Exploratory Data Analysis: Charging Events

Understanding Charging Events (2018 - 2019)

Pragnesh Anekal

May 2, 2024

Introduction

This report presents a comprehensive Exploratory Data Analysis (EDA) of the provided dataset on Charging Events. The analysis is structured into two main categories:

Categories of Analysis

- **Analysis by Time Events**: This category explores how charging events are distributed over the years 2018 to 2019.
- **Analysis by Charger Events**: This category delves into the characteristics of charging stations (chargers).

Report Format

The report follows a structured format that includes:

- **Visualization Explanation**: Each visualization is accompanied by an explanation of its significance and insights derived from the graph.
- **Code Representation**: Short snippets of code are included after some visualizations to demonstrate the process used to generate the visual output. The complete code is available in the Jupyter Notebook attached.

Dataset Preprocessing Assumptions

1. **Handling Null Values in Charger Name Column**:
    - Null values in the charger name column have been replaced with '**Unknown**' to indicate that the charger category is unclear.
    - Clustering techniques were used to group these values with chargers exhibiting similar performance, providing a rough estimation of the unknown charger identities.
2. **Identifying and Handling Charger Malfunctions**:
    - Charger events where the **total duration** is **zero** indicating the charging process has not begun but the **meter total** has a **high value**, or **vice versa**, is assumed to indicate that the charger did not function properly.

- Visualizations in the report exclude these events.
- However, setting the parameter `remove_assumed_errors` as `False` in the attached Jupyter Notebook allows generation of visualizations including these values.

Analysis of Time Events

This section includes visualizations and statistics related to daily, weekly, or monthly charging patterns of charging events.

Monthly Occurrences of Charging Events

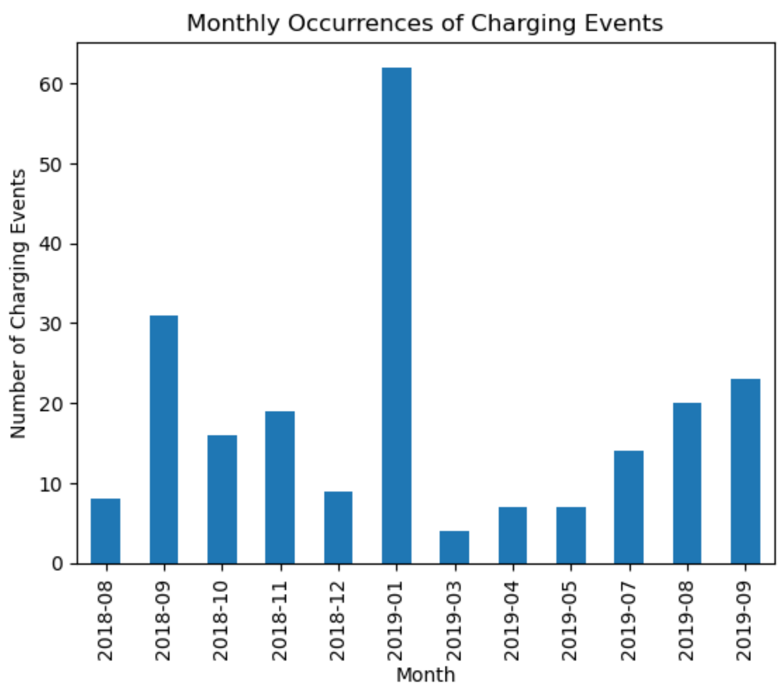

Figure 1: Monthly Occurences

- Notable surge in charging events in January 2019: likely due to the new year rush in demand
- Cyclical trend with peaks around the Fall season (**September**) - Seasonal Factors

Statistical Overview of Charger Usage

Table 1: Descriptive Statistics of Charging Data

	Meter Start (kWh)	Meter End (kWh)	Meter Total (kWh)	Total Duration (h)
Average	414.128 20	421.512 63	7.384 43	25.388 948
S.D	380.376 547	377.449 748	13.324 386	84.966 383
Median	232.778 00	239.262 980	2.287 920	3.088 889
75%	753.246 00	753.246 00	7.871 760	21.574 303
Max.	1204.911 00	1204.911 00	126.350 920	839.003 056

- The mean and median differences in 'Meter Total (kWh)' and 'Total Duration (h)' show that while most charging sessions are relatively short (2.29 kWh and 3.09 hours), a few outliers with high usage or exceptionally long durations skew the average significantly.
- The maximum 'Total Duration (h)' of 839 hours indicates a significant outlier that could suggest an error in recording, as a duration of 839 hours (or about 35 days) is highly abnormal for a charging session.

Table 2: Charger Usage Summary

Chargers	
Unique	17
Top	Charger 4
Freq	67

The table summarizes charger usage; among the 17 unique chargers (including 'unknown' chargers), 'Charger 4' is the most frequently used, with 67 occurrences.

