ISO NE Energy Data Analysis 06/25/2025-06/29/2025

Devasya Mitra

1 Data Cleaning and Collation

Using packages like `pandas` and `numpy`, I employed the following method to weed out the data points that would have caused problem and extrapolated values for missing data:

1.1 Inspection through heatmap

I inspected which all columns can be omitted. I observed the following heatmap:

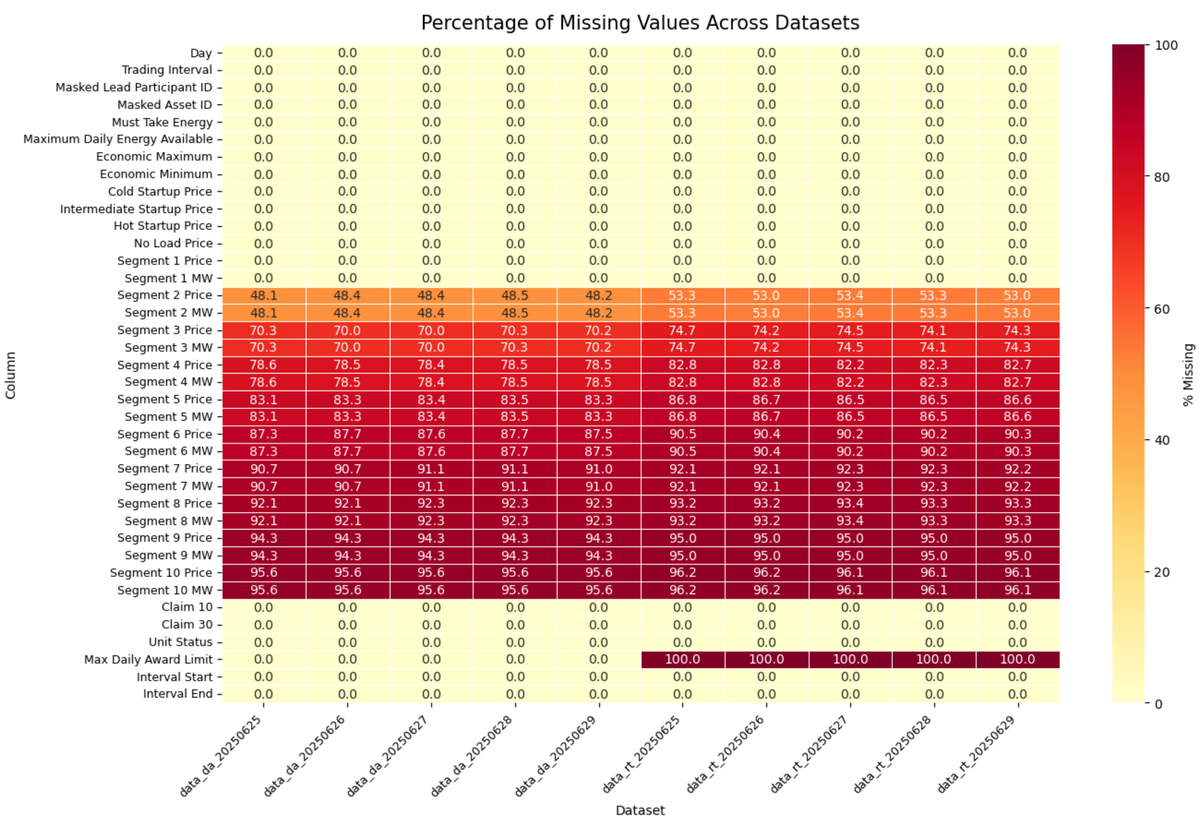

Figure 1: Heatmap of Missing Data

Darker shades indicated more missing values, like the column `Max Daily Award Limit`. Since real-time markets function in real-time, segment-wise declaration of maximum power generation might not be relevant. Thus, no values were found in the real-time market. To be on a safer side, I have omitted this variable altogether so that we can carry out nuanced analysis for other variables.

