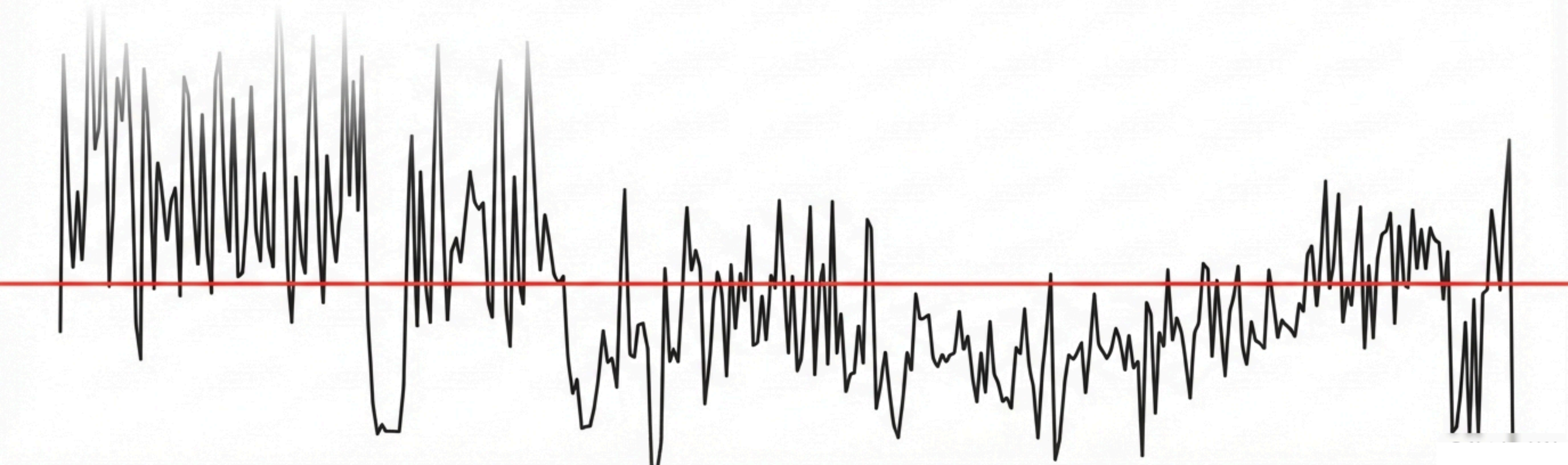
Household Energy Consumption: Patterns & Anomalies

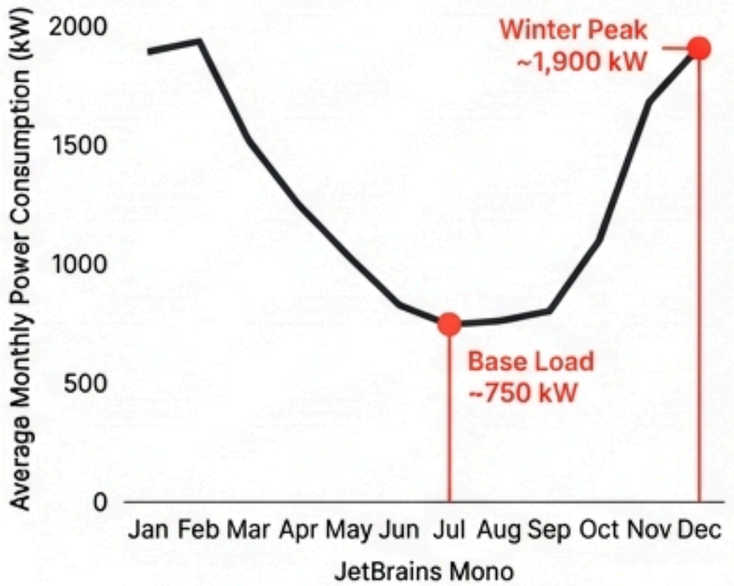
An analysis of daily power usage trends, seasonality, and peak events (Dec 2006 – Nov 2007).