Charging Events Over Time

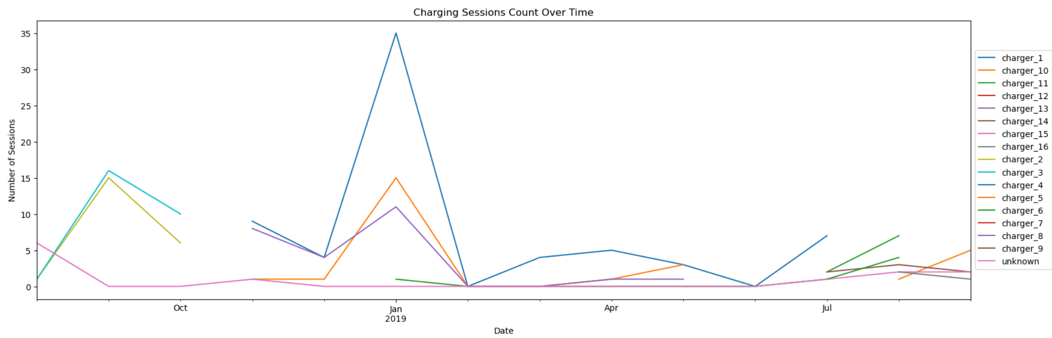

Figure 2: Events Counts for Different Chargers

- Several chargers exhibit usage only during specific periods, followed by extended periods of inactivity.

Rolling Average of Meter Total (7-Day)

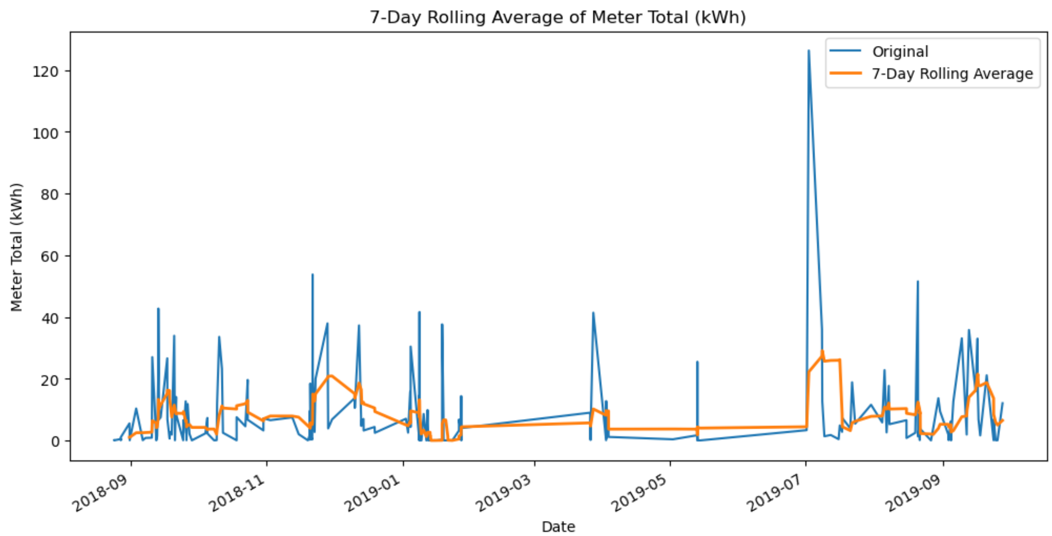

Figure 3: Rolling Avg: 7-Day

- The rolling avg. oscillates between 0 and 20 kWh, might reflect that the EV fleet are dominated by passenger cars with lower energy requirements.
- Occasional peaks might suggest few fleet of trucks, with higher energy consumption.

Meter Total For Each Month

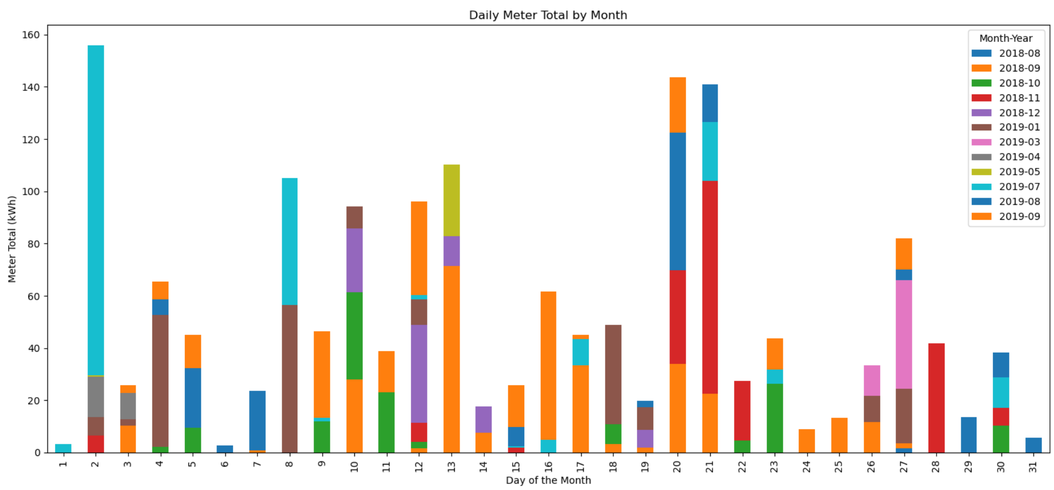

Figure 4: Meter Total Over 30 Day Span For Each Month

The graph displays the daily total energy consumption (kWh) for electric vehicle charging sessions, segmented by month, highlighting specific days with notably high usage.

```
def plot_daily_meter_month(self):
    # Extract day of the month and month-year as separate columns
    self.df['Day'] = self.df.index.day
    self.df['Month-Year'] = self.df.index.to_period('M')

    grouped = self.df.groupby(['Month-Year', 'Day'])['Meter Total (kWh)'].sum()

    pivoted = grouped.unstack(level=0)

