AI-informed Holistic Electric Vehicles Integration Approaches for Distribution Grids

D3.3

Multilevel EVs flexibility services

May 2026 (M24)

Co-funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101160665.

DELIVERABLE DETAILS					
Project	AHEAD	Grant Agreement	101160665		
Deliverable Type (R, DEM, DEC)	R	Deliverable Title	Multilevel EVs flexibility services		
Status	Final	Month due	M24	Date	29/05/2026
Lead author	Ana Catarina Morais (IST ID)				
Co-author(s)	Sheikh Mostafa (IST ID), Lucas Pereira (IST ID), João Góis (IST ID), Lenart Ribnikar (EG), Blaž Dobravec (EG), Marzio Barresi (POLIMI)				
Peer reviewer(s)	Lisa Calearo (SPIRII)				
Dissemination level	PU - Public				

HISTORY OF CHANGES				
Document history	Date	Version	Submitted by	Comments
	23/02/2025	V0.1	IST-ID	Initial outline of the work.
	24/02/2026	V0.2	IST-ID	Compact Draft version shared with the participants of the task for comments.
	30/03/2026	V0.3	IST-ID	Draft version sent for internal review.

	15/05/2026	V0.4	IST-ID	Revision of the document based on comments made.
	29/05/2026	V1	POLIMI	Uploaded document.

LEGAL DISCLAIMER:

This project has received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No. 101160665.

EXECUTIVE SUMMARY

This deliverable aims to develop multilevel electric vehicles (EVs) flexibility services for grid integration to support planning and operations by grid operators. This is achieved through characterisation of flexibility services offered by EVs using AI-based clustering and forecasting methods applied to EV charging profiles and charging infrastructure. The methodology identifies patterns in charging behaviour and associated flexibility potential based on key session attributes (energy and idle time). It then forecasts their temporal occurrence for operational planning and decision-making.

Robustness is ensured through multiple AI-based methods. Clustering is performed using K-Means and BIRCH, while forecasting is based on Exponential Smoothing, and selected machine learning models. K-Means and Random Forest show the best overall performance.

The methodology is validated five EV charging datasets covering kerbside, residential, and workplace/campus charging, including real-world, hybrid, and synthetic data. For e-bus and vessel applications, only preliminary analyses are performed due to data constraints.

Results indicate clear differences in flexibility potential across charging contexts. Residential charging has the highest flexibility potential, followed by workplace and campus charging, and lastly kerbside. The later shows somewhat stable structural patterns across different geographical contexts.

The framework provides a scalable and transferable approach for EV flexibility assessment and forecasting, supporting system operators in grid and market integration.

The deliverable was made possible through the collaboration of all task participants (IST ID, DTU, EDP, SPIRII, KEMPOWER, AMINA, UPT, EG, POLIMI, and MOBIE), particularly those providing data (EG, MOBIE, POLIMI, ATM Group, SPIRII, and DTU).

Table of Contents

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Scope and Objectives

Energy systems have been transitioning from highly centralised structures to more decentralised networks with increasing integration of distributed energy resources (DER) [2]. The growth of renewable generation, the emergence of prosumers, and the rapid adoption of electric vehicles (EVs) introduce new challenges for distribution grids, including load uncertainty, potential peak increase demand, and infrastructure reinforcement needs [2–5]. At the same time, these developments create opportunities to enhance grid flexibility and support the transition towards smart grids. In this context, EV charging sessions have the potential to offers flexibility services of several natures. Naturally, in addition to charging settings, grid infrastructure conditions, and regulation, the flexibility that EVs can offer are inherently dependent on the behaviour of EV owners and their EV model choices [6]. Understanding and forecasting charging patterns is therefore essential to effectively assess and utilize EV flexibility.

Against this background, ***Task 3.3 - EVs flexibility clustering, characterisation, and classification*** of the AHEAD project has as main objective to develop multilevel flexibility services related to EVs grid integration for use in planning and operations by grid operators. This will be achieved through clustering, characterisation and classification of flexibility services offered by EVs, their charging profiles and charging stations via AI-based algorithms. Additionally, forecasting tools for operational planning will be developed. In other words, this deliverable uses machine learning forecasting techniques and clustering algorithms to predict where, when, and for how long EV charging sessions can potentially provide flexibility for ancillary grid services.

Note that AI is a broad discipline concerned with the development of intelligent systems capable of learning and problem-solving, whereas machine learning is a subfield that centres on data-driven approaches. Although machine learning falls under the broader umbrella of AI, both terms are used interchangeably in this report for convenience, given the data-driven nature of this analysis.

Authors such as [7] have proposed various flexibility modelling approaches based on the amount of available data, with some approaches incorporating parameters such as battery capacity and degradation rates. In practice, however, due to data privacy requirements and other constraints, charging session datasets often contain little to no information about EV owners or the vehicle models they use. As a result, the attributes of the charging session itself become the primary source of information for EV charging analysis. Accordingly, this deliverable aims to develop a comprehensive methodology to accurately group the flexibility potential of a session while minimising data requirements.

1.2. Relation to Other Deliverables and Work Packages

This work builds on knowledge from previous deliverables of the project as well as other ongoing EU-funded consortiums such as EV4EU, FLOW, and SCALE. Within the AHEAD project, the deliverable D3.3. basis its methodology on 1) the charging behaviour descriptive analysis from deliverable D1.1; 2) the outline of the actors and their roles on the e-mobility ecosystem of deliverable D3.1; and 3) the envelope representation of flexibility of deliverable D3.2.

The outputs of this deliverable can be used in subsequent tasks of this project, in particular:

- Forecasting and clustering profiles can be used as data inputs for *T3.4 – Flexibility simulation tools*.
- These flexibility profiles can also support the prequalification recommendations in *T3.5 – Flexibility contracting and pre-qualification*, by improving the predictability and reliability assessment of aggregated EV flexibility and identifying bottlenecks or market products where EV participation may be unsuitable.
- Integration of relevant parts of the methodology into the operational tools developed in *WP4 - Grid and spatial mapping models* and *WP6 - AI tool for planning of grids and charging infrastructure*.
- Due to its profiling component, the developed methodology, can potentially contribute the analysis of the results of the demo WPs (*WP7 – Nordic Europe demo*, *WP8 – Southern Europe Demo*, *WP9 – Western Europe Demo*).

1.3. Document Organisation

This document is divided in 5 sessions, including this introductory one. The reminder of the deliverable is organised as follows:

- Section 2 – Methodology: Outlines the main steps of the methodology, including data processing, clustering, and forecasting techniques employed.
- Section 3 – Data collection: Provides a characterisation of the datasets used to evaluate the methodology developed in this deliverable.

- Section 4 – Results and discussion: Presents the descriptive, clustering, and forecasting results for each dataset, followed by a discussion of the performance of the different models and a comparison of results across datasets.
- Section 5 – Conclusion: Summarises the key findings, limitations, and progress achieved.

2. METHODOLOGY

This methodology clusters EV charging sessions based on their flexibility potential, excluding timestamp variables such as parking start and end times. Such temporal timings are introduced afterward to profile the clusters and to forecast the number of sessions (active or new) per start hour. This approach keeps the analysis clear and allows meaningful insights from both clustering and demand prediction stages.

Given the central role of flexibility in this report, it is important to define it before detailing the methodological steps. In general, electrical flexibility refers to adjusting power generation or consumption at a specific time and location, within a limited duration, and without causing unplanned disruptions [3,10–12]. Previous literature, including the previous WP3 deliverable D3.2, represents this concept through feasible power trajectories P_t bounded by time-coupled envelopes $[P_t^{min}, P_t^{max}]$ consistent with State of Charge (SoC) dynamics and charger ratings.

Alternatively, because SoC information is not always available in datasets, this envelope can also be represented in terms of delivered energy. In this case, flexibility is represented as the cumulative energy between two time intervals, where the rate of energy change is limited by the maximum charging power [13–19]. Under the assumption that the maximum charging rate is known and that the SoC cannot fall below the initial SoC, the total flexibility potential of a charging session (v) can be visualised as the area of the parallelogram shown in Figure 1, or mathematically as the integral of the difference between the energy at the maximum charging rate and the minimum charging energy at time t between the arrival and departure times *(1)*:

$$\boldsymbol{Flex_v^{Energy}} = \int_{t_v^{arrival}}^{t_v^{departure}} \boldsymbol{E_{max}(t) - E_{\min}(t)\, dt} \qquad \textbf{(1)}$$

Which, when arrival and departure are known can also be expressed as the area of the parallelogram *(2)* from Figure 1:

$$\boldsymbol{Flex_v = Energy_{deliverad} * Idle\ time} \qquad \textbf{(2)}$$

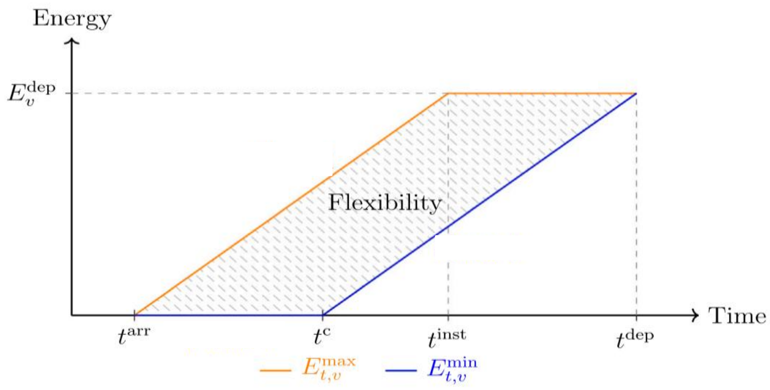


Figure 1 Flexibility envelope area.

The methodology was designed to analyse the total flexibility potential with minimal data requirements per charging session, making it easily adaptable to other research projects. This is especially relevant as datasets often lack detail or contain missing information [20,21]. For example, none of the datasets used in this deliverable included EV driver information, preventing a direct analysis of consumer behaviour patterns. Similarly, only one dataset, hybrid Madeira dataset contains information about the EV model.

To overcome these limitations, the proposed methodology focuses on core flexibility attributes of charging session: idle time and energy delivered for clustering, and time of the day for forecasting. Energy delivered and idle time can be calculated from timestamps, SoCs, charging versus parking durations, and power, increasing the general applicability of the approach. When available, additional variables such as maximum and average charging rates and Electric Vehicle Supply Equipment (EVSE) location are used for profiling purposes.

In contexts such as bus depots or ports, where charging is tied to a fixed set of routes and vehicles, additional considerations are required. Depot and port operations typically follow deterministic schedules, so it is important to assess whether the dataset is representative of the full set of routes and vehicles over a sufficiently long period. The data samples for these vehicle types in this deliverable, however, are small and include short or interrupted time spans, which can compromise the interpretation of results.

Although the main methodology was not fully tested due to limited entries, a K-Nearest Neighbour (KNN) approach for handling missing information in the Milanese dataset was applied. This approach provides preliminary insights into the potential flexibility of sessions even when idle time data is missing and represents a different application of clustering. Additionally, the results of the clustering step of the main algorithm of the methodology for both depot and port datasets can be found in the Results and Discussion section of this report.

The flowchart in Figure 2 illustrates the main steps of the methodology, beginning with the clustering stage and followed by the forecasting stage. The clustering step includes data cleaning and feature engineering, clustering modelling and evaluation, and the profiling of the resulting clusters. The forecasting step then involves constructing time series based on the obtained clusters, performing a preliminary analysis of the series, and running and testing forecasting models. The methodology was applied individually to each dataset. Although the main steps, such as clustering and forecasting modelling, remain the same across datasets, their different characteristics and varying levels of available information require different analytical and data processing options within the developed algorithm (Figure 2).

