Enhancing Electric Vehicle Charging Hub Efficiency through Data Analytics

Team: Team 15 - Remainder Team 2
Authors: Ann Weitz, Berra Üngör, Kevin Schomaker, Nico Bröhl and Nils Hornstein
Supervisor: Univ.-Prof. Dr. Wolfgang Ketter
Co-Supervisor: Janik Muires
Link to Repository: GitHub Repository Team 15

Department of Information Systems for Sustainable Society
Faculty of Management, Economics and Social Sciences
University of Cologne

February 7, 2024

Executive Summary

Due to climate change and the associated new challenges, the focus is also shifting to private transport as a major greenhouse gas emitter. Electric vehicles (EVs) are therefore becoming increasingly popular as a sustainable alternative. Efficient utilization of EV parking spaces as charging hubs is therefore an essential component of a more sustainable future in private transport.

The available data includes observations of the charging processes at the charging hubs and corresponding weather data for the period from early 2018 to late 2021. However, as the emergence of the coronavirus pandemic has severely disrupted life and thus influenced the data, the time frame has been cut off at March of 2020 for predictive purpose.

Analyzing Key Performance Indicators (KPIs) for EV charging stations, including site utilization, session duration, idle time, and general demand, uncovers a consistent morning peak in activity that aligns with conventional work schedules. This pattern, consistent among both registered and unregistered users, highlights the broad appeal of the stations and the potential for platform improvement. The analysis also indicates that the private car park experiences fewer weekend sessions and longer average session times, while the publicly accessible car park sees more weekend sessions and shorter average session times. The clustering analysis of EV charging station data reveals distinct patterns in user behavior, differentiated by the timing of charging sessions within the week, the amount of energy requested, and the efficiency of these sessions. Further analysis of temporal patterns and seasonal trends shows most connections occur around 6 AM, with a higher frequency of connections on weekdays, particularly from Tuesday to Thursday, and a notable decline during weekends. There are no monthly trends in 2018 and 2019, but a significant decrease in early 2020 due to the coronavirus pandemic. Seasonal patterns are not apparent.

The analysis of KPIs and clustering for EV charging stations reveals two key insights: a consistent demand pattern aligned with work schedules and significant variability in user charging behaviors and efficiency. These findings highlight opportunities for targeted improvements in station operations and user engagement strategies to optimize efficiency and satisfaction.

To enhance EV charging station efficiency, focus on operational readiness during early morning peaks and tailor engagement to user preferences with personalized incentives and demand-orientated pricing. Address the lower weekend usage by offering special promotions to encourage consistent use throughout the week. Ensure operational flexibility and continuously monitor trends to adjust services proactively, meeting changing user demands effectively.

Contents

List of Figures

List of Tables

List of abbreviations

Abbreviaton	Meaning
EV	Electric Vehicle
KPI	Key Performance Indicator
MAE	Mean Absolute Error

1 Problem Description

Due to climate change and the associated new challenges, the focus is also shifting to private transport as a major greenhouse gas emitter. Electric vehicles (EVs) are therefore becoming increasingly popular as a sustainable alternative. Efficient utilization of EV parking spaces as charging hubs is therefore an essential component of a more sustainable future in private transport.

Accordingly, we define key performance indicators that provide us with relevant information on performance of the public and private parking sites and display them in a comprehensible manner. Moreover, we use the data for both sites from the previous years' charging sessions to predict future utilization demand to improve operational performance. We use this to provide an improved user experience and pave the way towards sustainable mobility.

2 Data Preprocessing

The original data consists of two datasets: Hourly weather data from the Burbank Airport in California, US (Table 2) spanning from 01 January 2018 until 01 January 2021 and data from over 60.000 charging sessions from two different sites near the Burbank Airport in California, US spanning from April 2018 until September 2021 (Table 3).

To prepare the charging dataset for further analysis and prediction, a two-step process was used. First, the dataset was cleaned to remove duplicate columns and rows and erroneous data. Some columns were combined to an aggregate of both. Additional columns relating to date, time and duration of the charging events were extracted.

In a second step, the refined data was transformed to reflect the hourly utilization of both sites. For prediction purposes, this period only spans from April 2018 until March 2020. We argue that the data after this cutoff point is influenced by onset of the Corona pandemic and not usable for predicting on current events. There is a clear recess beginning at that point, lasting until the end of the year 2020. Although the data begins to normalize afterward, we lack any weather data beyond this point and attempts to supplement the weather dataset were unsuccessful.

3 Descriptive Analytics

3.1 Temporal Patterns and Seasonality

Next, we want to take a closer look at temporal patterns and seasonal trends, starting with the connections over the course of a day. As seen in Figure 4, most of the connections take place after midday, fade out during the night and are very low in the morning. This behaviour seems strange. Especially if you consider the case of the private site used by the employees of a company. It should be noted that the times in this dataset are measured in GMT. However, Los Angeles is in a different time zone, GMT-8. By recalculating the connection time, we get a much more reasonable graph.

