# Predicting power consumption for electric cars by Watt The Data

General data information

We have a dataset containing information on approximately 100,000 electric car charging sessions in Norway. The data spans from January 2022 to June 2024 and focuses on sessions using fast chargers by Eveny. The chargers in use have nominal power ratings ranging from 50 DC to 400 DC, providing a diverse set of data points for analysis.
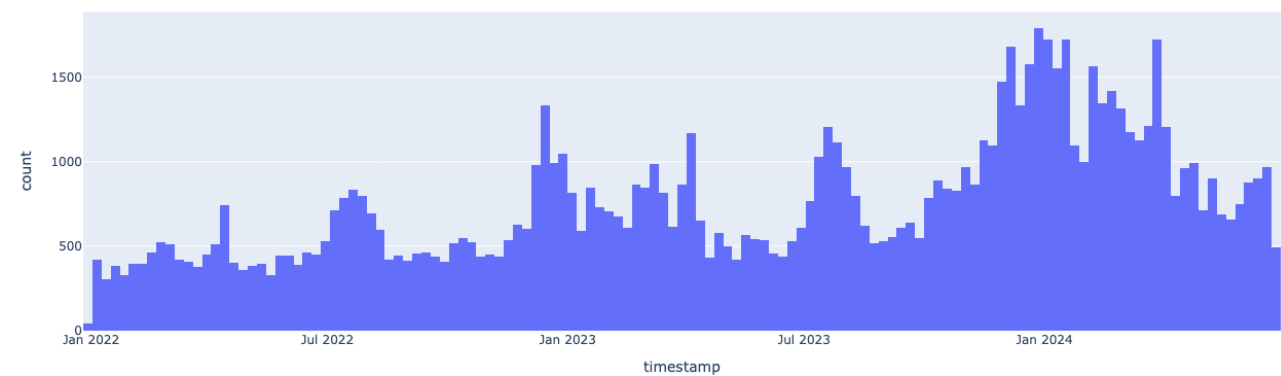

Histogram of Location IDs with Categorized Nominal Power
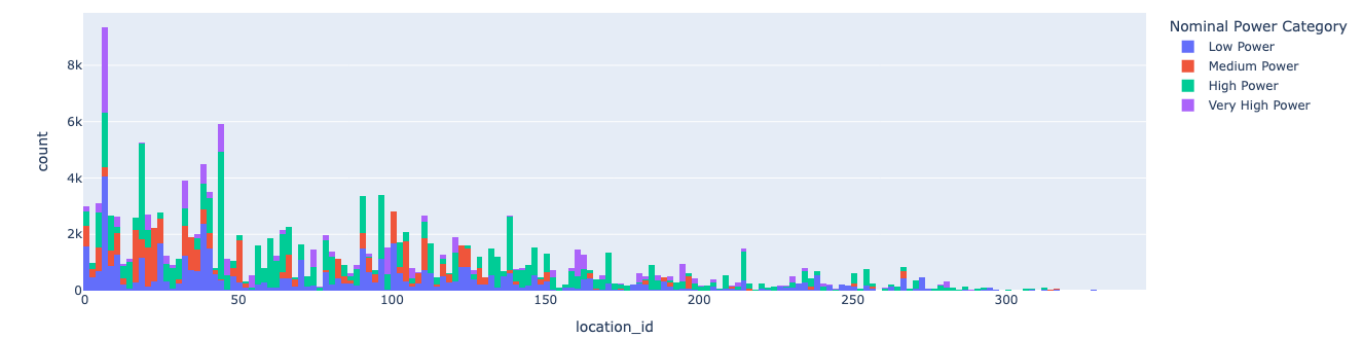
A general trend we see is that the charging power reduces when the SOC is increasing. Specifically around when reaching the 80% state of charge level.

