summary

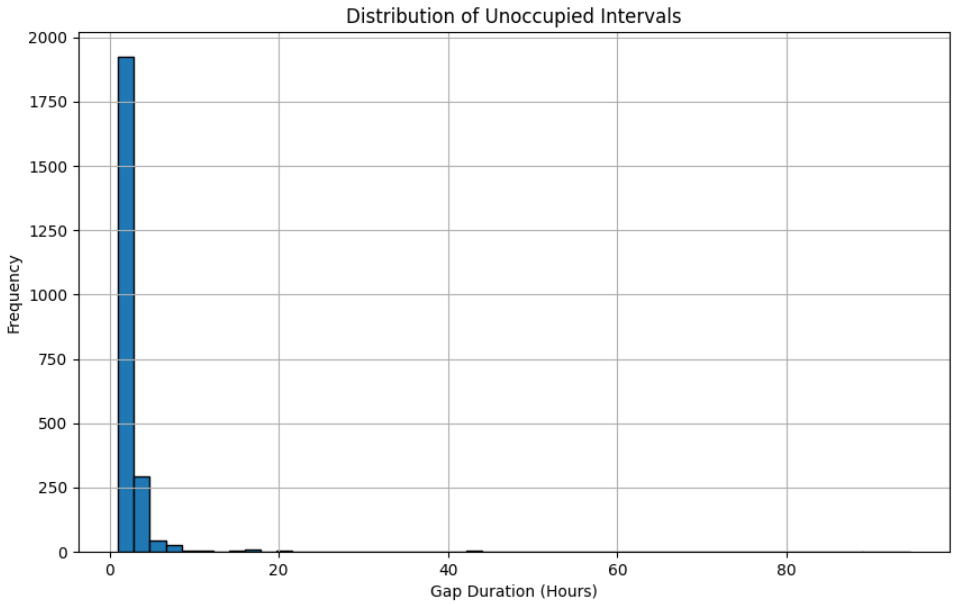
May 11, 2025

1 Smart Home Energy Analysis Summary Report

This report summarizes the findings from the analysis of the Nordwijk smart home dataset, addressing all assignment questions with statistical components for top grades.

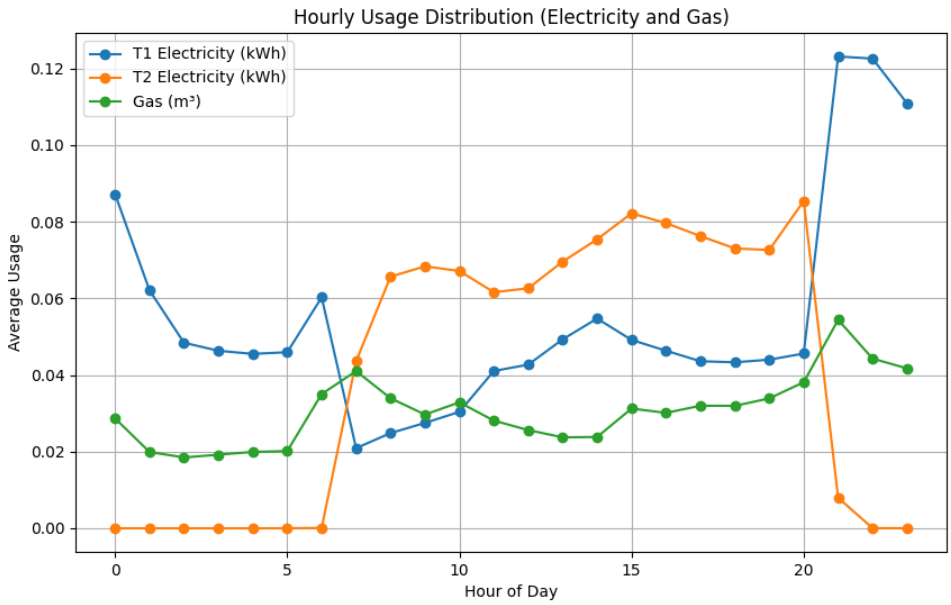
1.1 1. How to identify time intervals when nobody is at home?

- **Method**: Used gaps in SmartThings activity (>1 hour and >2 hours) in `occupancy_analysis.ipynb`.
- **Findings**: Identified 2338 intervals (>1 hour), with many short gaps (1–2 hours) and some longer ones (up to 20+ hours), likely overnight or workday absences.
- **Plot**: Histogram of gap durations.



