- E-Mobility -

Data Science Capstone Report

David Bobek
Antonino Rossi
Bertold Vincze

Table of Contents

Introduction

In this report, we (Bertold, Antonino and David) will try to collect all the insights, struggles, successes and failures that came about this Data Science Capstone Project.

We picked a dataset that describes the electrical activity of an electric vehicles (EV) charging station (data courtesy of Siemens®).

During this semester, the team met multiple times a week to practice what we learned about Data Analysis, Data Visualization, Data Preparation and Machine Learning, in order to take this (well-formed) heap of data and try to make some sense out of it, with the ultimate goal of gaining insights over the power consumption by EVs.

During the discussions presented in this report, we will explain why we made the choices we made to give context to this project, as well as the high-level implementation details and several plots of our results.

We decided to develop this project in the Python programming language.
This was done because we felt more confident with it, being the main focus of our course of study, and because there now exists a large ecosystem of libraries and frameworks to work with Artificial Intelligence which is likely never to be ported in R.

Objectives and Scope

In this report, we'll go over the objectives we established for our data science capstone project, which examined wallbox data for EV charging. Our main goals were to identify charging events, group them to make them easier to comprehend, and look for any relationships between charging patterns and the days of the week. We wanted to learn important things about how EV charging works so we could tailor the charging infrastructure. To do this, we mostly looked at wallbox data.

- Detection of charging events

  Our project's main objective was to create trustworthy algorithms and procedures for precisely identifying charging events in the wallbox data. We tried to recognize and categorize charging sessions by looking at spikes and dropdowns in power utilization. To better analyze use trends and improve the charging infrastructure, we would need to be able to detect charging events.

- Clustering of charging events

  After the charging events had been correctly identified, our next goal was to group them according to several factors, including length, energy usage, and charging rates. We were able to identify unique trends and acquire understanding into the various charging behaviors displayed by EV consumers by grouping charging occurrences. We sought to also try to find a pattern between different electrical vehicle brands and models, and their charging behaviors.

- Finding correlation between days and charging events

  Utilizing wallbox data, our project's ultimate objective was to look into any potential relationships between charging habits and the days of the week. We analyzed the data to see if there were any trends related to weekdays vs weekends, or whether specific days of the week had distinctive billing practices. Understanding these patterns might offer useful insights for improving charging infrastructure and creating plans for the effective use of resources.

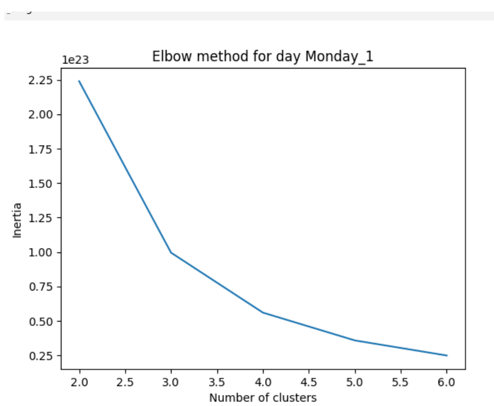
Methods

Day-wise K-means Clustering

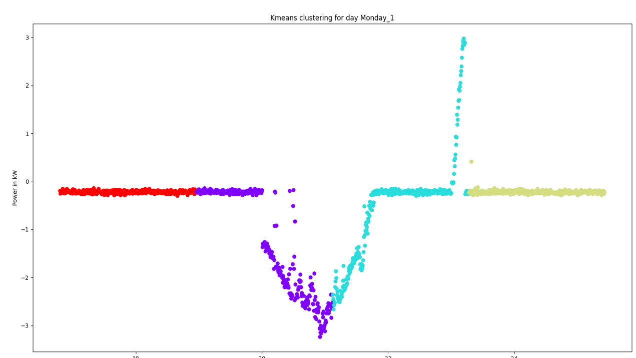
One approach that we took at the beginning of the project was to try and cluster different sections of the days to decide if some relevant pattern was there.
This was done in the interest of exploratory analysis, as well as a way to find indicators for future and more targeted clustering approaches.