1.2 Extrapolation algorithm

No treatment was applied when all ten segments were filled. If some segments were empty, assets for which the difference between the total offers across all segments and the economically maximum possible unit was less than 0.1 megawatts were excluded from the analysis. However, if this difference exceeded 0.1 megawatts and there was only one empty segment, the entire difference was assigned to that segment, and its price was estimated using Ordinary Least Squares (OLS). When there were more than one but fewer than nine empty segments, the difference was distributed evenly among them, and the prices were again estimated using OLS. In cases where nine segments were empty, or only one segment was filled, only the next segment was filled, and the difference obtained in the previous step was recorded in the Segment 2 MW column, with its price set equal to that of Segment 1 Price. The flowchart shows the whole process.

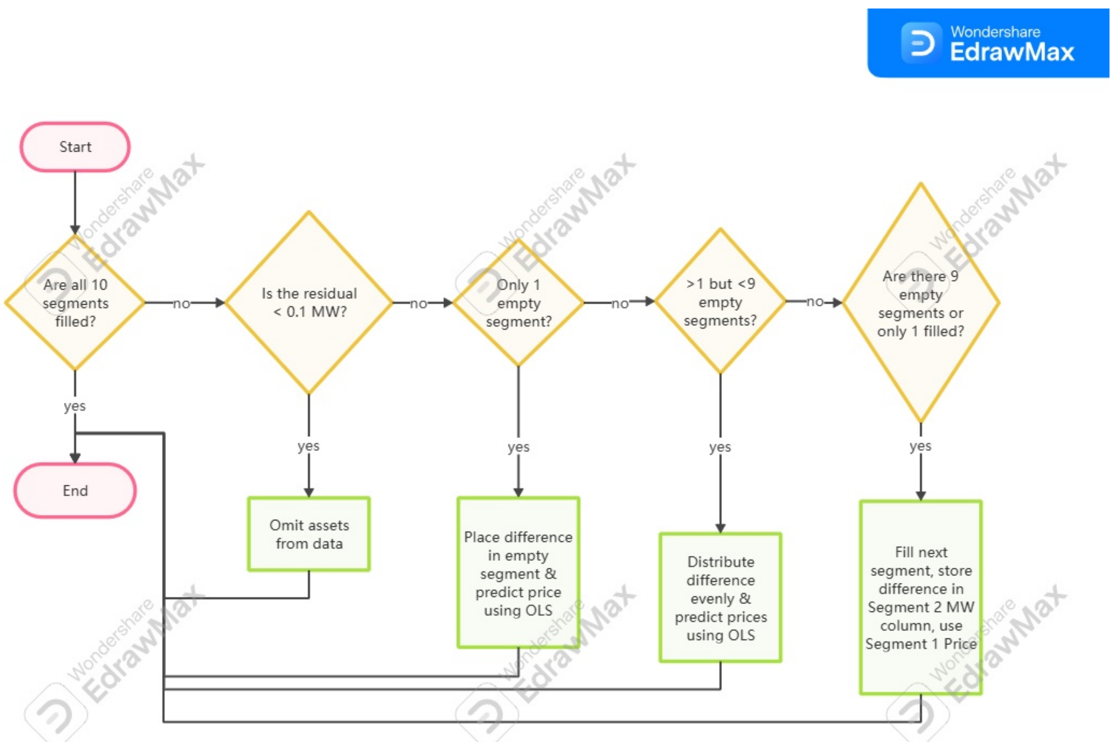

Figure 2: Flowchart showing the data cleaning and interpolation process (Courtesy: edrawmax.com)

After running the required codes, I ended up removing another column named `Unit Status` (which wouldn't have helped in our analysis) and initially came up with a result that contained negative energy units (measured in MW) post extrapolation. The following is that heatmap containing those negative unit values:

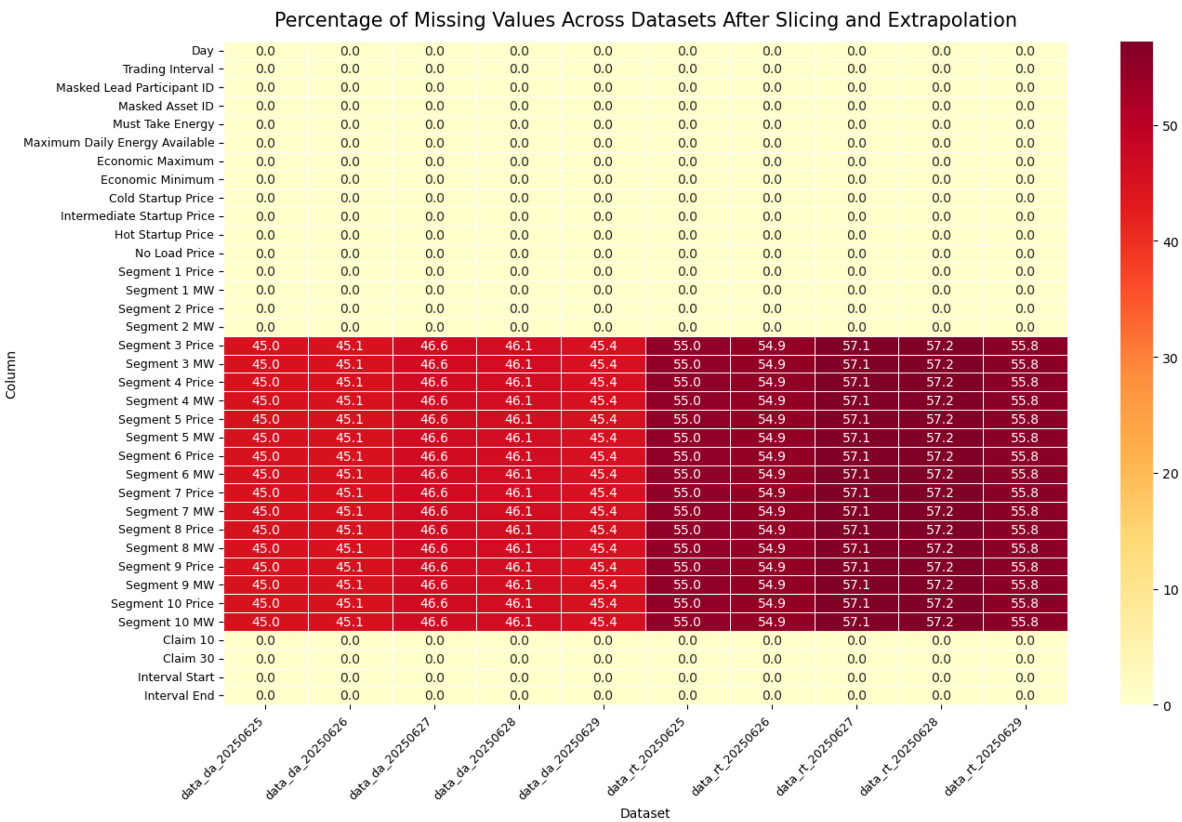

Figure 3: Heatmap: Prior to No Negative Offer Units

However, this would be against the law of physics. That's why I created another notebook, named `cleaning_no_neg.ipynb` and kept those units that had only positive units. Due to oversupply, generators could have negative prices. Also, on the account of removing the negative price assets, not only the number of assets have decreased but also the proportion of missing data has also decreased (missing value lies only because initially, there were some assets which only offered in the first segment and my algorithm did not make predictions due to lack of data). Refer to the heatmap in the next page.

