Math 189 Final report

Team member: Xuanrun Qu, Zixuan Rao, Roham Mehrabi

1. Statement of the Problem

The primary aim of this study is to determine the impact of fast charging compared to slow charging on electric vehicle (EV) battery degradation. Fast charging is perceived to be more convenient but potentially harmful to battery health due to increased stress from higher charging rates. This investigation seeks to understand whether this perception holds true and to what extent fast charging (DC) contributes to battery degradation in comparison to slow charging (AC).

2. Relevance and Inspiration

The rising popularity of electric vehicles establishes the important role of battery technology in transportation. However, the lifespan of electric car batteries varies significantly based on usage conditions, which could be challenging for consumers and manufacturers. Understanding battery life more not only enhances the economic value of electric vehicles, but also reduces environmental impact by decreasing the frequency of battery replacements. Other than social factors, this topic is interesting to our team, since we are very interested in automobile, sustainability applications. By analyzing the factors that could affect battery life, our study will contribute insights that could lead to more durable battery designs and usage strategies, promoting more reliable and environmentally friendly electric vehicles.

Our primary data source for this project is a comprehensive dataset described in the study by Zhongwei Deng. This dataset consists of detailed charging data from 20 electric vehicles, collected every 7 seconds over approximately 29 months. The data includes various parameters essential for our analysis such as battery current, voltage, state of charge, and temperature metrics. This dataset provides a good opportunity to examine the impacts of usage conditions on battery lifespan. The dataset is explored in the original paper in order check whether there is a valid prediction model of future battery capacity based on time series inputs of charging parameters. This differs from our report, in which we discover the effect of some user choices over the degradation in full battery capacity.

4. Exploratory Data Analysis

Step 1: Descriptions of the Raw Data

We firstly obtained 20 datasets from the GitHub page mentioned in our data source. These datasets were combined into a single dataframe for further analysis. We used dask and parquet file format for efficient handling of large datasets.

In [3]:
```
data = dd.read_csv('data.csv')
print(data.head())
```

```
   Unnamed: 0     record_time   soc  pack_voltage (V)  charge_current (A)  \
0           0  20190726200235  27.2             328.2           -52.20001
1           1  20190726200243  27.6             328.5           -52.20001
2           2  20190726200251  27.6             328.6           -52.20001
3           3  20190726200259  27.6             328.6           -52.20001
4           4  20190726200307  27.6             328.8           -52.20001

   max_cell_voltage (V)  min_cell_voltage (V)  max_temperature (℃)  \
0                 3.656                 3.640                   41
1                 3.663                 3.645                   41
2                 3.665                 3.647                   41
3                 3.666                 3.649                   41
4                 3.666                 3.649                   41

   min_temperature (℃)  available_energy (kw)  available_capacity (Ah)
0                   38                  12.40                     37.28
1                   38                  12.44                     37.39
2                   38                  12.47                     37.51
3                   38                  12.52                     37.64
4                   38                  12.56                     37.76
```

Using the normal pandas and loading all data from the 20 vehicles into RAM crashes the online notebook the group is sharing. To enable successful processing of data, we used dask on deepnotes to enable some data to be stored on the hard drive with Parquet format.

Step 2: Summary Statistics

To analyze the structure and key characteristics of the dataset, we printed out the data shape, data types, and summary statistics.

Data Shape: The dataset consists of about 16 million records and 11 columns.

Data Types: The data types of the columns are mostly floats and integers. The record_time is an integer and need to be converted to a datetime format for the following analysis steps.

Summary Statistics: soc (State of Charge): The mean state of charge is around 64%, with a standard deviation of 22%. This indicates a moderate level of variability in the state of charge across the dataset.

pack_voltage (V): The mean pack voltage is approximately 352V, with a standard deviation of 16.2V. The minimum and maximum values range from 312.8V to 386.2V, suggesting a big range in the voltage levels.

charge_current (A): The mean charge current is -52.2A, with a standard deviation of 17.3A. The negative mean value indicates that the dataset includes both charging and discharging events, as charging currents are positive.

max_cell_voltage (V) and min_cell_voltage (V): The mean maximum and minimum cell voltages are 3.93V and 3.90V, respectively, with standard deviations of approximately 0.18V.

max_temperature (°C) and min_temperature (°C): The mean maximum and minimum temperatures among all cells in a battery pack are 33.9°C and 32.0°C, respectively, with standard deviations of about 5.6°C. The temperatures range from 0°C to 49°C.

available_energy (kw) and available_capacity (Ah): The mean available energy is 27.7kW with a standard deviation of 9.6kW, and the mean available capacity is 83.3Ah with a standard deviation of 28.9Ah.

To ensure the completeness and reliability of our dataset, we checked for missing or null values, and the result shows that there are no missing values in any of the columns of the dataset.

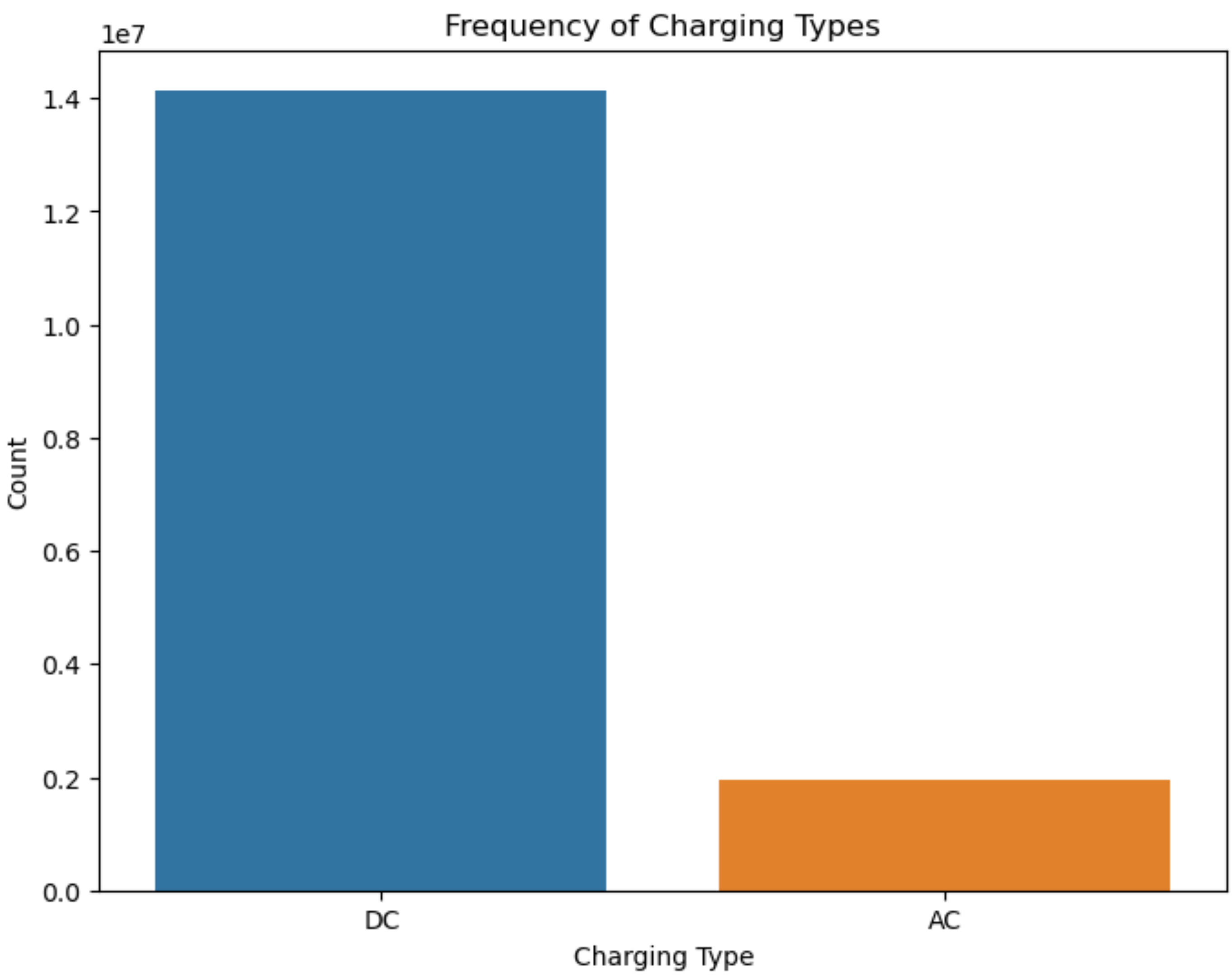
To ensure that all data types are appropriate, we converted the record_time column from an integer format to a datetime format. This conversion will allow us to do time-series analysis.

Step 3: Categorizing Fast and Slow Charging

In this step, we categorized charging sessions into fast (DC) and slow (AC) charging. This categorization helps us differentiate between the two types of charging methods and analyze their respective impacts on battery degradation.

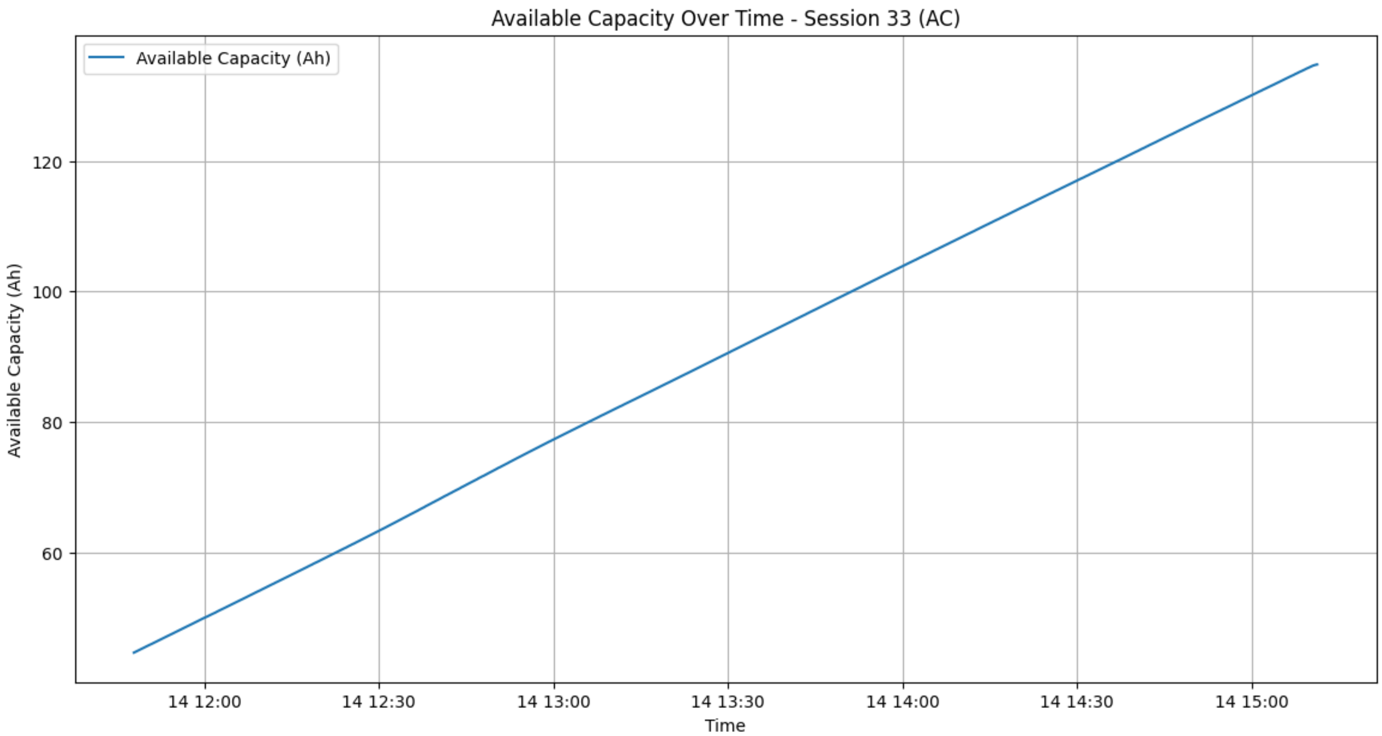
We calculated the time differences between consecutive records and used a threshold of 600 seconds of inactivity to identify new charging sessions. This helps in grouping continuous charging events together. Creating the new_session column identifies distinct charging sessions, facilitating session-wise analysis.

