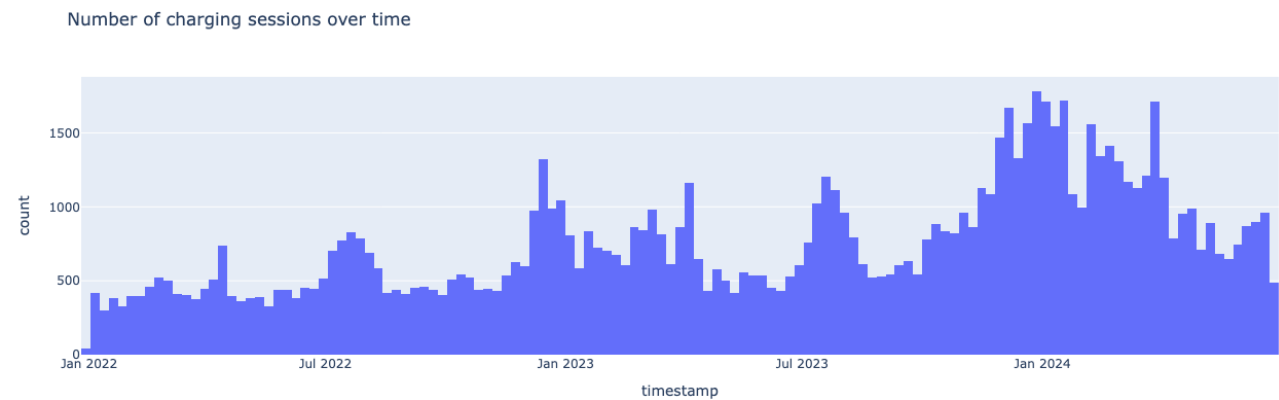


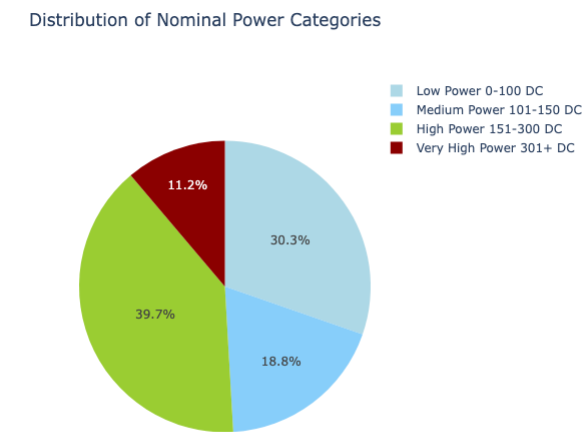
# Predicting power consumption and SoC for electric cars by Watt The Data

## Part 1: Exploring, processing, gaining insight from the data

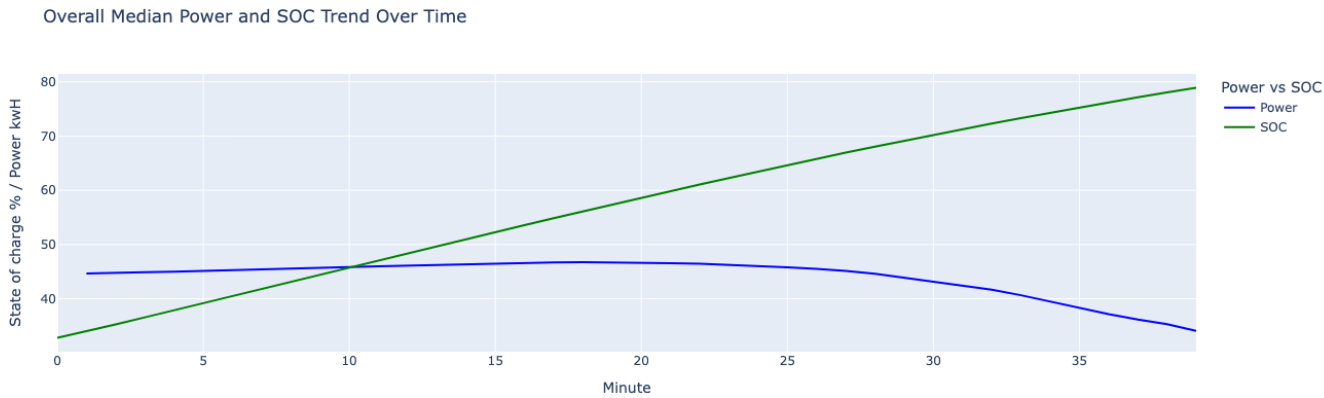
We have a dataset containing information on approximately 100,000 electric car charging sessions in Norway. The data spans from January 2022 to June 2024 and focuses on sessions using fast chargers by Eveny. The chargers in use have nominal power ratings ranging from 50 DC to 400 DC, providing a diverse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