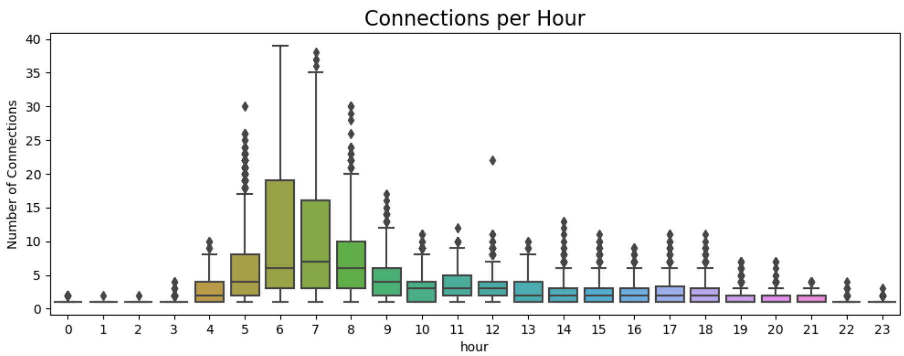


Figure 1: Connections per hour, corrected for time difference

Most connections take place at 6 o'clock and are relatively even throughout the rest of the day. During the night, however, there are zero or very few connections. A pattern can also be recognised over the course of a week (Figure 5). The number of connections is highest during working days, especially from Tuesday to Thursday. One possible explanation for the slight, yet significantly lower values on Mondays and Fridays is probably the higher popularity of working from home on these days. The number of connections drops sharply at the weekend. We should consider opening our view to a yearly perspective when looking at trends on a monthly basis (Figure 6). In both 2018 and 2019, we can observe many ups and downs in the number of connections per month. But they are all quite regular around 2000 connections, except for an outlier in October 2018. Suddenly, at the beginning of 2020, we can observe a huge downward trend that does not seem to recover until 2021. The coronavirus pandemic began in 2020, and the first lockdown in the US began in March of that year. However, no seasonal trends can be observed in relation to the seasons (Figure 7).

3.2 Key Performance Indicators (KPIs)

Below are the three main KPIs used to measure charging hub performance and utilization. Further supplementary KPIs are listed in the Appendix A.3.

Site utilization is crucial for optimizing e.g. the resource allocation of EV charging stations. Both sites exhibit a parabolic pattern with peak utilization, site 1 at 10AM and site 2 at 12PM. There is significant decline during the Coronavirus pandemic in 2020. This understanding is vital for resource allocation, growth planning, and site performance monitoring.

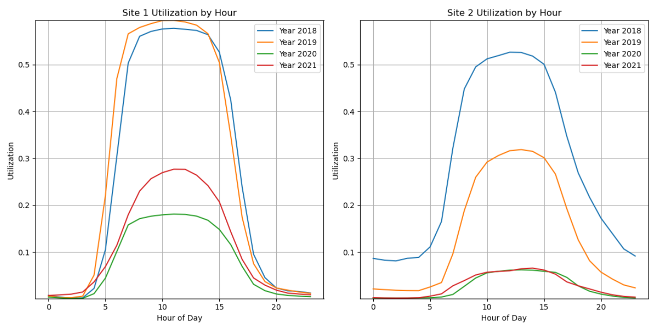


Figure 2: KPI 1 Site Utilization

General Demand tracks hourly station usage trends (Figures 8, 9, 10). Similar patterns are seen for registered and unregistered users, with peak usage at 6-7 AM. Unregistered users show a 5 PM spike, diverging from registered users' nighttime decline.

Average Session Duration reveals session length and station turnover. Sessions often exceed charging times, suggesting cars remain parked longer than needed, especially overnight (Figures 11, 12).

3.3 Site Characteristics

The given data set contains observations from two distinct charging sites. One is a private company car park, the other is publicly accessible at a university. It is necessary to identify which of the sites is public or private, as we can assume that they are used and demanded differently. As a key differentiating feature, we assume that the location of the company can be identified based on the use of working hours and the working week. It can be seen that site 1 of the data set exhibits far less sessions on weekends compared to site 2. As further evidence, we also consider the average occupancy time of the car parks by a customer. We can see that the average session lasts about 7 hours 12 minutes at site 1, which corresponds to a typical working day, and 5 hours 52 minutes at site 2, reflecting the irregular and short-term nature of a public location. We conclude that site 1 is the private parking site and site 2 in the data set comprises the public site.

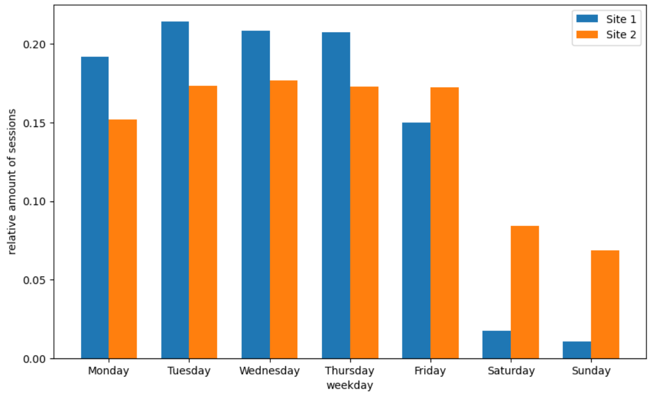


Figure 3: Distribution of sessions on weekdays

4 Data Analytics

4.1 Cluster Analysis

The results of the first clustering (Figure 22a and 22b) show that many people tend to charge their car at the beginning of the week, but not fully. Some of these people and others charge their car for the first time or recharge towards the middle or end of the week. A small proportion charge their car almost full regardless of the week. Three groups with exemplary names, Monday Chargers, Midweek and weekend Rechargers and Full-Throttle Chargers can be formed to summarize people or typical charging processes.