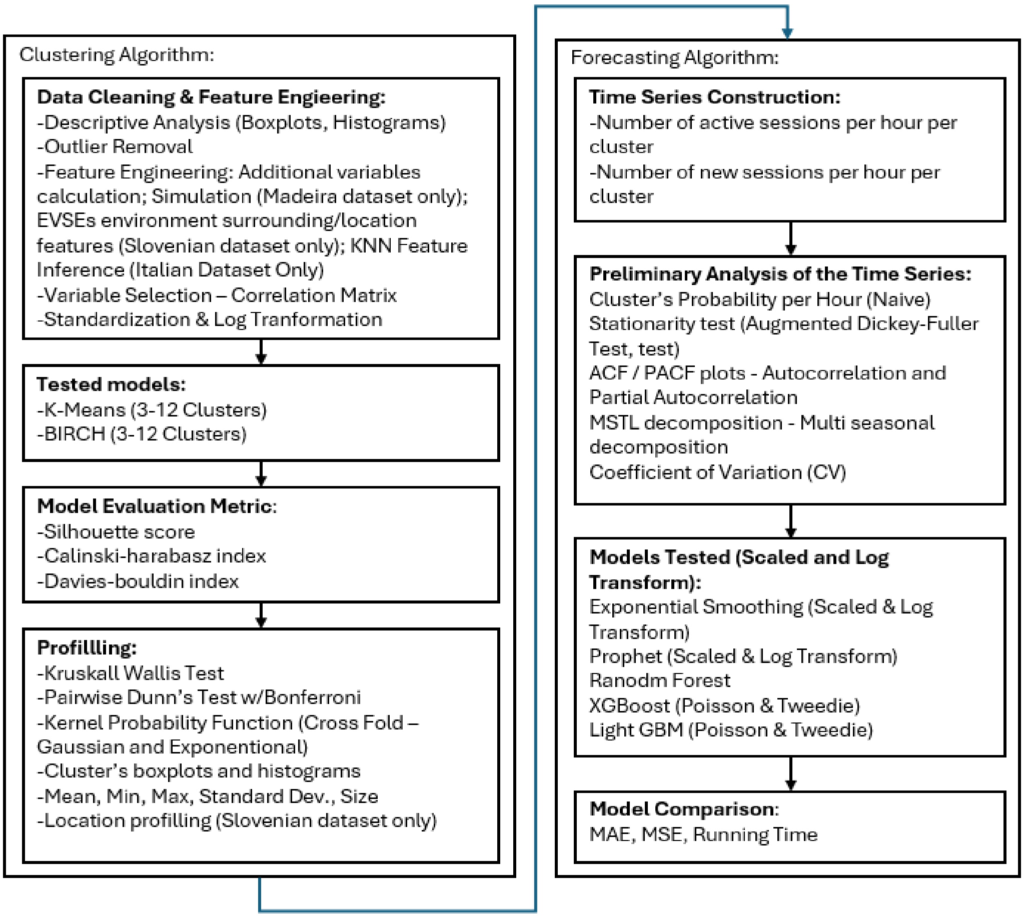


Figure 2 Methodology flowchart

2.1. Data Cleaning and Feature Engineering

Before clustering the datasets, to guarantee robust results, the data was cleaned and processed. The integrity of the charging sessions datasets was checked by looking for outliers, duplicated values, and missing data [22–24]. The outliers were calculated based on threshold values related to current battery sizes of EVs, type of chargers on the datasets, and the type of charging occasion. In addition to duplicated columns, the algorithm looked for the following outliers:

- Zero energy and charging times inferior to 1 minute.
- Parking durations that lasted more than 4 days. While longer sessions might be theoretically possible in some locations, it is unlikely that charging on kerbsides, for instance, would exceed this duration.
- Charging time above a defined threshold (24h residential / 18h others).
- Boundaries for energy delivered were established between 0.6 kWh and 100 kWh. The lower bound of 0.6 kWh corresponds approximately to the energy delivered to a Plug-in Hybrid Vehicle (PHEV), which, for example, would limit hybrid charging to about 3.6 kW for a 10-minute session. For e-buses and vessels the upper boundaries was superior.
- Exclusion of sessions where the average power delivered in greater than the maximum EVSE power (when the information is available).
- Exclusion of sessions with charging or total parking durations shorter than 10 minutes. For Type 2 chargers (240 V, 7.6 kW), it is highly unlikely that a vehicle would complete charging in less than 10 minutes. Moreover, energy markets typically operate in 10–15-minute intervals, making the flexibility of sessions shorter than 10 minutes negligible.
- Data entries with missing datetime features were not excluded if at least two other variables were available, as their values could be calculated or estimated. For example, if a session had start of parking and charging duration but no end of parking, the end of parking was assumed to coincide with the end of charging.

The above are pre-set limit values used to determine whether a data entry should be considered an outlier. However, the code allows users to adjust them if needed, for example, in the case of a vessel analysis. After, excluding outliers, additional energy/power and time features were calculated (list below), followed by data normalisation/standardisation. A logarithmic transformation was employed in skewed data cases [23–25].

- Average Session Power (when maximum power was not available): $E = P * Charging\ Time * Rate$
- COVID-19 dummy on datasets that comprise time windows between 2019 and 2021.

- Dummies for a national holidays and weekends, and overall days off (holiday plus weekends) were created.
- The following time-related variables were calculated:
  - Start/End Month, Year, Day of the Week.
  - Start/End Time in hours, seconds, and minutes.
  - Charging Time/Total Parking Time in hours, seconds, and minutes.
  - Cos/Sin time representation.
- Flexibility potential metrics:
  - Idle time.
  - Flexibility envelope area as per described in equation (2).

The process described above was applied to all datasets used in this analysis. Nevertheless, the Portuguese (EVs) and Slovenian datasets required additional processing steps. The datasets as well as their specific feature engineering steps are described in Section 3. In addition to session-level and vehicle-level variables, spatial context was incorporated into the dataset through features derived from OpenStreetMap (OSM). These features represent the area surrounding each EVSE within a certain distance, capturing the intensity and type of nearby urban activity, including the number of shops, amenities, leisure facilities, land use, road infrastructure, buildings, and religious sites.

By measuring these attributes within a 400 m radius of each charging station, the analysis captures how the surrounding area affects charging demand. Commercial and amenity density act as indicators of destination attractiveness, while highway proximity and road classification reflect accessibility and traffic exposure. Building density serves as a proxy for population concentration.

2.2. Clustering Methods

2.2.1. Clustering Models

Clustering algorithms are among the most widely used machine learning techniques across industries, being the energy and mobility sectors not an exception. Previous works on EVs and EVs charging behaviour have used clustering for profiling user behaviour, charging sessions profile, and defining new charging locations [6,25,27–29]. Moreover, authors have also reviewed and conceptualised the most common clustering techniques and their uses within EV adoption, charging behaviour, and EV flexibility [5,30,31]. In this task, two different methods were tested, the K-Means and the BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) algorithms. The K-Means algorithm was used, as it is among the most used clustering technique and has shown robust results in several study fields [5,30,31].

The second algorithm is the Birch hierarchical clustering, which is a technique designed to handle “big-data” datasets [32,33].

K-Means is a partitional algorithm that minimises the distances between data points and their respective centroids [5,30]. The centroids are not instances of the dataset and each data point is assigned solely to one cluster (hard clustering) [30].

BIRCH is a clustering algorithm designed to deal with scalability, running time, and memory issues of big-data datasets, while being robust to outliers and easily parallelised [32,34,35]. It is a multistep procedure for numerical data, that relies on Clustering Feature (CF) - Trees to aggregate the data into a small and compact summary structure much smaller than the original data. This condensed representation is then clustered instead of the original, large dataset [35].

2.2.2. Clustering Parameters and Model Selection

In addition to the models themselves, the current methodology tested the clustering methods for different numbers of groups (between 3 and 12 for both K-Means and BIRCH) and for different ways of quantifying flexibility, either using the envelope value (based on equation 2) or treating envelope variables separately. Each dataset was analysed separately. Four different variable variations were tested to determine the most accurate way of grouping the datasets according to their flexibility potential (Table 1).

Table 1 Definition of the algorithm configurations.

Configuration	Clustered Dataset	Variables Used for Clustering
Configuration 1	Full dataset	Energy and Idle Time (2 variables)
Configuration 2	Full dataset	Flexibility Envelope (Energy * Idle Time)
Configuration 3	Dataset without <10 min idle time sessions*	Energy and Idle Time (2 variables)
Configuration 4	Dataset without <10 min idle time sessions*	Flexibility Envelope (Energy * Idle Time)
*These sessions with an idle time smaller than 10 min were later added on the profiling and forecasting steps as a predefined “No Flexibility” group.		

The model and clustering configuration evaluation was done using the Silhouette score, Calinski-Harabasz index, and Davies Bouldin index [36]. These indexes measure the cohesion within the clusters and how far apart they are from each other. Specifically, the Silhouette score measures how similar each data point is to its own cluster compared to other clusters, with higher values indicating better separation. The Calinski-Harabasz index assesses the ratio of between-cluster dispersion to within-cluster dispersion, where higher values reflect more distinct and well-separated clusters. And finally, the Davies-Bouldin score quantifies the average similarity between each cluster and its most similar cluster, with lower values indicating better clustering performance. Complete information on how these metrics are calculated can be found on [36]. Code-wise these three scores are calculated by default, but an option to not calculate the Silhouette score is available, as it is the most computational expensive calculation. The algorithm identifies the top three optimal model-cluster configurations for each data sample. However, users, have access to a table with the scores of all the variations and can choose to keep running the algorithm with an alternative setting.

2.3. Profiling

During the profiling step, the clusters were tested for statistical significance using the Kruskal-Wallis test, followed by Dunn's pairwise test with Bonferroni correction to identify whether and which clusters were significantly different.

In addition to visualisation tools, mean, maximum, minimum, and standard deviation, a Kernel Density Estimation (KDE) algorithm was used better describe the main features of each (energy, charging tine, parking time, energy, location - Slovenia, etc.). Both exponential and Gaussian probability density functions were tested, and an h parameter tunning was done through cross-folding.

When additional geospatial information about the surroundings of the EVSEs was available, it was utilised to further profile the clusters. These variables were analysed at the session level rather than at the individual EVSE level, since the clusters are based on session flexibility. Note that the same EVSE can host sessions from different clusters, however the goal is to understand where charging sessions with relevant features occur, not to classify locations themselves. Session-weighted totals and Z-Scores of the clusters were used to show where distinct flexibility potential levels are most common, and highlight which surroundings are represented in each cluster.

Finally, the naive probability of a certain cluster having an active/new session per hour and day of the week was calculated. These probabilities can be calculated based on the total timespan of the dataset or based on a defined time interval (e.g., last two years).

2.4. Forecasting

After clustering, the methodology presented in this deliverable (Figure 2) includes a forecasting step. Accordingly, this section briefly describes the forecasting methods used in this research. This forecasting methodology follows a bottom-up approach, meaning that the total number of active/new charging sessions is the sum of the sessions of each cluster. The analysed data had an hourly frequency, with the goal of producing hourly predictions for the next 24 hours. The size of the analysed time window was the same across datasets (three months), and when relevant, sliding windows of roughly two months were used. This window length was chosen as a balance between short-term variations and long-term trends. Thus, providing enough data for stable predictions without being overly sensitive to noise while remaining responsive to recent changes in charging behaviour. The use of sliding windows is consistent with established time series forecasting practices [37,38].

To find the most robust method, and similarly to what occurs on machine learning research and competitions [39–42], different algorithms were tested on the different clusters charging time series. Five different models were tested, an exponential smoothing (statistical classical model) [39], three machine learning algorithms (Random Forest [43], XGBoost [44], and Light GBM [44]), and Prophet, a deep learning Generalised Additive Model (GAM) developed by Facebook [45].