Overall Median Power and SOC Trend Over Time
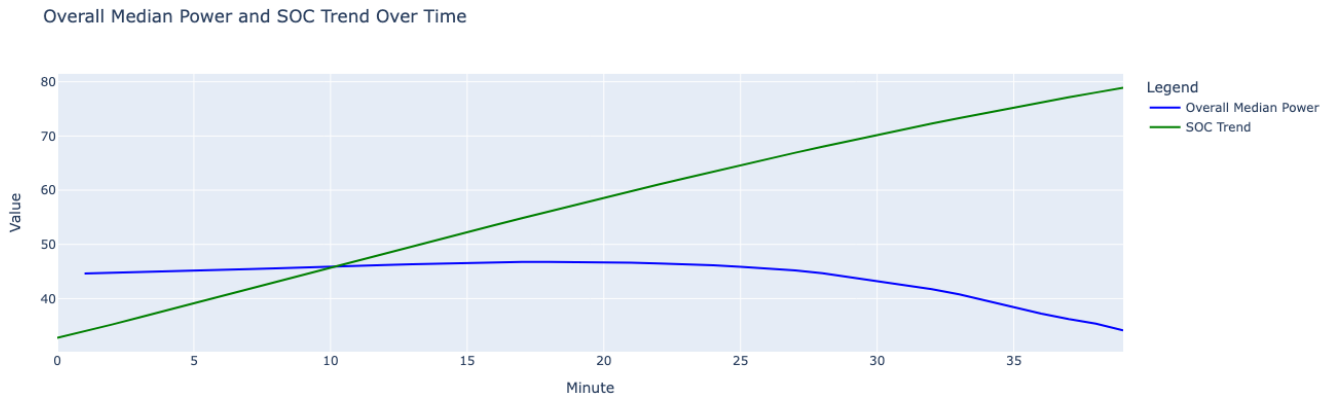
For all the different nominal powers, the power consumption to the car follows the same pattern that it slowly declines over time.

Median Power Trend for Different Nominal Powers
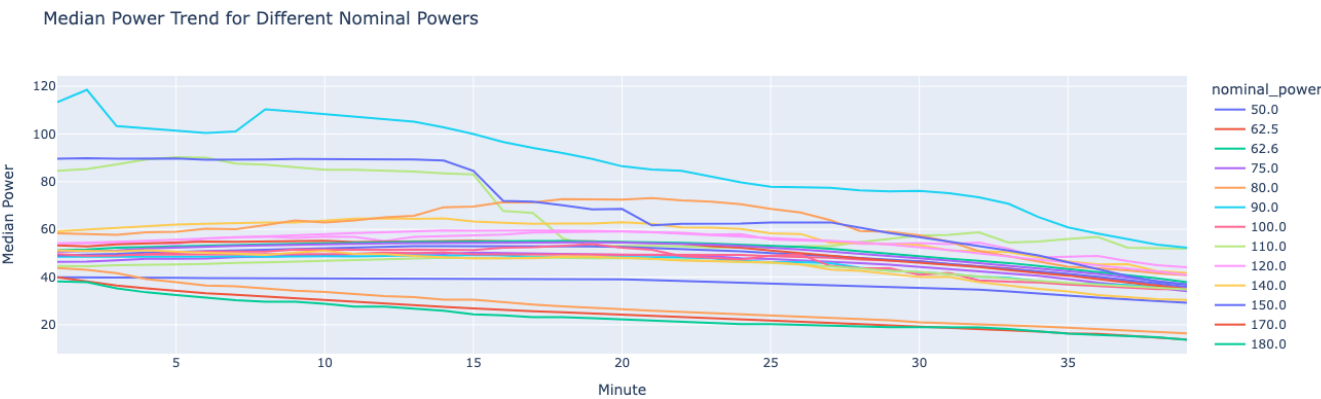
When we also include the State of Charge trend we can see the same pattern as earlier.