Clustering the amount of energy requested and the amount of energy delivered showed that there are 2 clusters, one describes the charging sessions in which a low to medium amount of energy is requested and delivered. The other cluster contains all sessions in which more or less energy is supplied than requested or in which a lot is requested and supplied. The people or sessions in cluster one could be called Gentle Chargers. The other group could include clusters such as Power-Hungry Globetrotters, Modest Plug Users and Electric Optimizers. The results are shown in Figures 23a and 23b.

Clustering the time of the week and efficiency reveals that there are sessions with high and low efficiency, both at the beginning and at the end of the week. The clusters in Figure 24a and 24b could be named as follows: Efficient Monday sessions, Monday overcharging sessions, Efficient midweek and weekend sessions and Inefficient midweek and weekend sessions.

The results of the cluster analysis can be used for a variety of purposes. In addition to marketing purposes, in combination with the distribution of the clusters they can give a charging station operator a guideline on how resources, in this case energy, can be allocated.

4.2 Utilization Prediction

4.2.1 Predictive Models

To guarantee comparability between the different models, we split the utilization data into train, test and validation splits of 60%, 20%, 20% respectively using a common seed. After this, data was normalized to values between 0 and 1, to improve stability of the machine learning models. As features for the predictive models, we used the hour, weekday and month of the utilization data point, as well as whether it is a holiday. We also used temperature and precipitation from the weather dataset. The target variable is the utilization of the specific charging site.

To compare our models, we mainly rely on the performance measures of mean absolute error (MAE) and r^2. The MAE in particular can be interpreted well with regard to the effects on utilization score we want to predict.

As our first predictive model, we decided to use a decision tree. We also tested linear regression as well as polynomial regression with L2 regularization. However, these performed significantly worse. We trained two distinct models for the private and the public site. Both are based on the same foundation: A decision tree regression with hyperparameter selection via GridSearch. To obtain the best possible models, we tested the use of different hyperparameters to control the complexity of the model and improve its results on performance measures. Figures 25 and 26 show the decision trees resulting from the best hyperparameters. It is immediately apparent that the tree for the public site is slightly more complex, while it performs worse in terms of the R^2 and the MAE value. Minor improvements could only be achieved by increasing the complexity and thus a poorer generalisation of the model. Nevertheless, it is logical that the MAE and R^2 values are better for the private location. As mentioned in the data analysis, the variance within the private site is much lower due to the use of car parks according to regular working hours. To improve performance, ensemble methods such as random forests and bagging were also included and tested. However, this resulted in a limitation with regard to the hardware. Ensemble methods with simpler hyperparameter settings, which also terminated in a reasonable time, did not lead to any significant improvements in terms of performance measures.

As the backbone of the neural network, Tensorflow's Keras was utilized. A few experiments lead us to an architecture of a simple feedforward neural network of 5 dense layers, with a dropout layer after each of the first two (Figure 27). Training and validation loss in this model decreased up until epoch 150, with no remarkable divergence between the two (Figure 28 and 29).

We experimented with more and less complex models, but the results proved unsatisfactory. While the more complex model achieved comparable results to our best model, complexity, training and inference time increased. The less complex model achieved slightly worse results (Table 4).

Both models, the neural network and the decision tree, show a relatively similar performance in terms of MAE. Compared to the neural net, the decision tree performs slightly worse on both the private and public site. Nevertheless, both models are able to explain at least 86% of the variance of the data on the private site and 70% on the public site. While the performance for both could be improved with additional hyperparameter testing or changes to the architecture, those small and negligible improvements come with a bigger time investment.

	Site 1		Site 2	
	MAE	R^2	MAE	R^2
Decision Tree	0.068	0.861	0.081	0.708
Neural Net	0.063	0.881	0.074	0.751

Table 1: Performance of machine learning models

With the outcome of both models being nearly the same, our recommendation would be to use the decision tree model. While neural networks are a black box, where we don't have any explainability about how the model actually works, the opposite is the case for the decision tree model. They are much easier for people to interpret, even with little technical background, which is suitable in this use case.

4.2.2 Business Case

Possible business scenarios resulting from the ability to predict utilization are demand-oriented pricing for the public site and the dynamic opening of the private site to the public. The latter makes it possible to generate additional income using unused space. One risk this would carry, would be a divergence between predicted and actual utilization of the parking sites. This could result in the site not being available for its original use as an employee car park, causing inconvenience and dissatisfaction with the company as our client. Demand-orientated pricing could lead to higher revenues. By taking predicted utilization and the resulting demand into account, we can adjust our prices accordingly and further in advance.

