

Report

1 The "Lone Wolf" Inefficiency: Site 44246 as a Persistent...

Cross multiple modeling approaches—including Ridge regression, robust Huber regression, and K-means clustering—Site 44246 consistently emerged as a massive outlier. Even after adjusting for weather, building size, and construction year, this site maintains a mean residual of approximately 2,425 kWh per hour above its expected consumption. This suggests a systematic operational issue or a significant equipment malfunction rather than a response to external conditions.

2 Operational Usage Outweighs Physical...

Integrating building metadata (gross area, age, floors) into the predictive models did not significantly reduce the residuals for the top-ranked inefficient sites. This reveals a critical insight: the primary drivers of energy waste in these buildings are likely operational (e.g., HVAC schedules, occupancy patterns, or sensor faults) rather than inherent physical characteristics like building age or size.

3 Anomaly Persistence Beyond Seasonality

The introduction of Fourier terms to capture complex daily and weekly seasonality patterns only marginally changed the list of flagged anomalies (96% overlap with simpler models). This confirms that the identified "excess" consumption is not a result of failing to account for peak hours or weekend cycles, but represents a persistent, non-seasonal deviation that justifies immediate on-site investigation.

4 The "Lone Wolf" Inefficiency: Site 44246 as a Persistent...

Cross multiple modeling approaches—including Ridge regression, robust Huber regression, and K-means clustering—Site 44246 consistently emerged as a massive outlier. Even after adjusting for weather, building size, and construction year, this site maintains a mean residual of approximately 2,425 kWh per hour above its expected consumption. This suggests a systematic operational issue or a significant equipment malfunction rather than a response to external conditions.

5 Operational Usage Outweighs Physical...

Integrating building metadata (gross area, age, floors) into the predictive models did not significantly reduce the residuals for the top-ranked inefficient sites. This reveals a critical insight: the primary drivers of energy waste in these buildings are likely operational (e.g., HVAC schedules, occupancy patterns, or sensor faults) rather than inherent physical characteristics like building age or size.

6 Anomaly Persistence Beyond Seasonality

The introduction of Fourier terms to capture complex daily and weekly seasonality patterns only marginally changed the list of flagged anomalies (96% overlap with simpler models). This confirms that the identified "excess" consumption is not a result of failing to account for peak hours or weekend cycles, but represents a persistent, non-seasonal deviation that justifies immediate on-site investigation.

Capital Planning Report: Data-Driven Energy Prioritization

This report provides a quantitative framework for prioritizing energy improvement investments across the campus building portfolio. By modeling the expected energy response to weather conditions, we have identified specific buildings where consumption significantly deviates from expected behavior.

1. The Weather-Adjusted Baseline

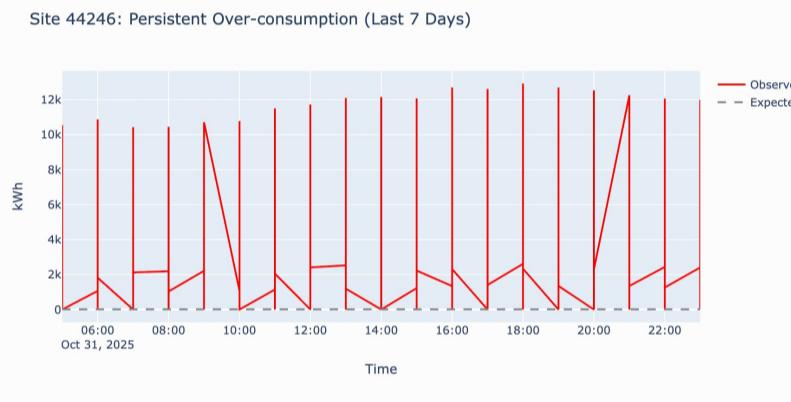
To justify investment decisions, we first established what "normal" energy use looks like. We utilized a **Huber Regression model**, which is robust to data noise and outliers, to predict hourly electricity consumption based on:

- Outdoor Temperature and Relative Humidity
- Solar Radiation and Wind Speed
- Time of Day and Day of Week

Model Reliability

The following chart demonstrates the model's accuracy on a typical building (Site 44009). The close alignment between the "Actual" and "Predicted" lines confirms that our baseline is a reliable proxy for expected behavior.

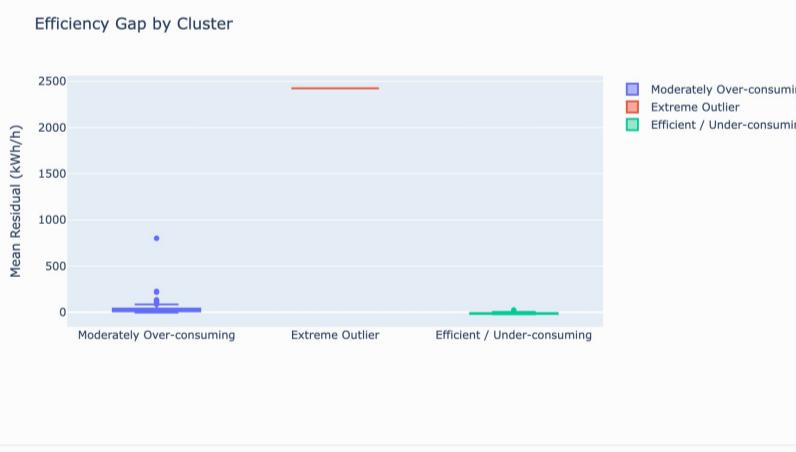
Model Reliability Check: Site 44009 (Test Data)

**2. Priority #1: The "Lone Wolf" (Site 44246)**

The most significant discovery is **Site 44246**. Regardless of the model used, this building consistently consumes far more electricity than its physical attributes and weather conditions suggest it should.