2.4.1. Preliminary time series investigation

Before running the machine learning models mentioned earlier, a preliminary investigation of the time series was done. This analysis allowed a better understanding of data characteristics such as seasonality, trend, and stability, which in turn are relevant when defining window features, lags, and other parameters of the machine learning models. The following test/analysis were performed: 1) Stationarity test (Augmented Dickey-Fuller Test); 2) Autocorrelation and Partial Autocorrelation plots (ACF, PACF) in which a second derivative was used to suggest the recommend lags and seasonality in addition to the initial plot interpretation; 3) Multi seasonal decomposition (MSTL) plot; and finally a 4) Coefficient of Variation (CV) as suggested in [38].

2.4.2. Forecasting models

Exponential Smoothing is a classical statistical approach when it comes to time series. Future values are recursively predicted based on weighted average of past observations. These weights decrease exponentially as the data gets older, thus giving more importance to recent patterns. While the basic model does not account for trends or seasonality, there are exponential smoothing model variations capable of dealing with such components of the time series. Different smoothing methods are available, offering flexibility in the way these aspects of the time series are included in the model (e.g., additive, damped, or multiplicative). When compared to other statistical benchmarks such as the Autoregressive Integrated Moving

Average (ARIMA) family, exponential smoothing is faster and, in some cases, has outperformed machine learning models [39,40,46].

As a widely used method, there are several Python packages calculating this model and its variations. In particular, ETS is an algorithm which automatically selects the model optimal parameters that has been extensively used in forecasting academia over recent years [39].

Prophet is a fast-performing Deep Learning GAM. Due to the GAM formulation, it decomposes easily and accommodates new components of the time series [45,47]. It as an automatic, flexible model that does not need regularly spaced measurements. It also has already embedded on its algorithm the capacity to deal with holidays as well as daily, weekly, monthly, and yearly seasonality [45,48]. Authors such as [47] and [48], have used this model on EV charging datasets.

Random Forest (RF) is an extension of the bagging method that combines the output of multiple decision trees by averaging them to improve accuracy and reduce bias [49].This method, that has been utilised in several EV related studies [43] holding robust results, utilises both bagging and feature randomness to create an uncorrelated forest of decision trees [49].

The last type of machine learning models tested were the gradient boost ensemble algorithms XGBoost and Light GBM. XGBoost stands for eXtreme Gradient Boosting while Light GBM means Light Gradient Boosting Machine. Both regress gradient boosted trees in a non-linear, sequential way [40,44]. On one hand, Light GBM is a popular method on machine learning competitions because it only expands to the best performing leaves making it faster than other GBMs and can handle multiple features of different natures without pre-processing [40].On the other hand, instead of a leafwise tree growth approach, XGBoost uses a level-wise optimisation strategy [49]. According to [40,44,49], while more computationally expensive, this model is high performing and efficient, having shown robust results on EV flexibility forecasting studies. These models allow the use of different optimisation functions, in this work, Poisson and Tweedie were tested. They were selected for their ability to handle possible sparsity and zero inflation of hourly flexible charging counts [50–52].

These models where compared on their running time as well as on their Mean Absolute Error (MAE) and Mean Squared Error (MSE), two mainstream metrics for model comparison.

The forecasting models were run in Python recurring to two main packages. ETS and Prophet were run with the assistance of the DARTS package [53], and for the other methods the SKForecast package [54] was used. Since DARTS offers optimisation features to the traditional exponential smoothing model, a cyclic encoder and holiday feature were added to the model parameters in the additive version of the method.

2.5. Code Structure

The methodology was developed in Python following a package-like structure. A version of the code with its application to the open access datasets is available in AHEAD project Git repository [54]. The deliverable package is divided into two main folders (Figure 3), namely: 1) Flexcluster which contain the modules related to the clustering algorithm and initial data modulation; and 2) Flexforecast that comprises the modules related to the forecasting algorithm and time series transformations.

```
evflex_33/
│
├── main.py
│
├── flexcluster/
│   ├── __init__.py
│   ├── clust_step.py
│   ├── preprocess_utils.py
│   └── visuals.py
│
├── flexforecast/
│   ├── __init__.py
│   ├── fcast_utils.py
│   └── forecast_step.py
│
├── old/
│   └── (old version of the code)
│
├── data_inputs/
│   └── (charging sessions input files)
│
└── data_outputs/
    └── (output files: excel and pickle files)
```

Figure 3 Deliverable python code structure.

Flexcluster is divided into three modules (Figure 3):

- [Visuals] has all the functions used for the pre-processing and profiling visualisation.
- [Preprocess_utils] contains all the data cleaning and features engineering functions as well as utils such as time conversions and holiday calendar.
- [Clust_Step] contains core function to clustering and clustering evaluation.

Flexforecast is divided into two modules (Figure 3):

- [Fcast_utils] has functions that support the forecasting algorithm.

- [Forecast_step] contains core functions to forecasting and forecasting evaluation.

3. DATA COLLECTION

To test the robustness of the methodology proposed on the following section, several datasets were considered with different locations, characteristics, vehicle types, and timeframes. In total, the proposed clustering and forecasting methodology was tested on five EV charging datasets, three of them concern kerbside data, one semi-private (office/university), and another one is based on residential charging profiles. From the tested datasets, two are US, real-data open-source resources (Caltech campus and national research lab – California [8] and Boulder – Colorado [9]), and the other three are from European countries (Denmark, Slovenia, and Portugal). Both Danish and Portuguese datasets rely on simulated or hybrid data.

For the Portuguese EV-charging dataset, parking duration, EVSE charging power, as well as parking start and end timestamps are empirical variables, while total energy charged and charging duration are derived through a simulation constrained by the real-world inputs (as summarised in Section 3.1). Because total energy and charging duration are key features in the clustering step, it is important to clarify that the objective of the clustering analysis, for this dataset, is methodological rather than interpretative. In particular, clustering is applied to evaluate whether structurally coherent patterns emerge under controlled but physically realistic charging assumptions. Regarding the forecasting step, because the resulting groups reflect structurally feasible charging intensity patterns, forecasting is used to evaluate whether behavioural segmentation improves the modelling of real-world charging demand.

Regarding the Danish dataset, although the dataset is simulated, it is generated using probabilistic functions derived from empirical data - described in Deliverable 1.1 - and therefore preserves key behavioural characteristics. Thus, the clustering profiles can be meaningfully analysed in addition to supporting methodological validation. Forecasting, however, is included primarily for methodological validation, as the simulated daily volumes are not necessarily representative of real-world demand.

Additionally, since e-buses and vessels are included in the AHEAD project demos, a preliminary analysis of the flexibility potential of an e-bus charging depot in Milan, along with a vessel charging simulation at the Madeira port (Table 1). The full methodology is not applied due to the current incompleteness of these datasets, which is expected to be resolved as the demos progress. In particular, the forecasting step of the methodology was not performed due to small sample size, sparsity of timestamps (Milan), and short time span (one week for

Madeira B). These limitations would prevent a reliable interpretation of the results. For the e-bus depot, given the high ratio of missing data relative to the full sample, a dataset-specific feature engineering process is also documented to understand possible viable routes for flexibility analysis in such circumstances.

In the case of the Madeira port, the dataset consists of one week of simulated maritime touristic operations, assuming 10 electrified vessels. Each vessel departs twice per day during weekdays on a fixed schedule, in the morning and early afternoon. The simulation considers a plug-and-charge scenario, assuming there is enough installed capacity at the port to meet the charging demand. For more information, the reader is referred to D9.2: *Port Digital Twin – Western*.

The datasets include different charger types with varied capacities. The Slovenian and ACN datasets include both AC and DC charging, whereas the Danish residential, Madeira electrified vessels, and Boulder datasets contain only AC chargers (<22 kW). The Milanese e-Bus depot dataset is the only one with exclusively DC chargers, which is unsurprising given the battery sizes of e-buses. Although the Madeira EV sample includes chargers within the dataset, it is limited to AC charging power. DC chargers are usually connected to medium-voltage (MV) grids, whereas residential chargers are directly connected to low-voltage (LV) grids. This has further implications for flexibility activation and the types of services that can be provided.

Table 2 Main characteristics of the analysed datasets.

Dataset/Location	Charging Sessions	Start/End Dates	Main Charging Occasion	Available Attributes
Adaptive Charging Network (ACN): Caltech campus, research lab – California (US)	65,062	2018.04.25 – 2021.09.13	On campus, study, work	Object Id; User Inputs; Session Id; Station Id; Space Id; Site Id; Cluster Id; Start Park (connect time); End Park (disconnect time); Energy; End Charge; Time zone; User ID.
City of Boulder – Colorado (US)	148,136*	2018.01.01 – 2023.11.30	Kerbside/public charging locations	Station Name; Address; City; State/Province; Postal Code; Start Park; Start Time Zone; End Date Time; End Park; Total Parking Duration; Charging Time; Energy; GHG Savings; Gasoline Savings; Port Type; Object Id.

Dataset/Location	Charging Sessions	Start/End Dates	Main Charging Occasion	Available Attributes
Slovenia (across the whole country)	65,190	2025.01.01 – 2025.07.06	Kerbside/public charging locations	Object Id; Start Park; End Park; End Charging Time; Total Parking Duration; Energy; Session Status; Payment Status; EVSE Id; EVSE Type; Latitude; Longitude; Charger Max Power; Foreign Location.
Madeira Island – Portugal (A)	12,691	2025.04.10 – 2025.06.28	Kerbside/public charging locations (hybrid dataset)	EVSE Id; Location Id; Connector Standard; Connector Format; Max Voltage; Max Amperage; Charger Max Power; Start Park; End Park; Terminated by the User; Total Parking Duration; Duration Formatted; Coordinates; Tariff Type; Tariff Price; SoC Start; SoC End; Energy, Start Charging Time; End Charging Time; Start Time Day; End Park Day; Session Id; EV Model.
Denmark	345,851	2025.01.01 - 2025.12.31	Residential (simulated)	Case; EVSE Id; Object Id; Start Park; Start Charge; End Charge; End Park; Energy; EVSE Max Power; Total Parking Duration; Charging Time.
Milan – Italy	1,831 (285**)	2023.10.19 – 2024.07.19 (8 interrupted intervals)	Depot e-Bus charging (85 buses)	Observation Id, Vehicle/bus Id, Event, Start Charge Time; End Charge Time, SoC Start, SoC End, Charger Type; Charger Max Power, Energy, Departure from Depot**, Return to Depot**, Service Duration**, Expected Delta SoC**; Min Delta SoC**; Max Delta SoC**; Observed Delta SoC**
Madeira Island – Portugal (B)	6,720 15-min timestamps (122 sessions)	2025.09.01 - 2025.09.07 (1-week)	Port vessel charging (simulated, 10 vessels)	Boat Id, Timestamp, Charger Max Power; Charger Effective Power; SoC %; State
*This value contains several duplicated sessions; **Only 285 out 1832 entries have information on **features				

The Milanese data collection tracks the charging sessions and service shifts of a sample of e-buses operating in the city. The dataset covers eight observation periods collected between October 2023 and July 2024 (Table 2, Table 3). Most observation windows span only three consecutive operating days, with the exception of the combined period from April to July 2024, which includes six monitored days. Across all observation periods, the dataset covers 27 unique operating days, with 44–67 distinct e-bus routes observed (up to 85 bus IDs). In total, 1,831 charging sessions were recorded, with individual periods capturing 112–315 events depending on fleet activity (Table 3). Only 285 sessions contain sufficient data to calculate idle time-related variables (Table 2).

Table 3 Summary of observation periods in the e-bus charging dataset.