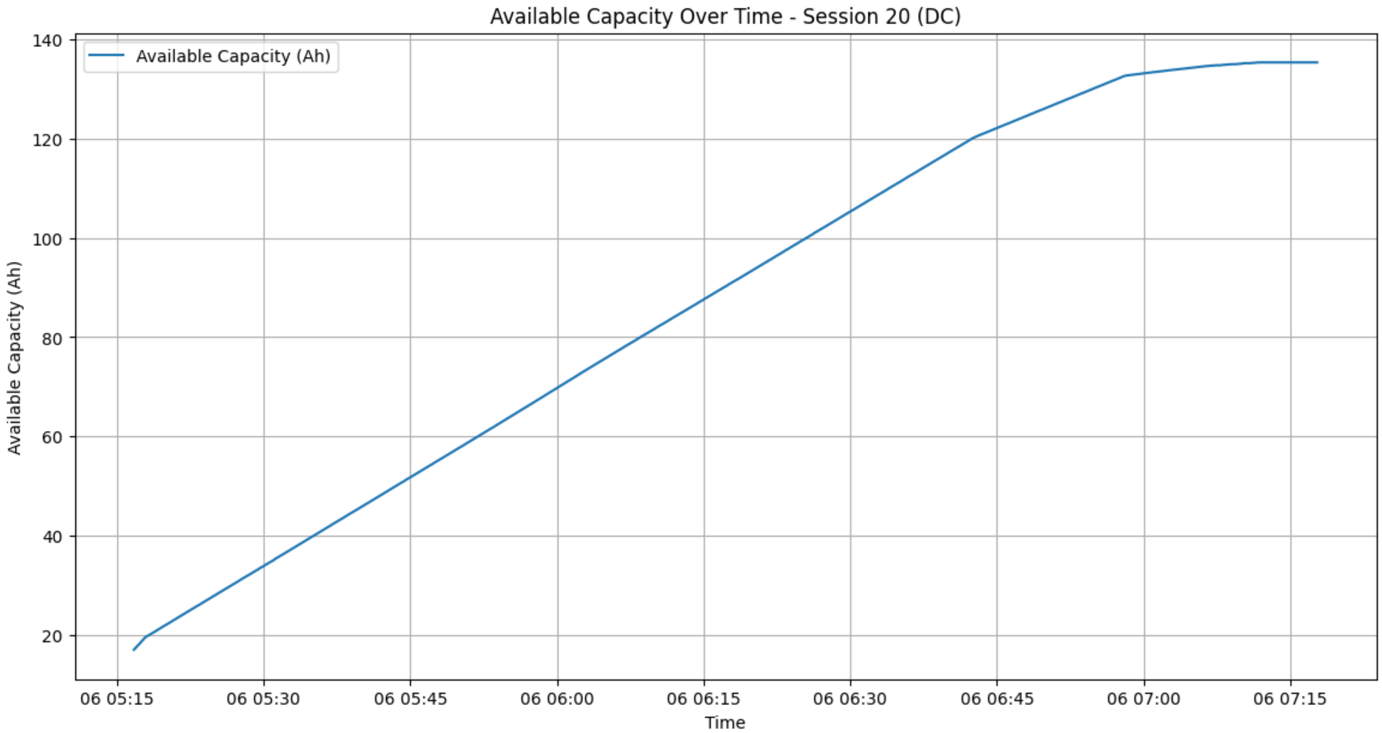
By applying a threshold on the minimum charge current (-40A), we classified sessions into DC (fast charging) and AC (slow charging). For AC charging, we expect the currents to have a a smaller magnitude. This classification helps us distinguish between the different charging methods. After doing so, we have DC and AC categorized. There are 14,142,983 DC charging sessions and 1,957,745 AC sessions. The mean charge current for DC sessions is -56.45A, indicating higher power levels, whereas for AC sessions it is -27.74A. The standard deviation for DC sessions (14.99A) is higher than for AC sessions (10.93A). The minimum charge current for DC sessions is -400.02A, much lower than the -40.00A for AC sessions.



Step 4: Validate that the Labeling is Correct

We validate the correctness of labeling by randomly sampling a few charge sessions and view how battery capacity change during that session. This helps us to catch any obvious mistakes in categorizing these two charging types.



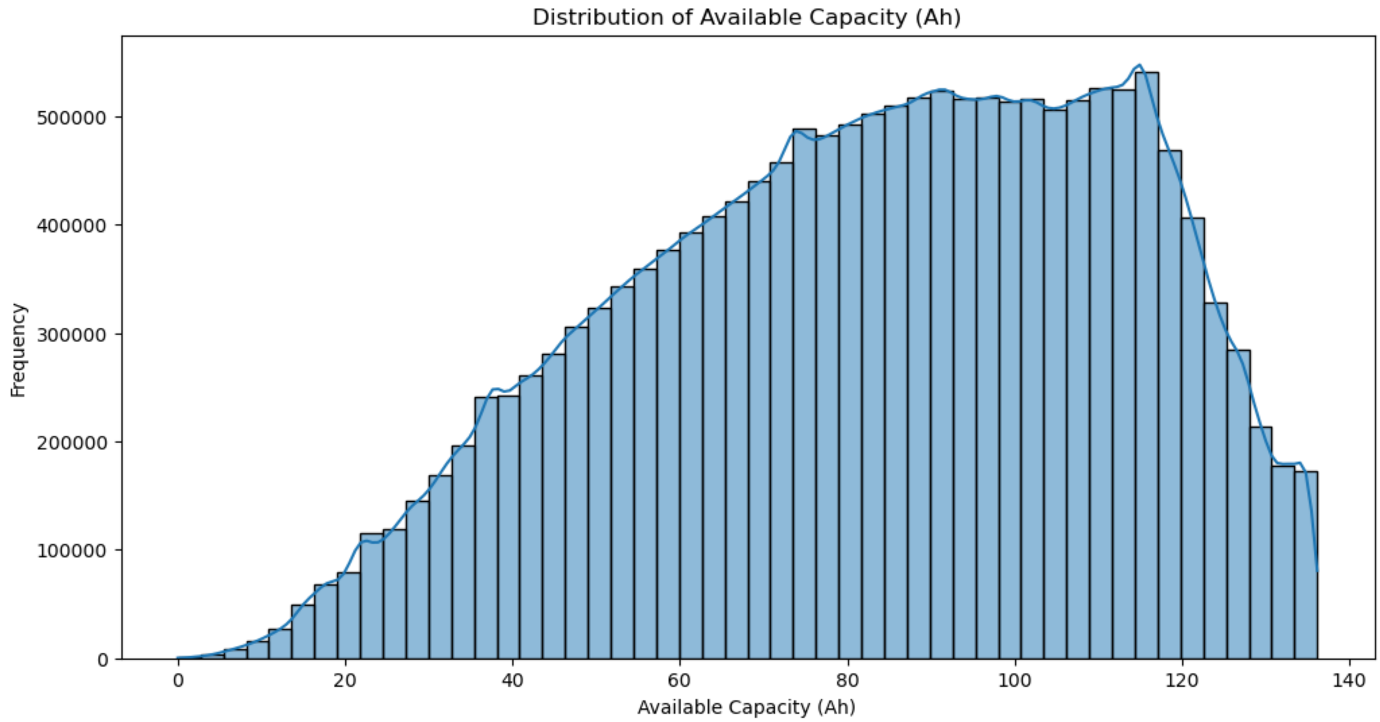


For DC charging, the rate of charge is expected to be slower nearing the end. For AC charge, the rate of charge should be stable across the session. Since all samples behave similar to what is shown in the graph. The labeling is correct.

Step 5: Understand the Distribution

In this step, we used multiple visualizations of the distributions to better understand the dataset.

1. Histogram for Battery Lifespan (available_capacity (Ah))

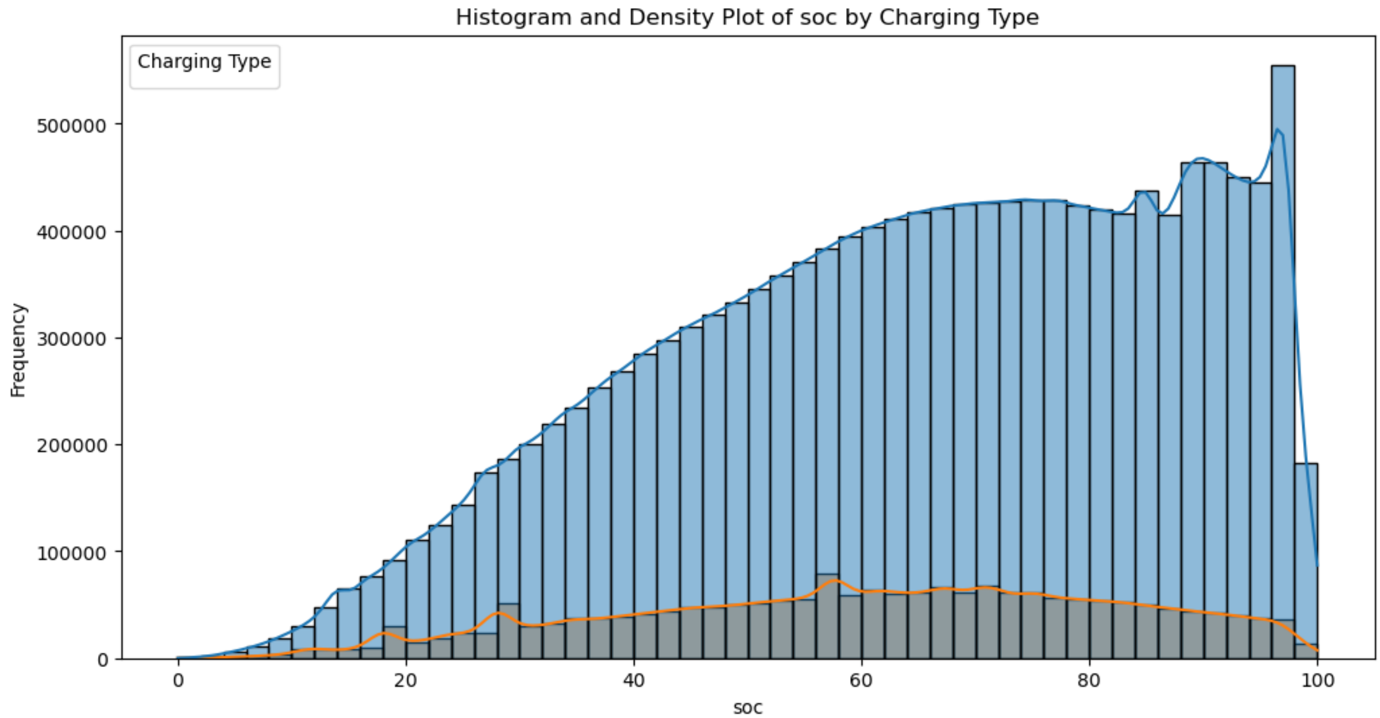


The histogram shows a gradual increase in frequency up to around 120 Ah, after which it sharply declines.

There is a noticeable peak around the 80-120 Ah range, indicating that battery is charged to this range of capacity most of the time.

Then, we visualized the distribution of key metrics (SOC, pack voltage, charge current, and available capacity) by charging type (DC vs. AC). These visualizations help us understand how these metrics differ between fast and slow charging sessions.