A Appendix

A.1 Data Description

field	type	description
id	string	Unique identifier of the session record
connectionTime	datetime	Time when the EV plugged in
disconnectTime	datetime	Time when the EV unplugged
doneChargingTime	datetime	Time when of the last non-zero current draw recorded
kWhDelivered	float	Amount of energy delivered during the session
sessionID	string	Unique identifier for the session
siteID	string	Unique identifier for the site
spaceID	string	Unique identifier of the parking space
stationID	string	Unique identifier of the EVSE.
timezone	string	Timezone of the site. Based on pytz format
userID	string	Unique identifier of the user, if provided
userInputs	list	Inputs provided by the user

Table 2: Charging Sessions Data

field	type	description
city	string	Name of the city
timestamp	datetime	Point in time of weather measurement
temperature	float	Temperature
cloud_cover	float	Degree of cloud cover
cloud_cover_description	string	Description of the degree of cloud cover
pressure	float	Air pressure
windspeed	float	Wind speed
precipitation	float	Amount of precipitation
felt_temperature	float	Perceived temperature

Table 3: Weather Data Burbank Airport

A.2 Descriptive Analytics

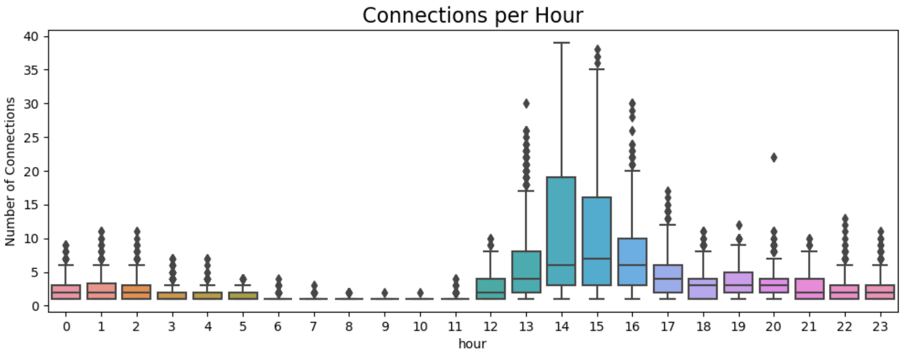


Figure 4: Connections per hour

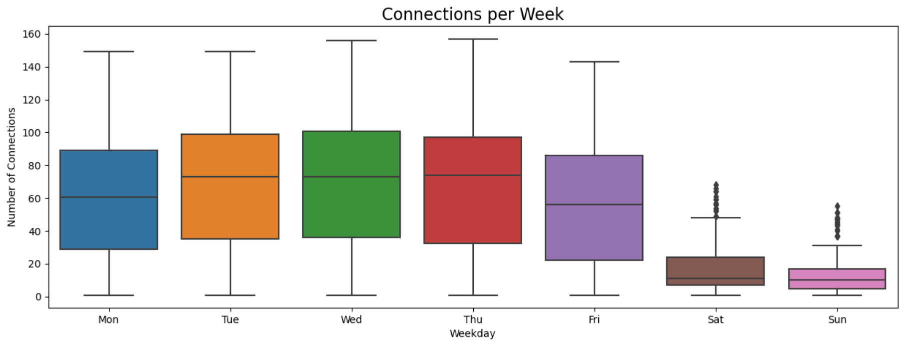


Figure 5: Connections per Week

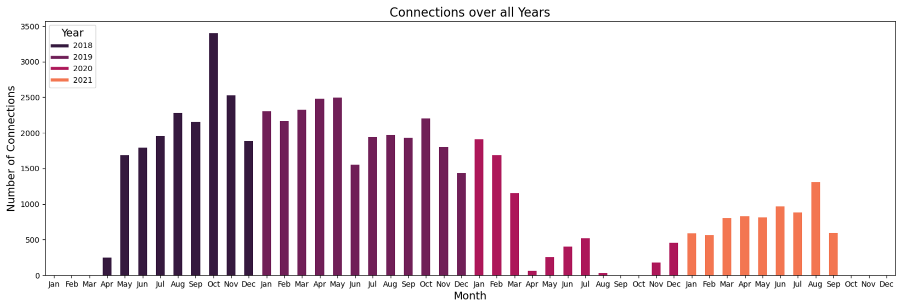


Figure 6: Connections over all Years

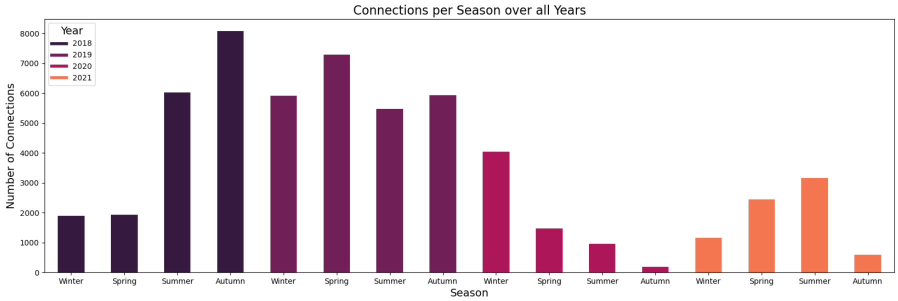


Figure 7: Connections per Season over all Years

A.3 Further Key Performance Indicators (KPIs)