Observation ID	Observation period	Unique days	Unique buses	Charging sessions
20ott23	19–21 October 2023	3	48	161
22ott23	21–23 October 2023	3	44	112
14gen24	13–15 January 2024	3	63	203
15gen24	14–16 January 2024	3	63	234
17apr24	16–18 April 2024	3	67	315
21apr24	20–22 April 2024	3	66	232
14lu24	April and July 2024 operating days	6	62	315
18lu24	17–19 July 2024	3	62	259
8 (Total)		**27 (Total)**	**Max 67, Min 44**	**1,831 (Total)**

3.1. Madeira (EVs) Data Simulation

The EV charging session records collected in Funchal (Madeira) provide temporal boundaries for each connection event and the rated power of the associated charging point. However, they do not include energy delivered, charging curves, or SoCs. These gaps make it impossible to directly compute minimum charging duration, idle time, or the flexibility envelope of a session. Considering these limitations, and so the methodology can be tested, a simulation was necessary. The simulation comprises three different types of features, namely: 1) variables derived directly from the dataset; 2) Features introduced to represent realistic EV characteristics; and 3) Probabilistic and assumption-based features.

The first kind of variables are taken directly from the observed records and include arrival and departure timestamps, as well as the maximum charging power of the EVSE. These features define the connection window and the upper power limit for each charging session. In other words, they form the temporal and technical boundary conditions for all later calculations.

Subsequently, the following technical characteristics were calculated: 1) Battery capacity (kWh); 2) Onboard AC charger rating (kW); 3) AC charging power limit (kW); and finally, 4) Charging power as a function of SoC, defined over the 0 to 100 percent SoC range. To approximate real-world vehicle behaviour, a market weighted sampling approach was adopted to calculate such variables. EV models were sampled according to their relative presence in the Portuguese market [26], with consideration of regional relevance for Madeira. The simulated fleet consisted of ten representative vehicles, including five Battery Electric Vehicles (BEVs) and five PHEVs.

Only AC charging was modelled in this dataset. Therefore, manufacturer specified AC onboard charger ratings were used to constrain the maximum effective charging power. For each session, the effective charging rate was determined by the minimum of the EVSE rating and the vehicle onboard AC limit.

Since the dataset does not include battery state information, additional variables were generated using probabilistic assumptions (third variable type). The assumptions are:

- SoC Initialisation: The arrival SoC was sampled from a predefined distribution assuming typical public charging behaviour. Target SoC levels were also defined to reflect common user charging patterns.
- Charging Power Approximation for PHEVs: Plug-in hybrid vehicles were modelled with simplified charging power behaviour. Given their smaller battery capacities and typical charging characteristics, an approximate SoC dependent charging profile was applied.

These assumptions allowed the reconstruction of energy demand and charging duration for each observed connection window, enabling the calculation of flexibility envelopes.

3.2. Slovenian Location Surrounding Features Calculation

In addition to session-level and vehicle-level variables, spatial context was incorporated into the dataset through features derived from OpenStreetMap (OSM). These features represent the area surrounding each EVSE within a certain distance, capturing the intensity and type of nearby urban activity, including the number of shops, amenities, leisure facilities, land use, road infrastructure, buildings, and religious sites.

By measuring these attributes within a 400 m radius of each charging station, the analysis captures how the surrounding area affects charging demand. Commercial and amenity density act as indicators of destination attractiveness, while highway proximity and road classification reflect accessibility and traffic exposure. Building density serves as a proxy for population concentration.

3.3. Milan (e-Bus) Data Processing and Considerations

In the operation of e-bus systems, the timing of charging activities relative to scheduled service plays a central role in ensuring both efficiency and reliability. To capture this temporal margin, two complementary measures are defined: before flexibility and after flexibility. Together, they represent the available buffer time around each charging event.

Before and after temporal flexibility measure the idle time of an e-bus at the depot surrounding a charging session. Before (idle time) flexibility is the interval between the return of a bus and the start of charging. After (idle time) flexibility is the interval between the end of charging and scheduled departure, indicating the slack available for schedule adjustments or vehicle reallocation.

The Milanese dataset contains significant information gaps. Charging sessions are unevenly distributed across vehicles. Some buses have many observations, while others have very few, suggesting that the sample may not fully represent depot operations. Moreover, only a small portion of the data includes the minimum information required to estimate the flexibility envelope (Equation 2). The time before and after charging is rarely recorded within the same operational cycle, and one of these components is often missing. As a result, total idle time and flexibility envelopes cannot be directly computed for most observations. Consequently, data pre-processing is essential, including filtering incomplete records and addressing missing values, to avoid biased results and ensure a more representative analysis of fleet behaviour.

To address missing and incomplete information, a feature selection process was carried out using correlation analysis to identify variables most relevant to charging temporal flexibility. Two correlation matrices were analysed. One focused on battery and charging characteristics (e.g., state of charge and energy delivered), and the other on temporal variables (e.g., start time, end time, and calendar features). Based on these results, a subset of variables capturing both operational and temporal patterns was selected for modelling.

A KNN-based approach was then applied to estimate missing temporal flexibility (before and after idle time) values by identifying charging sessions with similar characteristics based on selected features, under the assumption that similar sessions will exhibit comparable idle time patterns. The optimal number of clusters used to structure the data was initially evaluated using the silhouette method. Nevertheless, a balance between cluster quality and data availability led to the selection of a reduced number of clusters, ensuring sufficient observations within each group. This approach provides a reliable imputation of missing temporal flexibility (idle time) values and demonstrates an alternative use of clustering to study EV-based flexibility potential, even with incomplete operational data.

4. RESULTS AND DISCUSSION

This section presents the main results for each dataset and discusses the similarities and differences between them. The results are divided into three parts: descriptive analysis, clustering results, and forecasting results. In the clustering part, the cohesion and separation quality of the clusters obtained in the different model-cluster configurations is evaluated using the scores described earlier. This is followed by the profiling of the groups obtained under the selected optimal clustering settings. For the forecasting results, a brief preliminary analysis of the time series is presented, after which the MAE values obtained for the different model settings are analysed.

4.1. Boulder Dataset

4.1.1. Descriptive Analysis

As mentioned earlier this is an open access kerbside dataset from Boulder in Colorado (US). 50% of the EVs were parked for roughly 1h 47min, being that 1h 37min was for charging. Only about 12% of the charging sessions had an idle time superior to 1 hour and an additional 5% had idle times between 30 minutes to 1 hour.

On average, 9.83 kWh were charged per session with energy consumption ranging from 0.60 kWh to 92.61 kWh. Regarding power, the larger number of sessions had a charging rate of around 7.4 kW, followed by 3-4 kW and > 8 kW.

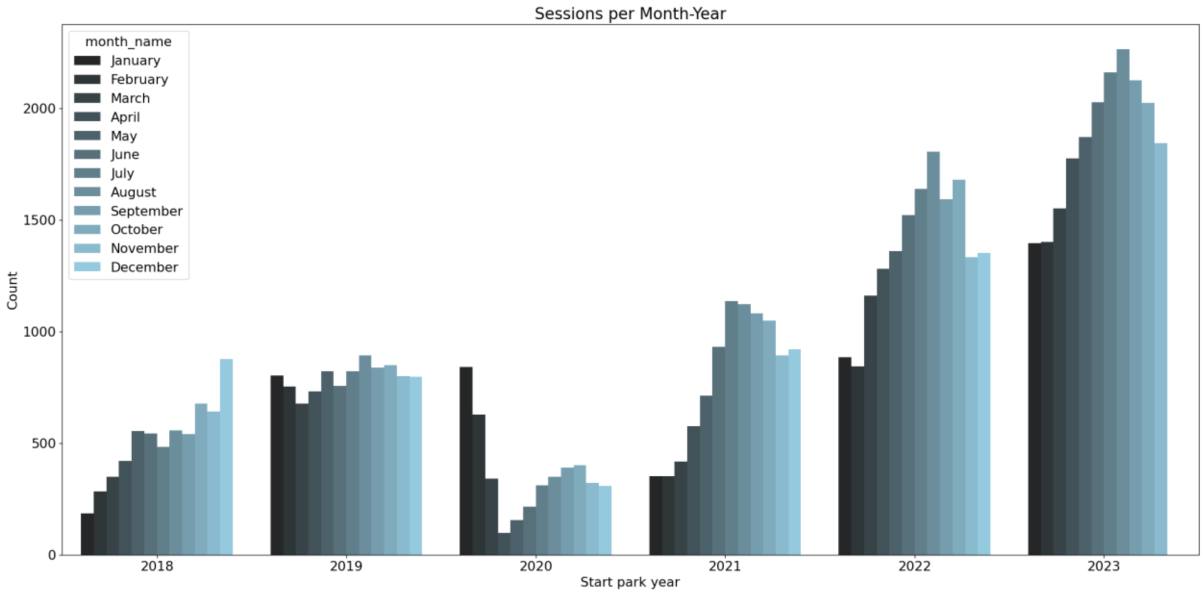


Figure 4 Number of sessions per month and year (Boulder).

As visible on Figures 4, 5 a), and 5 b), the number of sessions rapidly increased after COVID-19, with summer months summing having a higher number of daily sessions, on average. There seems to be a larger number of sessions of charging sessions on Friday (int 4) and Saturday (int 5), however, overall, there is not a pronounced difference between the number of charging sessions on a day off and on a working day.

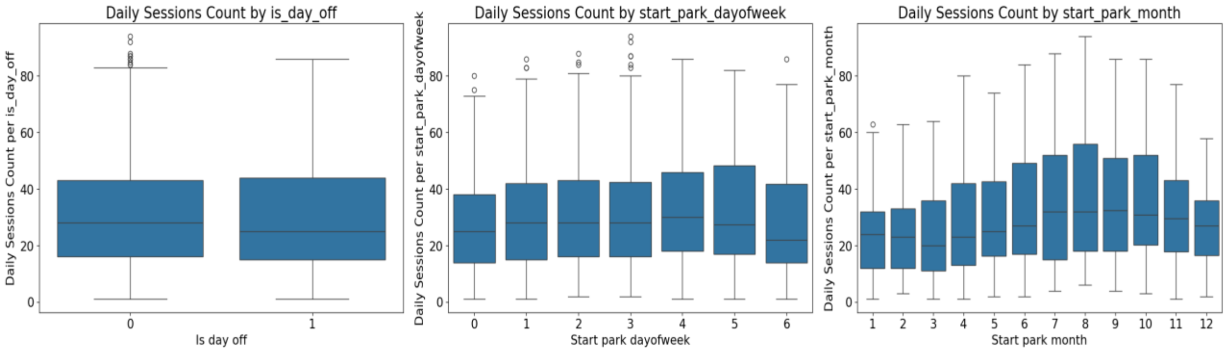


Figure 5 Boulder: (a) Daily sessions boxplot per day off dummy; (b) boxplot per day of the week; (c) boxplot per month.

The sessions seem to occur mostly between 8h and 19h, with a big overlap between the arrival and departure time (Figure 6). This is aligned with the reduced idle time of the sample. After the morning peak slump, the number of sessions rises around lunchtime (12:00–13:00) and increases again at the end of the day (18:00–19:00). During the night and early morning, only a small number of sessions start and end (Figure 6).