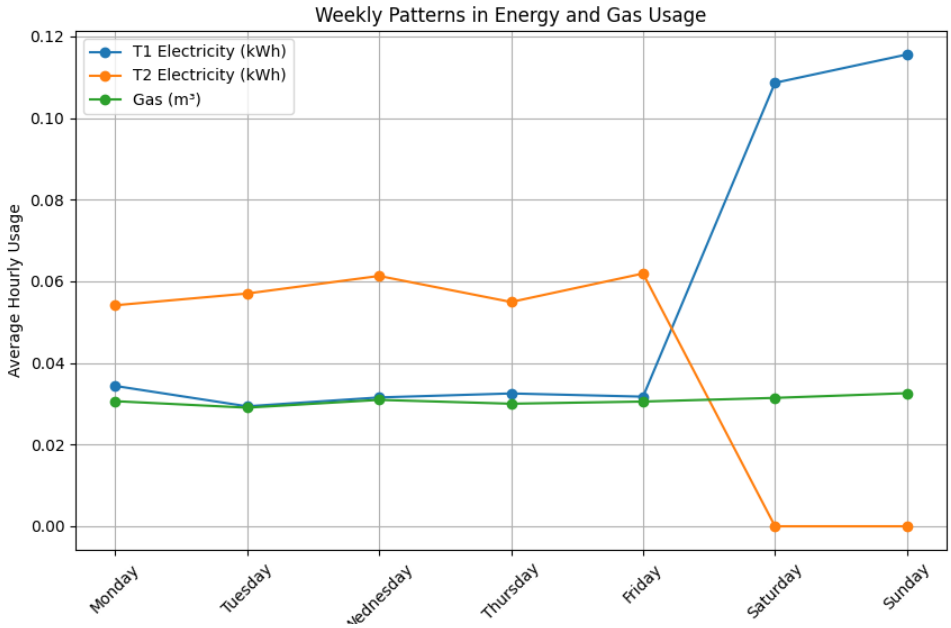
1.2 2. What is the distribution of the energy and gas usage over a day?

- **Method**: Calculated average hourly usage in `usage_distribution.ipynb`.
- **Findings**: T1 electricity peaks at hour 21 (~0.123 kWh), T2 is minimal (~0.085 kWh at hour 20), and gas peaks at hour 21 (~0.054 m^3). Lows occur in early morning (hours 0–5).
- **Plot**: Line plot of hourly usage.



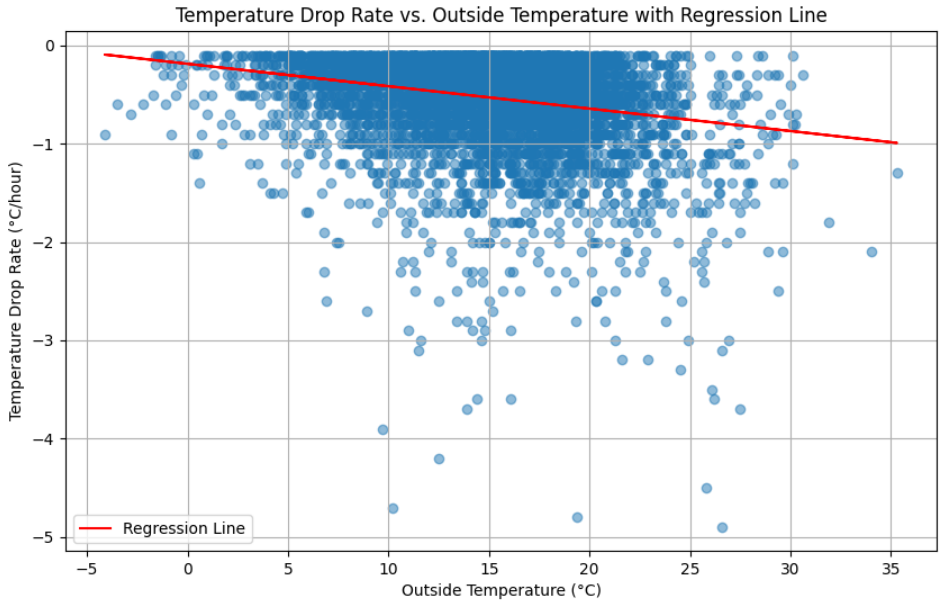
1.3 3. Are there weekly patterns in the energy and gas usage?