Key Findings: Seasonality and Lifestyle Drive Demand

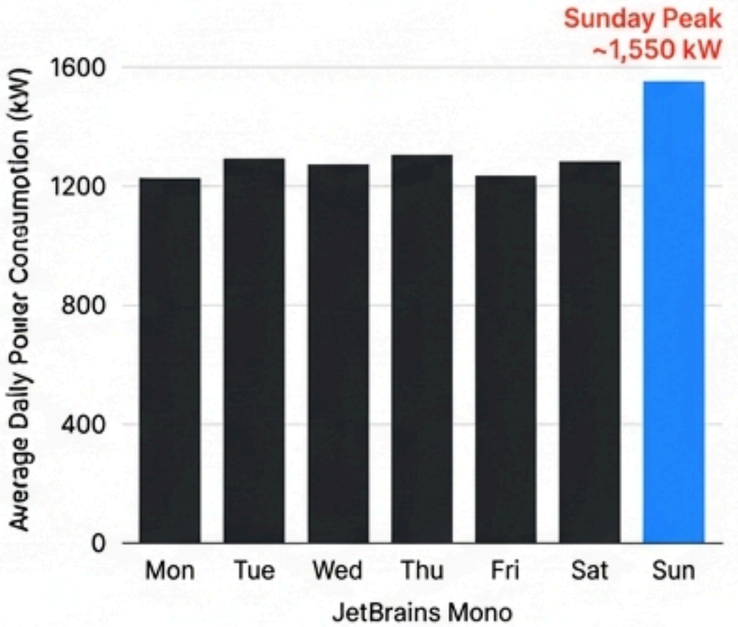
High Seasonality

Winter consumption dominates the dataset. Usage peaks in December and January, driven principally by heating loads, before dropping to a stable 'base load' plateau in July and August.



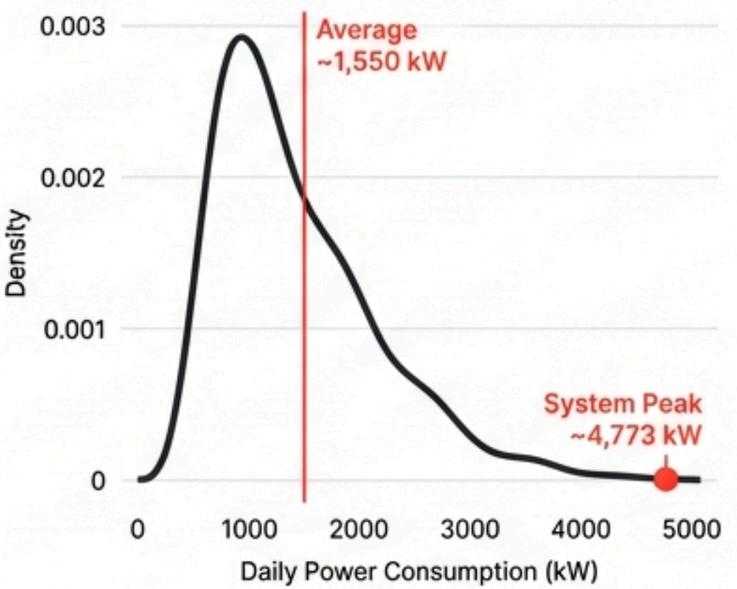
The Weekend Effect

Sundays exhibit the highest average power consumption of any day of the week, suggesting that occupancy and household chores (cooking, laundry) are major load contributors.



Volatility & Capacity

While the daily average is ~1,550 kW, the system must withstand peaks up to ~4,773 kW. The distribution is right-skewed, meaning extreme high-load events are rare but significant.



```
data <- read.csv("household_power_consumption.txt",
                 sep=";",
                 header=TRUE,
                 na.strings="?",
                 stringsAsFactors=FALSE)

# Inspect dataset structure
str(data)
# Output excerpt:
# 'data.frame': 474160 obs. of 9 variables:
# $ Date                 : chr "16/12/2006" ...
# $ Global_active_power : num 4.22 5.36 5.37 5.39 3
# $ Voltage              : num 235 234 233 234 236 .
# ...
```

Data Profile & Methodology

Source: household_power_consumption.txt

Total Observations: 470,227 (post-cleaning)