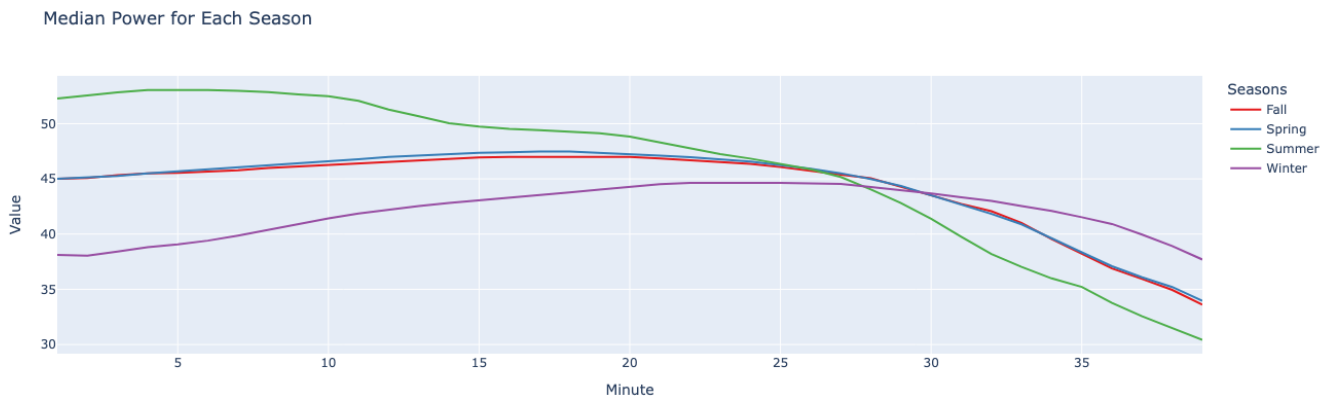
Overview of the number of charging sessiong over datasets time span, showing a general increasing trend. The figure illustrates how charging activity has changed over time, potentially indicating shifts in demand, seasonality, or other influencing factors.



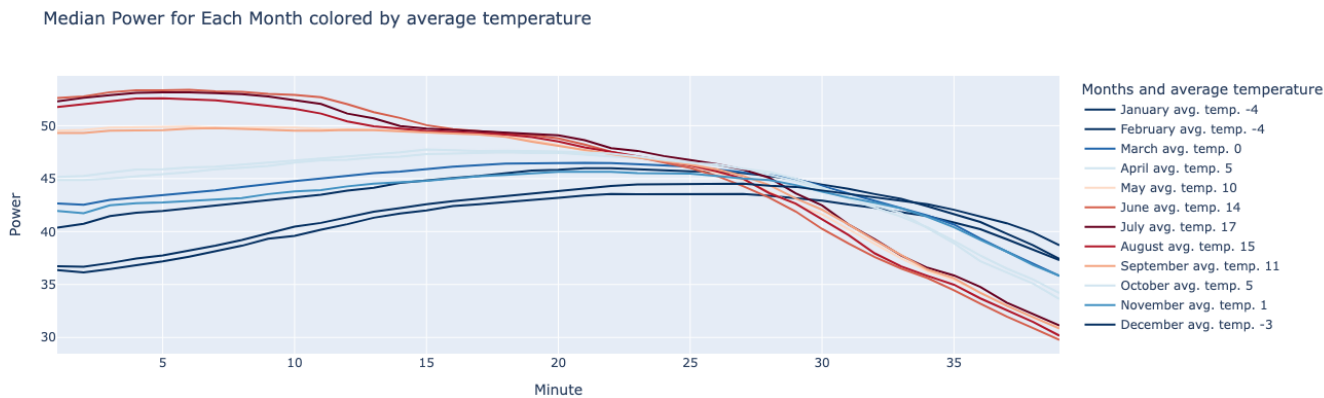
The pie chart illustrates the distribution of nominal power categories across the dataset.