Overall Median Power for Each Nominal Power and SOC Trend Over Time
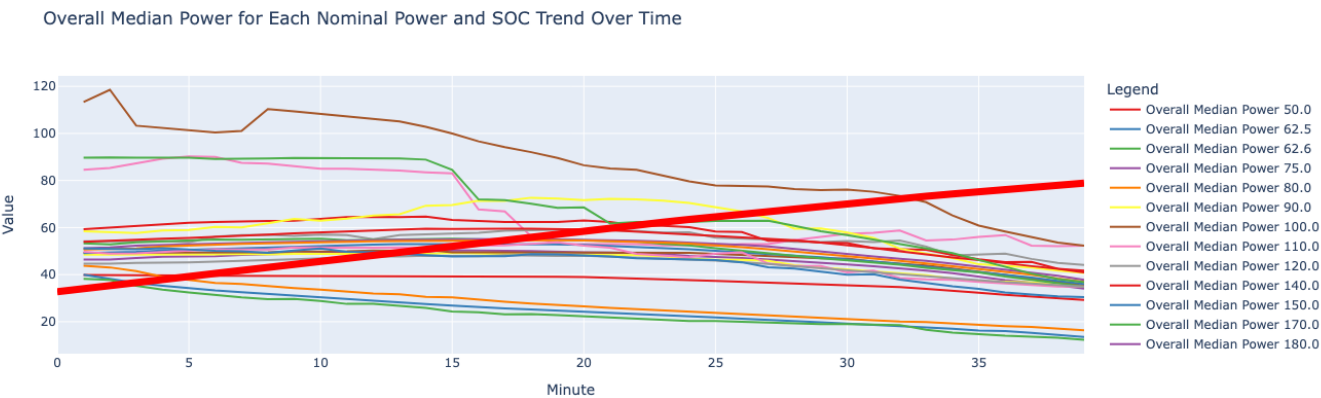
The power is also greatly impacted by the month of the charging. Likely because of weather conditions.
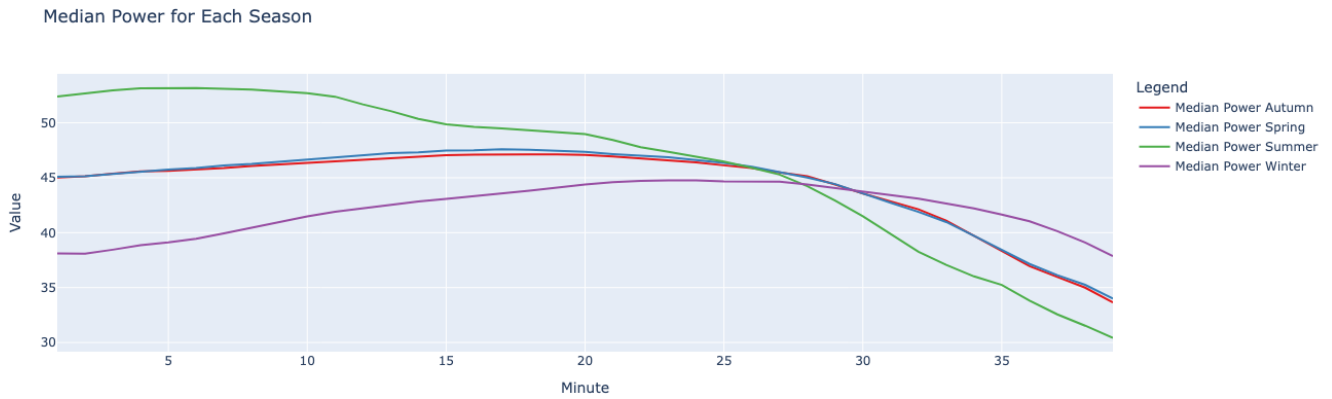
Specifically we see that the colder seasons have a slower charging rate. When we look at each month seperately we can see that some month correlate highly together. The