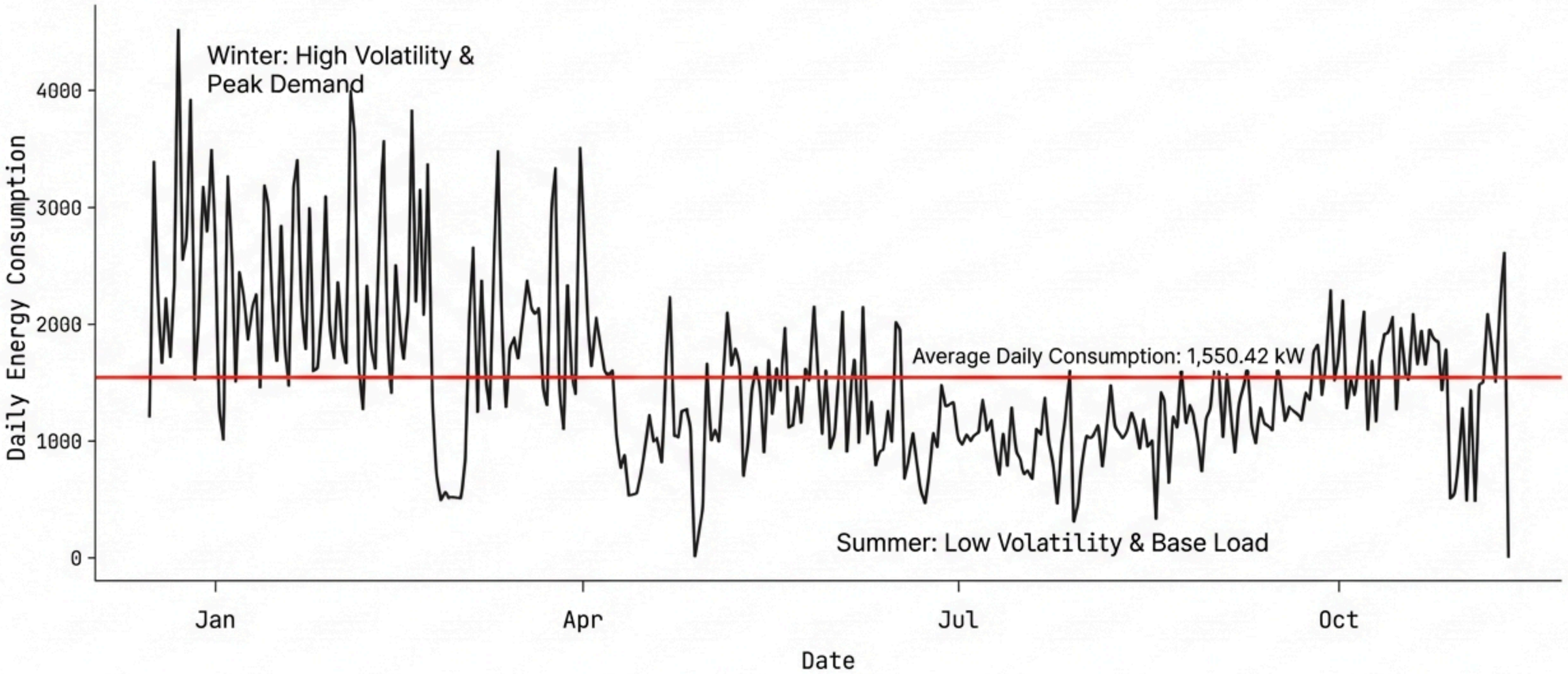
Date Range: 16/12/2006 – Nov 2007

Primary Metric: Global Active Power (kW)