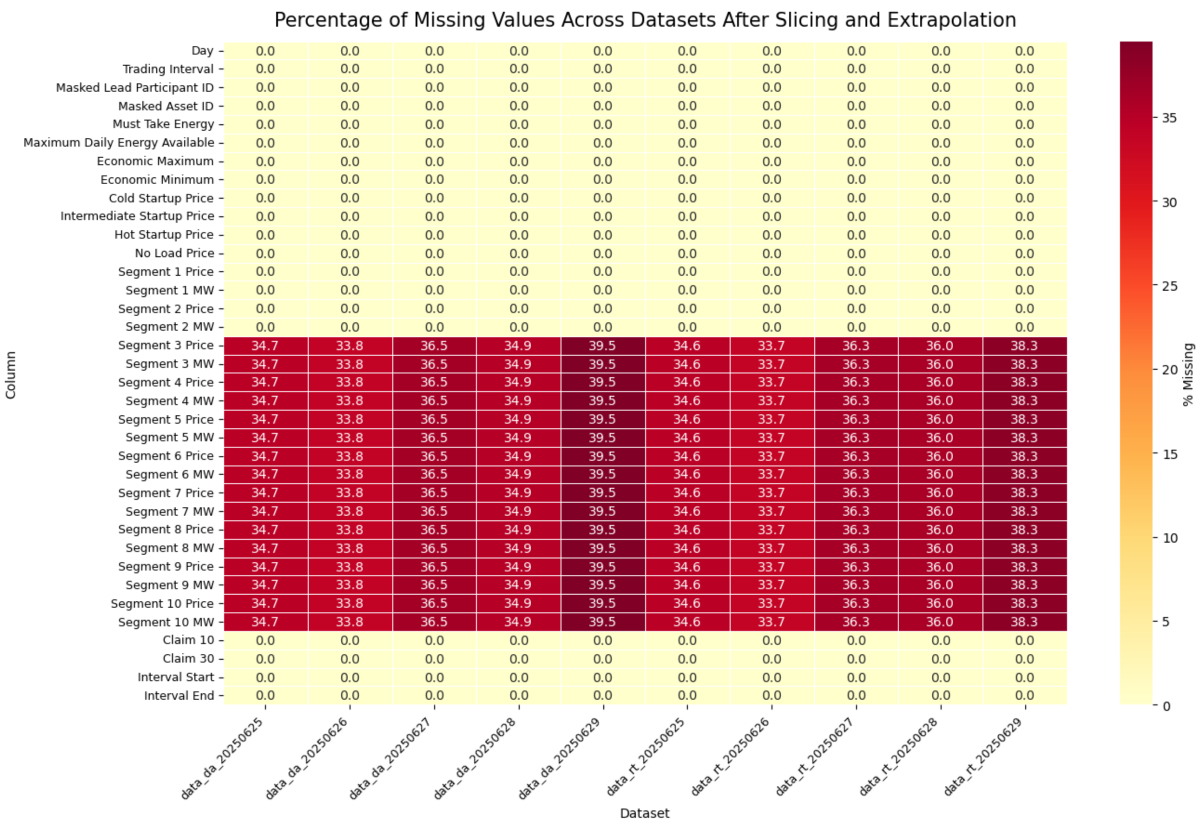

Figure 4: Heatmap After Data Cleaning: No negative Offer Units

2 Exploratory analysis

2.1 Price time series and MW-weighted series

Simple average prices are higher in DA than RT. The simple arithmetic mean makes DA look more expensive, but this effect weakens when we weight by MW. This implies that the higher simple DA average is largely driven by smaller-volume, higher-price offers; the main MW blocks show much smaller differences between markets. For operational decisions and welfare assessment, MW-weighted metrics are therefore recommended.