colder the month, the longer time the charger uses to reach optimal power. January and December are the slowest as they are the coldest months. November, February and March also have similar powar as they have roughly the same temperature. The months June, July and August have the fastest time to reaching optimal charging power.
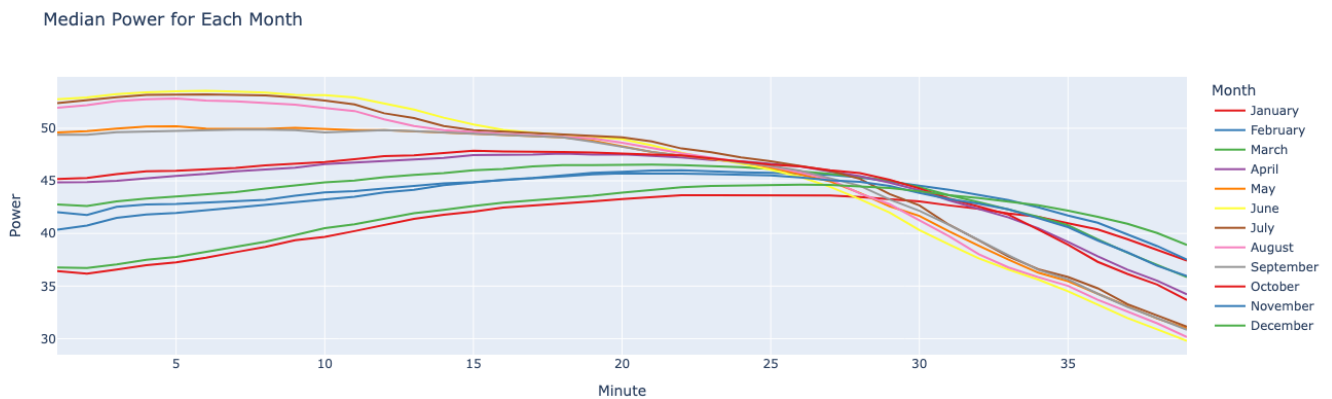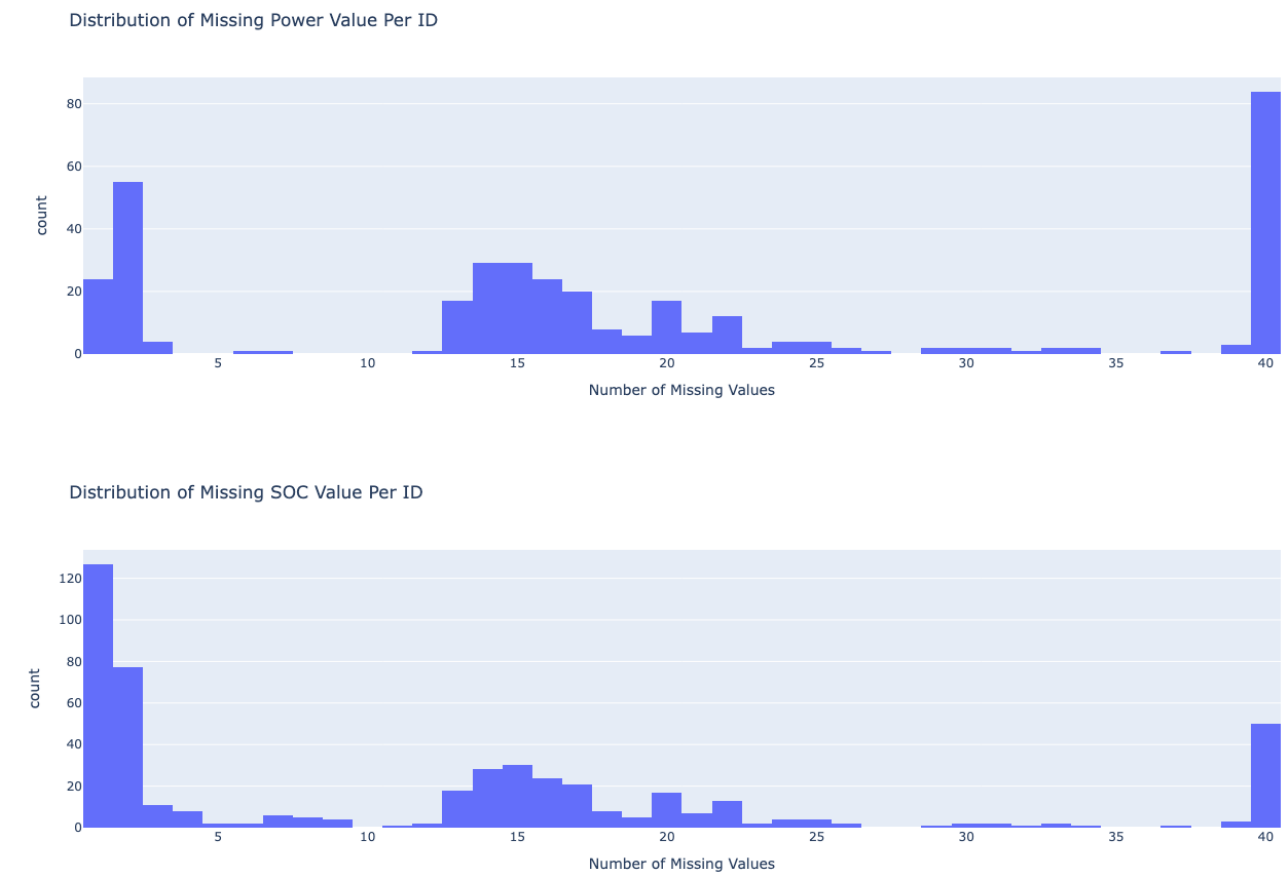