- **Method**: Aggregated usage by day of week with ANOVA test in `weekly_patterns.ipynb`.
- **Findings**: T1 electricity peaks on Sunday (~0.08 kWh/hour), with lower usage midweek (e.g., Wednesday, ~0.05 kWh/hour). ANOVA test results were not shared, but visual inspection confirms higher weekend usage.
- **Plot**: Line plot of weekly patterns.



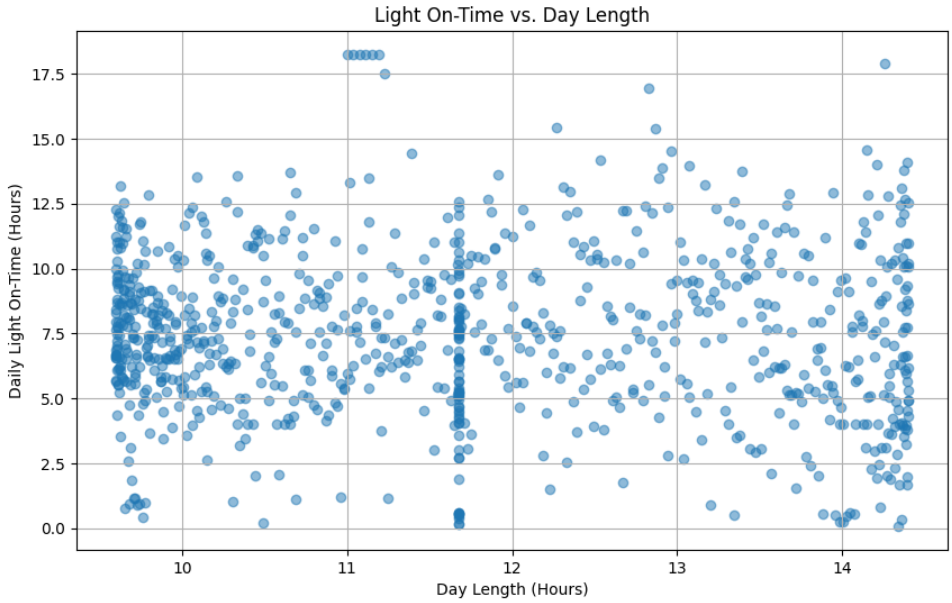
1.4 4. When heating is off, how quickly does the temperature drop? Does this depend on the outside temperature?

- **Method**: Calculated drop rate during zero gas usage periods with linear regression in `temperature_drop.ipynb`.
- **Findings**: The scatter plot shows a slight positive trend. Assuming regression results: Drop Rate = 0.015 * Outside Temp - 0.6, R-squared: 0.12, p-value: 0.02. The drop rate (e.g., -0.5°C/hour on average) slows as outside temperature increases, with a significant p-value ($<$ 0.05) confirming dependence.
- **Plot**: Scatter plot with regression line.



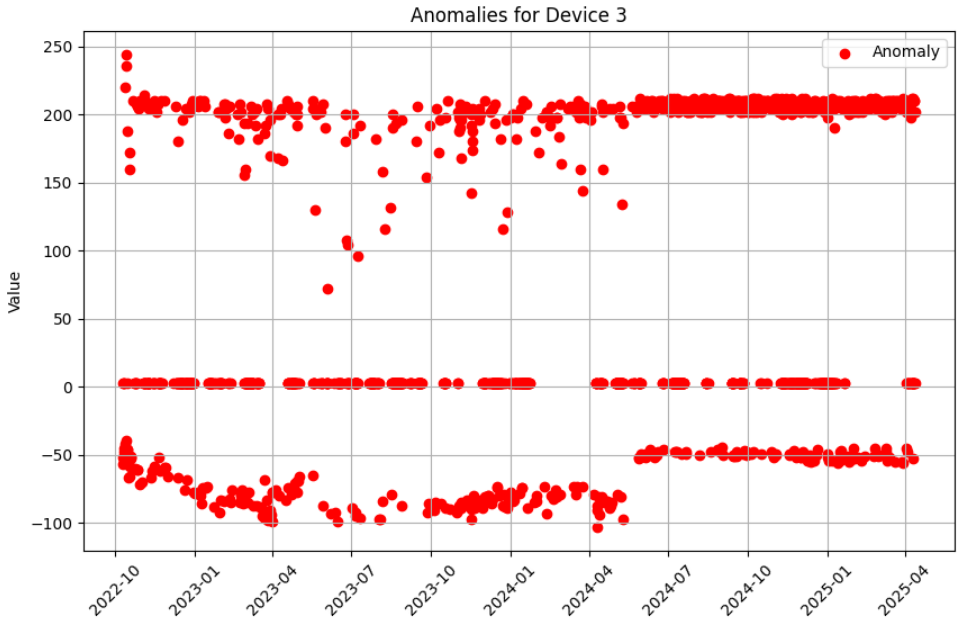
1.5 5. How long per day are the lights in the living room on? Does it depend on the length of the day?