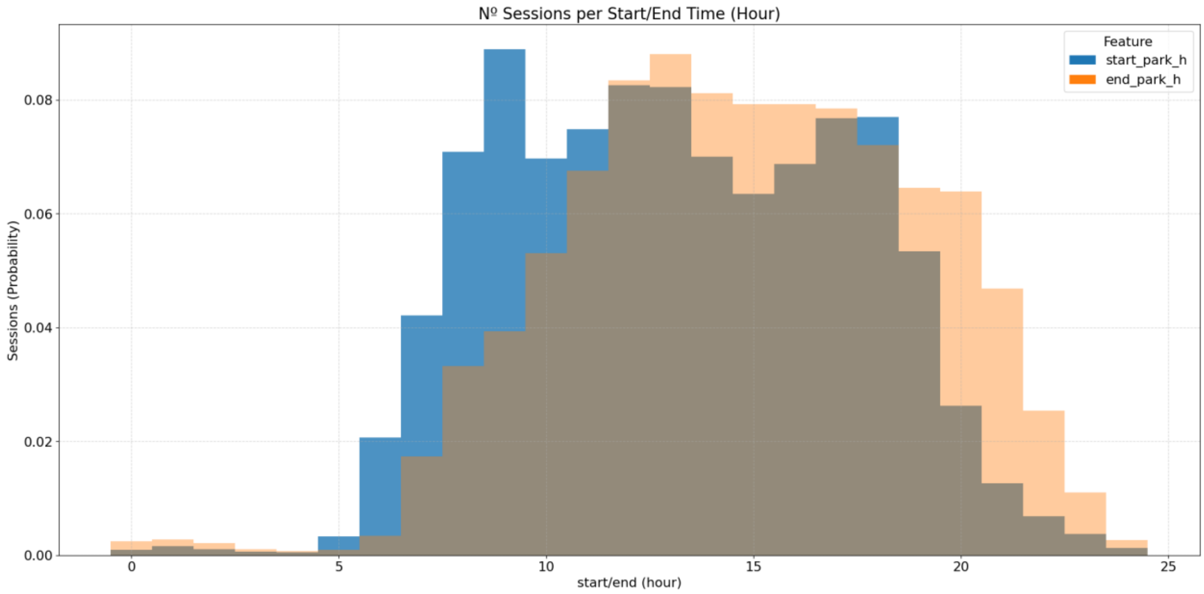


Figure 6 Frequency of sessions per hour (Boulder).

4.1.2. Clustering Results

After feature selection, different clustering configurations were explored. Among these, the configurations using the single-variable flexibility envelope showed higher Silhouette scores and Calinski-Harabasz indices, as well as lower Davies-Bouldin indices, indicating better-defined clusters than the remaining configurations (Table 3). Configuration 1 performed better than Configuration 4, but worse than Configuration 2. However, the scores of Configuration 4 were calculated on a sub-sample of Configuration 2, including only sessions with idle times greater than 10 minutes. When the indices were recalculated for the full sample for Configuration 4, Configuration 2 still showed better overall performance. Therefore, Configuration 2 with four clusters was selected for the subsequent stages of the methodology.

Table 4 Cluster evaluation (Boulder).

Clustering Model	Parameters	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index	Configuration
K-Means	3	0.85	374,744.37	0.43	Configuration 2
K-Means	**4**	**0.84**	**477,972.54**	**0.46**	**Configuration 2**
K-Means	5	0.83	553,919.06	0.48	Configuration 2
K-Means	**3**	**0.57**	**48,473.57**	**0.52**	**Configuration 4**
K-Means	4	0.55	55,042.32	0.53	Configuration 4
K-Means	5	0.54	65,577.54	0.54	Configuration 4
BIRCH	(3, 0.2)	0.50	21,695.11	0.80	Configuration 1
BIRCH	(4, 0.2)	0.48	22,458.84	0.86	Configuration 1

From the total sample, 51,747 sessions (most of the dataset) were allocated to Cluster 1 – No Flexibility (Figure 7a). This cluster has no flexibility potential because their idle time is, on average, zero making its average flexibility envelope value zero too. The second largest cluster was Cluster 0 – Minimal Flexibility. With 7,156 sessions, this cluster averaged 9.26 kWh of energy delivered per session and 40 minutes of idle time (Figure 7, Table 5). Cluster 2 – Mid Flexibility had 4,776 sessions with an average idle time of almost 2h20 min. Finally, Cluster 0 – High Flexibility had the highest amount of energy delivered as well as the longest idle time (Figure 7). This is the smallest cluster with 2,774 sessions. Note that this algorithm does not assume the discharging bellow the initial SoC is possible, meaning that, in principle more energy results in more flexibility.

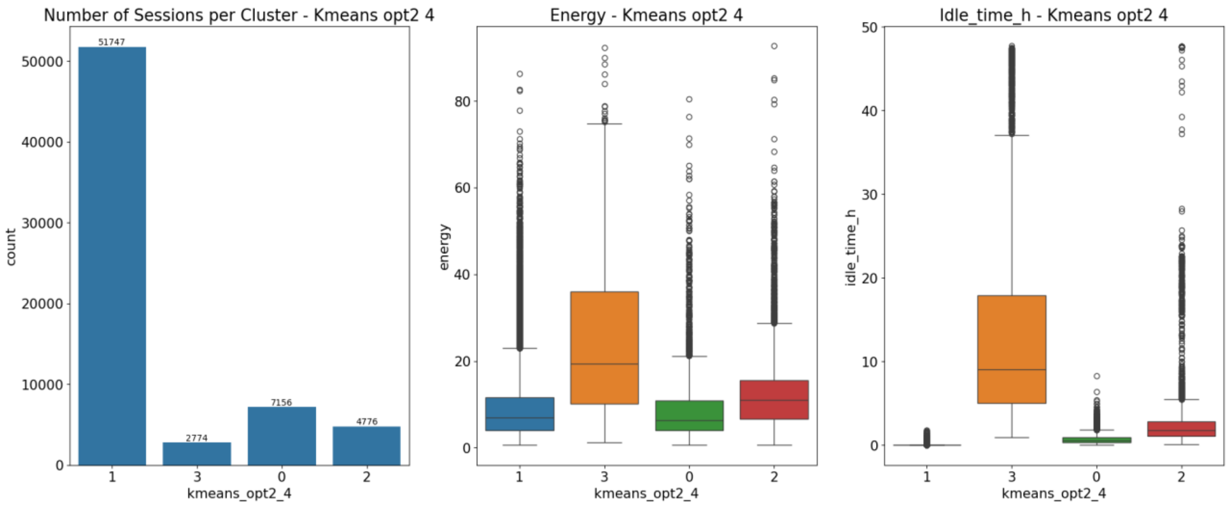


Figure 7 Boulder: (a) Clusters' size; (b) Energy clusters' boxplots; (c) Idle time clusters' boxplots.

Regarding energy, Cluster 0 – No Flex and Cluster 1 – Min Flex correspond to sessions with less energy delivered (Table 4 and Fig. 7b). In terms of total duration, both Cluster 2 and Cluster 3 include longer sessions, whereas Cluster 0 and Cluster 1 (Table 4) tend to be shorter. The Kruskal-Wallis and subsequent Dunn's test with Bonferroni correction showed the clusters were statistically different between each other.

Table 5 Clusters' profiling (Boulder).

Clusters	Variable	Min	Max	Mean	Std	Kurtosis	Skewness
0	AVG Charg. Power (kW)	0.34	9.17	5.39	1.85	-1.29	0.13
1	AVG Charg. Power (kW)	0.33	10.26	6.43	1.78	-0.82	-0.67
2	AVG Charg. Power (kW)	0.20	8.98	5.75	1.75	-0.94	-0.23
3	AVG Charg. Power (kW)	0.58	9.01	5.70	1.90	-0.85	-0.45
0	Charging time (h)	0.11	13.49	2.11	1.42	9.26	2.37
1	Charging time (h)	0.10	15.36	1.82	1.40	9.12	2.31

2	Charging time (h)	0.23	15.65	3.00	1.83	5.80	1.95
3	Charging time (h)	0.26	15.97	5.68	2.89	0.48	0.88
0	Energy (kWh)	0.64	80.46	9.26	7.97	12.75	2.90
1	Energy (kWh)	0.60	86.30	9.45	8.00	7.60	2.22
2	Energy (kWh)	0.64	92.69	13.93	10.40	6.99	2.22
3	Energy (kWh)	1.25	92.25	27.02	17.70	-0.14	0.75
0	Idle time (h)	0.02	6.39	0.67	0.58	9.86	2.43
1	Idle time (h)	0.00	1.80	0.02	0.07	161.38	10.85
2	Idle time (h)	0.11	47.20	2.30	2.79	77.39	7.05
3	Idle time (h)	0.90	47.74	12.43	11.84	1.85	1.69
0	Total duration (h)	0.91	13.55	2.78	1.26	11.04	2.55
1	Total duration (h)	0.17	15.36	1.84	1.40	9.22	2.32
2	Total duration (h)	2.26	48.00	5.29	2.86	58.27	5.65
3	Total duration (h)	6.21	48.00	18.10	11.47	1.77	1.70

The cluster intensity throughout the day also varied (Figure 8). Clusters 1 and 0 more closely resembled the overall sample distribution, with Cluster 1 - Min Flex being slightly more concentrated in the morning than in the afternoon. Cluster 2 - Mid Flex shows the highest probability of occurring during the morning, particularly between 7:00 and 10:00. In contrast, Cluster 3 - High Flex is more frequent in the afternoon, with the most probable arrival time occurring around 17:00.

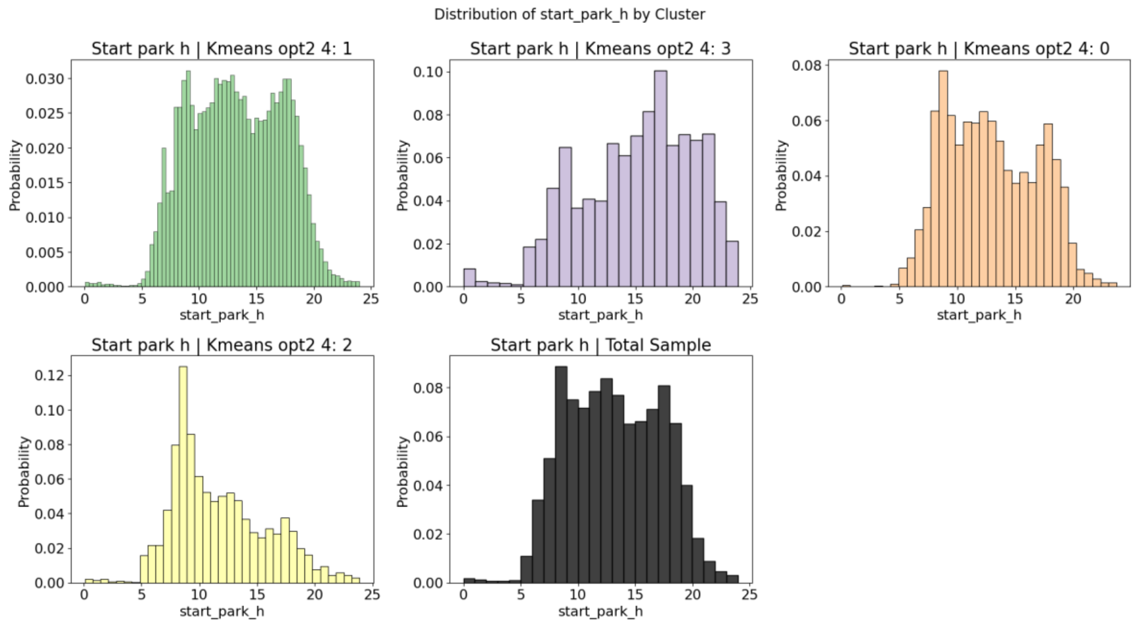


Figure 8 Start parking hour per cluster (Boulder).

4.1.3. Forecasting Results

After clustering the charging sessions of Boulder, Colorado, a time series forecasting analysis, in which several methods were tested, was conducted. The forecasting methods were applied to four hourly time series, one for each respective cluster, thus following a bottom-up approach. While the results presented below regard forecasting the number of active sessions of a certain cluster at a given hour of the day, the algorithm can also calculate the number of sessions that start per hour.