```

```
    ax = pivoted.plot(kind='bar', stacked=True, figsize=(15, 7)
)
```
Listing 1: Python code for plotting stacked bar plot

Weekday vs Weekend Charging Events

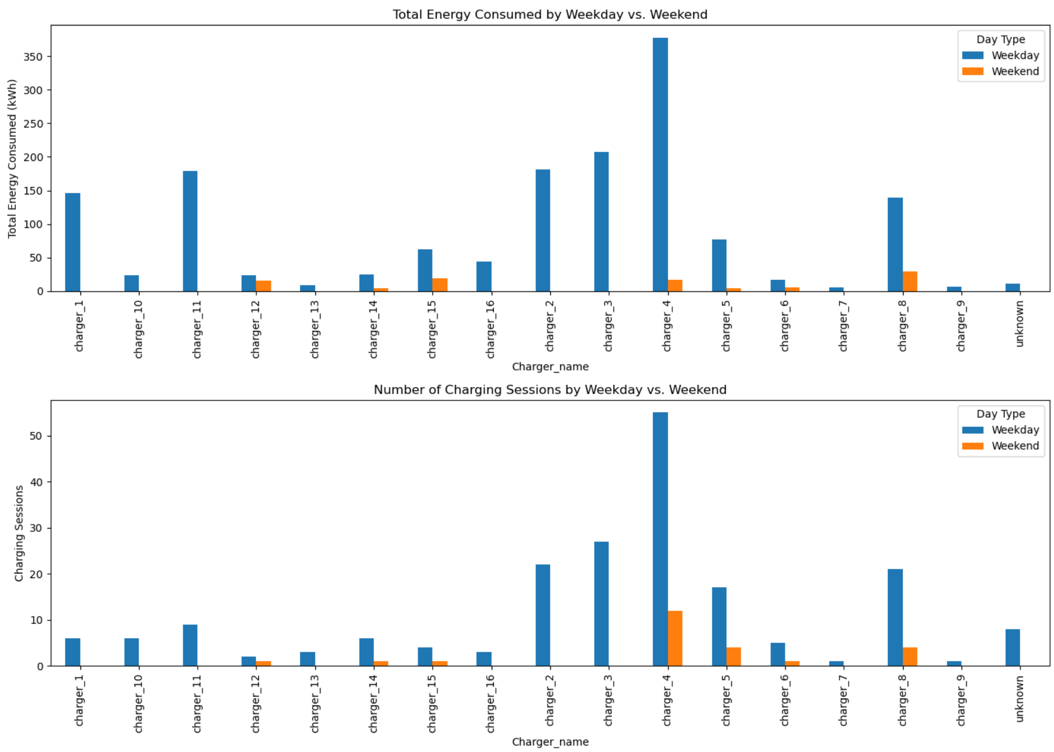

Figure 5: Weekday vs Weekend Charging Events

- The data reveals that most charging events occur on weekdays across almost all chargers.

Analysis of Charger Events

This section examines metrics such as usage frequency, duration of charging sessions, and charger popularity, providing insights into the utilization and performance of different chargers.