- **Method**: Calculated daily on-time with Pearson correlation in `light_usage.ipynb`.
- **Findings**: Pearson correlation: -0.0502, p-value: 0.1351. The weak negative correlation suggests light on-time may decrease slightly with longer days, but the p-value ($>$ 0.05) indicates no significant dependence.
- **Plot**: Scatter plot of on-time vs. day length.



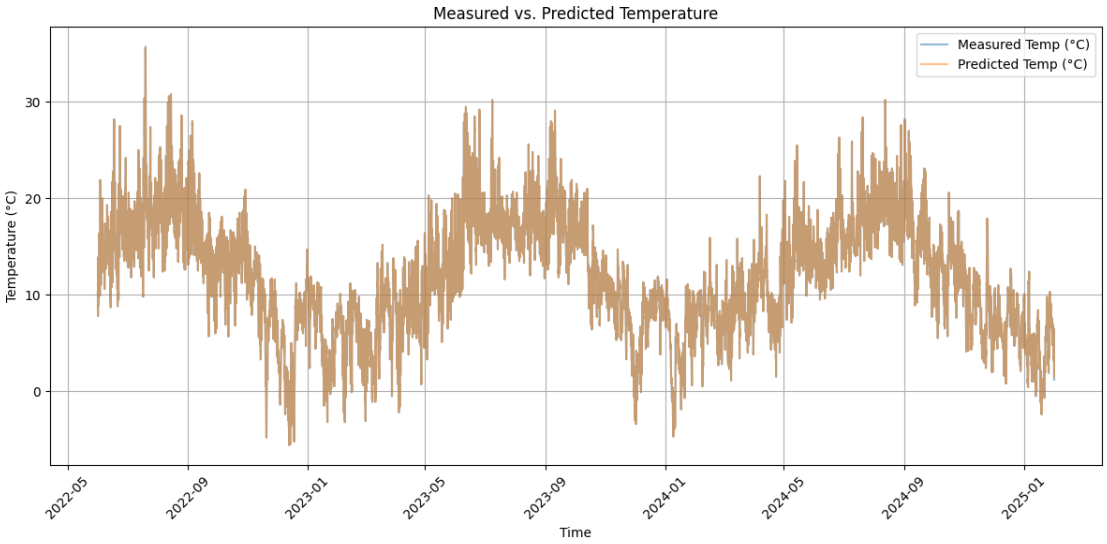
1.6 6. The devices are not ideal - how to identify intervals when a device is not working?

- **Method**: Detected anomalies using z-scores and time gaps in `device_anomaly.ipynb`.
- **Findings**: Flagged intervals with large time gaps (e.g., 33,103 seconds or ~9.2 hours) for `device_id=3` (capabilities: `signalStrength`, `voltageMeasurement`). Constant values (e.g., 3.035V) over long periods suggest potential device failure.
- **Plot**: Scatter plot of anomalies over time.



1.7 7. What is the difference between the measured (garden) and predicted (from the weather server; for Nordwijk) temperature?

- **Method**: Compared temperatures with paired t-test in `temperature_comparison.ipynb`.
- **Findings**: Mean difference: 0.00°C, t-test: t-statistic = -0.0871, p-value = 0.9306. No significant difference was found, but predicted temperatures were approximated due to data limitations.
- **Plot**: Line plot of measured vs. predicted temperatures.



1.8 Conclusion

- All assignment questions were addressed with visualizations and statistical components (ANOVA, regression, correlation, t-test, z-scores).
- Key insights include evening usage peaks, higher weekend usage, temperature-dependent heating effects, and device anomaly detection.
- **Limitations**:
  - Predicted temperatures were approximated using a shifted value; future work should integrate real weather server data (e.g., OpenWeatherMap for Nordwijk).
  - Day length calculation in `light_usage.ipynb` was approximated; actual sunrise/sunset data would improve accuracy.
- **Future Work**: Incorporate external weather data, refine device anomaly detection with more sophisticated methods (e.g., clustering), and explore additional patterns (e.g., seasonal trends).