Based on preliminary analysis, all the time series are stationary. However, the series for Clusters 1 and 3 are stable, whereas those for Clusters 0 and 2 are more volatile. All the models, at least the machine learning ones, ran with same lag features and window features for all the clusters of a dataset. A sliding window of 3 months was used, with a forecast horizon of 24h. Every 24h, the models are retrained. Based on ACF and PACF lags of 1, 2, 3, and 24 were used across the clusters. The window features comprised weekly, daily, and 12-hour moving average, as well as the weekly coefficient of variation used to deal with the seasonality of the data. Additionally, to cope with higher frequencies of zero values for smaller clusters, the model also included a binary lag (1 if there were sessions on *h-1* and 0 if not).

Table 6 Forecasting - MAE results (Boulder).

MAE	**Clusters**				
	1	**2**	**3**	**0**	**AVG**
ETS w/encoders	1.78	1.03	1.04	0.86	1.17
ETS w/encoders (log transform)	1.82	1.02	1.06	0.84	1.19
Light GBM – Poisson	1.90	1.21	1.00	1.07	1.29
Light GBM – Tweedie	1.85	1.11	0.99	1.00	1.24
Prophet	1.84	1.04	1.42	0.84	1.29
Prophet (log transform)	1.83	1.02	1.43	0.82	1.28
Random Forest	1.82	0.89	1.02	0.78	1.13
XGBoost – Poisson	1.93	1.18	1.51	0.92	1.38
XGBoost –Tweedie	1.94	1.23	1.74	1.00	1.48

Cluster 1 – No Flex was the cluster with larger errors across the models. XGBoost had the worse results across clusters (Table 5). Overall, RF was the best performing model (Figure 9) with exponential smoothing with encoders as the second-best performing method.

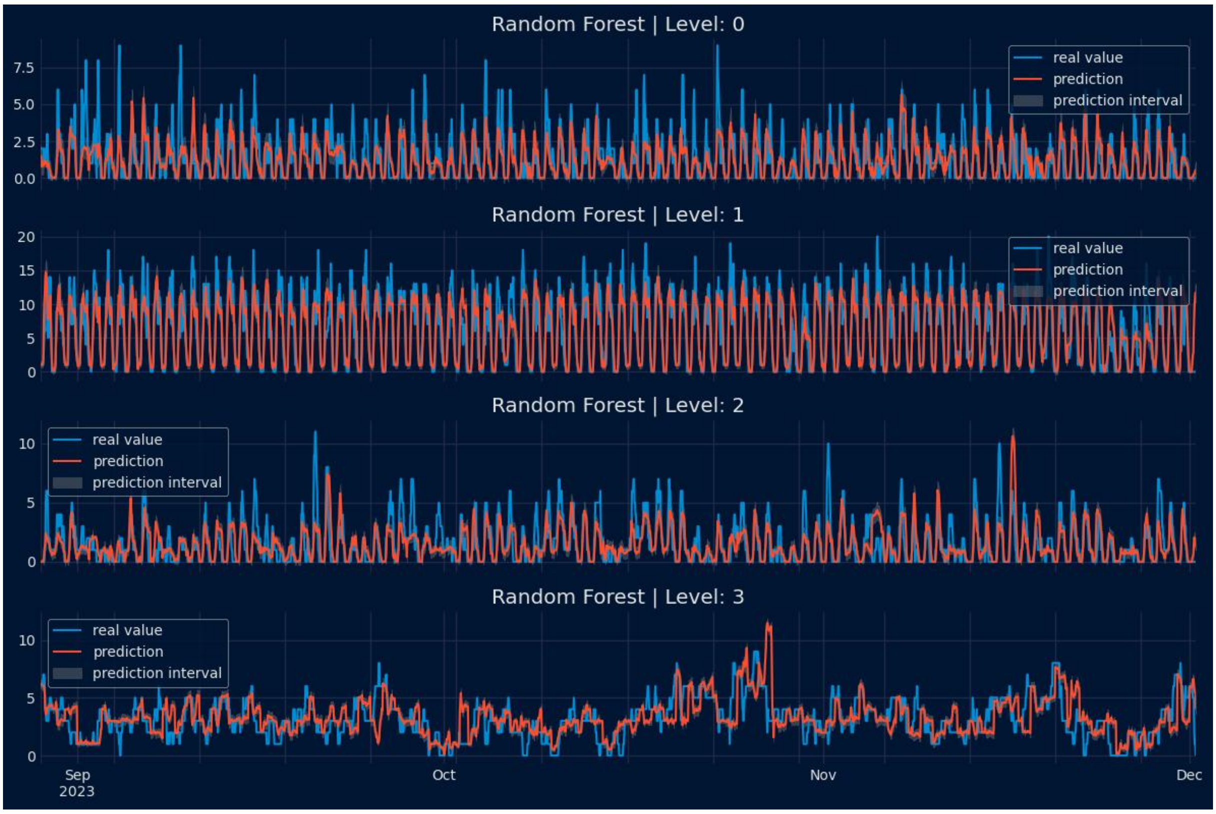


Figure 9 Random Forest per cluster (Boulder).

4.2. Californian Adaptive Charging Network (ACN) Dataset

4.2.1. Descriptive Analysis

The ACN dataset is also an US-based open access dataset, which have been used by research works such as [31]. It is regarded as a complete and good quality set [20,30]. Like the Boulder data, this dataset is also US-based, but from a different state and charging occasion. Instead of kerbside sessions, this dataset contains charging sessions which occurred at a university campus and at an office parking site. Thus, even though, the university charging location is open to the public, the peak of start and end of the sessions seems to coincide with a 9-to-5 like schedule (Figure 10). Both office (code 91109) and university sessions (code 91125) have a similar distribution shape. However, the morning peak at the office location happens slightly earlier than at the university campus, between 7:00 and 8:00 in the morning (Figure 10).

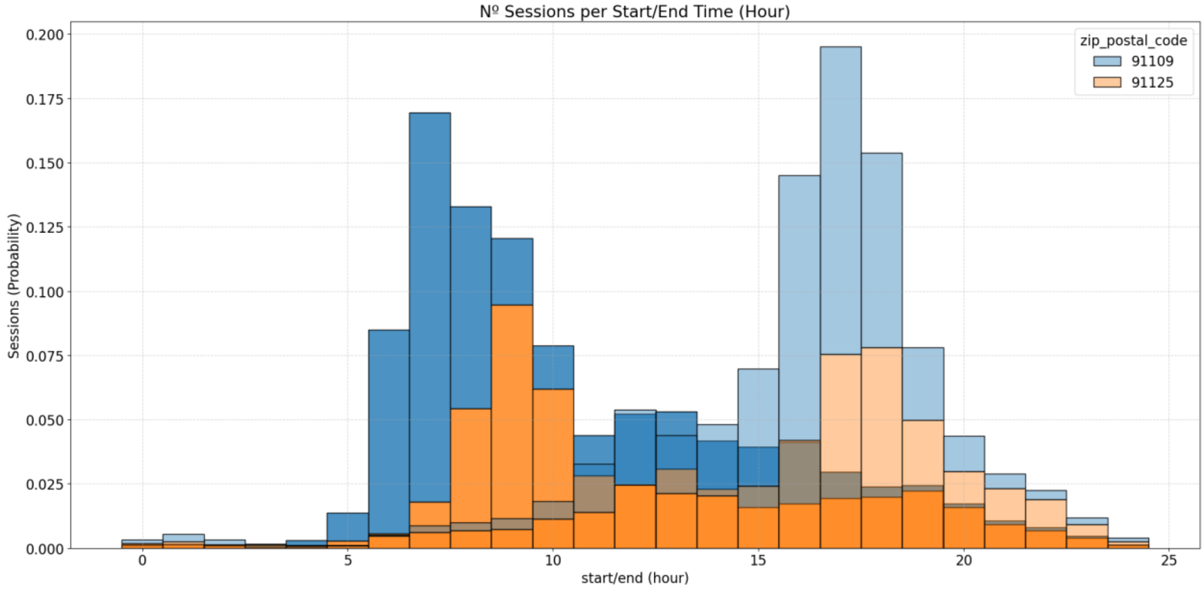


Figure 10 Sessions per hour (ACN).

When looking at the volume of sessions (Figure 11), most likely due to the COVID-19 pandemic, there is a sharp decrease by March 2020 [31]. Nonetheless, although at a lower volume than pre-pandemic number, by the time of the last recorded charging session on the 19th of September of 2021, the number of sessions was steadily increasing (Figure 11).

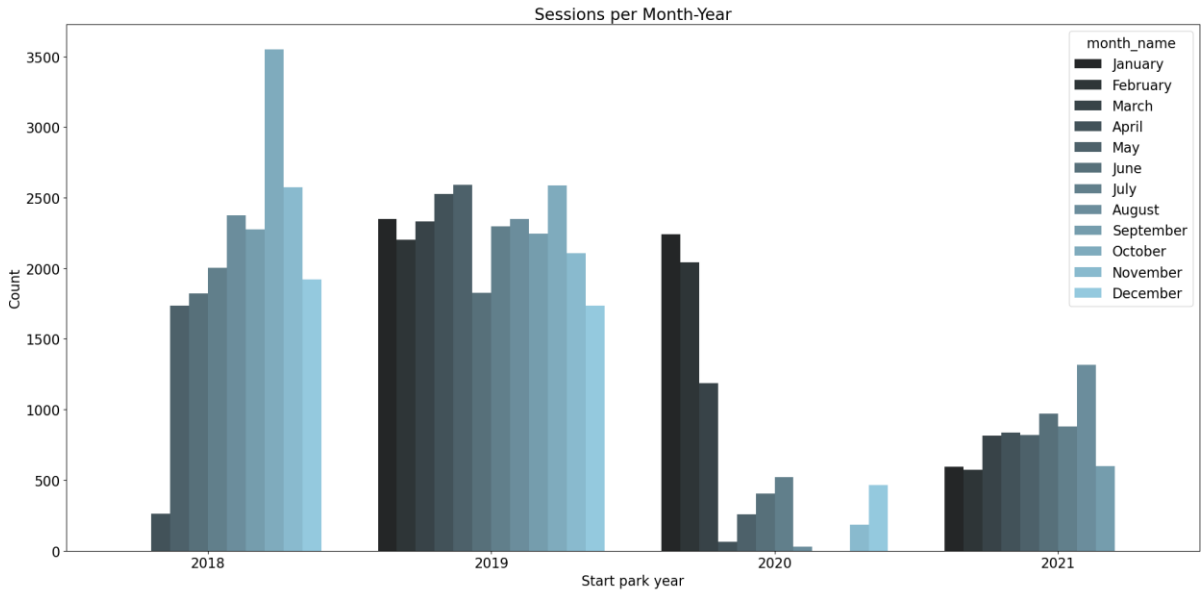


Figure 11 Sessions per month per year (ACN).

The average parking time was around 6h 7min, being 3h 30min of that time dedicated to charging. Around 50% of the sample had an idle time superior to 1h 30 min and, on average, 11.63 kWh were charged per session with the maximum energy consumption being 75.53

kWh. There is a clear different between the daily number of sessions between week/working days and weekends/public holidays, as well as a monthly fluctuation of the intensity of the charging activity (Figure 12).