Charger Lifetime

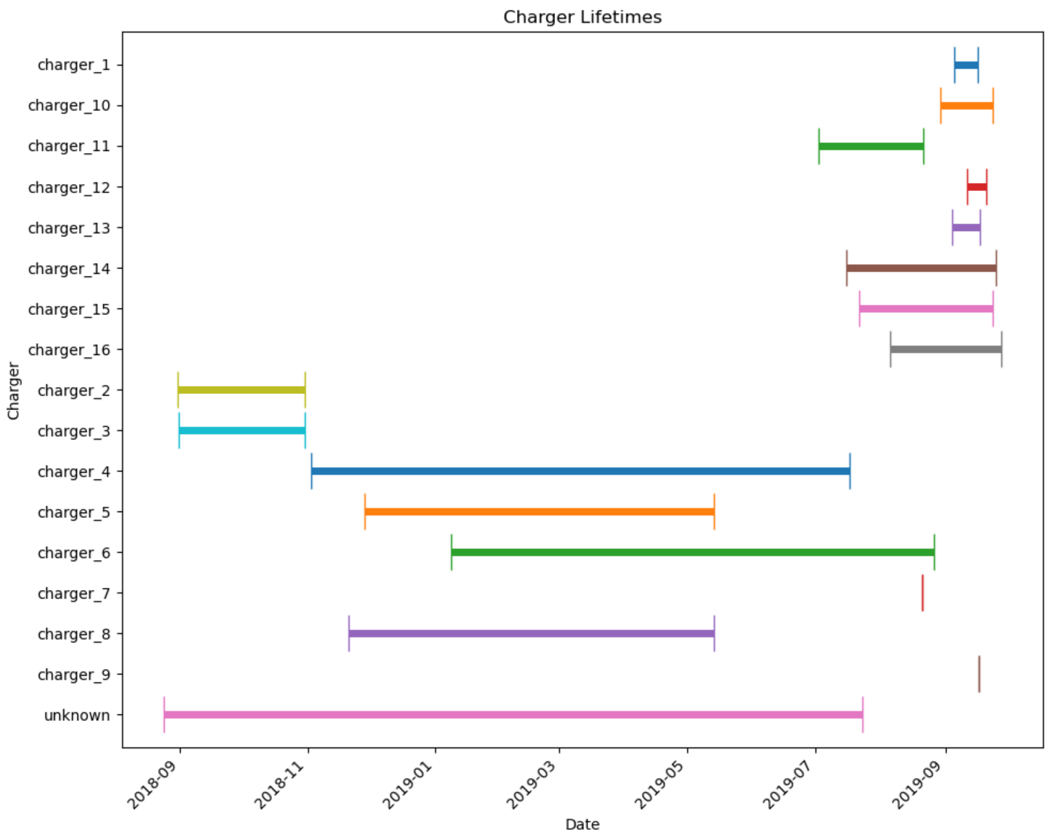

Figure 6: Lifetime of Chargers

Some chargers have been in use consistently over the entire period shown, while others have been active for shorter durations.

Weekly Active Period

Figure 7: Active Period of Each Charger

This graph adds detail to the charger lifetime by showing specific periods of charger activity within their overall operational span. It shows that while a charger may have a long lifetime, it may not be in continuous use.

Problematic Charging Events

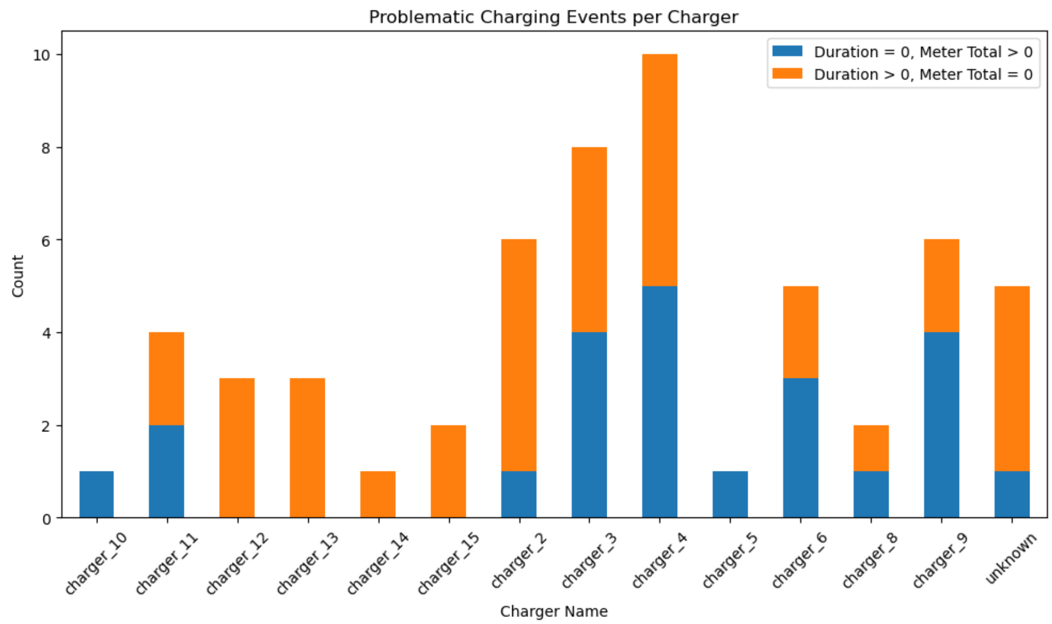

Figure 8: Enter Caption

The stacked plot categorizes problematic charging events into two groups: events with zero duration but nonzero meter total, and events with nonzero duration but zero meter total.

- Chargers 2, 4, and 6 show more events with recorded duration but zero meter total, indicating issues with energy recording or charging efficiency.
- Chargers 4 and 8 have events where energy is dispensed without recorded charging time, suggesting problems in duration tracking or charging initiation.

Power Comparison for Different Chargers

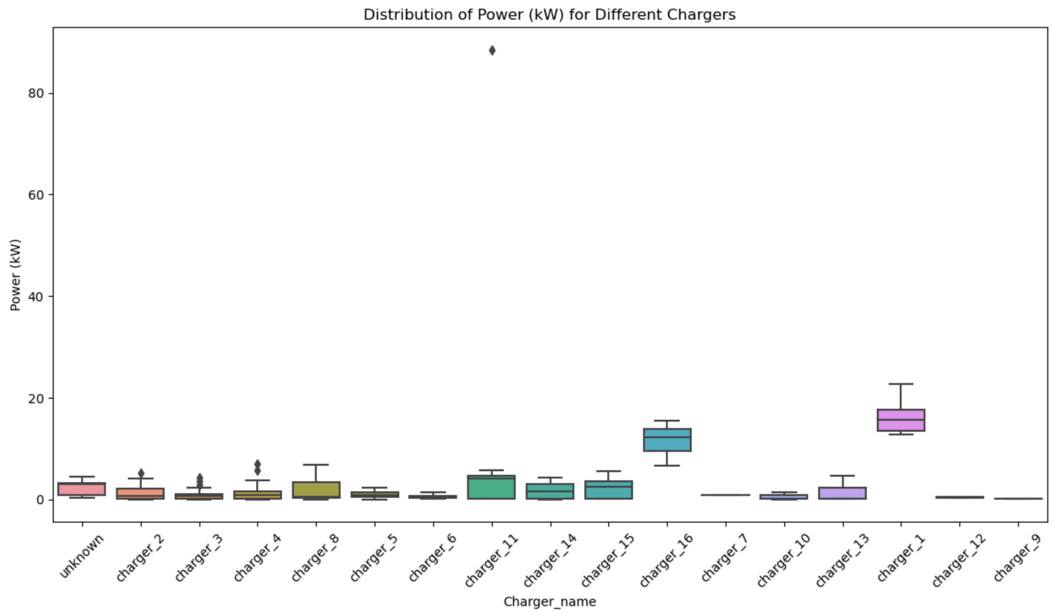

Figure 9: Power vs Types of Charger

- Chargers 16 and 1 deliver a lot of power during charging, but they are used less often compared to other chargers, potentially indicating they are reserved for specific purposes or underutilized.
- This high power output allows for faster charging, which may contribute to a more efficient charging process overall