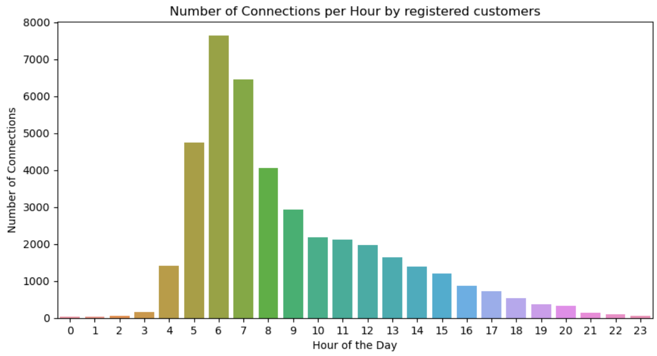


Figure 8: KPI 2 General Demand of registered users

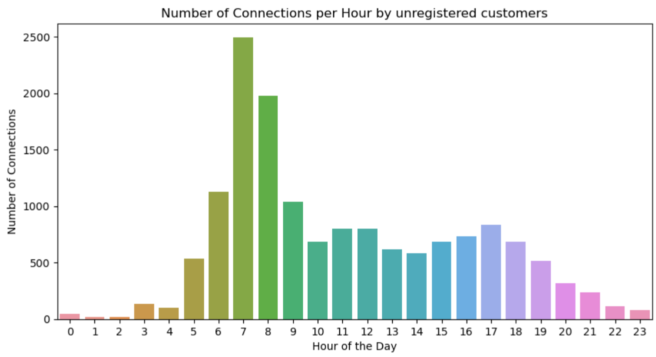


Figure 9: KPI 2 General Demand of unregistered users

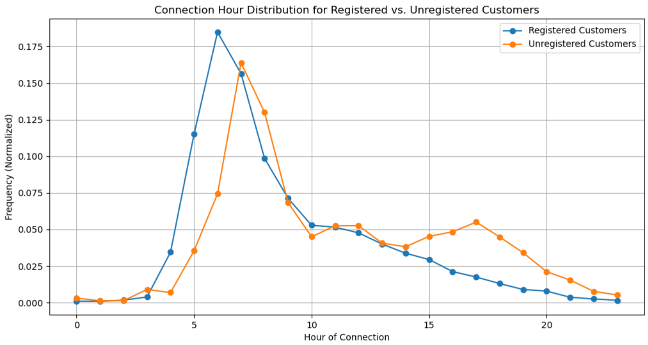


Figure 10: KPI 2 General Demand

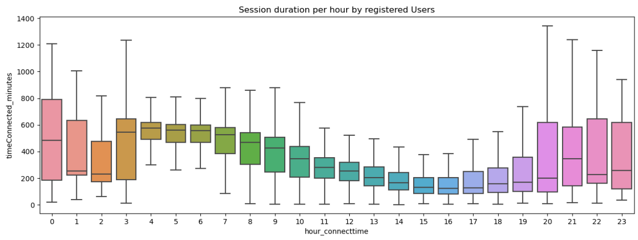


Figure 11: KPI 3 Session Duration of registered users

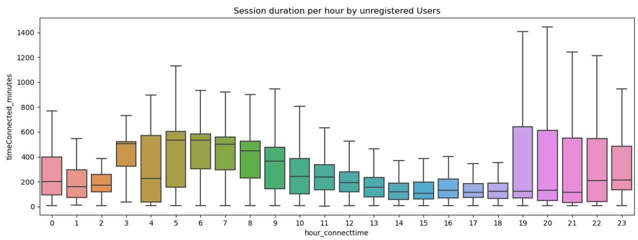


Figure 12: KPI 3 Session Duration of unregistered users

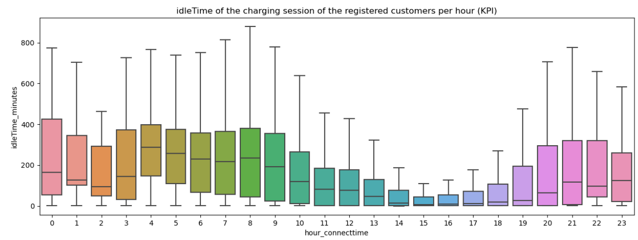


Figure 13: KPI 4 Idle Time of registered users

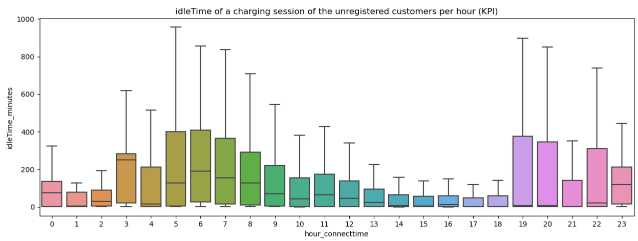


Figure 14: KPI 4 Idle Time of unregistered users