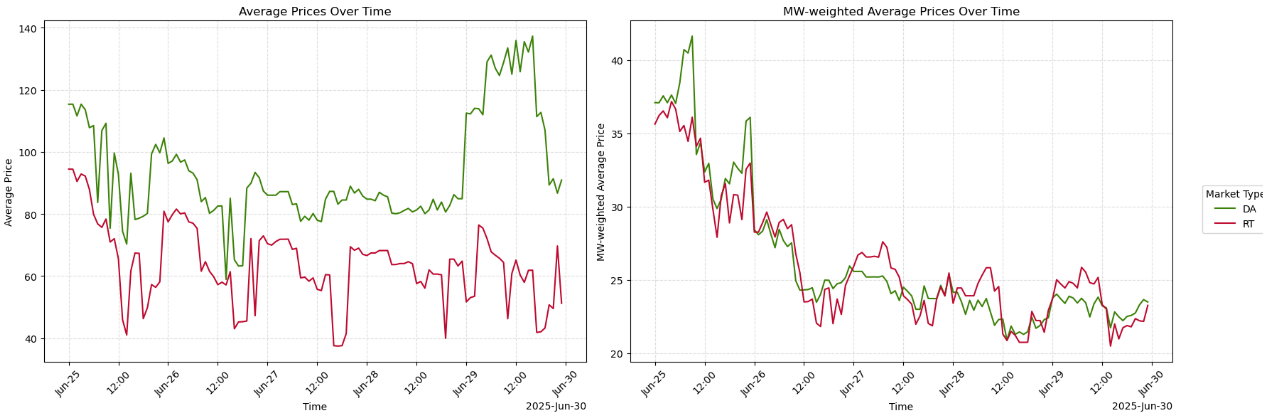

Figure 5: Left: Simple average prices by market type. Right: MW-weighted average prices by market type.

2.2 Negative-price behaviour

Negative price events behave similarly in DA and RT. The parallel, step-like negative-price curves indicate system-wide drivers (for example, oversupply or operational constraints) that affect both markets in tandem, rather than uniquely affecting one market. This suggests that mitigation or policy responses should focus on system-level drivers rather than market-specific idiosyncrasies.

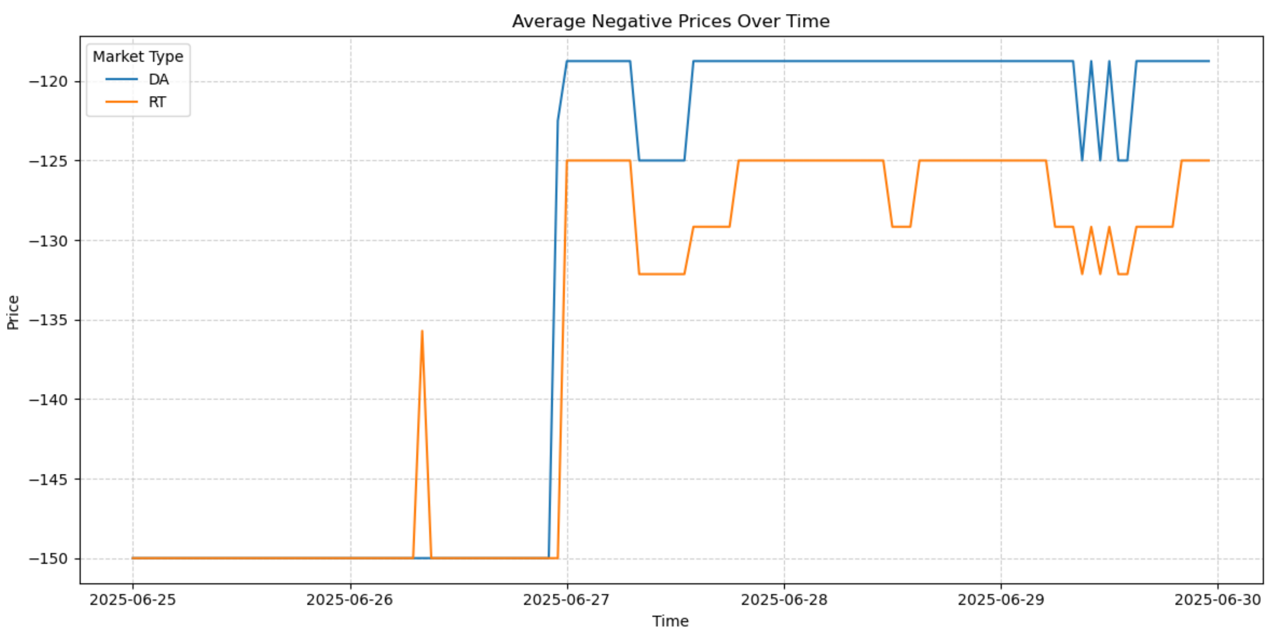

Figure 6: Average negative prices over time for DA and RT (shows parallel, step-like curves).