Total Duration vs Meter Total (By Charger)

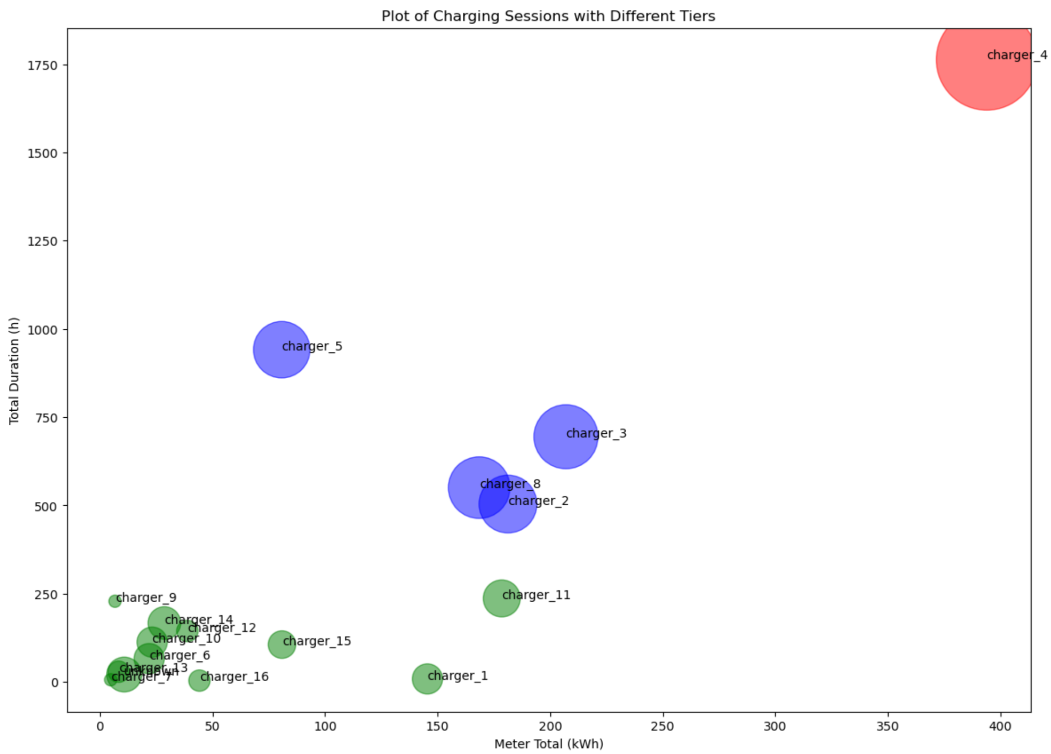

Figure 10: Charger Total Duration vs Charger Meter Total

- Charger 4 stands out with the highest meter total and duration, indicating prolonged and energy-intensive charging sessions.
- Most chargers are in the bottom-left quadrant, showing less frequent and less intensive use.
- The 'unknown' charger is clustered with tier 1 chargers, suggesting it is similar in type and may serve similar vehicles.
- Chargers like charger 4 may require more frequent maintenance due to higher usage, while others could be evaluated for efficiency or repurposing based on demand.

Peak Usage Time for Different Chargers (7-D Window)