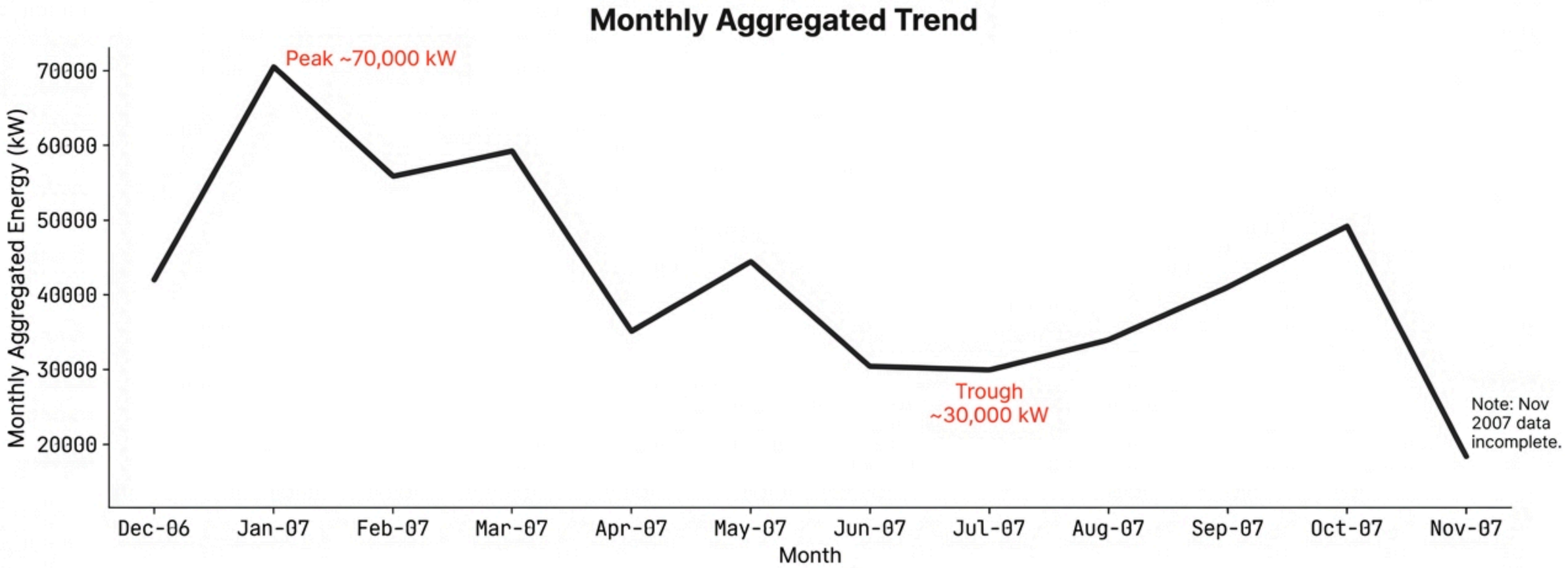
Methodology

Raw data contained 474,160 observations. Missing values were removed via `na.omit`, character dates were converted to Date objects, and data was aggregated by daily sum to reduce noise.

The Annual Heartbeat: Consumption follows a strict seasonal rhythm



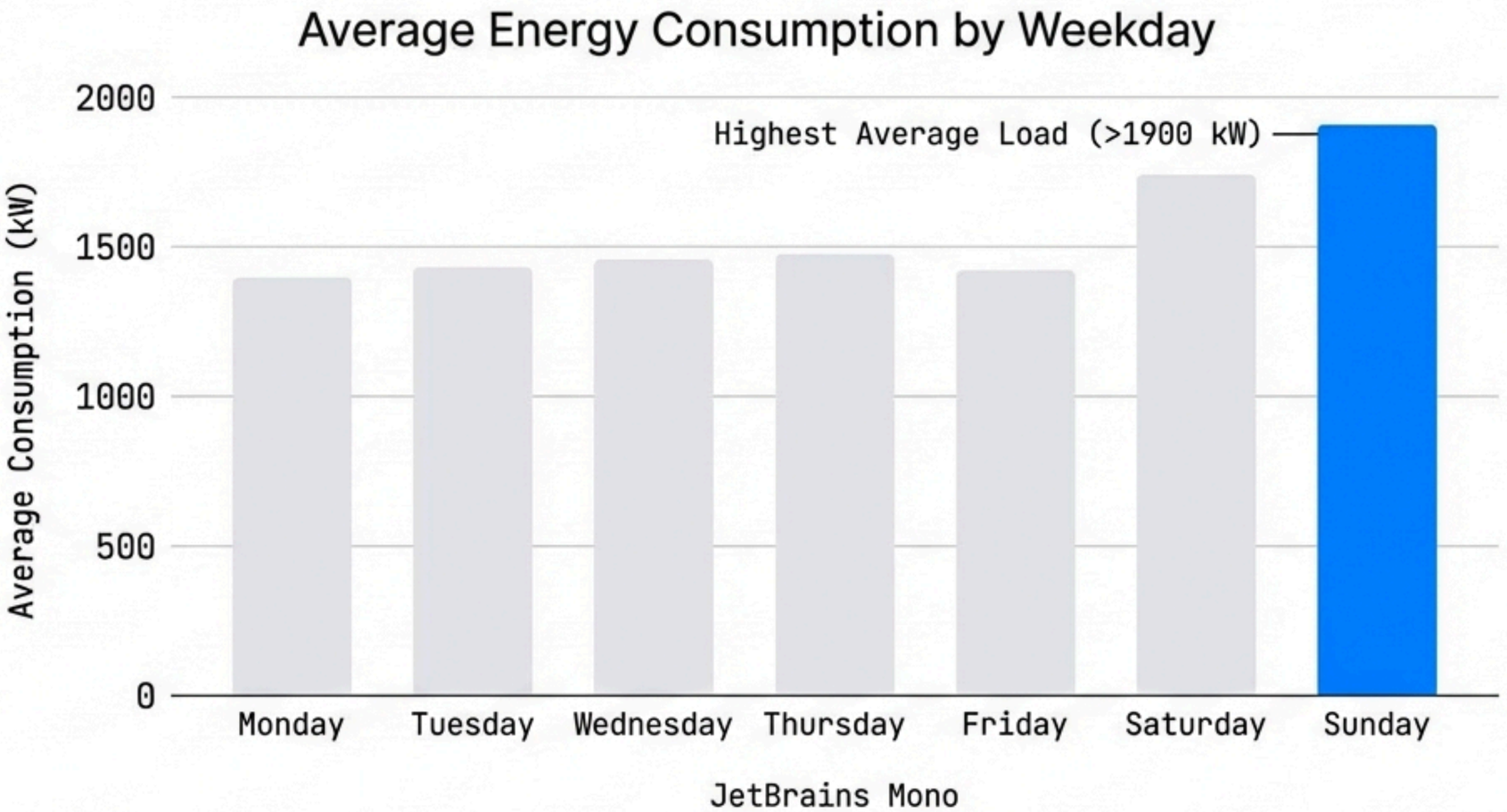
Winter demand is nearly double that of Summer