The wallboxes dataset was split in 8 parts, one per each Wall box, showing electricity usage over a 24-hour period, with the X-axis representing time and the Y-axis representing power in kilowatts. We investigated K-means' appropriateness for this task and propose explanations for the observed creation of 4 clusters.

In order to identify the optimal number of clusters we have decided to use the elbow method. By observation we can see that the elbow is at 4 clusters.



Example of the Elbow Method computed on a Monday

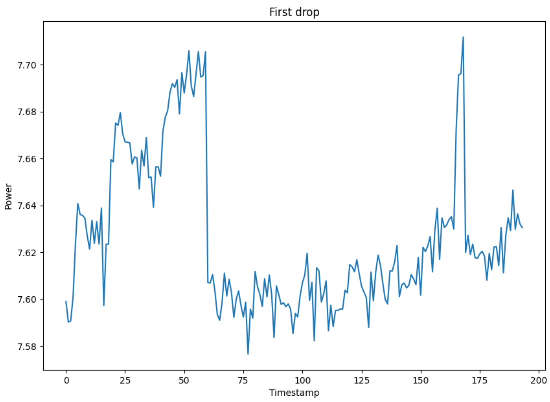


A day divided in 4 clusters, as the elbow suggests

Spike detection

In order to be able to charge events we have to at first spot their occurrences. On the large overview we could see the charging events by a rapid change in kW consumption. Just by simply taking all the occurrences where the kW consumption change would not be sufficient as there were a lot of false positives in our data. By a false positive we can declare a burst that is very short in comparison to a charging event. These bursts were usually less than 1 timestamp long (1-minute). After careful analysis on these short bursts we could confidently say that they were not charging events. The way we can spot the real charging events is by looking and zooming into them. On the Y-axis we can see the power consumption in kW and on the X-axis there is a duration in minutes. This charging event is varying in power consumption from 7.58 kW up til 7.7kW. The baseline of the wall box when there was no charging event was roughly 0.

We can see that throughout the 200 minutes this charging event was going on there have been movements of power. Initial phase of rapid charge and tan around 60 minutes in a rapid drop to a more stable activity. In later stages of this report we are going to look into clustering of these events. In order to detect this spike we needed to be trying to find an increase of power consumption from baseline but on the other hand take account of the false positives that were occurring. We also need to be aware that each of the wallboxes we are going to be examining is running on different power and therefore if we would have a uniform rule for detecting charging events we might not be able to detect them in wall boxes with low power when using settings for high power. After careful examination and testing I have decided to create the following configurations. The first index represents at what value of power do detect the charging event, known as lower threshold and the second index is representing at what value is the charging event over. We can easily compare that the wallbox Raptor 50 is being activated on more than 30 kW meanwhile KEBA boxes on 10-times less. This does not directly mean it is 10 times more powerful.



Example Drop detected in the Dataset

```
configurations = {
    DataWallboxesColumns.PowerKEBA1: [2, 2],
    DataWallboxesColumns.PowerKEBA2: [3, 0.5],
    DataWallboxesColumns.PowerKEBA3: [3, 0.5],
    DataWallboxesColumns.PowerLadebox1: [3, 0.5],
    DataWallboxesColumns.PowerLadebox2: [10, 6],
    DataWallboxesColumns.PowerLadebox3: [10, 6],
    DataWallboxesColumns.PowerDeltaWallbox: [15, 10],
    DataWallboxesColumns.PowerRaption50: [30, 25],
}
```

Wallboxes Configuration

Spike-wise K-means Clustering

As our last step, we decided to combine the two previous methods to classify charging events in clusters.

We found that the best way to classify charging events would be Dynamic Time Warping.

Dynamic time warping (DTW) is a technique for comparing sequences with varying speed or timing, useful when dealing with sequences of different lengths or with time shifts/distortions. It is commonly used in signal processing, pattern recognition, and data mining.