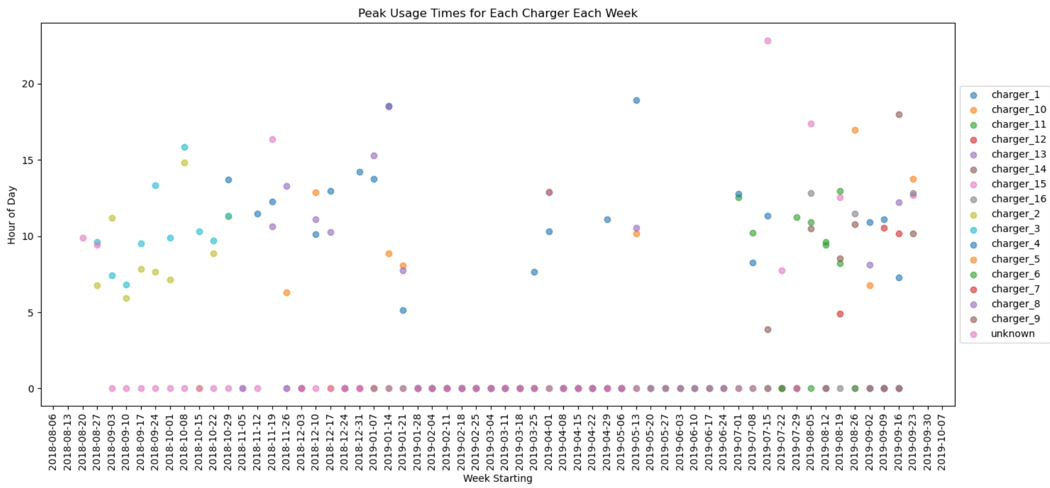

Figure 11: Peak Usage Time per Charger

- The data indicates a **peak charging period around 10 AM** across all chargers, suggesting a common trend where users prefer to start charging in the late morning.
- This consistent pattern in peak times throughout the weeks highlights a regular charging habit among users.
- This information can be leveraged to **plan maintenance**, **optimize power supply, and ensure charger availability** during these busy hours.

```
def calculate_weekly_statistics(self):
        # Group by charger name and resample weekly
        grouped = self.df.groupby('Charger_name').resample('W')

        average_duration = grouped['Total Duration (h)'].mean()
        median_duration = grouped['Total Duration (h)'].median()
        weekly_energy = grouped['Meter Total (kWh)'].sum()
        peak_usage_times = self.df.groupby(['Charger_name', pd.Grouper(freq='W')])['Power (kW)'].idxmax()
        charging_frequency = grouped.size()
        energy_variance = grouped['Meter Total (kWh)'].var()
```

```python
        average_energy_per_session = grouped['Meter Total (kWh)'].mean()

        summary_df = pd.DataFrame({
            'Average Duration (h)': average_duration,
            'Median Duration (h)': median_duration,
            'Total Energy Dispensed (kWh)': weekly_energy,
            'Peak Usage Time': peak_usage_times,
            'Charging Frequency': charging_frequency,
            'Energy Consumption Variance': energy_variance,
            'Average Energy Per Session (kWh)': average_energy_per_session
        })

        summary_df.fillna(0, inplace=True)

        return summary_df
```

Listing 2: Python code for weekly summary statistics by charger

Impact of Initial Meter Value on Time Taken to Accomplish Meter Total

In this section, the aim is to examine whether the time taken to achieve a specific meter total during charging sessions is influenced by the initial meter value and the type of charger.

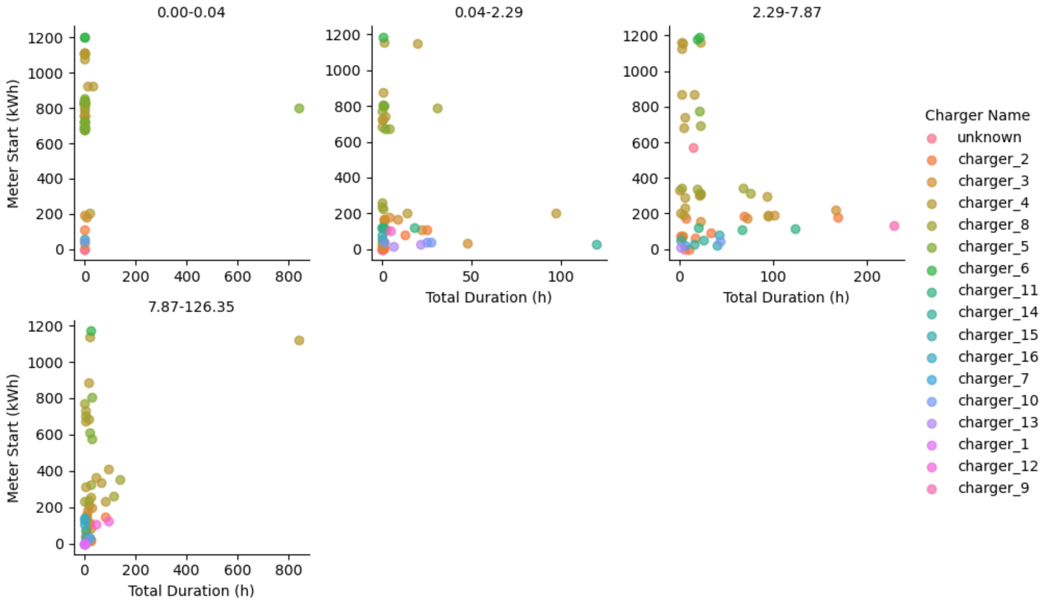

Figure 12: Meter Start vs Total Duration for Various Meter Total Ranges

- Each plot corresponds to a specific range of total energy used during charging events.
- There is no clear linear correlation, implying that longer sessions do not necessarily start with higher initial readings.

Hence ANOVA tests were done to see if there is a statistical significance.

Table 3: ANCOVA Results for Meter Total Bins

Source	Sum Sq	df	F	PR
ANCOVA for Meter Total Bin: (2.288, 7.872]				
Charger_name	55570.438876	13.0	2.690784	0.008179
Meter Start (kWh)	495.058880	1.0	0.311627	0.579795
ANCOVA for Meter Total Bin: (7.872, 126.351]				
Charger_name	56804.802901	12.0	0.350848	0.973136
Meter Start (kWh)	84071.477106	1.0	6.231093	0.016666

- In the (2.288, 7.872] 'Meter Total' bin, 'Charger_name' exhibits an F-statistic approaching significance, with a potential effect on meter totals.
- In the highest bin (7.872, 126.351], 'Meter Start' demonstrates a potential influence with a p-value below the 0.05 threshold, indicating that higher starting readings may affect the total meter reading.

