CEDAS Hackathon - Part 1

Team AIKI

February 2, 2025

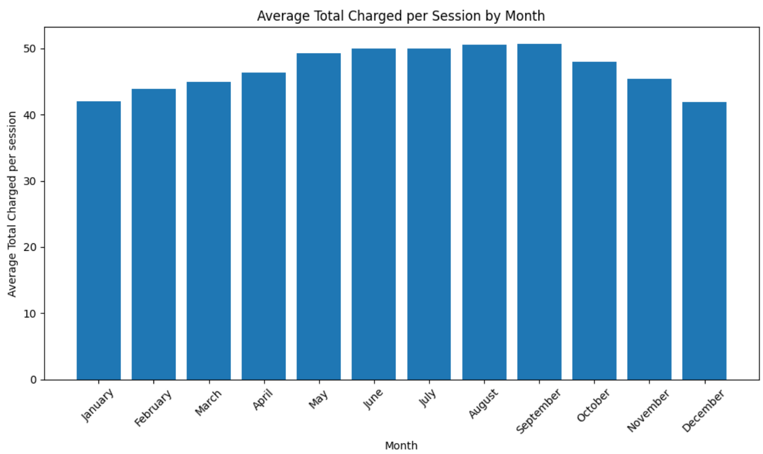
Introduction

This report contains our main results and insights from the exploratory data analysis performed on the charge curves dataset.

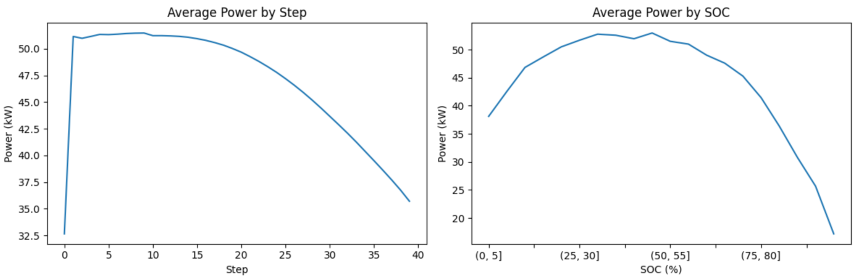
Data Cleaning and Processing

Our first step was removing data points with nan values, and ensuring that the data used further in the analysis was based solely on complete, reliable data. We removed charging sessions with less than 5 loggings, since testing on 15 random samples showed that these generally were abnormal and consisted of faulty/missing measurements.

We assumed that the seasons might affect the charging rate. To investigate this, we calculated average charging speed in kW by dividing the total increase in state of charge (from the start to the end of the session) by the session's duration (converted from seconds to hours). Looking at the mean of charging rate per month, reveals that the colder months generally charge significantly slower than the hotter months. These results suggest that using the month as a feature—potentially using a sine function to account for the cyclical nature of time, could be a logical approach.



Insights



Plots

Left Plot: Average Power by Step
The power starts at a relatively high level (50 kW), remains steady for the first 10-15 steps, then steadily declines until the end of the session (step 40). This trend suggests a typical charging curve where power is reduced after reaching a certain level of battery capacity. The early plateau (steps 5-15) indicates the charger operates at near-nominal power initially, before tapering off.

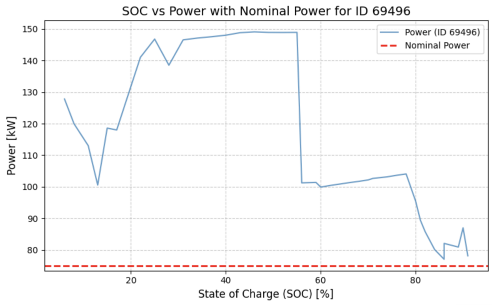
Right Plot: Average Power by SOC
Power increases with SOC up to (50%), peaks between 50-55% SOC, and declines afterward. The rise in power for lower SOC aligns with faster charging rates when the battery is less full.

The decrease after (55%) SOC suggests a protective mechanism to avoid overheating or overcharging the battery as it approaches full capacity.

Identification of Outliers and Problematic Charging Stations

Our analysis of the charging data has revealed multiple charging stations and machines that exhibit anomalies in their power production. These anomalies are defined as instances where the actual charging power (power) exceeds the nominal charging capacity (Nominal Power). Such occurrences may indicate technical faults, calibration errors in measurement instruments, or incorrect reporting of nominal_power values.

One specific example is machine ID 69496 (and many others), where the nominal power is recorded at approximately 78 kW, but the actual power output reached peaks of 150 kW during certain charging cycles. This discrepancy highlights the need for a thorough inspection of the machine's measuring systems and actual operating capacity.



Further analysis at the station level identified charging locations with notably high outlier occurrences. Location ID 100 and ID 38 stand out with 898 and 570 outliers, respectively, marking them as critical areas of concern. These locations should be prioritized for further investigation to uncover any systematic technical issues.

	location_id	Antall outliers
38	100	898
15	38	570
42	110	568
6	10	466
48	122	458
10	25	376
14	37	345
0	0	320
49	125	320
3	5	300

Furthermore, we observed that in many cases, the reported Nominal Power values were incorrectly calibrated, often significantly lower than the actual machine output. This mismatch suggests the need to recalibrate nominal power values to reflect the true operating capacity of the machines.

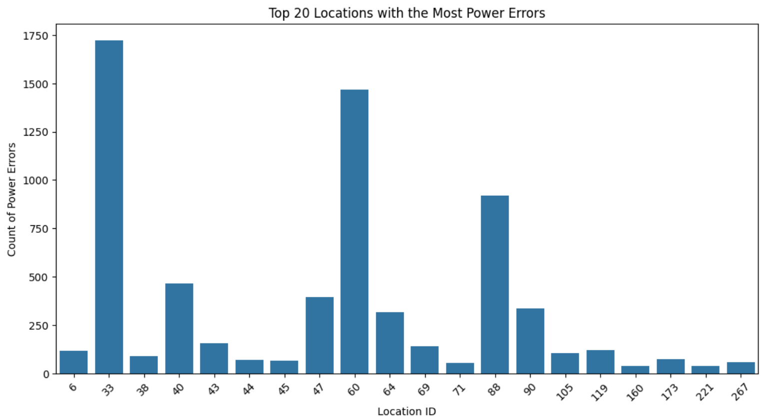
As a conclusion, our analysis highlights clear patterns of anomalies at specific charging stations and machines that warrant further investigation. We recommend a comprehensive technical inspection of the high-outlier locations, particularly Location ID 100 and ID 38, and implementing real-time monitoring systems to ensure accurate operation and prevent future issues.

Top 20 location with the most errors

The bar plot visualizes the results after we implemented a data validation step to identify and flag records with potential errors or outliers in power and SOC values.

Power Errors:
A Boolean flag was created to identify records where the power value falls outside the valid range of 0 to 350 kW. If a record's power value is less than 0 or greater than 350, it is considered invalid, and the flag is set to True for that record.

SOC Errors:
Similarly, a Boolean flag was created to identify records where the State of Charge (SOC) value is less than 0% or greater than 100%. Such records are flagged as invalid with the flag set to True.



Location ID 33 stands out with the highest number of power errors (approx: 1750), far exceeding other locations. Other significant locations include 60 and 88, each with over 1000 errors. Error Patterns: The majority of the errors seem concentrated in a small subset of locations, suggesting that these locations might have: - Faulty equipment. - Consistently unreliable data reporting. - Operational constraints (older chargers).