Data preperation

We have set each id to be a row. So each id is a charging sessions of 40 minutes. We also have a charging session with an invalid State Of Charge above 100%, specifically 104%. This oberservation has therefor been removed. We also have some power data that has power over nominal power of that charging station. Some have minimal differences ex. 0.1/0.2, we have decided to keep as we believe minimal differences won't impact model performace. But we have a lot of misreadings on location_id 6. There we have powers from 500 to 2400 with a nominal charging of 350. None of the session iDs has multiple of these invalid reading. We therefor set only that observation to NaN for later imputation. We have decided to set a treshold that if power is 1 over the nominal powers we discard that observation, and do later imputation if

the other observations for that charging session is correct.

Count of location IDs with inconsistencies between power and nominal power
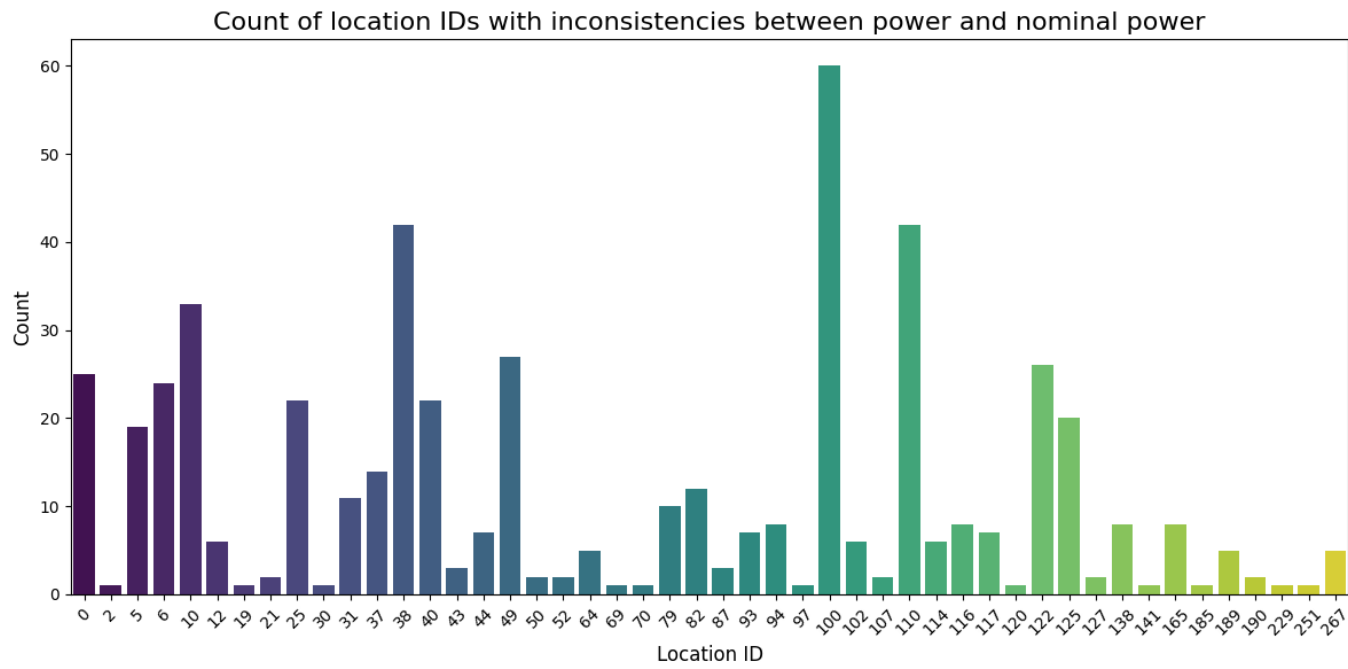
Missing values

State of Charge (SOC) and Power are the only variables that have missing values.

Distribution of Missing Power Value Per ID

Distribution of Missing SOC Value Per ID

We have some charging sessions that have a lot of missing data. For simplicity for imputation we have removed charging sessions that has more than ... missing values.