Plot of Total Duration vs Charger for Meter Total (2.288, 7.872]

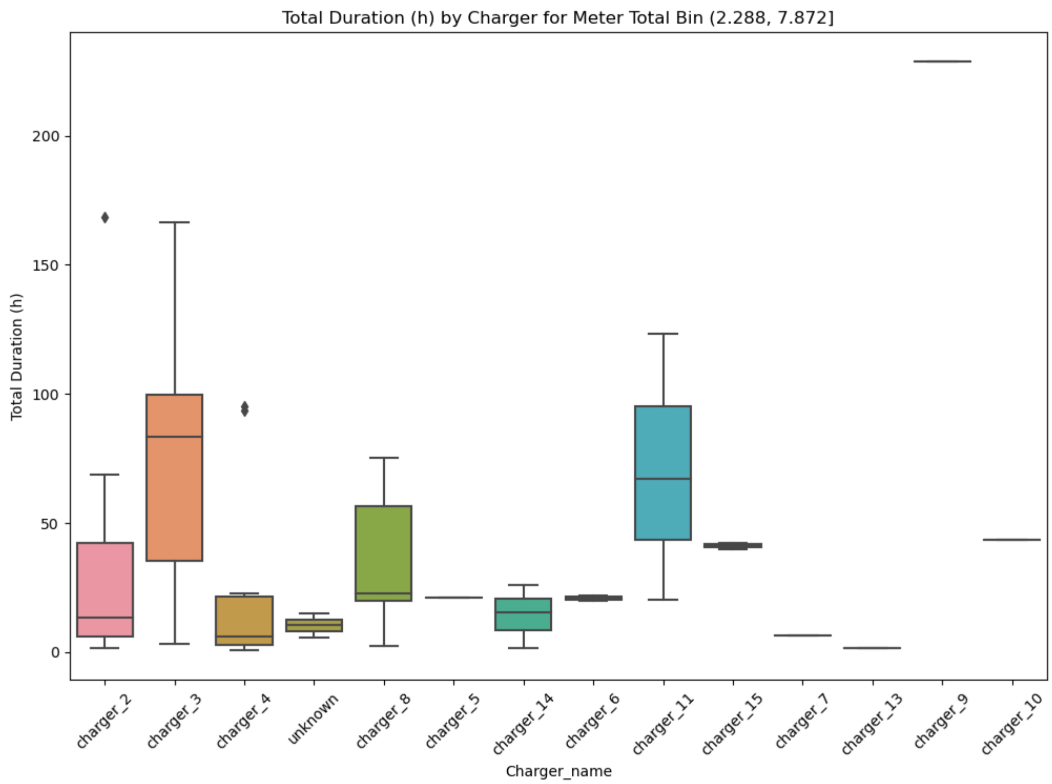

Figure 13: Total Duration vs Charger Type

Charger 4 shows the most **time-efficient performance** with a low **median** total duration, while **Charger 2**, although generally efficient, shows **greater variability** in charging session durations, with occasional longer charging times.

30-Day Power Analysis (kW) of Top 5 Most Used Chargers

The 5 most used chargers where extracted based on the total duration that the chargers have run.

```
def plot_30_day_performance(self):
        charger_usage = self.df.groupby('Charger_name')['Total Duration (h)'].sum()
        top_chargers = charger_usage.nlargest(5).index
        top_chargers_data = self.df[self.df['Charger_name'].isin(top_chargers)]
\begin{figure}
            \centering
            \includegraphics[width=0.75\linewidth]{30_day_performance_c5.png}
            \caption{Enter Caption}
            \label{fig:enter-label}
        \end{figure}
                top_chargers_data.sort_values(by=['Charger_name', 'Start Time'], inplace=True)
        top_chargers_data['30_day_Performance'] = top_chargers_data.groupby('Charger_name')['Power (kW)'].rolling('30D').sum().reset_index(level=0, drop=True)

        for charger in top_chargers:
            charger_data = top_chargers_data[top_chargers_data['Charger_name'] == charger]
            plt.figure(figsize=(20, 5))
            charger_data['30_day_Performance'].plot(title=f'30-Day Performance of {charger}', kind='area')
```

Listing 3: Python code for 30 day performance analysis

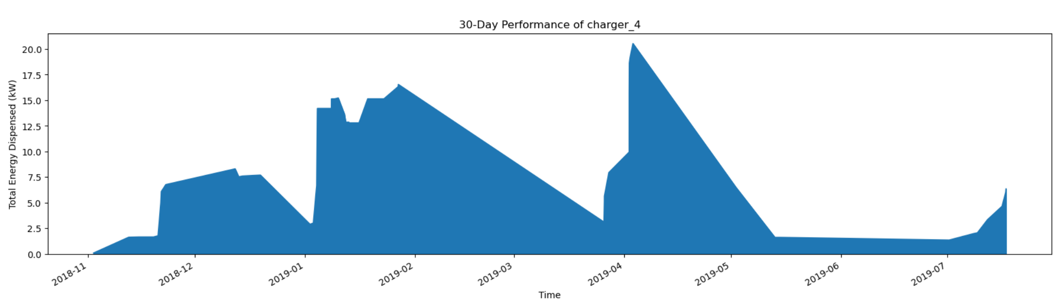

Figure 14: Charger 4

Charger 4's usage pattern varies noticeably. Recent data shows a rising trend, suggesting a potential increase in usage, marking the start of another period of higher activity.