2.3 Price distribution and skewness

Price distributions are heavily skewed towards low prices for both the markets. Bids are tightly clustered at minimal price levels with a long right tail of high prices. This skewness reduces the usefulness of simple means and argues for robust statistics (median, trimmed mean) or MW-weighted summaries when describing price behaviour.

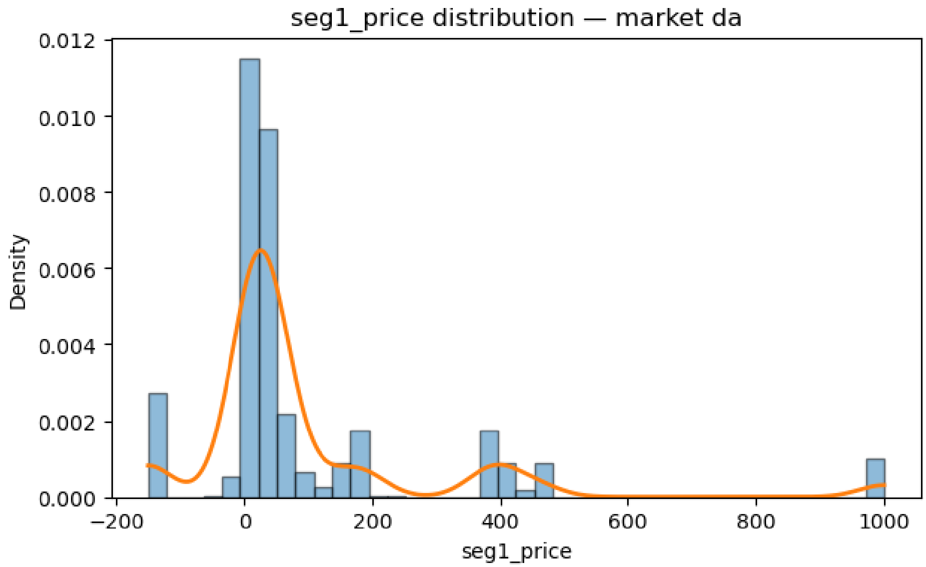

Figure 7: Distribution of segment-1 prices for DA market (histogram with KDE).

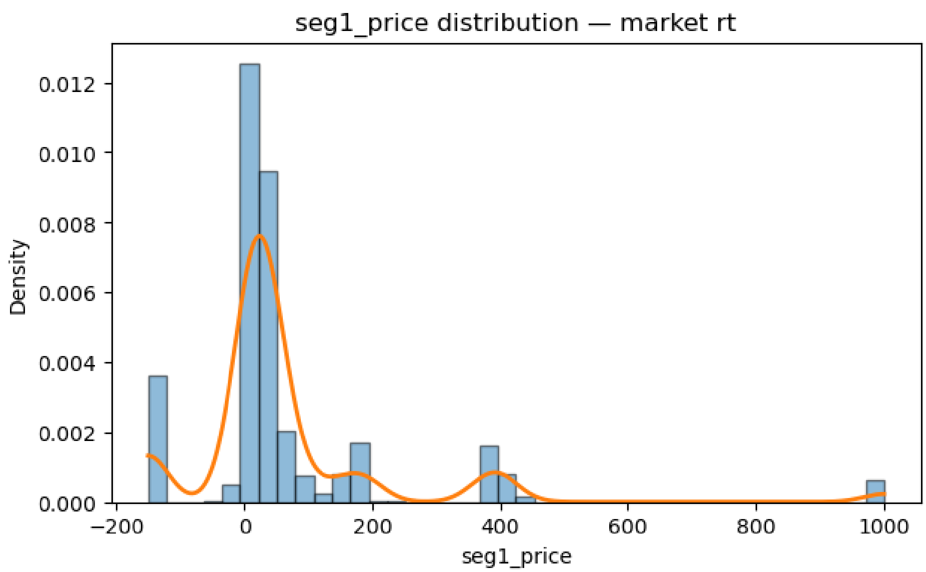

Figure 8: Distribution of segment-1 prices for RT market (histogram with KDE).