2. Histogram and Density Plot of SOC by Charging Type:

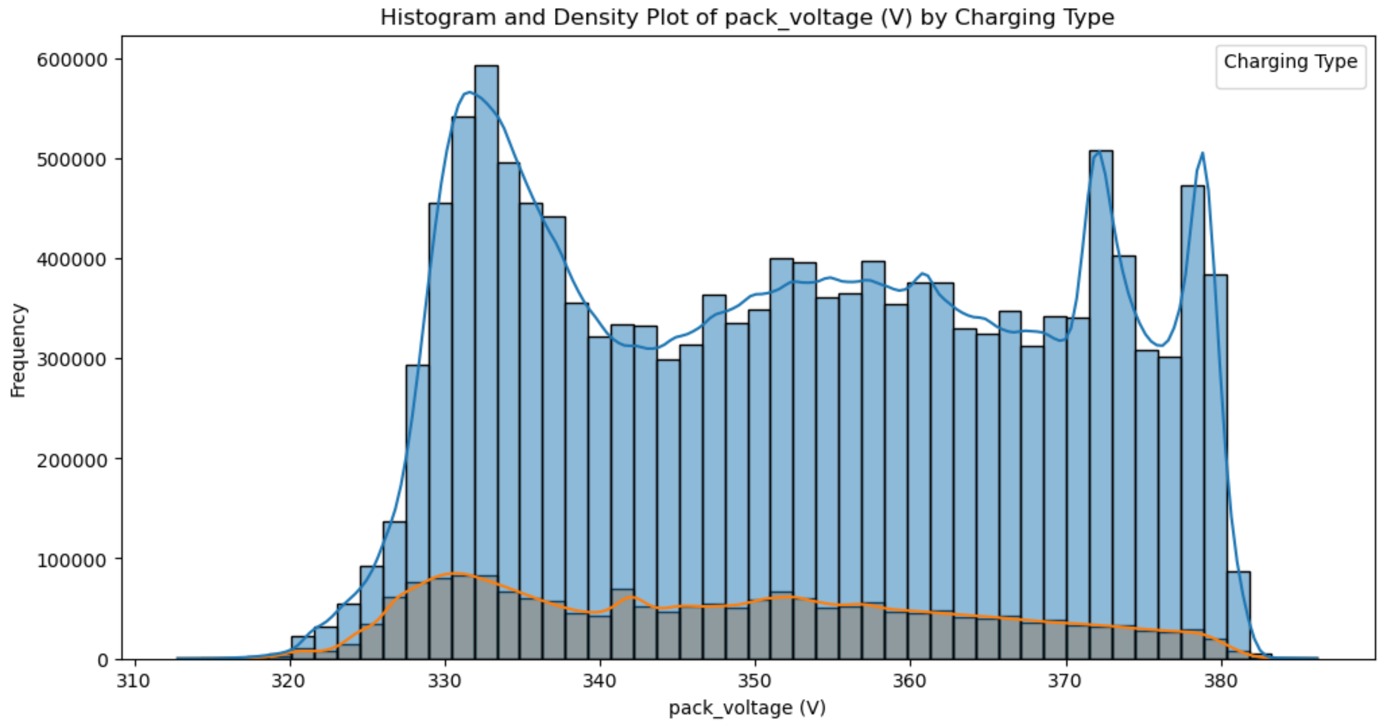


The SOC distribution is more spread out for DC charging compared to AC charging.

DC charging shows higher frequencies at the extremes (near 0% and 100% SOC), indicating more frequent deep discharges and full charges.

AC charging has a relatively flat distribution, suggesting more consistent SOC levels.

3. Histogram and Density Plot of Pack Voltage (V) by Charging Type:

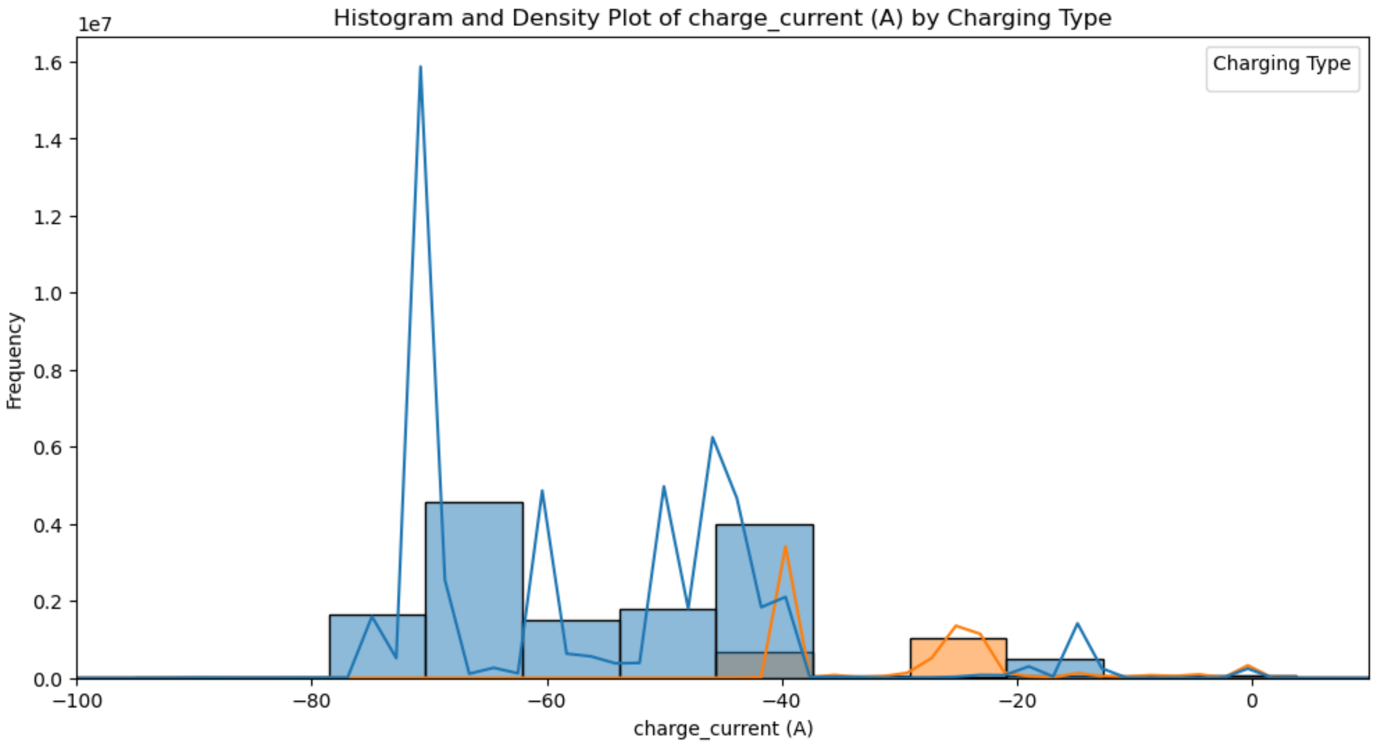


The pack voltage distribution is bimodal for DC charging, with peaks around 330V and 370V.

AC charging shows a more uniform distribution with a slight peak around 330V.

The higher voltages associated with DC charging suggest its use for faster charging cycles, which typically operate at higher voltages.

4. Histogram and Density Plot of Charge Current (A) by Charging Type:

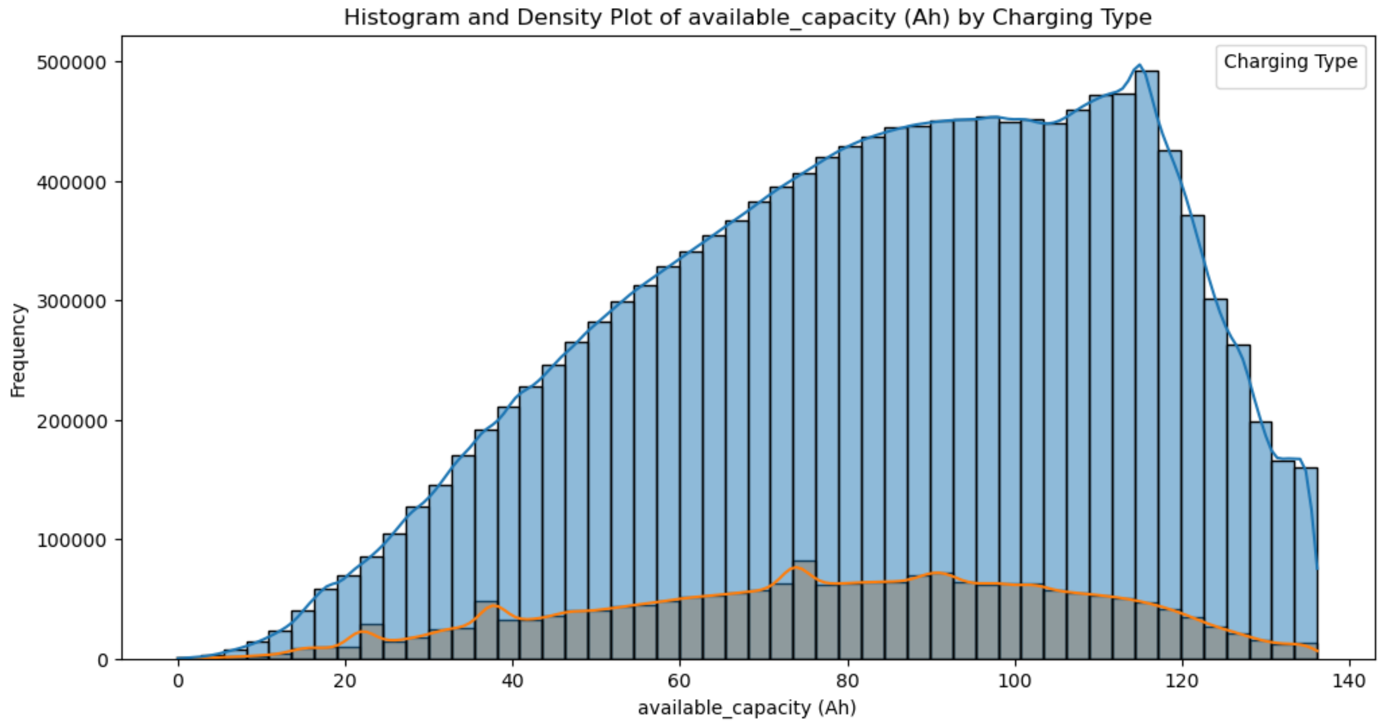


DC charging has a significantly wider range of charge currents, with a peak around -50A.

AC charging currents are generally lower, with a peak around -30A.

The higher and more variable charge currents for DC charging reflect its use for rapid energy transfer.

5. Histogram and Density Plot of Available Capacity (Ah) by Charging Type:



The available_capacity distribution shows that DC charging sessions tend to occur across a wider range of capacities.

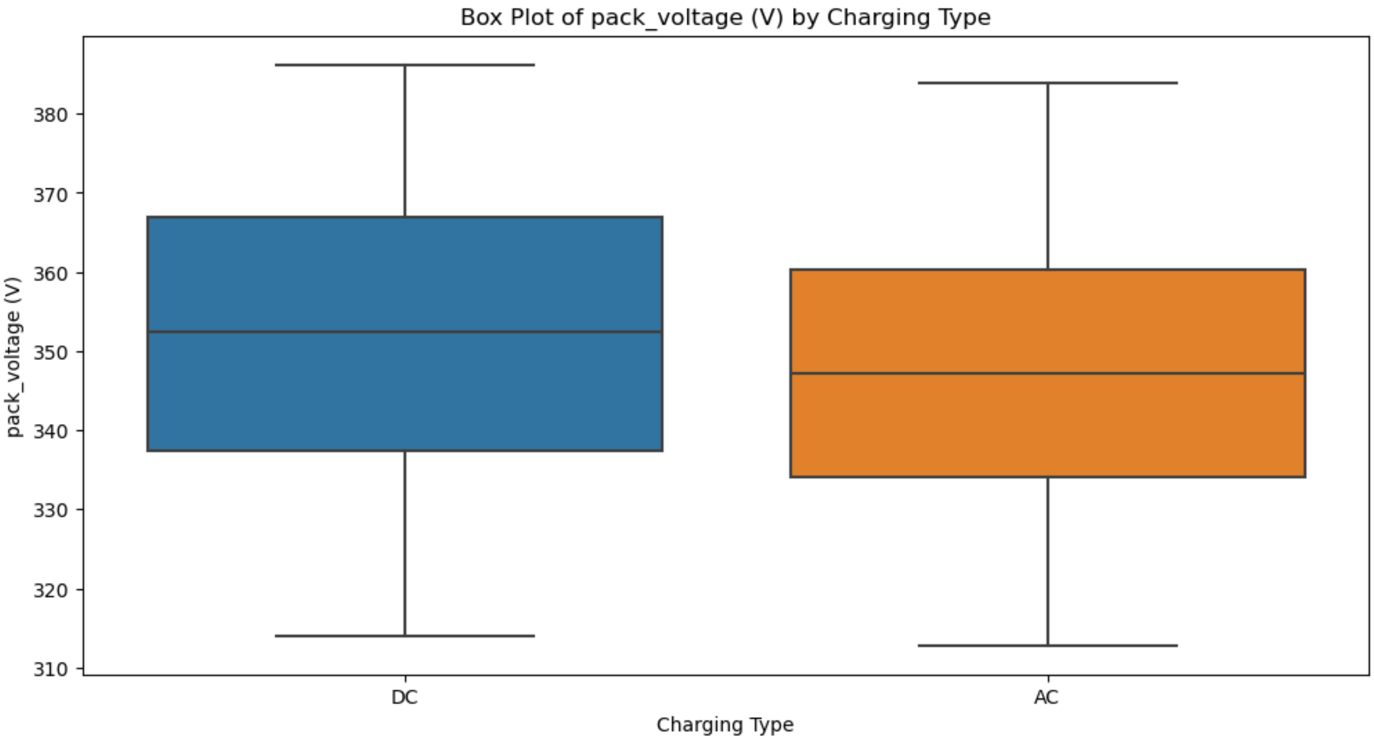
AC charging has a relatively flat distribution with a slight peak around 80-100 Ah.