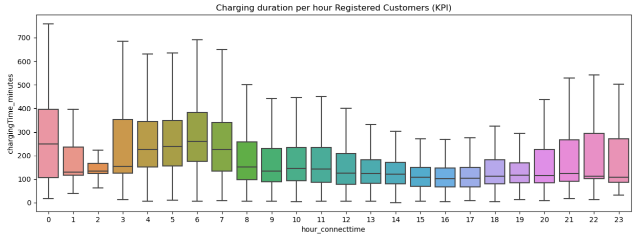


Figure 15: KPI 5 Charging duration of registered users

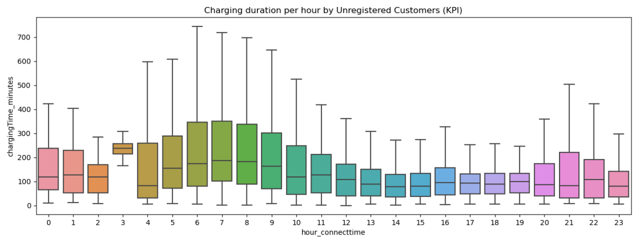


Figure 16: KPI 5 Charging duration of unregistered users

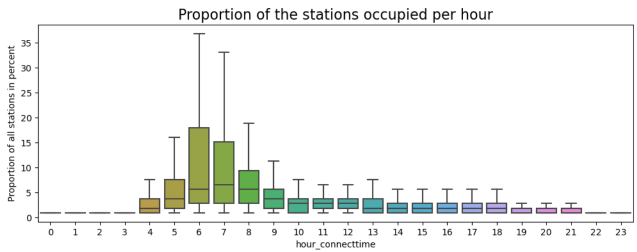


Figure 17: KPI 6 Propotion of the stations

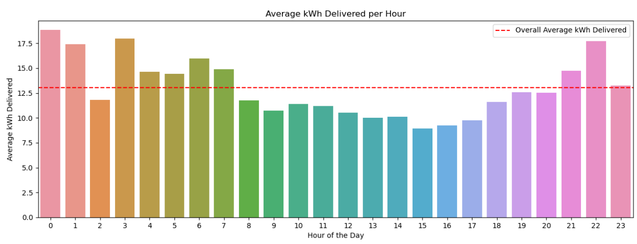


Figure 18: KPI 7 Average kWh delivery

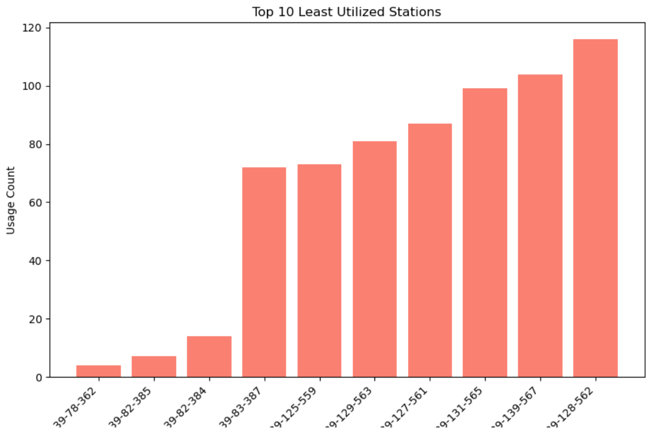


Figure 19: KPI 8 Top 10 least utilized stations

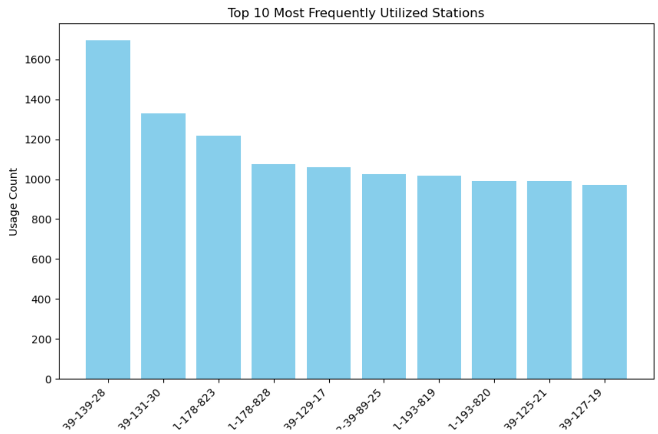


Figure 20: KPI 8 Top 10 most utilized stations

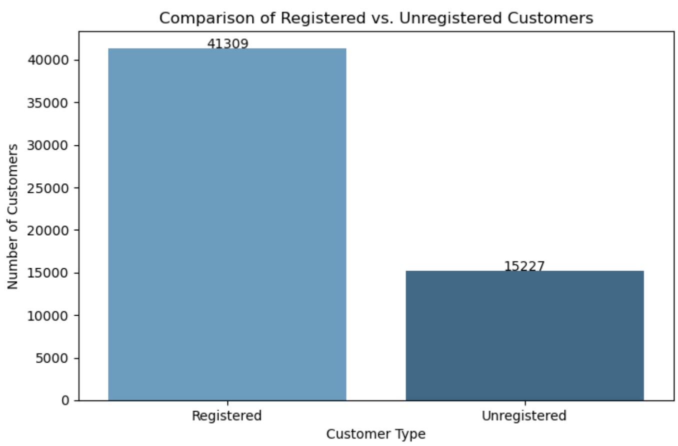
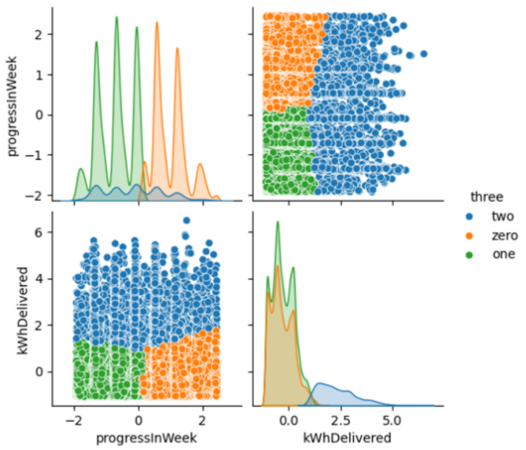
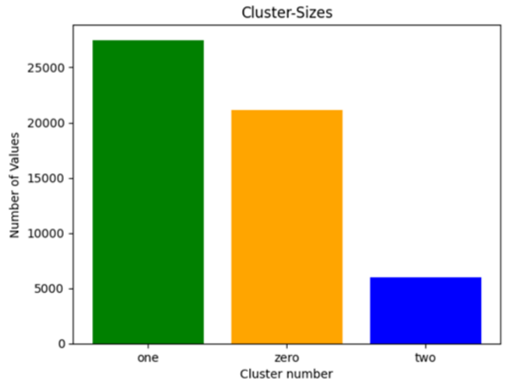