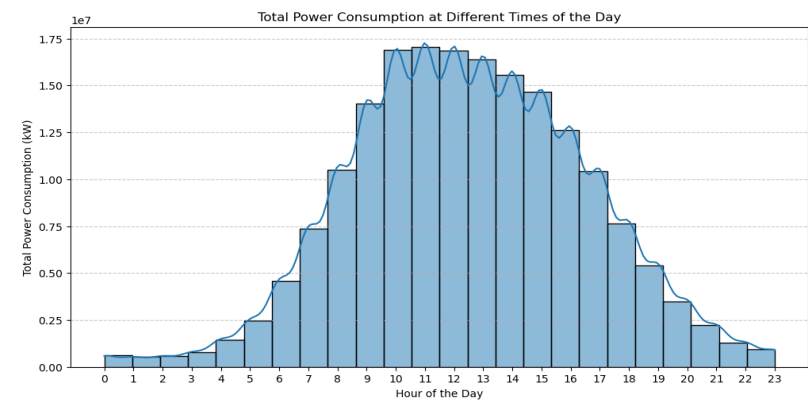
The line plot visualizes the relationship between state of charge (SoC) and charging power over time. A general trend emerges, showing a linear increase in SoC while power gradually declines. Notably, the charging power reduction becomes more pronounced as SoC approaches 80%.



This line plot displays the median power at each minute, categorized by season. We observe that winter starts at the lowest power level and follows a flatter curve, while summer begins at the highest and declines more steeply over time. Interestingly, spring and autumn exhibit nearly identical curves, suggesting similar charging power behavior in these transitional seasons. Likely because of weather conditions.



To investigate the impact of seasonality on power curves, we visualize average power by month in a line plot, colored by average temperature. A clear seasonal trend emerges, where warmer months start with higher power and gradually decline throughout the session. In contrast, colder months begin with lower power, increase to a peak, and then decrease toward the end.




This histogram shows total power consumption by the hour of the day. We observe an aggregated consumption pattern peaking around midday, forming a distribution resembling a normal curve.

Does the location ID matter, are there patterns here????

Errors


The descibe function reveals that the maximum State of Charge (SOC) value is 104%, which is invalid since SOC cannot exceed 100%. We identified and removed a charging session with this incorrect SOC value.

To determine whether errors are evenly distributed across locations, we examined power readings that exceeded the nominal power of the charging stations. While minor deviations (e.g., 0.1–0.2 above the nominal power) were retained, as they are unlikely to significantly impact model performance, we explored clipping power readings to the nominal power. This approach had little effect, so we decided to discard all observations where the power exceeded the nominal power threshold.

 Inconsitencies nominal power

The histogram below illustrates the distribution of inconsistencies across all location IDs, providing insight into the frequency of these anomalies.

Missing values

 Missing Power Values State of Charge (SOC) and Power are the only variables that have missing values.The histograms below illustrate the extent of missing data across charging sessions. To simplify imputation, we removed charging sessions with more than ... of missing values.

Data preperation

We reshaped data frame so each row represented a charging sesiong of 40 minutes. This restructuring facilitates for an easier modelling process.