The broader distribution for DC charging could indicate more varied usage patterns compared to AC charging.

The histograms and density plots by charging type reveal distinct differences in the distribution of key metrics between DC and AC charging sessions. DC charging sessions show higher variability and more extreme values in SOC, pack voltage, and charge current, reflecting their role in rapid energy transfer. AC charging sessions exhibit more consistent and moderate values, indicating steadier and potentially less aggressive charging behaviors.

After that, we used box plots to compare key metrics (SOC, pack voltage, charge current, and available capacity) between DC (fast charging) and AC (slow charging) sessions.

6. Box Plot of Pack Voltage (V) by Charging Type:

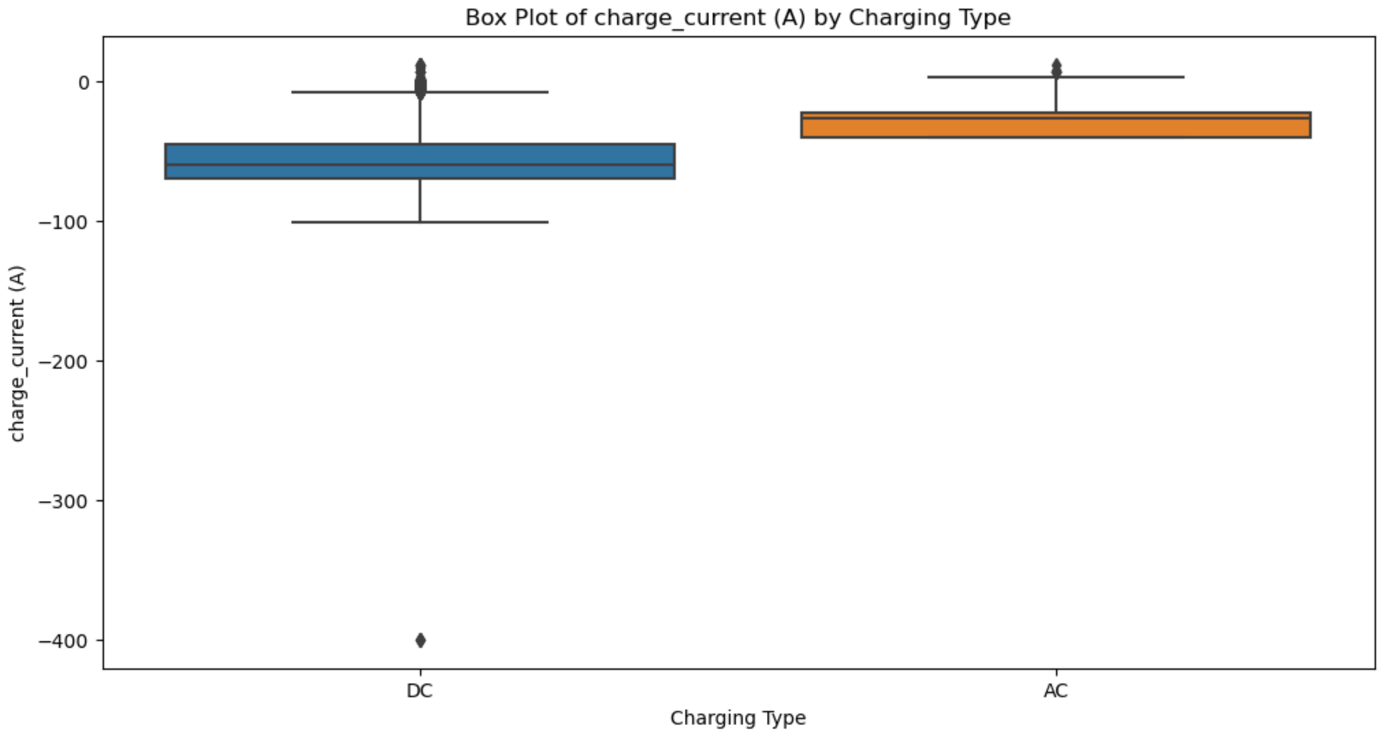


DC charging exhibits a wider range of pack voltages compared to AC charging.

The median pack voltage is higher for DC charging sessions.

The wider IQR for DC charging indicates more variability, consistent with its higher power levels.

7. Box Plot of Charge Current (A) by Charging Type:

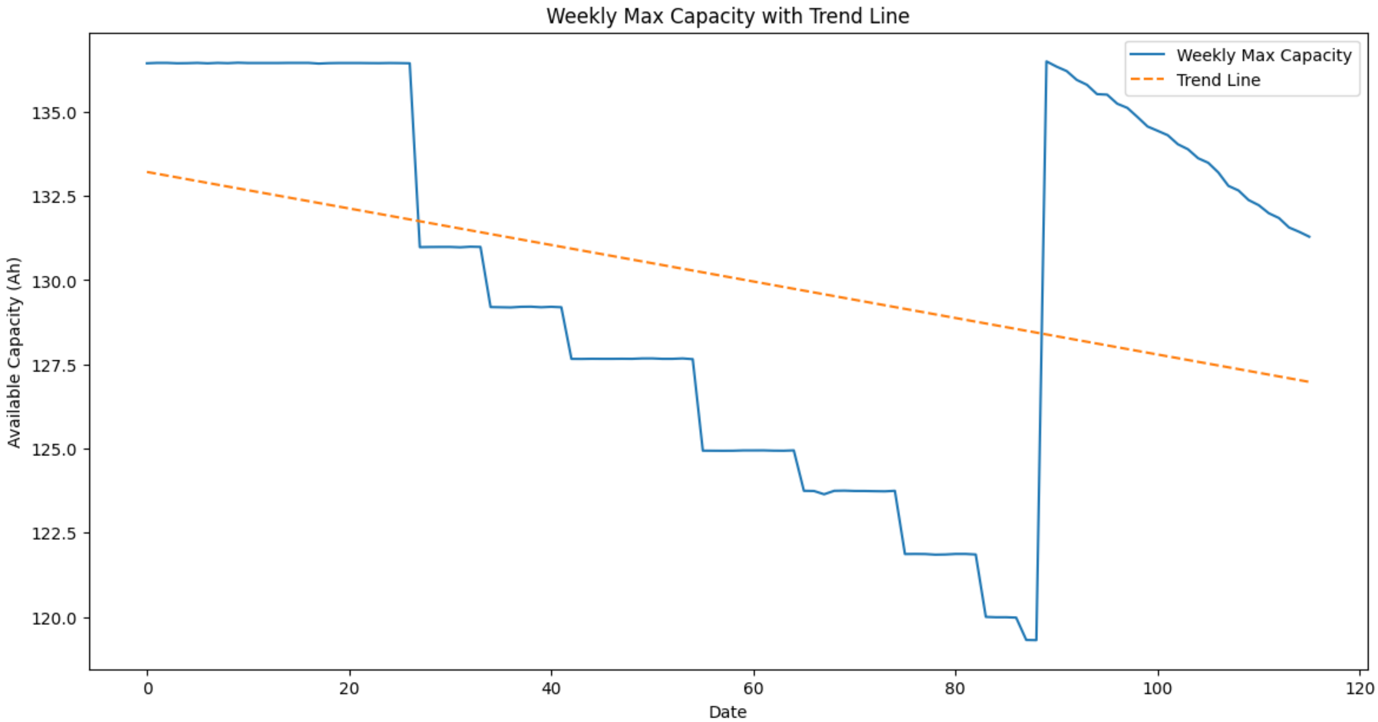


DC charging shows a significantly wider range of charge currents, including more extreme negative values, indicating higher power usage.

AC charging currents are more consistent and less negative, reflecting slower charging rates.

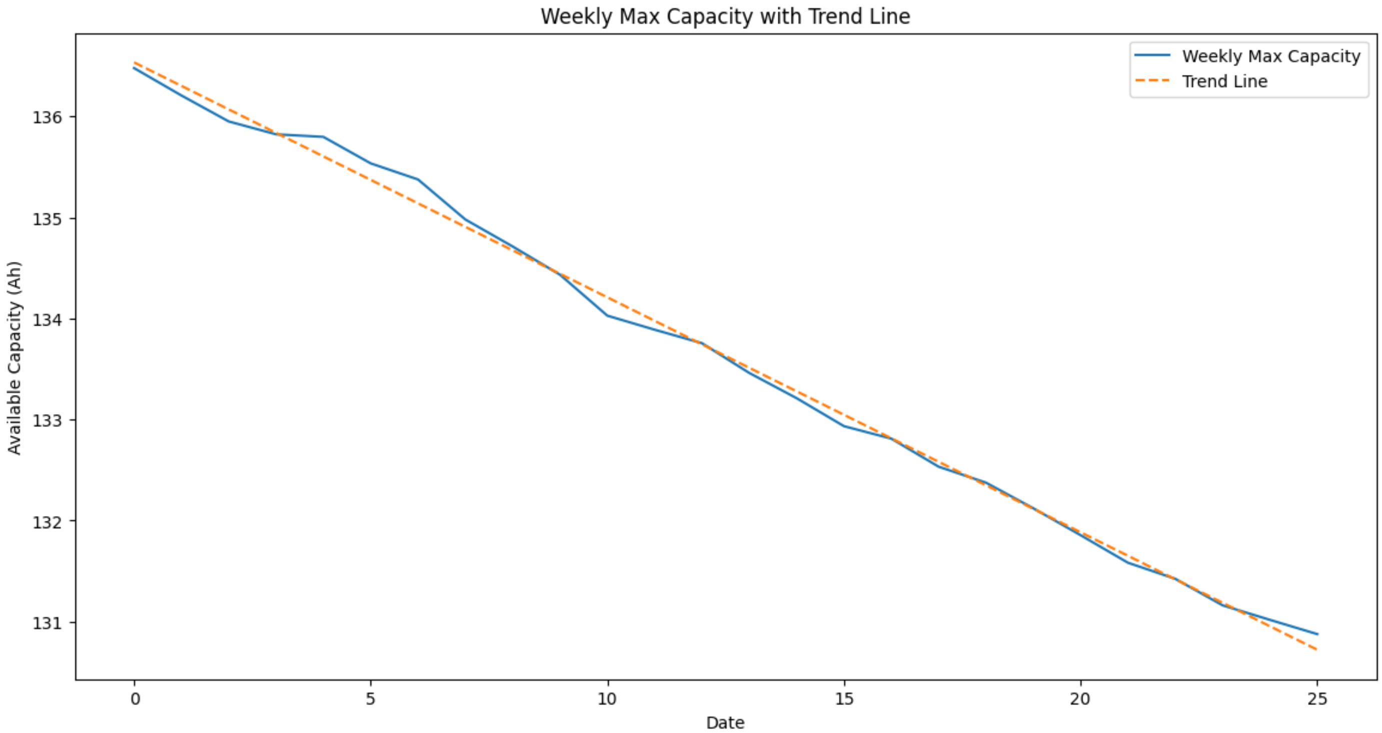
The presence of outliers in DC charging indicates occasional very high current draw, which is characteristic of fast charging sessions.