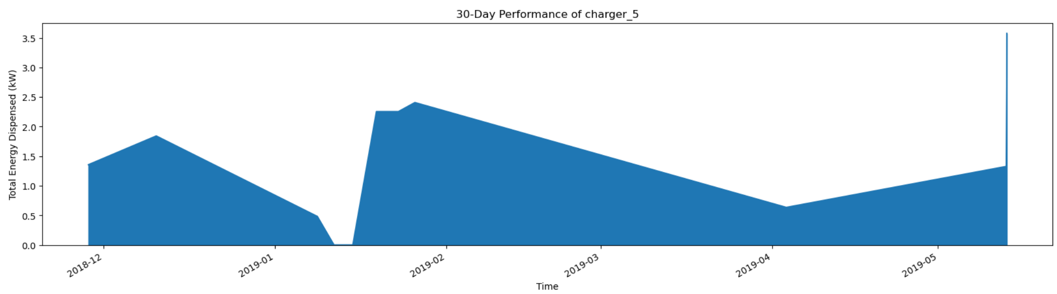

Figure 15: Charger 5

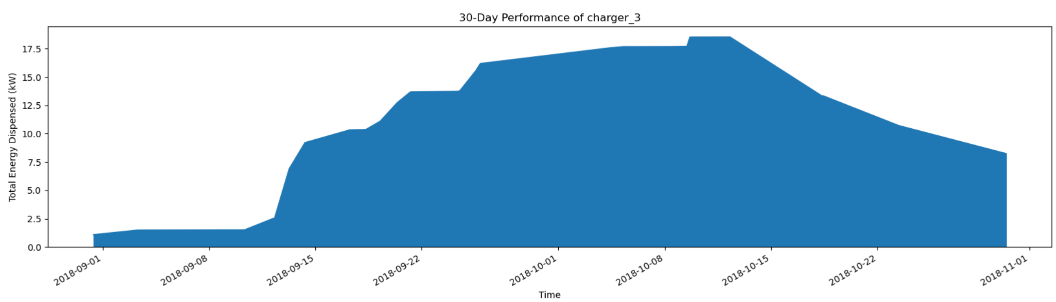

Figure 16: Charger 3

Charger 3 demonstrates a more **consistent usage pattern** compared to Charger 5, without sharp decreases or spikes during the timeframe.

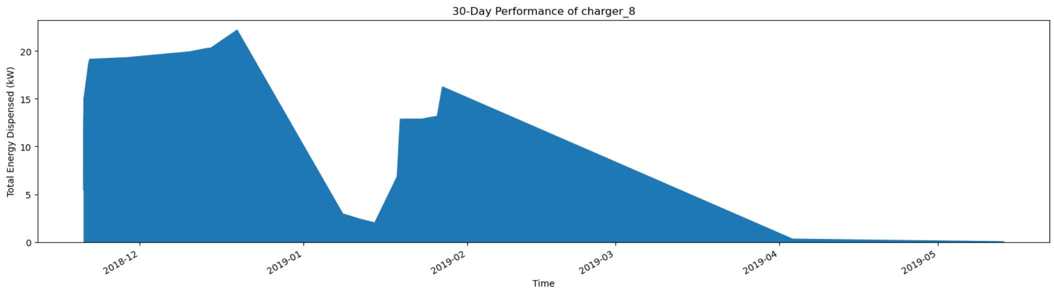

Figure 17: Charger 8

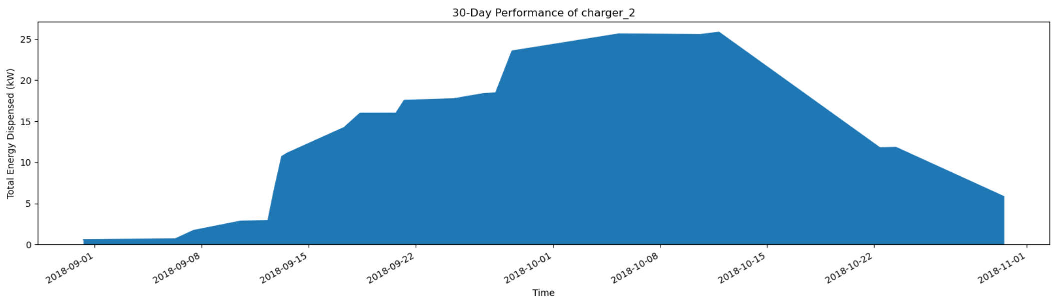

Figure 18: Charger 2

Charger 2 maintains a **consistent level of performance** over time, despite the gradual decrease after reaching a peak.

Conclusions

- Weekdays show a significant increase in charging events compared to weekends, reflecting fleet vehicle usage over private vehicles.
- Charger 4 is consistently used for energy-intensive charging sessions, indicating reliable performance.
- Differences between median and mean values in meter totals and durations suggest outliers, possibly due to system errors or unusual charging sessions.

- Peak charging times around 10 AM reveal a consistent energy consumption pattern, useful for operational planning and charger maintenance scheduling.
- Charger usage variability over time suggests different lifecycle stages and operational patterns.

Additional Note

Other visualizations were also created during the analysis process, including histograms of individual variables and scatter plots for Power. Please refer to the Jupyter Notebook for all visuals.

References

- Code: Jupyter Notebook