Figure 21: KPI 9 General ratio between registered and unregistered customers

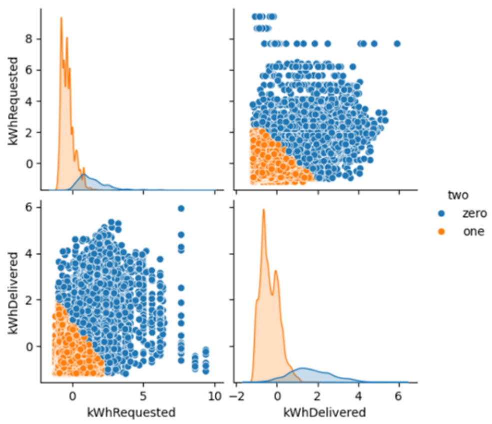
A.4 Clustering Results



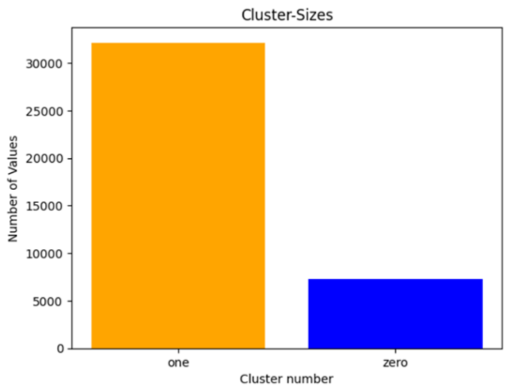
(a) Clustering progressInWeek - kWhDelivered



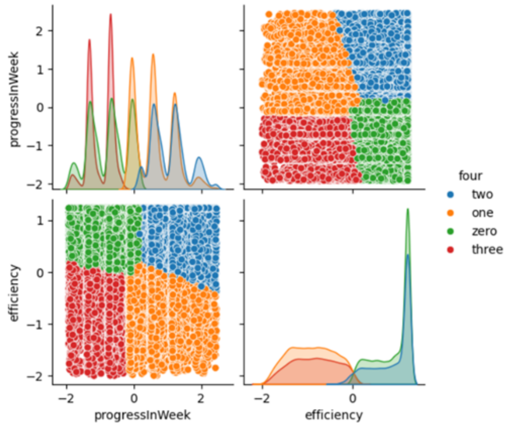
(b) Distribution efficiency - progressInWeek



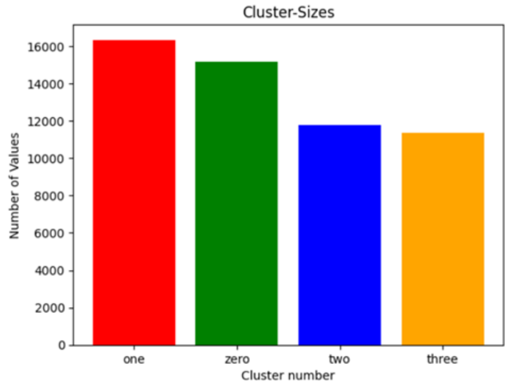
(a) Clustering kWhRequested - kWhDelivered



(b) Distribution efficiency - progressInWeek



(a) Clustering efficiency - progressInWeekk



(b) Distribution efficiency - progressInWeek

A.5 Predictive Analytics

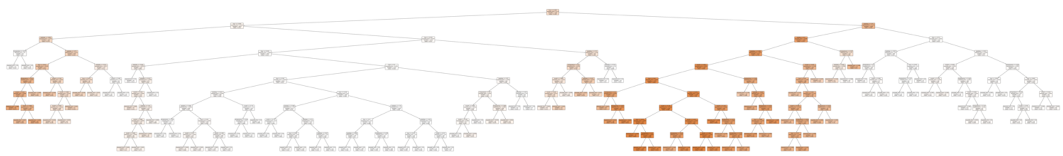


Figure 25: Decision Tree Site 1

Figure 26: Decision Tree Site 2

	Site 1		Site 2	
	MAE	R^2	MAE	R^2
Model_1	0.068	0.873	0.169	-0.51
Model_2	0.065	0.875	0.169	-0.51
Model_3	0.063	0.881	0.074	0.751
Model_4	0.069	0.869	0.170	-0.53
Model_5	0.094	0.835	0.160	-0.31