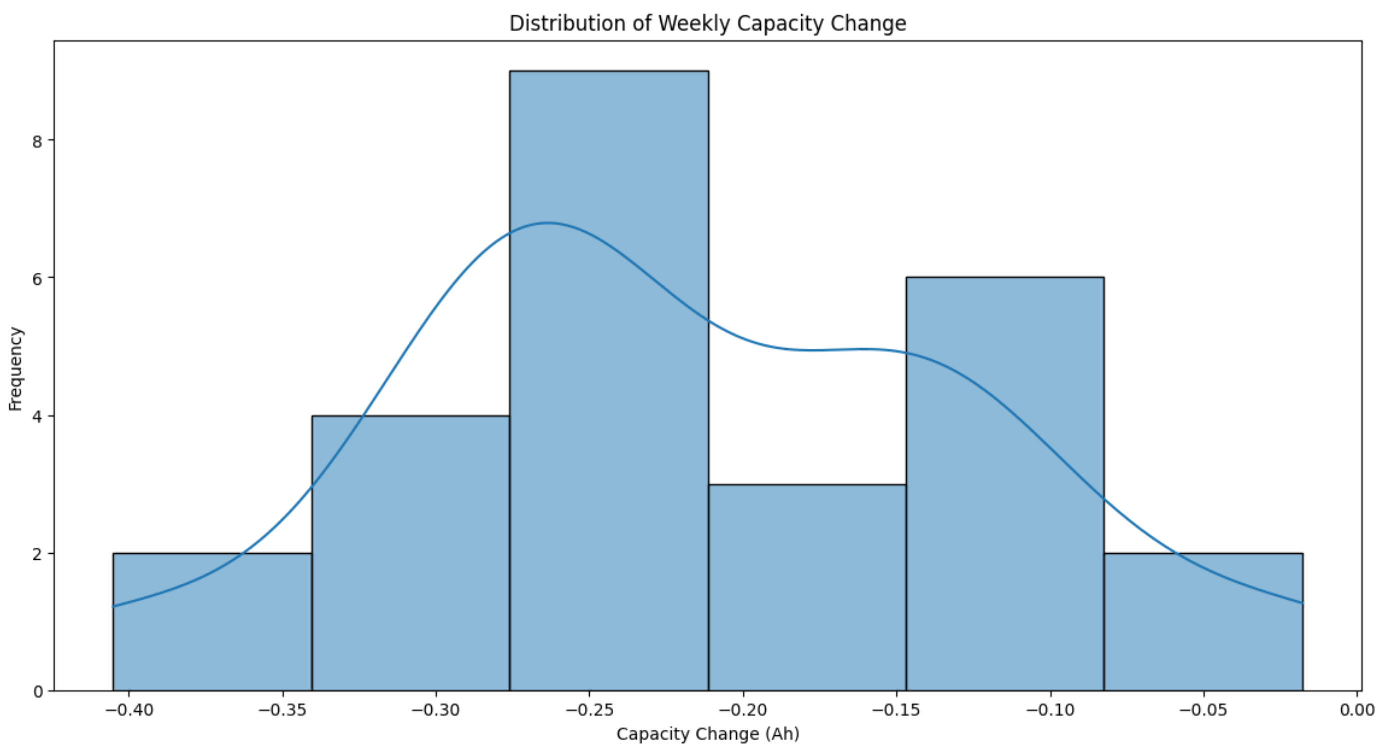
Step 6: Variation of Battery Capacity over Time



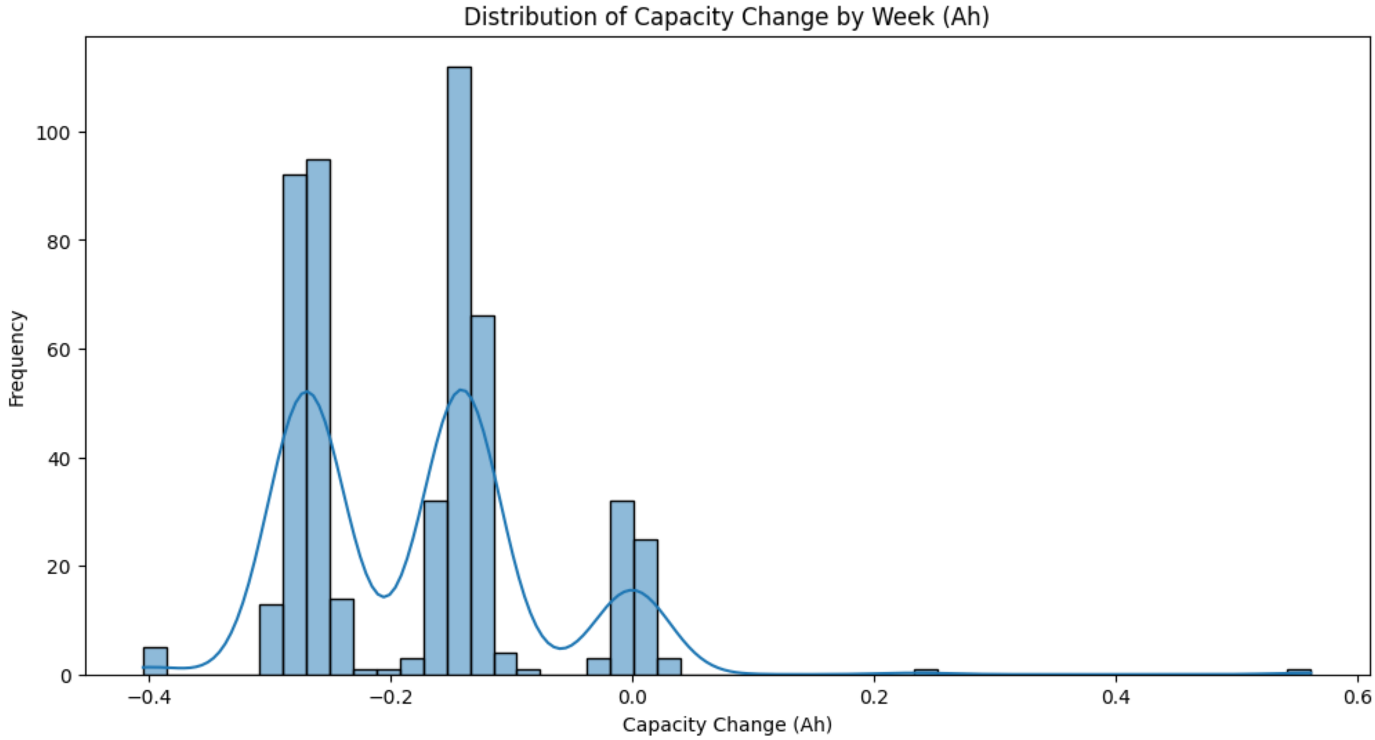
From the graph we can see that the battery was initially over 135Ah, but over time decreased to 120Ah in capacity. Later on, the battery seems to be replaced and regained the capacity back to 135Ah. Together with the increase is a more frequently updated capacity metric. We located the point of battery change and filtered the data to only have rows that are after the battery change. This way, we have data that are easier to work with.

After the revision, we can visualize that the capacity change of the sample vehicle is in bell curve, and the weekly max capacity metric decreases steadily over time.



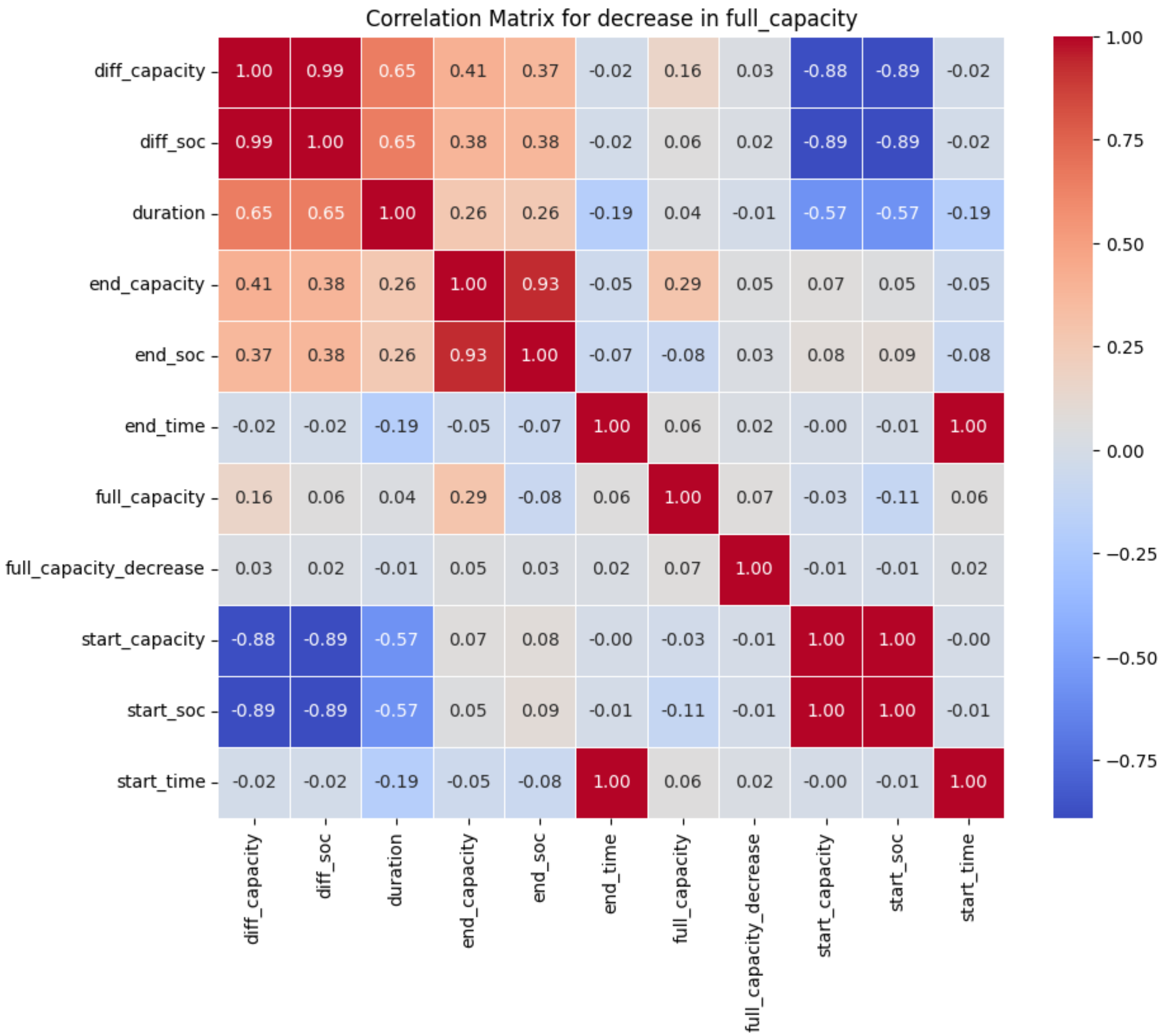


By combining data from all vehicles, we can see that the capacity change by week is roughly in 3 different normal distributions.



Step 7: Correlation Matrix

We used a correlation matrix to oberserve the linear relationships between numerical features in the dataset and how they influence the decrease in full capacity.



The correlation between full_capacity_decrease and duration is very low (0.02), indicating that the duration of charging sessions does not have a significant direct linear relationship with the decrease in full capacity.

State of Charge and Capacity Changes:

The correlations between full_capacity_decrease and diff_capacity (0.03) and diff_soc (0.03) are also very low, suggesting that changes in capacity and SOC during individual charging sessions are not significant factors in determining the decrease in full capacity.

Starting and Ending Capacities and SOCs:

The correlations between full_capacity_decrease and start_capacity (0.07) and end_capacity (0.05) and the correlations between full_capacity_decrease and start_soc (0.02) and end_soc (0.03) are very low, suggesting that the starting and ending capacity and SOCs are not significant predictors of the decrease.

The correlation between full_capacity_decrease and full_capacity is also very low (0.07), indicating that the initial full capacity of the battery does not have a strong linear relationship with the decrease in full capacity.

The correlation between full_capacity_decrease and most other variables is very low. This suggests that the decrease in full capacity is not directly influenced by these individual factors in a significant linear manner. This may be due to a delay in the response variable. Changes to the battery full capacity is not reflected on the data until later.