This curve suggests that heating is the primary variable load for this household.

Editorial Data Journalism

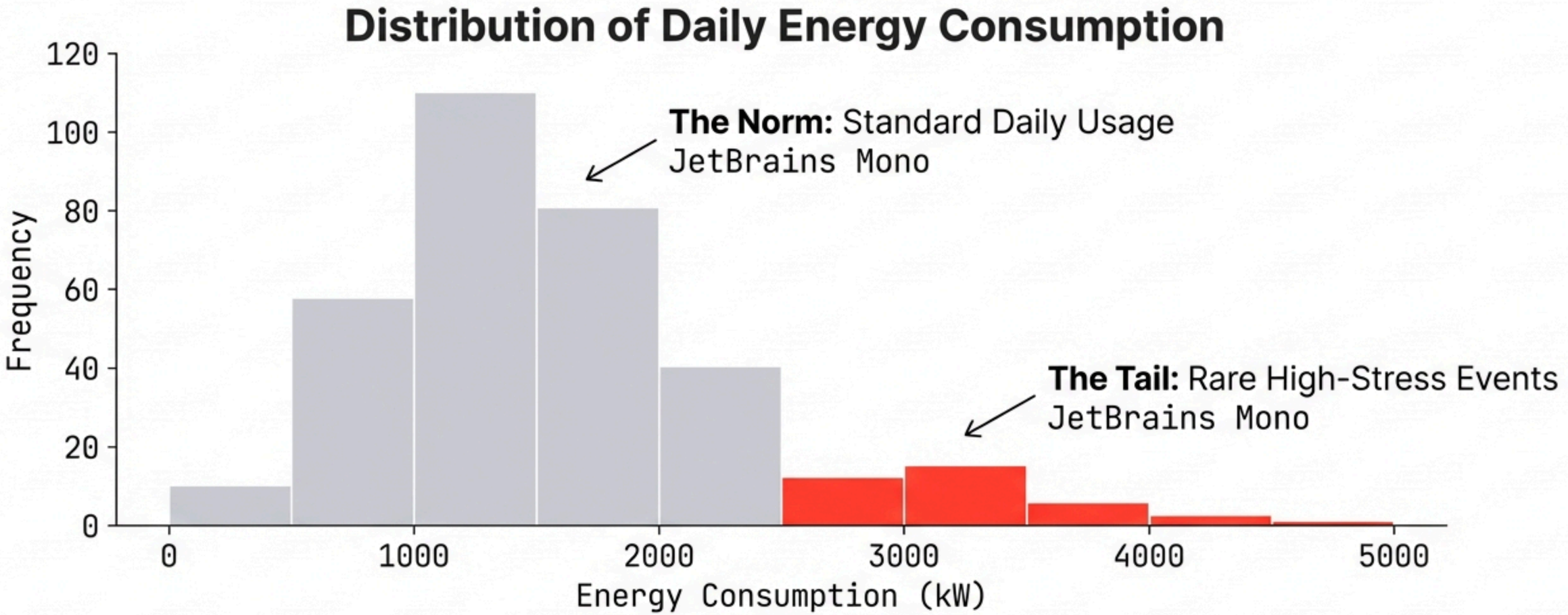
The ‘Sunday Roast’ Effect: Weekends drive the highest loads



Energy efficiency measures targeting ‘lifestyle’ usage (laundry, cooking, heating while home) will have the highest impact on weekends.