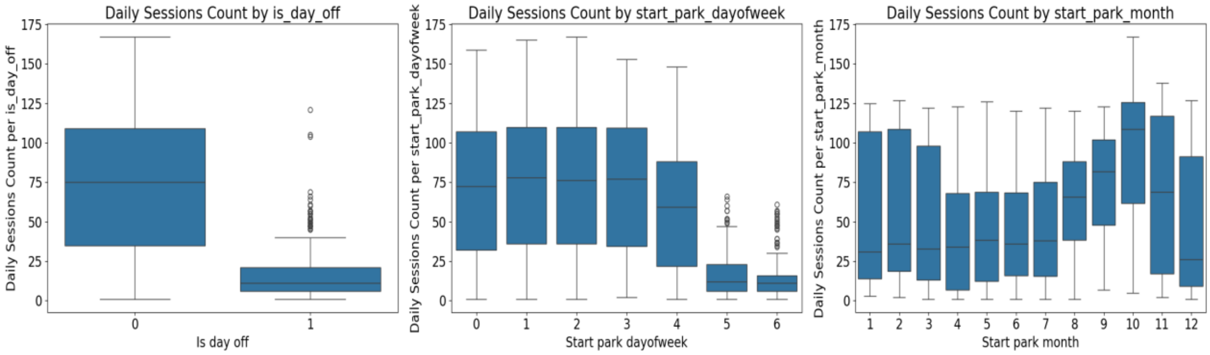


Figure 12 ACN: (a) Daily sessions boxplot per day off dummy; (b) boxplot per day of the week; (c) boxplot per month.

4.2.2. Clustering Results

Across the clustering variations, and similarly to Boulder, the single flexibility envelope option achieved better evaluation scores than the other configurations. Based on these scores, Configuration 2 with 6 clusters was the best performing (Table 6).

Table 7 Cluster evaluation (ACN).

Clustering Model	Parameters	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index	Configuration
K-Means	3	0.68	370,171.44	0.47	Configuration 2
K-Means	5	0.64	508,190.59	0.50	Configuration 2
K-Means	**6**	**0.64**	**584,316.90**	**0.50**	**Configuration 2**
K-Means	10	0.63	905,025.97	0.49	Configuration 2
K-Means	4	0.63	431,474.92	0.51	Configuration 2
BIRCH	(3, 0.1)	0.58	53,907.94	0.44	Configuration 4
K-Means	3	0.54	109,605.12	0.56	Configuration 4
K-Means	5	0.53	148,630.11	0.54	Configuration 4
K-Means	7	0.53	182,923.28	0.54	Configuration 4
K-Means	4	0.52	126,505.28	0.57	Configuration 4

The ideal number of clusters for this dataset (six clusters) was superior to the Boulder case. Given the overall ACN sample showed a higher flexibility potential, this is somewhat expected. The six groups are statistically distinct according to the Kruskal-Wallis test. The Dunn's pairwise test results were also significant for all the clusters in terms of idle time and energy, except for Clusters 4 and 1 energy distribution.

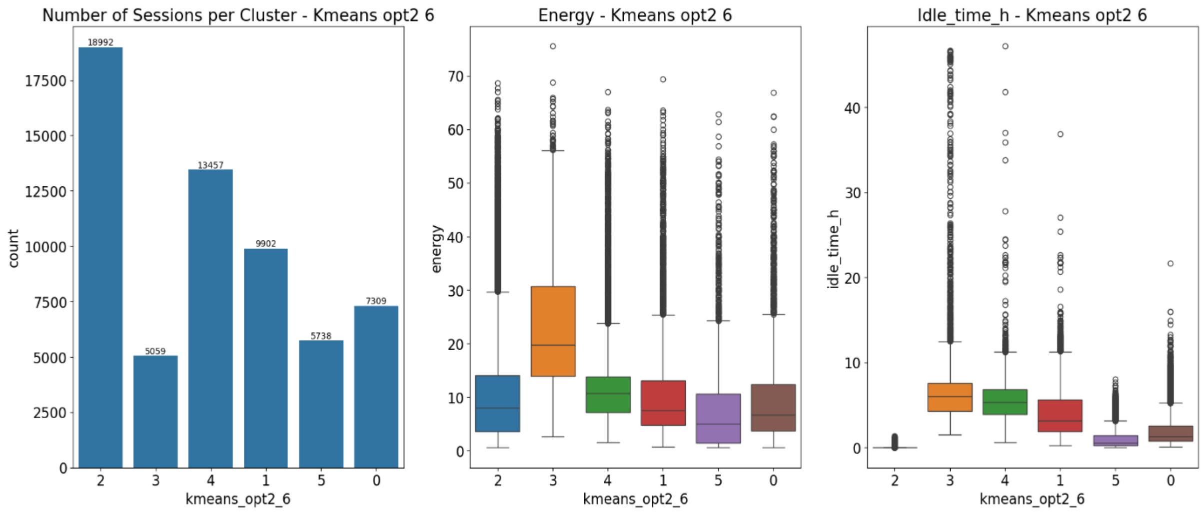


Figure 13 ACN: (a) Clusters' size; (b) Energy clusters' boxplots; (c) Idle time clusters' boxplots.

The largest cluster was Cluster 2 – No Flex (18,992), followed by Cluster 4 – Mid Flex (idle time) and Cluster 1 – Mid Flex (Figure 13). The little flexibility clusters were smaller than the latter, with Cluster 0 – Min Flex (idle time) having 7,309 sessions and Cluster 5 – Min Flex, 5,738 data entries. Cluster 3 – High Flex was the smallest group with 5,059 charging sessions (Figure 13).

While Cluster 4 and Cluster 1 deliver similar loads to EVs during a charging session, Cluster 4 has, on average, an additional 1h 30min of idle time, thus making it more flexible than Cluster 1, but less than Cluster 3 (Table 7). When looking at total duration, Clusters 3 and 4 EVs are parked on average for more than 8h, whereas Clusters 2 and 5 seem to be associated with either morning or afternoon shifts with an average parking time of roughly 3h 30 min and 4h, respectfully (Table 7). These different profiles align with an office/university schedule.

Table 8 Clusters' profiling (ACN).

Cluster	Variable	Min	Max	Mean	Std	Kurtosis	Skewness
0	AVG Charg. Power (kW)	0.11	31.98	4.22	2.06	13.17	1.64
1	AVG Charg. Power (kW)	0.17	26.28	4.57	2.04	7.92	1.39
2	AVG Charg. Power (kW)	0.08	45.99	4.83	2.52	9.15	0.45
3	AVG Charg. Power (kW)	2.18	34.12	6.51	1.90	63.17	4.84
4	AVG Charg. Power (kW)	1.32	32.04	4.51	1.87	19.01	2.29
5	AVG Charg. Power (kW)	0.11	30.60	3.62	2.34	9.64	1.46
0	Charging time (h)	0.10	13.45	3.04	2.52	1.43	1.48
1	Charging time (h)	0.12	12.89	3.27	2.19	1.14	1.33
2	Charging time (h)	0.02	15.97	3.48	2.77	0.30	1.06

3	Charging time (h)	0.20	13.97	4.70	1.90	0.34	0.60
4	Charging time (h)	0.24	12.12	3.70	1.73	0.98	0.97
5	Charging time (h)	0.05	14.98	2.86	2.67	1.23	1.42
0	Energy (kWh)	0.61	66.93	9.58	8.48	5.68	2.06
1	Energy (kWh)	0.71	61.73	11.39	9.13	4.89	2.08
2	Energy (kWh)	0.61	68.61	12.75	12.16	1.70	1.41
3	Energy (kWh)	3.68	64.40	23.97	10.80	-0.11	0.71
4	Energy (kWh)	1.70	66.97	12.98	8.59	6.67	2.41
5	Energy (kWh)	0.60	62.82	7.17	7.99	7.87	2.41
0	Idle time (h)	0.11	21.68	2.09	2.23	5.81	2.17
1	Idle time (h)	0.27	22.63	3.45	2.47	6.96	1.81
2	Idle time (h)	0.00	1.24	0.05	0.16	18.08	4.16
3	Idle time (h)	1.50	46.57	6.17	4.37	35.71	4.96
4	Idle time (h)	0.61	33.80	5.18	2.19	12.24	1.46
5	Idle time (h)	0.02	7.65	1.10	1.25	3.16	1.83
0	Total duration (h)	1.40	21.79	5.14	2.44	2.42	1.36
1	Total duration (h)	2.85	32.24	6.72	2.19	8.89	1.50
2	Total duration (h)	0.17	15.97	3.53	2.76	0.32	1.08
3	Total duration (h)	5.13	48.00	10.86	4.01	40.07	5.48
4	Total duration (h)	4.23	34.60	8.88	1.65	21.41	1.76
5	Total duration (h)	0.69	16.04	3.95	2.58	0.83	1.32

In terms of parking start time, morning arrivals, particularly before 10:00, are common in both the overall sample and the clusters with larger flexibility envelopes, such as Clusters 3, 4, and 1 (Figure 14). Cluster 1 - Mid Flex also shows some arrivals during late morning and early afternoon. The cluster with the lowest flexibility presents a bimodal distribution, with a primary peak in the morning and a smaller peak in the afternoon (Figure 14). The no-flexibility cluster has the flattest distribution throughout the day (Figure 14). Overall, activity between 20:00 and 5:00 remains low across all clusters and the full sample, which is consistent with the typical operating hours of office and campus environments.

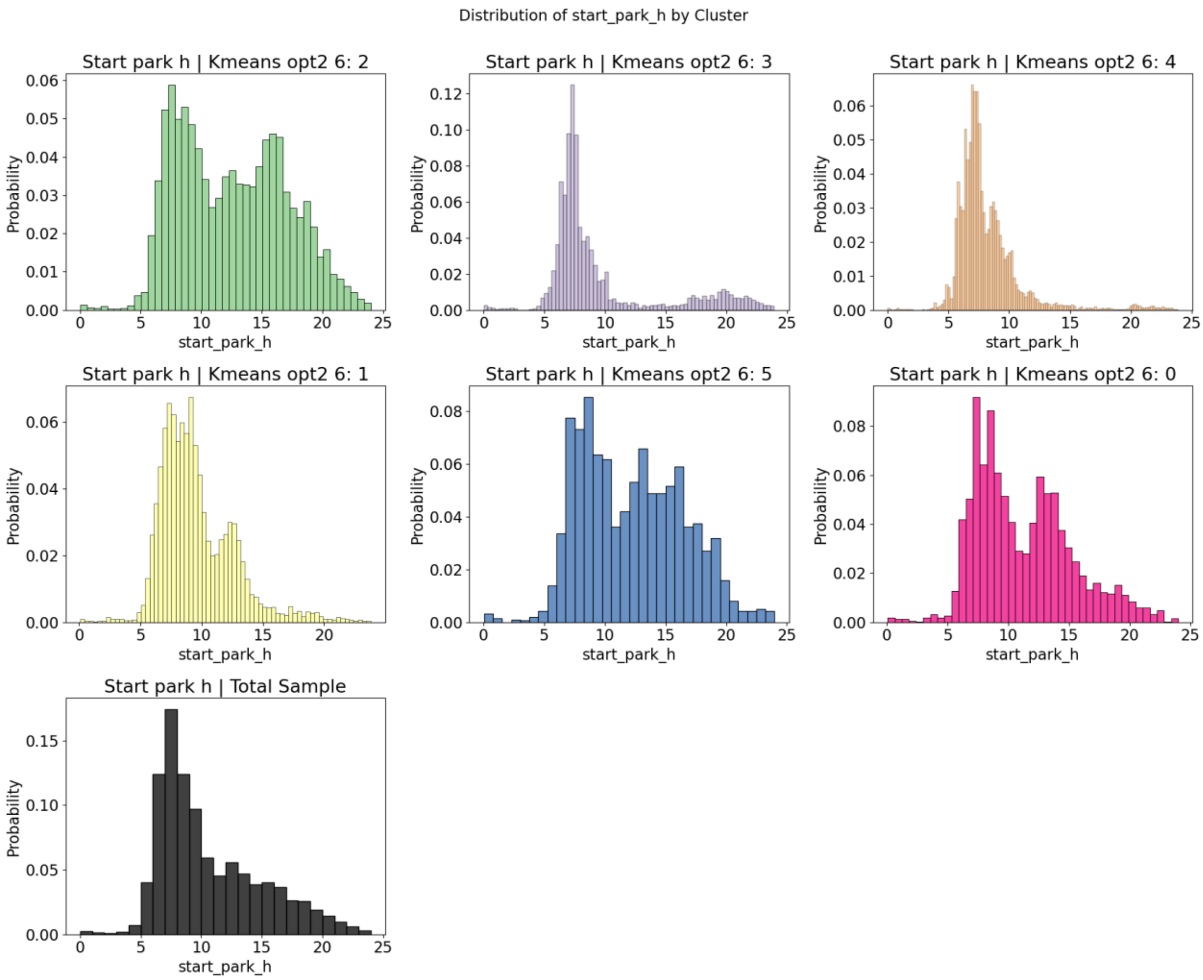


Figure 14 Start parking hour per cluster (ACN).