Step 8: Grouping the data points into charging sessions

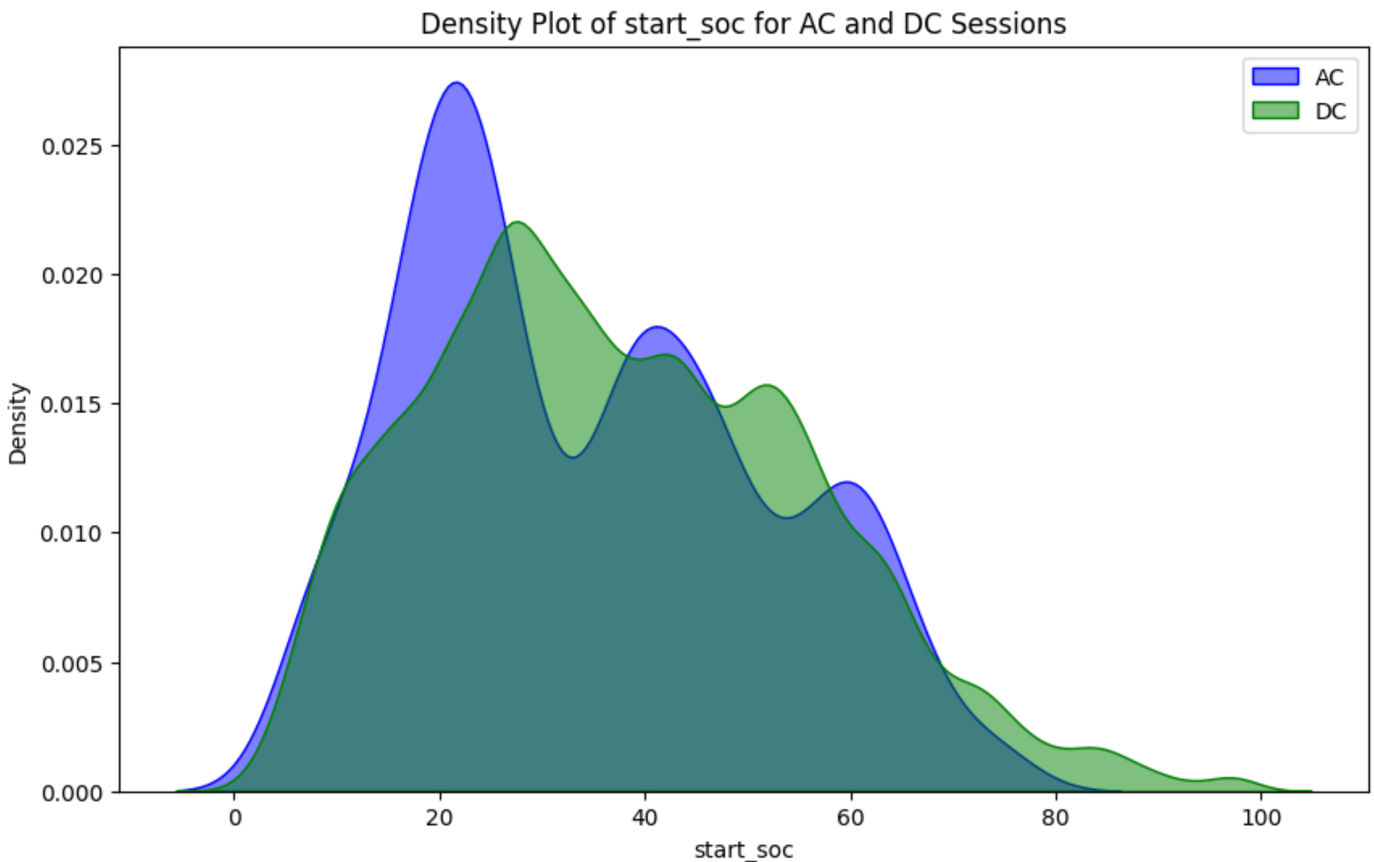
The previous data represents the data points generated from charging every couple of seconds. By grouping the data into charging sessions, we have a row for each time the battery is charged and can compare the results more easily.

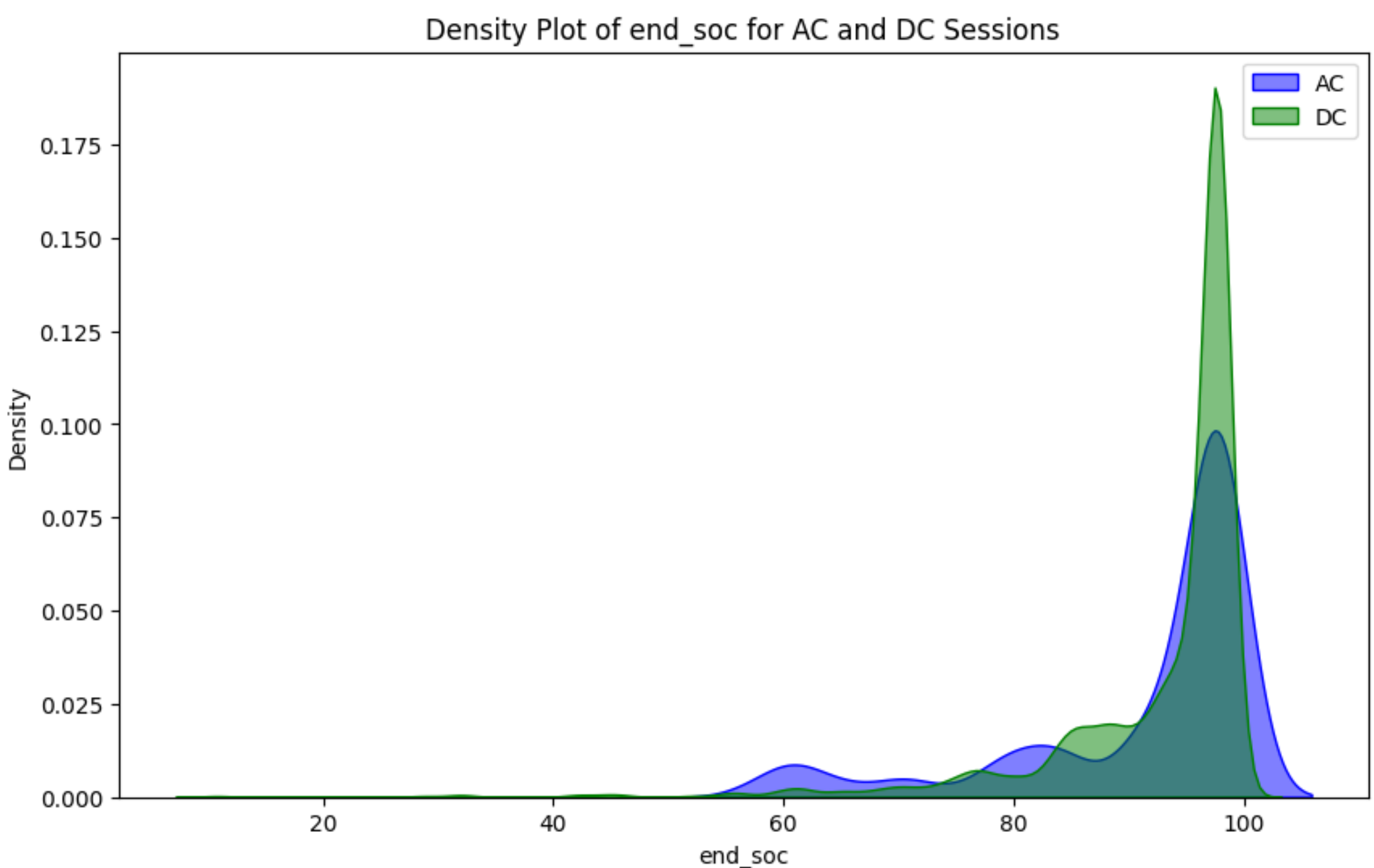
The previous data represents the data points generated from charging every couple of seconds. By grouping the data into charging sessions, we have a row for each time the battery is charged and can compare the results more easily.

""" start_soc end_soc start_capacity end_capacity \ new_session 100000276 6.400000 10.800000 8.770000 14.890000 100000277 11.200000 77.599998 15.120000 105.660004 100000278 15.600000 68.000000 21.040001 92.660004 100000279 68.000000 76.800003 92.709999 104.470001 100000280 76.800003 76.400002 104.290001 104.260002 start_time end_time source_file charging_type \ new_session 100000276 2019-12-13 15:17:57 2019-12-13 15:28:15 1.parquet DC 100000277 2019-12-13 15:54:01 2019-12-13 18:29:59 1.parquet DC 100000278 2019-12-14 01:30:33 2019-12-14 03:13:50 1.parquet DC 100000279 2019-12-14 04:54:20 2019-12-14 05:04:28 1.parquet DC 100000280 2019-12-14 07:05:51 2019-12-14 07:08:31 1.parquet DC diff_soc diff_capacity duration full_capacity new_session 100000276 4.400000 6.120000 0 days 00:10:18 137.870377 100000277 66.400002 90.540001 0 days 02:35:58 136.159805 100000278 52.400002 71.620003 0 days 01:43:17 136.264709 100000279 8.800003 11.760002 0 days 00:10:08 136.028641 100000280 -0.400002 -0.029999 0 days 00:02:40 136.465973 """

'start_soc', 'end_soc', 'start_capacity', 'end_capacity', 'start_time', 'end_time', 'source_file', 'charging_type', 'diff_soc', 'diff_capacity', 'duration', 'full_capacity', 'charging_type_DC', 'charging_type_AC', 'rolling_diff_soc', 'rolling_charging_type_DC', 'rolling_charging_type_AC', 'rolling_diff_soc_DC', 'rolling_diff_soc_AC', 'rolling_degraded_capacity'

Distribution of State of Charge (SOC) Before and After Charging

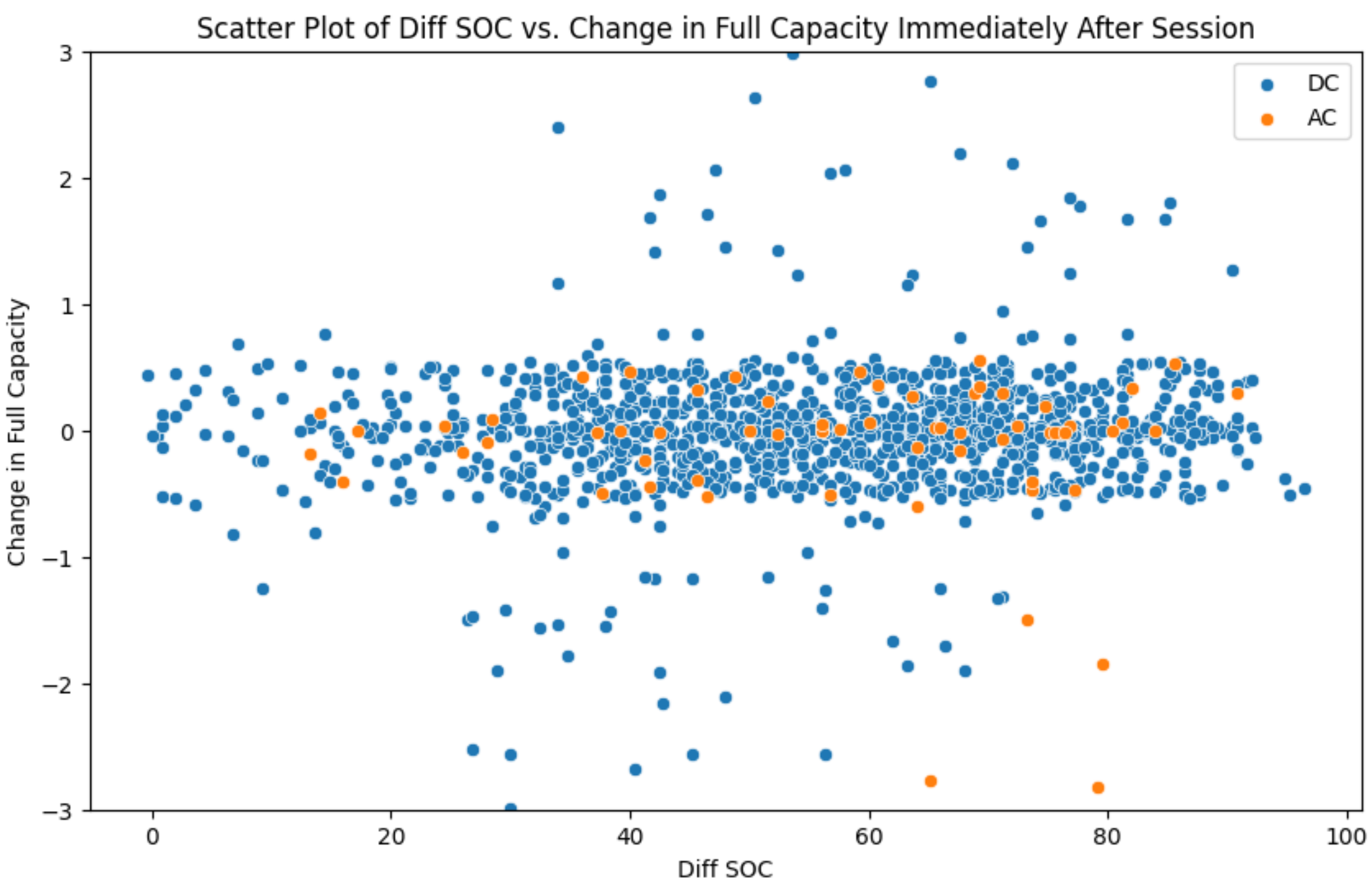

Density Plot of start_soc for AC and DC Sessions
AC
DC
Density
0.000
0.005
0.010
0.015
0.020
0.025
0
20
40
60
80
100
start_soc


Density Plot of end_soc for AC and DC Sessions
AC
DC
Density
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
20
40
60
80
100
end_soc

There is no significant difference in the distribution.

