Energy Data Analysis: 30-Day Review of Interval and Register Records

Over a 30-day period, data was collected for an apartment complex managed by Metergy Solutions and I analyzed interval and register data to check how complete and reliable the energy readings were. The goal was to identify any missing information, spot potential data quality issues, and draw out insights about daily energy usage. The data came from multiple JSON files, which I processed into three main datasets: one that included 15-minute interval readings (called *combined_df*), hourly interval readings (called *combined_interval_df*) and another that tracked daily energy totals from the registers (*combined_register_df*).

Missing Data and Time Coverage

To check if the data covered the full month as expected, I first converted all the timestamps into datetime format. Then I looked at how many unique dates were in the 15-minute interval dataset. While we were aiming for a complete 30-day window, only 27 days were accounted for. This suggests that three days of data might be completely missing, or that some days have only partial coverage. The interval data showed a similar pattern, which confirms the gap is likely consistent across both types of readings.

To pinpoint exactly which days were missing, I compared the available timestamps against a full calendar generated using *pd.date_range()*. This helped highlight the days that weren't captured at all (6/13-6/16). While having every single day isn't always necessary in practice, it's important to flag missing data early so we know what we're working with and can take it into account when analyzing energy usage patterns or generating reports.

Data Quality Review

A thorough review of data quality revealed that the dataset was structurally sound. Key fields such as timestamps, measurements, and source identifiers contained no missing values. All timestamps were valid and correctly converted, and no duplicate rows were found,indicating that the upstream data pipeline is functioning reliably.

However, a few data integrity issues were observed. Within the interval dataset, 90 records reported a measurement of exactly 0.0 kW. While this may reflect legitimate periods of

inactivity, it could also point to sensor malfunctions or transmission issues. These zero readings were dispersed, but their impact could be more significant if found concentrated around specific dates, such as weekends or known outage windows.
Outlier detection using the 99.9th percentile threshold flagged several abnormally high readings. These spikes may be valid, potentially triggered by HVAC systems or large equipment, but should be reviewed further to rule out logging anomalies. Additionally, two full days exhibited identical measurements across all intervals, a strong indicator of device freeze or data logging failure. These days should be considered unreliable for analysis.

Usage Pattern Observations

Despite a few anomalies, the dataset presents a consistent and interpretable picture of energy consumption over the month. Most days followed a recognizable usage curve: gradual increases throughout the day, peak activity during the late afternoon or evening, and declines overnight; typical of residential or mixed-use environments.
The interval trends aligned well with daily register totals, further reinforcing the reliability of the data. That said, flatlined or abnormally low-usage days deviated from this pattern and should be excluded from any modeling or forecasting unless adjustments or explanations are applied.
Overall, after accounting for a handful of suspect days, the dataset provides a reliable foundation for usage monitoring and decision-making.

Final Thoughts and Recommendations

All in all, the data from Metergy is in pretty good shape. It's well-structured, and the vast majority of it is valid and useful. That said, I recommend taking a few extra steps to make sure the analysis stays clean and reliable. Any missing days should be flagged and filled in if needed. Outliers and flatlined intervals should be reviewed or filtered out before doing deeper analysis. It would also help to build a few automated scripts that check for duplicates, validate timestamps, and flag unusual gaps so future data can be processed more smoothly.
By tightening up these quality checks and maintaining a consistent data foundation, Metergy will be in a strong position to keep delivering clear, transparent insights to clients. That kind of trust in the data is essential when it comes to making decisions about energy usage and sustainability.