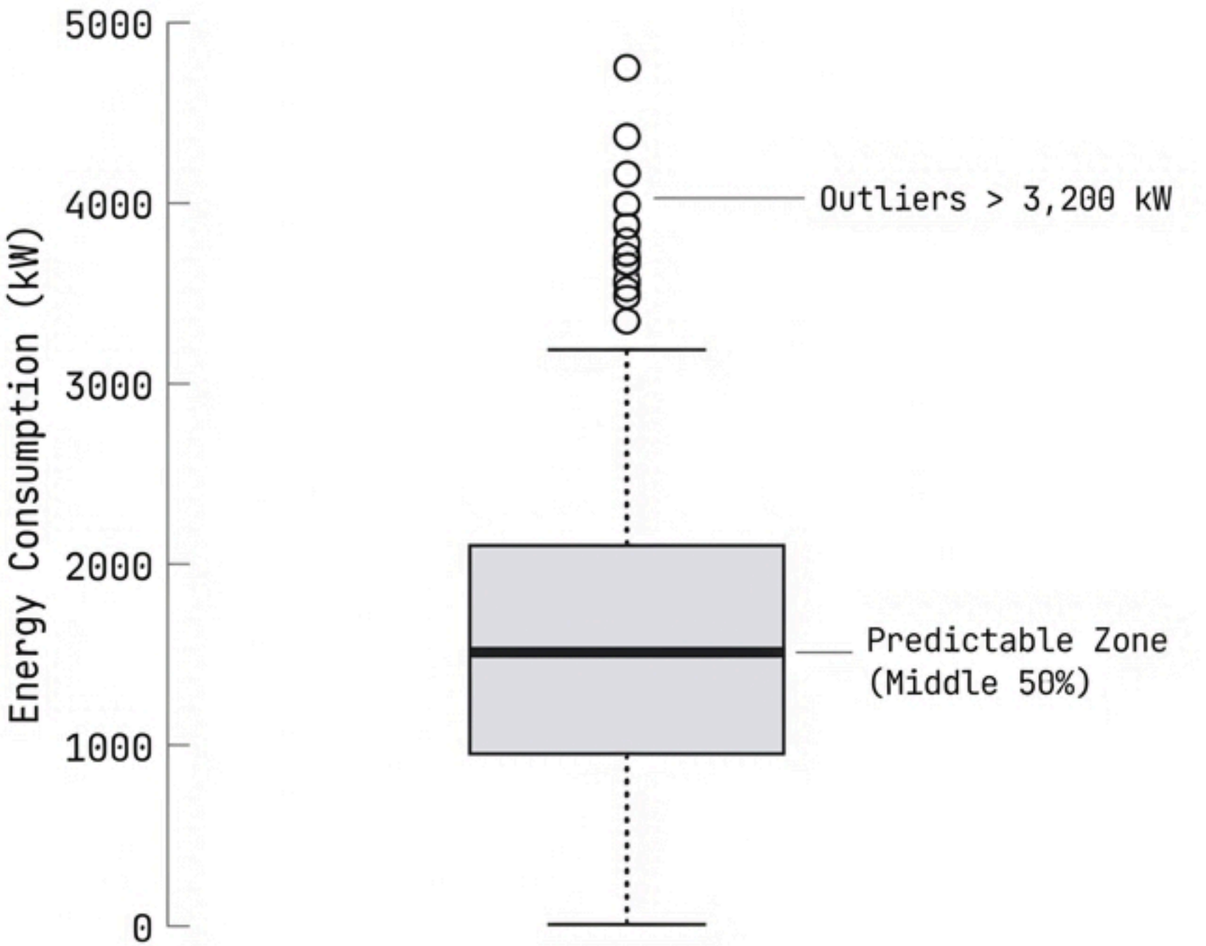
Inter Regular

Predictability vs. Risk: A Right-Skewed Distribution



The data is heavily right-skewed. While the mode is manageable, the long tail indicates that extreme consumption days are drastic in magnitude.