The Challenge: Delayed Effect on the Response Variable

One of the primary challenges in analyzing the effect of charging sessions on battery degradation is the delayed impact on the response variable (degraded_capacity). The degradation in battery capacity due to charging sessions may not be immediately reflected in the battery management system (BMS) updates. Instead, these effects might only become apparent in subsequent sessions, leading to a temporal misalignment between the independent variables (such as diff_soc and charging_type) and the dependent variable (degraded_capacity). This misalignment complicates the analysis and can obscure the true relationships between the variables.



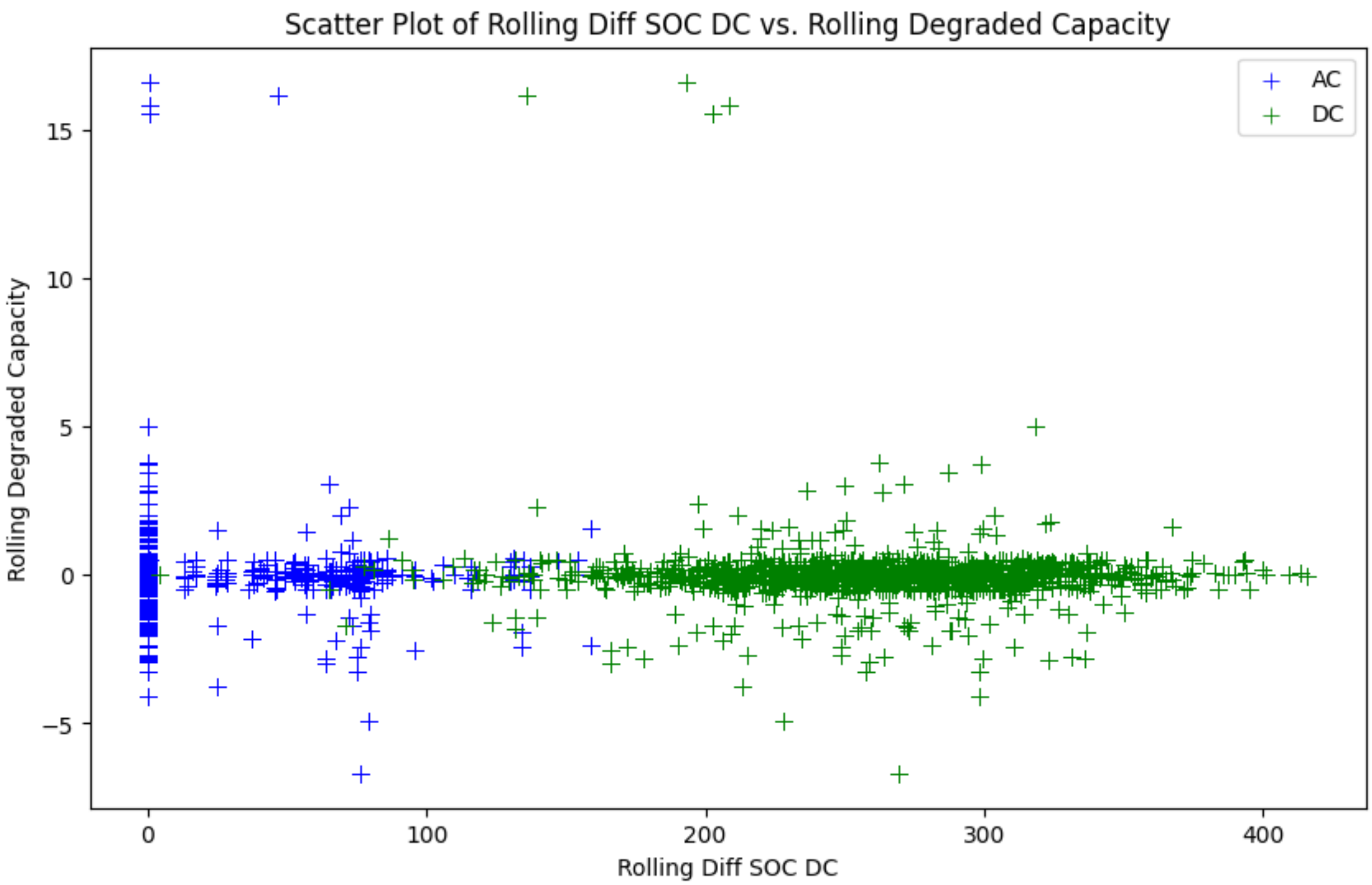
The graph shows no difference in effect. We presume that this is due to the delay of effect on the response variable.

Initial Approach: Grouping by Week

Initially, the data was grouped by weekly intervals. This approach aimed to capture the cumulative effect of all charging sessions within each week. By summing diff_soc and determining the predominant charging_type for each week, we hoped to correlate these weekly aggregates with the observed battery degradation. However, this method did not yield significant results. The primary reason for the lack of significant outcomes could be attributed to the rigid nature of the weekly grouping, which may not align well with the actual timing of BMS updates. The effects of charging sessions could span across weekly boundaries, leading to an inaccurate representation of their impact.

New Approach: Rolling Window

To better account for the delayed effects and provide a more flexible analysis, we adopted a rolling window approach. This method involves applying a rolling window to aggregate features over a specified number of sessions, thereby smoothing out the data and capturing the cumulative effects more accurately. By using a rolling window, we can account for the temporal lag in the response of degraded_capacity to the charging sessions.



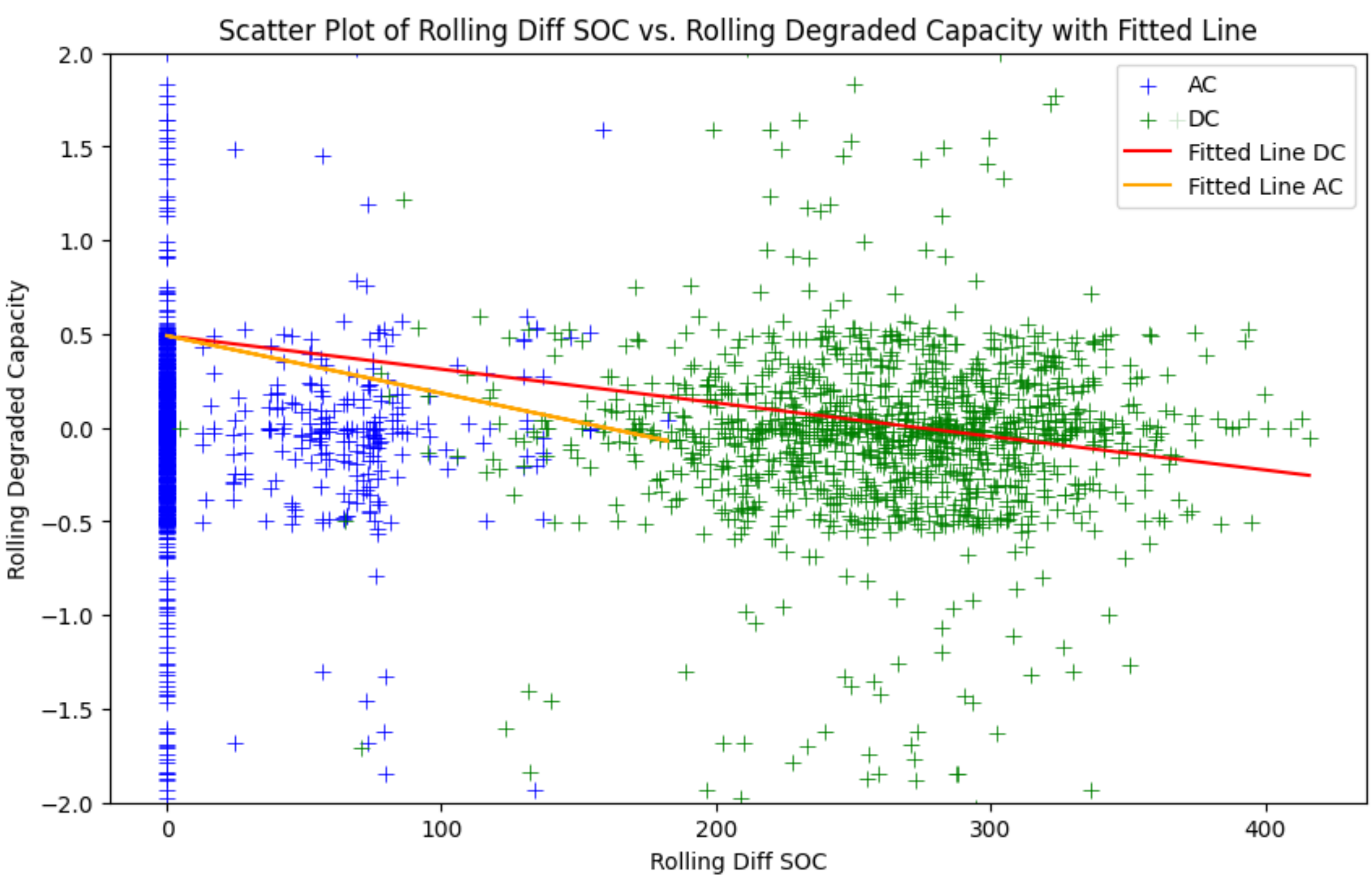
Some difference in effect can be observed.

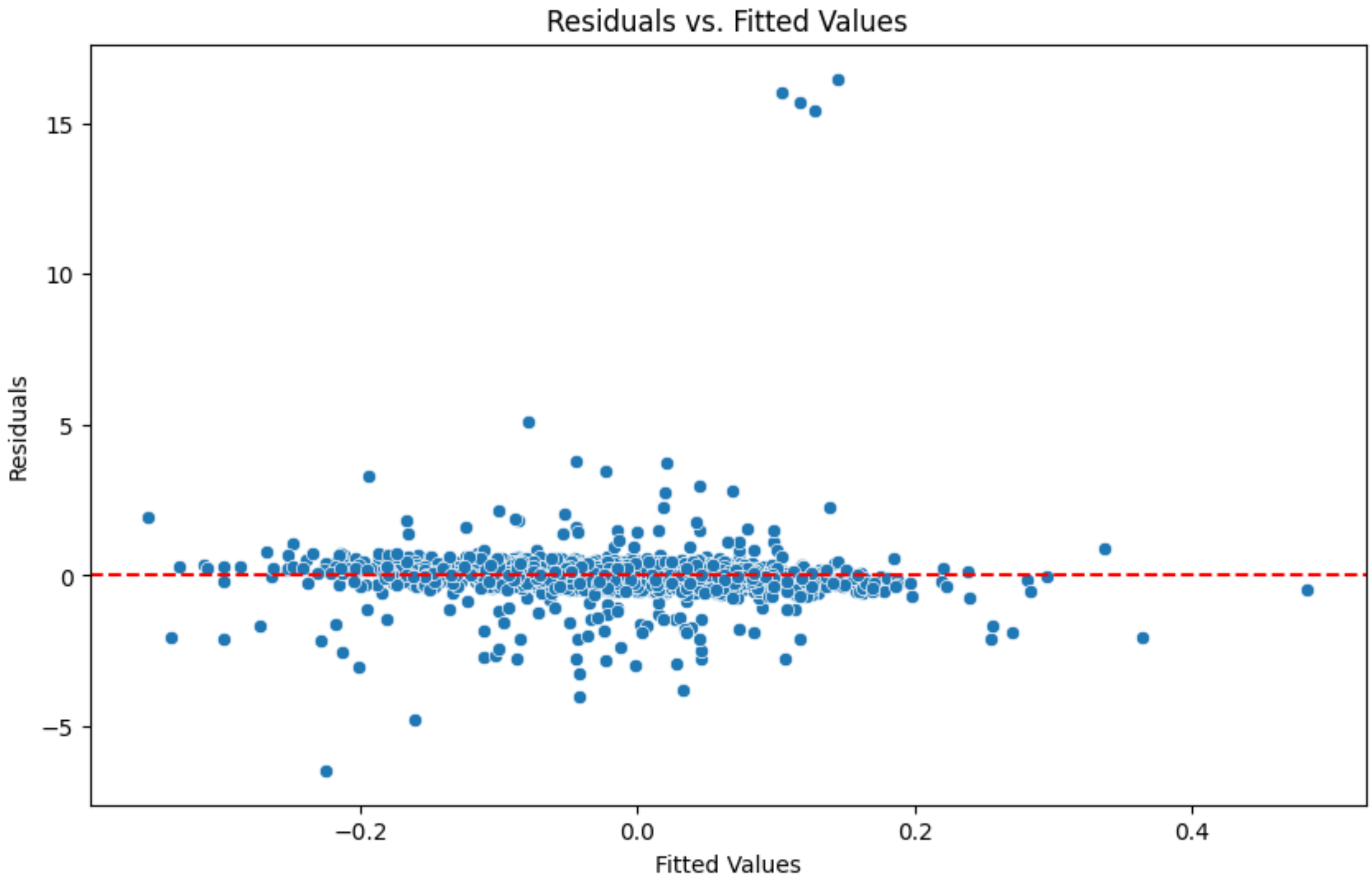
Linear Regression Results: 2 Variables

First, we fit a model to see if the amount charged by DC and AC each contributes to any degradation of the battery.