This required the following steps:

1. Find all charging events for all 8 charging stations
2. Transform the collected data into the right format (i.e. 3D numpy ndarrays)
3. Pad all events to the length of the longest event (so that we can apply DTW)
4. Find the right number of clusters
5. Plot the clustered events

Data Description

We were assigned 4 .csv files to take in exam, describing the following data over a span of 417 days (a snapshot is taken at every minute):

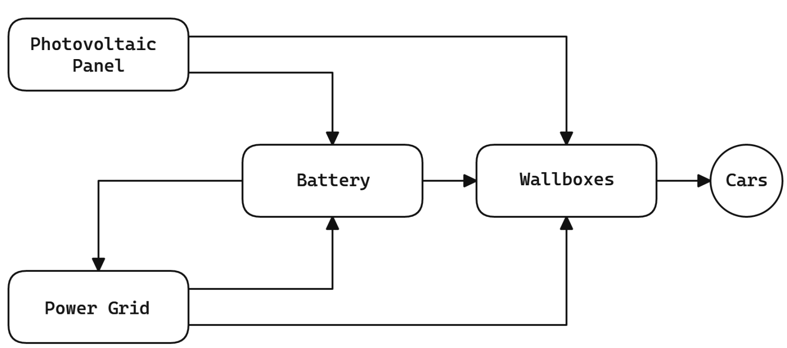
- Battery
  Timestamp, Total Power, Energy Charged, Energy Discharged, Battery Level

- Power Grid
  Timestamp, Power Provided

- Photovoltaic Panel
  Timestamp, Power Provided

- Wallboxes
  Timestamp, 3 Wall boxes provided by Keba, 3 Ladebox, 1 Delta and 1 Raption

Our analysis was heavily focussed on the Wallboxes, as we were trying to cluster Charging Events.



An exemplification of how the current circulates in the system

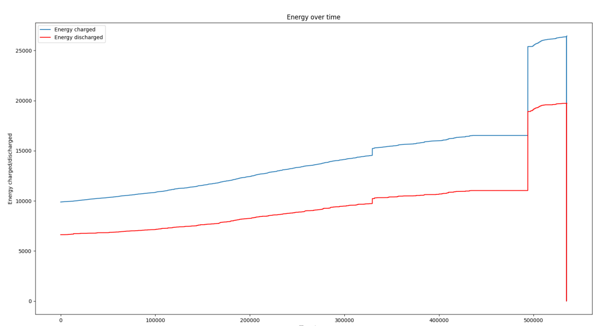
Missing Data

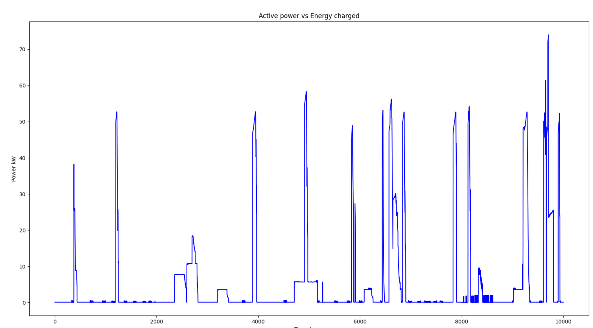
The Wallboxes dataset is remarkably complete; most days are fully documented. Since we had more than one year worth of data, and part of our interest was in time series, we decided to conduct most analysis on full days (233 in total).

EDA (Exploratory data analysis)

In order for us to get better insight into what the data provided mean we are going to be trying to visualize it and find meaningful information. In this first plot we can see 2 plots, the red being the energy discharged and blue being the energy charged. As we already know this architecture, whenever an energy is given away we need it to replace it in order not to run out of it. Therefore the blue line (Energy charged needs to stay above).

In the second chart we are going to be exploring the Wirkleistung-P which is the power. We can clearly see a lot of large spikes and a few little ones. These large spikes are representing the individual chagrin events. Data in this plot is gathered from the first 10000 minutes or almost 7 days. Each of these charging events need to be then categorized and evaluated.





Results and evaluation

General Wallbox statistics:
We looked at a total of 2,829 charging occurrences that were detected by the wallbox data throughout our observation period. These occurrences served as a useful starting point for our additional study and assessment since they shed light on the charging habits of electric cars (EVs). We looked at charging trends during the day and night as well as the distribution of charging events throughout the course of the weekdays and weekends.