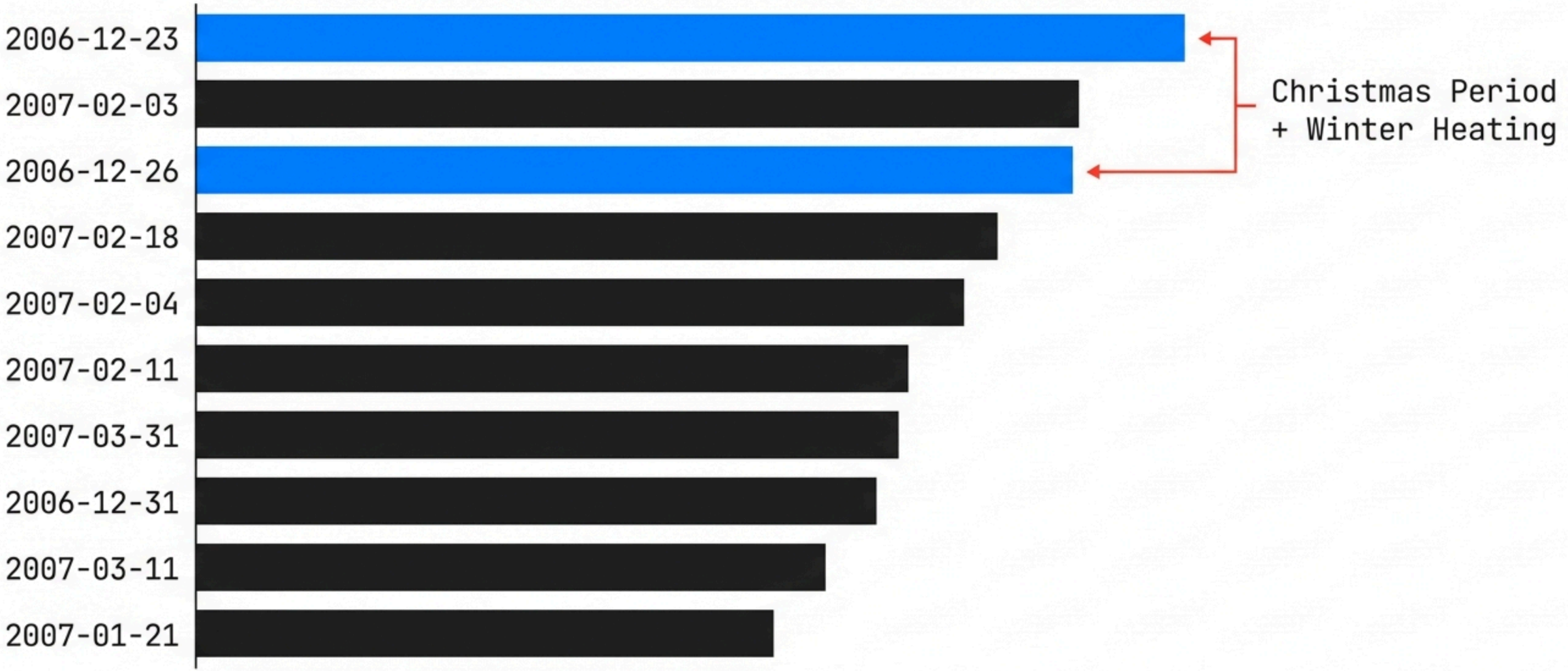
Visualizing the Extremes



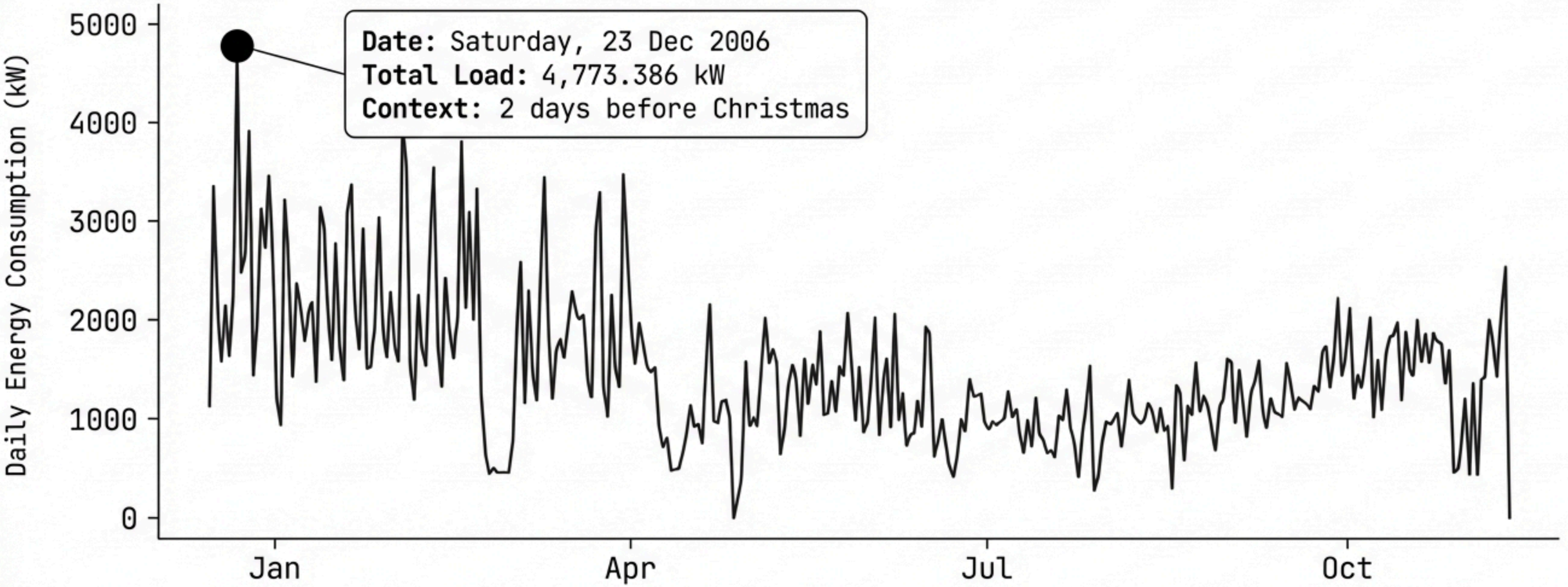
There are zero outliers on the low end, but a significant cluster of high-stress outliers on the top end. The risk is entirely one-sided.