Table 4: Performance of different Neural Network architectures

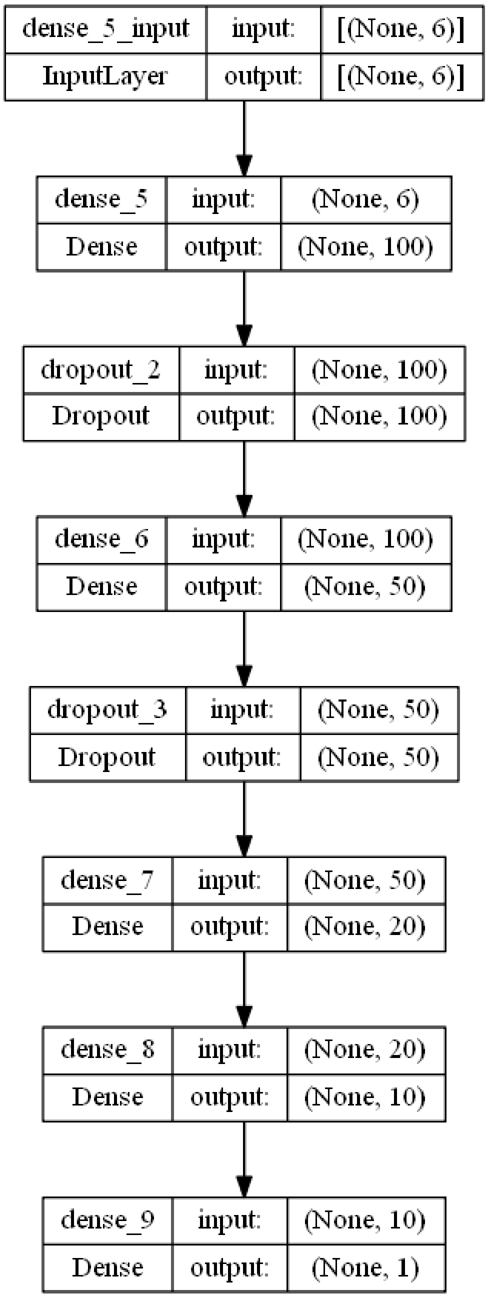


Figure 27: Visualization of the Neural Network architecture

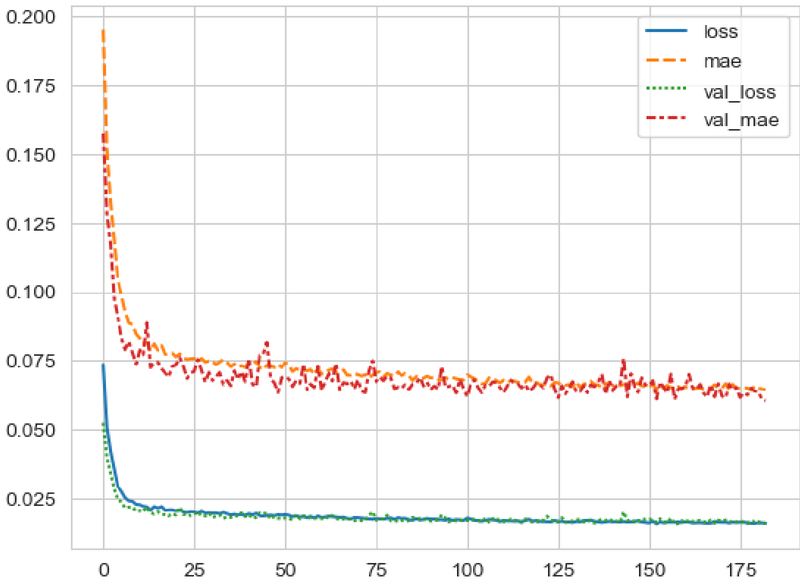


Figure 28: Training history of the neural net on site 1

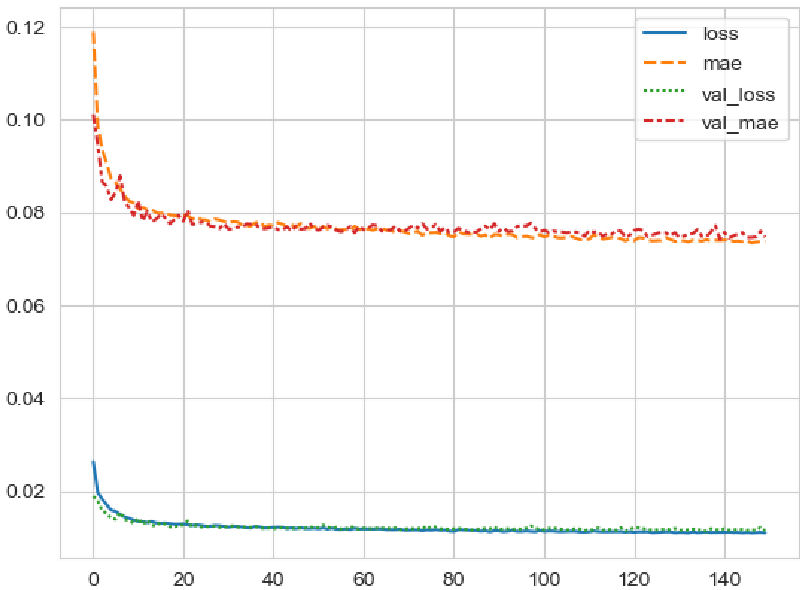


Figure 29: Training history of the neural net on site 2