Charging Events on Weekdays and Weekends: Of the total observed charging events, 2,096 took place on weekdays and 725 on weekends. This suggests that the bulk of EV owners often charge their cars during the weekdays, maybe to coincide with their usual commutes or work schedules. A very interesting part of this analysis is that the average charging event is 158 minutes or 2.6 hours.

When we examined the charge occurrences on particular weekdays,and the time they happen in a day, and, we discovered the distribution shown below:

```
Total_number_of_events: 2829
Events_on_working_days: 2096
Events_on_weekend: 725
Events_on_Monday: 445
Events_on_Tuesday: 414
Events_on_Wednesday: 430
Events_on_Thursday: 357
Events_on_Sunday: 330
Time_Range_Counts:
    Morning_(4-11): 1187
    Noon_(11-12): 168
    Afternoon_(12-18): 1053
    Evening_(18-22): 335
    Night_(22-4): 78
Average_length_of_events_(minutes): 157.51044538706256 minutes or 2.63 hours
Total_charging_time_(minutes): 445597.05
Total_charging_time_str: 445597.05 minutes or 7426.62 hours or 309.44 days
```

With Fridays having the most charging events, these figures shed light on the fluctuating demand for charging services during the workweek. This information can help with resource allocation and pricing infrastructure optimization for variations in demand over the workweek.

These figures suggest that EV charging is more common in the morning and afternoon, possibly correlating with usual commute times and the time when EV owners get home from work. It is possible that charging activities are less frequent throughout the noons, evenings, and nights based on the significantly smaller number of charging events during these times.
We picked two very performant wallboxes to see if there are relations between them and these were the results.

```
Wallbox statistics for LEM.Ladebox1.P

Wallbox: LEM.Ladebox1.P
Total number of events: 378
Events on working days: 324
Events on weekend: 53
Events on Monday: 64
Events on Tuesday: 70
Events on Wednesday: 61
Events on Thursday: 63
Events on Friday: 66
Events on Saturday: 27
Events on Sunday: 26
Time Range Counts:
    Morning (4-11): 221
    Noon (11-12): 25
    Afternoon (12-18): 101
    Evening (18-22): 17
    Night (22-4): 13
Average length of events (minutes): 298.4153439153439 minutes or 4.97 hours
Total charging time: 112801.0 minutes or 1880.02 hours or 78.33 days
```

```
Wallbox statistics for LEM.Raption_50.Wirkleistung_P

Wallbox: LEM.Raption_50.Wirkleistung_P
Total number of events: 829
Events on working days: 577
Events on weekend: 251
Events on Monday: 129
Events on Tuesday: 113
Events on Wednesday: 113
Events on Thursday: 103
Events on Friday: 119
Events on Saturday: 147
Events on Sunday: 104
Time Range Counts:
    Morning (4-11): 268
    Noon (11-12): 58
    Afternoon (12-18): 370
    Evening (18-22): 119
    Night (22-4): 13
```

Surprisingly, the Raption Wallbox wasn't the most used in terms of total charging time in our observation. It had a total charging time of 89,780.07 minutes, which is approximately 1,496.33 hours or approximately 62.35 days. On the other hand, LEM.Ladebox1 had a higher charging time of 112,801.0 minutes, which is approximately 1,880.02 hours or approximately 78.33 days. However, it is worth mentioning that the Raption Wallbox had a higher number of charging events compared to LEM.Ladebox1, more than double to be precise.
Additionally, Wallbox with the least total charging time: LEM.KEBA_P30_2 with 20792.0 minutes or 346.53 hours or 14.44 days

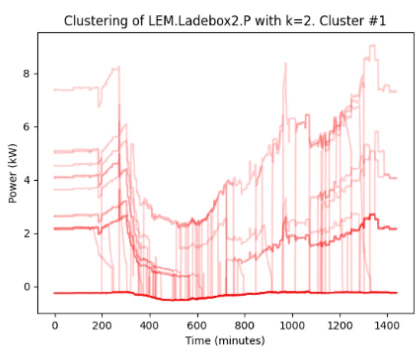
We can see the same things if we observe our other plots, as they show the utilization of batteries instead of the wallboxes. We can clearly see on the plots that batteries tend to be used more on weekdays, while they clearly seem to be recharging more from the grid and photovoltaic on weekends.