""" OLS Regression Results == Dep. Variable: rolling_degraded_capacity R-squared: 0.008 Model: OLS Adj. R-squared: 0.006 Method: Least Squares F-statistic: 5.214 Date: Mon, 10 Jun 2024 Prob (F-statistic): 0.00555 Time: 04:12:10 Log-Likelihood: -2041.3 No. Observations: 1344 AIC: 4089. Df Residuals: 1341 BIC: 4104. Df Model: 2 Covariance Type: nonrobust

==
================================== coef std err t P>|t| [0.025 0.975] -
----
---------- Intercept 0.4914 0.176 2.786 0.005 0.145 0.837
rolling_diff_soc_DC -0.0018 0.001 -2.839 0.005 -0.003 -0.001
rolling_diff_soc_AC -0.0031 0.001 -2.749 0.006 -0.005 -0.001
==
====================== Omnibus: 1999.734 Durbin-Watson: 0.858
Prob(Omnibus): 0.000 Jarque-Bera (JB): 935609.244 Skew: 8.593
Prob(JB): 0.00 Kurtosis: 131.109 Cond. No. 1.57e+03
==
====================== Notes: [1] Standard Errors assume that the covariance matrix of the errors is correctly specified. [2] The condition number is large, 1.57e+03. This might indicate that there are strong multicollinearity or other numerical problems. """





The fitted lines show a difference in effect between AC and DC. Note that for the scattered plot, the y axis value is the combined result of all AC and DC charging sessions during the rolling window. Thus, the fitted line is not expected to match the scatter plot.

With p-value slightly under 0.05, we conclude that the number of times being charged by Fast Charging (DC) is correlated with the battery degradation.

With p-value above 0.05, we fail to conclude that the number of times being charged by Slow Charging (AC) is correlated with the battery degradation.

Linear Regression Results: Using Interaction Term

degraded_capacity=$\beta 0+\beta 1 \cdot$ diff_soc$+\beta 2 \cdot$ charging_type_DC$+\beta 3$

$\cdot$ (diff_soc×charging_type_DC)+ϵ

Since we are applying the rolling window, we have

rolling_degraded_capacity=β0+β1 $\cdot$ rolling_diff_soc_DC+β2
$\cdot$ rolling_charging_type_DC+β3
$\cdot$ (rolling_diff_soc_DC×rolling_charging_type_DC)+β4
$\cdot$ rolling_diff_soc_AC+β5 $\cdot$ rolling_charging_type_AC+β6
$\cdot$ (rolling_diff_soc_AC×rolling_charging_type_AC)+ϵ

By fitting a linear model with interactive terms, we can compare the effect of different charging methods on the battery capacity degradation.

""" OLS Regression Results == Dep. Variable: rolling_degraded_capacity R-squared: 0.078 Model: OLS Adj. R-squared: 0.078 Method: Least Squares F-statistic: 378.5 Date: Sun, 09 Jun 2024 Prob (F-statistic): 0.00 Time: 20:45:53 Log-Likelihood: -39839. No. Observations: 26880 AIC: 7.969e+04 Df Residuals: 26873 BIC: 7.975e+04 Df Model: 6 Covariance Type: nonrobust == coef std err t P>|t| [0.025 0.975] -- Intercept -1.0823 0.228 -4.748 0.000 -1.529 -0.636 rolling_diff_soc_DC -0.0143 0.001 -10.359 0.000 -0.017 -0.012 rolling_charging_type_DC 0.2282 0.048 4.796 0.000 0.135 0.321 rolling_diff_soc_DC:rolling_charging_type_DC 0.0028 0.000 9.802 0.000 0.002 0.003 rolling_diff_soc_AC -0.0220 0.001 -26.021 0.000 -0.024 -0.020 rolling_charging_type_AC 2.3901 0.070 33.953 0.000 2.252 2.528

rolling_diff_soc_AC:rolling_charging_type_AC -0.0039 0.000 -8.475 0.000 -0.005 -0.003
==
==================== Omnibus: 34406.897 Durbin-Watson: 0.917 Prob(Omnibus): 0.000 Jarque-Bera (JB): 10861457.911 Skew: 6.905 Prob(JB): 0.00 Kurtosis: 100.504 Cond. No. 4.81e+04
==
==================== Notes: [1] Standard Errors assume that the covariance matrix of the errors is correctly specified. [2] The condition number is large, 4.81e+04. This might indicate that there are strong multicollinearity or other numerical problems. Fitted Linear Regression Formula: rolling_degraded_capacity = -1.0823 + (-0.0143 * rolling_diff_soc_DC) + (0.2282 * rolling_charg """

This shows that to the contrary of what assumed, fast charging (DC) seems to contribute less to the battery degradation than slow charging (AC).

R-squared: 0.078 this indicates that 7.8% of the variance in rolling_degraded_capacity is explained by the model.

F-statistic: 378.5 with a Prob (F-statistic) of 0.00 implies that the model is statistically significant overall.

Intercept: -1.0823 the battery capacity degrades by 1Ah each week in average.

rolling_diff_soc_DC: -0.0143A < 0 indicates that an increase in the amount charged with fast charging is associated with a higher magnitude of degradation in battery capacity.

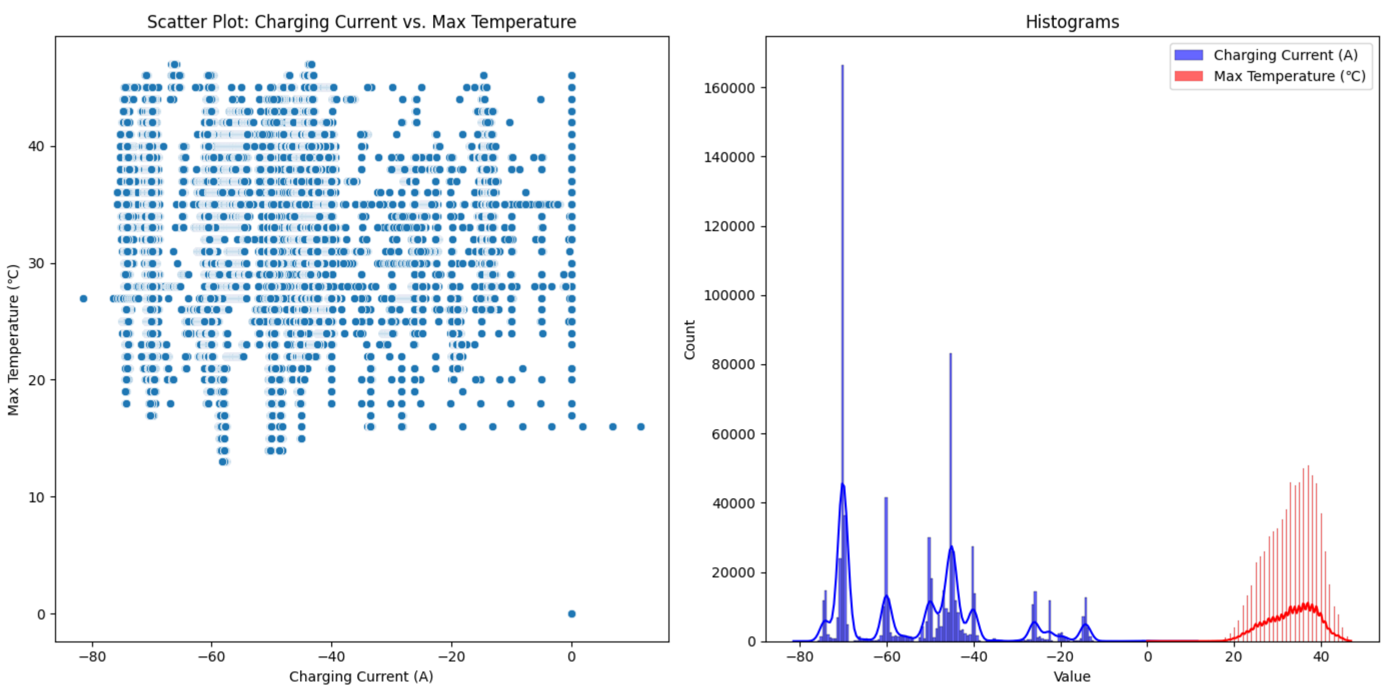
rolling_diff_soc_AC: -0.0220 < 0 indicates that an increase in the amount charged with slow charging is also associated with a higher magnitude of degradation in battery capacity.

rolling_diff_soc_DC:rolling_charging_type_DC: 0.0028 > 0 the interaction term shows that the degradation in battery capacity is less severe if Fast Charging (DC) is used in place of Slow Charging (AC). The 95% confidence interval of [0.002, 0.003] shows that this difference in effect is statistically significant.

rolling_diff_soc_AC:rolling_charging_type_AC: -0.0039 < 0 shows to the contrary of the above. The degradation is more severe when Slow Charging (AC) is chosen.

Correlation Results between Charge Current and Max Temperature

We want to make some attempts to explain the reasons behind the disparity in effect on the battery degradation. First thing that comes up to mind is the temperature. We want to see how the charging current (which is higher for Fast Charging) correlates to Max Temperature.



A descriptive analysis was done focusing on the relationship between charge current and maximum temperature. We wanted to show a potential correlation between them, hypothesizing that higher charging currents would lead to increased temperatures. Based on the result, the correlation coefficient between them was -0.04, indicating a very weak negative relationship. But, the correlation is statistically significant, shown by the extremely low p-value (2.22e-258), showing that higher charging currents are associated with lower temperatures, contrary to our hypothesis.

In working with the statistical inference, we realized that Fast Charging (DC) and Slow Charging (AC) differ in their contribution to the overall decrease in battery capacity.

The field data poses challenges in what conclusions we can draw. Since the full battery capacity is updated with a delay, we have to use strategies such as rolling window to account for the delay. However, the result shows with strong confidence that there is a disparity in the effect between Fast Charging and Slow Charging.

7. Potential limitations and shortcoming

1. Although users are typically not in control of when they have to use Fast Charging, their choices are far from resembling a natural experiment, the correlation between AC and DC to the battery degradation may not have causal link. The same factor that pushes the user to use Fast Charging (ie. cold ambient temperature) may help to mitigate the degradation on the battery, and thus invalidate our conclusions.

2. Rolling window as a solution to delayed response has limitations

3. The interaction term can contribute to multicollinearity or under-inflate the p-value

4. The effect on the response variable may not be linear.

5. The response variable, full battery capacity is calculated with the help of vehicle on-board computer, which calibrates the value both based on the actual performance of the battery (how much capacity is available in actual use) and on predictions of the capacity based on the charging data. This introduces inaccuracies in the findings and under-inflates the p-value.

Deng, Z., Xu, L., Liu, H., Hu, X., Duan, Z., & Xu, Y. (2023). Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles. Applied Energy, 339, Article 120954. (https://doi.org/10.1016/j.apenergy.2023.120954)

Deng, Z. (2023). Battery charging data of on-road electric vehicles [Dataset]. GitHub. Accessed May 3, 2024. (https://github.com/TengMichael/battery-charging-data-of-on-road-electric-vehicles)

Gao, Y., Jiang, J., Zhang, C., Zhang, W., Ma, Z., & Jiang, Y. (2017). Lithium-ion battery aging mechanisms and life model under different charging stresses. Journal of Power Sources, 356, 103-114. (https://doi.org/10.1016/j.jpowsour.2017.04.084)

Generative AI is used to assist with coding and editing texts.