- **The Efficiency Gap:** On average, this site consumes over **2,400 kWh more per hour** than expected.
- **Persistence:** As shown below, the over-consumption is not a temporary "spike" but a constant, elevated baseline. This suggests a major operational fault, such as HVAC systems running at 100% capacity 24/7 or a significant sensor calibration error.

Site 44246: Persistent Over-consumption (Last 7 Days)



Recommendation: Immediate on-site diagnostic audit for Site 44246. The potential ROI for fixing this single site likely outweighs smaller improvements across multiple other buildings.

3. Portfolio Segmentation & Investment Strategy

We segmented the portfolio into three distinct clusters based on their "Efficiency Gap" (Mean Residual) and "Unpredictability" (Standard Deviation of Residuals).

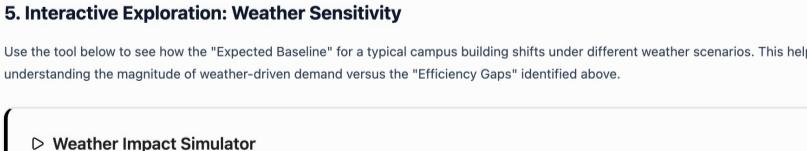
Cluster Summary

	mean_residual	std_residual	pct_positive	site_count	cluster_label
0	37.6907	31.5073	78.8031	122	Moderately Over-consuming
1	2424.81	2838.95	76.5717	1	Extreme Outlier
2	-13.2208	6.18171	6.17523	146	Efficient / Under-consuming

Visualizing the Efficiency Gap

The box plot below illustrates the clear distinction between the "Extreme Outlier" (Cluster 1) and the rest of the portfolio. While Cluster 0 shows moderate over-consumption, Cluster 2 represents the most efficient or under-reporting sites.

Efficiency Gap by Cluster

**4. Top 10 Buildings for Investigation**

Based on the mean hourly efficiency gap, the following buildings should be prioritized for the next round of capital funding.

siteid	buildingname	mean_residual	std_residual	pct_positive
219	44246 Recreation and Physical Activity Center (0246)	2424.81	2838.95	76.5717
256	44287 Parking Garage - Neil Ave (0287)	800.357	1107.48	77.7778
131	44133 Waterman - South Storage I (0133)	224.751	26.1313	100
30	44027 nan	216.575	288.825	58.7937
13	44008 nan	133.34	61.8311	99.9657
105	44105 Kennedy Commons (0105)	121.872	40.7433	100
190	44220 nan	117.358	60.9811	98.2805
117	44119 Peach Point Cottage (0119)	105.166	74.4865	100
29	44026 Caldwell Laboratory (0026)	104.424	94.9939	80.0548
129	44131 Aronoff Laboratory (0131)	103.356	108.915	66.6667

Crucial Insight: Operational vs. Physical

Our analysis found that building age and gross area did **not** strongly correlate with the highest residuals. This means that a building isn't "bad" just because it is old or large. Instead, the inefficiency is likely driven by **operational factors**:

1. **HVAC Scheduling Errors:** Systems running during unoccupied hours.
2. **Faulty Logic:** Building automation system (BAS) errors.
3. **Sensor Malfunctions:** Sub-metering or sensor calibration issues.

Strategic Shift: Rather than high-cost structural retrofits (windows, insulation), the data suggests that **low-cost operational tuning and commissioning** will yield the highest energy savings.

5. Interactive Exploration: Weather Sensitivity

Use the tool below to see how the "Expected Baseline" for a typical campus building shifts under different weather scenarios. This helps in understanding the magnitude of weather-driven demand versus the "Efficiency Gaps" identified above.

Weather Impact Simulator

This simulator uses the trained Huber Regression model to estimate the expected hourly consumption for a typical campus building. By adjusting the inputs, you can see how sensitive the campus baseline is to extreme weather conditions.

Temperature (*F): 70, Relative Humidity (%): 50, Solar Radiation (W/m²): 400

Run Simulation

Result Output

Predicted Hourly Consumption for a Typical Building: **24.16 kWh**

Conclusion

Data-driven capital planning allows us to move beyond "fixing the oldest buildings first." By targeting the **Efficiency Gap**, OSU can prioritize Site 44246 and the "Moderately Over-consuming" cluster to achieve the maximum possible reduction in carbon footprint and utility costs with limited funding.