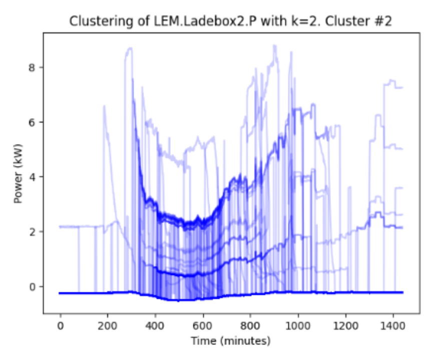
Implemented Approach

Time series clustering using Dynamic Warping

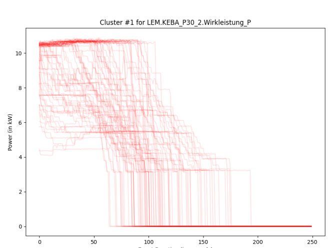
The elbow approach was used to calculate the optimal number of clusters, which we found to be two for all wallboxes. This suggests that the charging patterns may be efficiently divided into two different categories, indicating a distinct difference in the wallboxes' usage patterns.

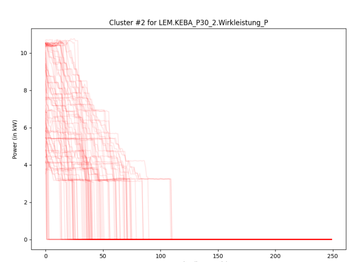
Consistent Usage Patterns
According to our data, wall boxes are used consistently over the course of many days, especially between the hours of 7 AM and 1 PM. This study implies that EV users typically follow a routine for charging during these times, which may coincide with their daily routines or typical commute patterns.

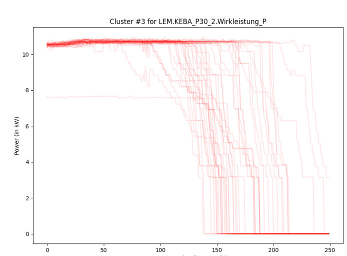




Instead, when applying the Elbow Method for charging events, we see that they cluster into 3 sections.







Find all clusters in the `plots/spike_clustering` directory.

Conclusion and Limitations

The results of our research can be summarized as follows:

- Highest utilization pick spans from 7.00 to 12.00.

- Days can be clustered into 2 main categories (roughly aligned in weekday vs weekend)

- Charging events exert different behaviors based on the charging station, with each charging station encountering 3 types of charging events.