Editorial Data Journalism

The Top 10 Heaviest Days are all in Winter



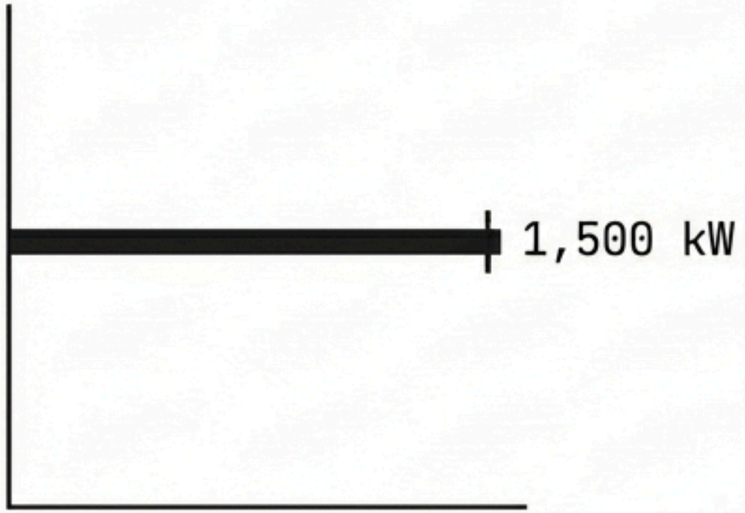
Anatomy of a Peak: 23rd December 2006



Occurring on a Saturday just two days before Christmas, this peak represents the ‘Perfect Storm’ of variables: Winter Seasonality + Weekend Lifestyle + Holiday Activity.

Editorial Data Journalism

Strategic Implications & Planning



Base Load Baseline

Planners can expect a reliable base daily load of **~1,500 kW**.

Capacity Buffer

Systems must be engineered to handle peaks of **4,800 kW** to avoid failure during winter extremes. A buffer of **3x the average** is required.

Targeted Efficiency

Efficiency campaigns should not be “always on.” They should target specific windows: Weekends and the Dec-Feb seasonal block.

Technical Appendix: Analysis Code

```r
# 1. Read dataset
data <- read.csv("household_power_consumption.txt",
                 sep=";",
                 header=TRUE,
                 na.strings="?",
                 stringsAsFactors=FALSE)


data$Date <- as.Date(data$Date, format="%d/%m/%Y")
data$Global_active_power <- as.numeric(data$Global_active_power)
data <- na.omit(data)
```

```r
# 4. Daily Aggregation
daily_consumption <- aggregate(Global_active_power ~ Date,
                               data=data,
                               sum)


# 5. Monthly Trend
daily_consumption$Month <- format(daily_consumption$Date, "%Y-%m")
monthly_data <- aggregate(Global_active_power ~ Month,
                          data=daily_consumption,
                          sum)
```

```r
# 1. Daily Trend (Line Chart)
plot(daily_consumption$Date,
     daily_consumption$Global_active_power,
     type="l",
     xlab="Date",
     ylab="Daily Energy Consumption (kW)",
     main="Daily Energy Consumption Pattern")

abline(h = mean(daily_consumption$Global_active_power),
       col="red", lwd=2)
```

Full reproducible R script available in source documentation.