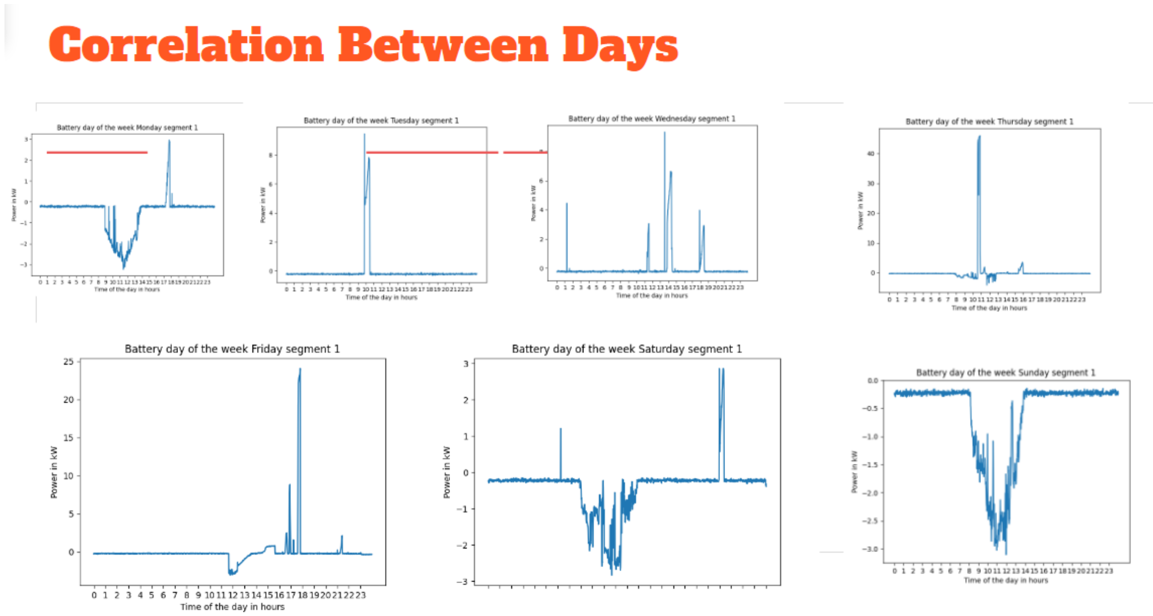
Many limitations were encountered during the process of development. Following, a list of approaches we tried but didn't yield the results we were expecting.

- Random Forest
  We first thought of Random Forest as a way of classifying charging events after detection. Sadly we couldn't achieve any meaningful results with it, even after several tries, so we decided to discard this approach.

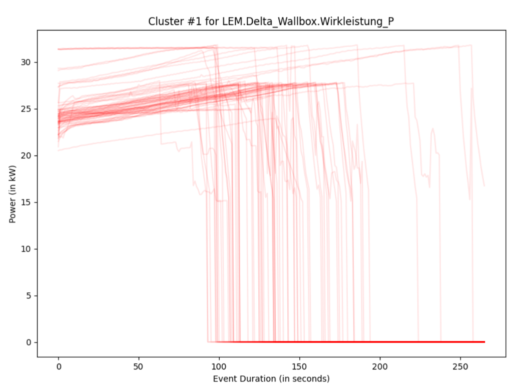
- ARIMA
  We used ARIMA as a first approach to forecasting. We thought that if we were to forecast future charging events we could generalize them and find some useful information about them. Sadly we couldn't find a reasonable way to do so, but it was insightful nonetheless.

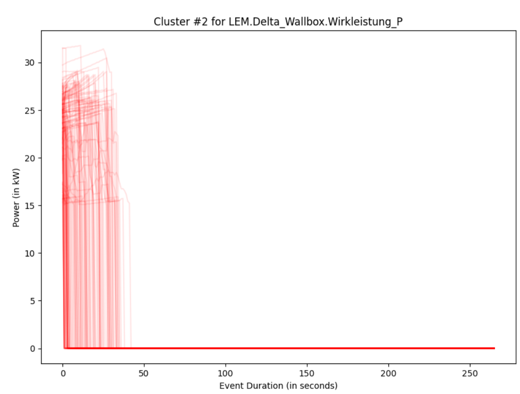
- OPTICS
  Finally, we also tried applying OPTICS to the time series describing charging events to see if it could be used for clustering.
  We sadly figured out that OPTICS on its own is not ideal. It could be used for anomaly detection, but charging events are not so unusual to be classified as such, and so we couldn't gather any useful information from it either.

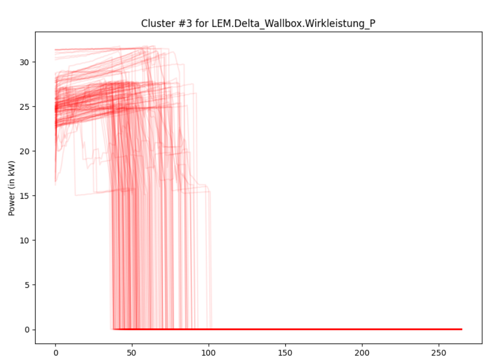
Appendix 1 - Extra Graphs



Correlation between days as obtained from clustering







Delta Wallbox Clustering Plots