4.2.3. Forecasting Results

Similarly to Boulder, the forecasting methods were applied to hourly time series associated to the groups found during clustering. The preliminary analysis showed all time series to be stationary and volatile. The sliding window and forecast horizon were the same as the previous dataset. The ACF and PACF suggested lags between 3 and 7 (depending on the cluster), so lags 1–6 and 24 were chosen for prediction. The window features and binary lag were the same of the Boulder. The MAE of each model is shown below per cluster (Table 8).

Table 9 Forecasting - MAE results (ACN).

MAE	**Clusters**						
Model	**1**	**2**	**3**	**4**	**5**	**0**	**AVG Tot**
ETS w/encoders	1.65	1.61	0.97	2.29	0.73	1.02	1.38
ETS w/encoders (log transform)	1.44	1.41	0.95	1.96	0.67	0.95	1.23
Light GBM - Poisson	1.43	1.37	1.13	1.77	0.93	1.10	1.29
Light GBM - Tweedie	1.29	1.31	1.13	1.62	0.83	1.06	1.21
Prophet	1.53	1.56	1.12	2.05	0.70	0.96	1.32
Prophet (log transform)	1.32	1.36	1.00	1.75	0.64	0.86	1.16

Random Forest	1.09	1.17	0.75	1.39	0.53	0.82	0.96
XGBoost – Poisson	1.58	1.44	0.95	2.07	0.72	1.17	1.32
XGBoost – Tweedie	1.66	1.61	0.99	2.23	0.74	1.18	1.40

Out of the clusters, Cluster 5 – Min Flex was the one with the smallest errors across the forecasting methods. On the other hand, Cluster 4 - Mid Flex (idle time), had the highest error, ranging from 1.39 on RF to 2.29 on ETS. Once again XGBoost was the worst method, and RF was the best performing (Figure 15).

Figure 15 Random Forest per cluster (ACN).

4.3. Slovenian Dataset

4.3.1. Descriptive Analysis

The Slovenian dataset is a 6-month country wide dataset of kerbside charging sessions (Figure 16). The dataset comprises both AC and DC chargers. In resemblance to what happens in Boulder, there is a bigger overlap between the start and end of the parking time of an EV (Figure 17). 50% of the EVs were parked for almost 2h, being that 1h 44 min was for charging. Only about 12% of the charging sessions had an idle time superior to 1 hour and an additional 4% had idle times between 30 minutes to 1 hour. On average, 17.51 kWh were charged per session with energy consumption ranging from 0.60 kWh to 99.28 kWh. Concerning power, 75% of the sessions had a charging rate equal or inferior to 10.82 kW, suggesting that DC charging (up to 48 kW) is less frequent than AC.

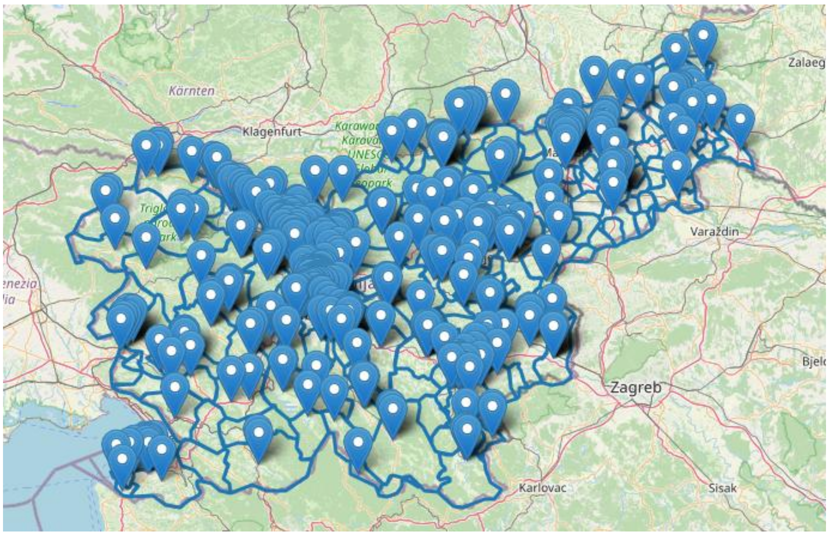


Figure 16 Slovenia EVSEs map.

Morning peak occurs between 9:00 and 11:00, and evening peak after 16:00. Since 17:00 until the following morning the number of sessions that start decreases sharply. End of parking times has a similar behaviour, but with only one peak around 12:00/13:00 (Figure 17).

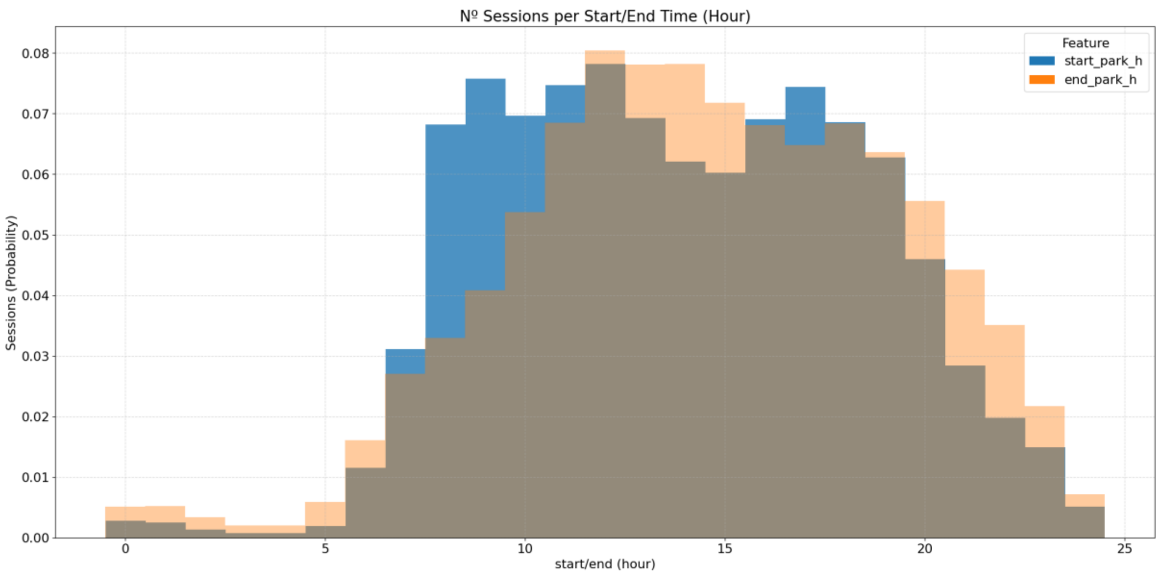


Figure 17 Sessions per hour (Slovenia).

There is a clear difference between the daily number of sessions across the days of the week, specially between weekdays and days off (weekends plus national public holidays) (Figure 18a). Fridays are the days registering more sessions, whereas Sundays the least (Figure 18b).

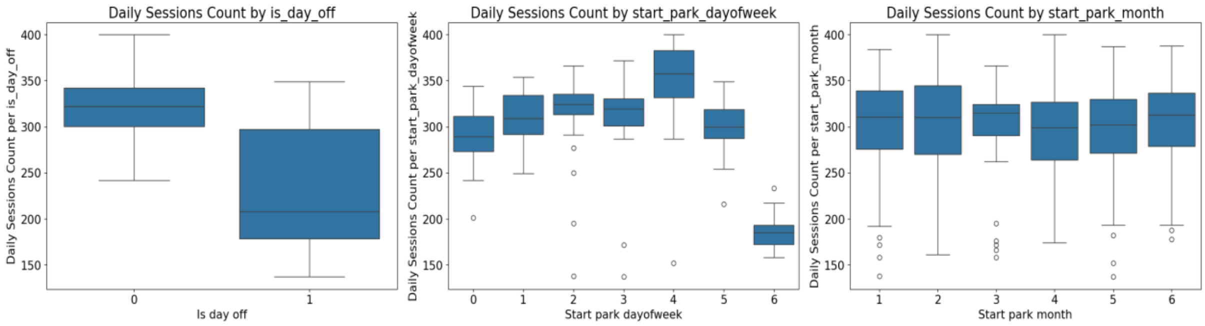


Figure 18 Slovenia: (a) Daily sessions boxplot per day off dummy; (b) boxplot per day of the week; (c) boxplot per month.

This dataset has sessions from both AC and DC charging types, which seem to be associated with different average idle times throughout the week (Figure 19). On average, sessions from DC chargers, which represent roughly 4% of the total sessions, do not have idle time. In contrast, AC chargers show average idle times ranging from slightly more than 30 minutes on Saturdays (day 5) to almost one hour on Sundays (day 6). This behaviour is consistent with the typical location and purpose of DC chargers, which are often installed in places such as highways where EV charging is expected to be brief.

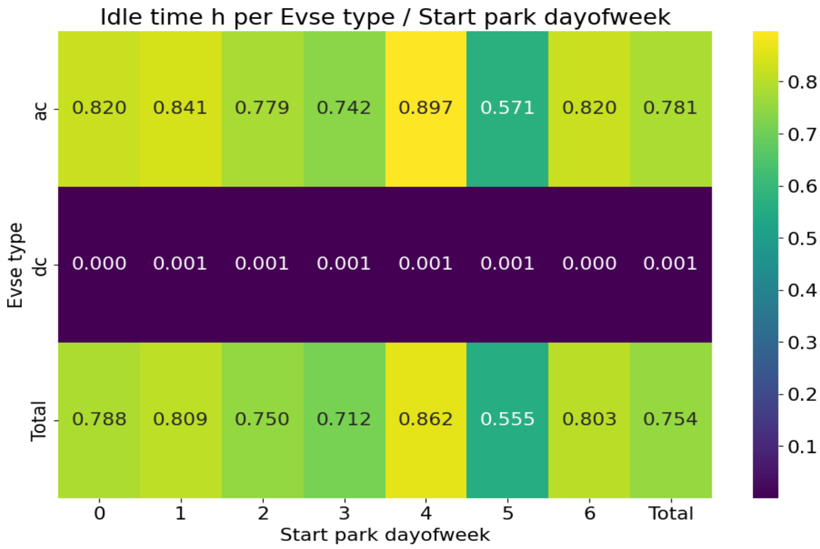


Figure 19 Average idle time per EVSE type (Slovenia).

Moreover, Figure 20 illustrates the relationship between OSM features and total charging sessions, separated into weekdays and weekends for the Slovenian dataset. On weekdays, positive correlations are observed for shop density, amenities, land use intensity, and highway proximity, suggesting that charging activity increases in commercially active and well-connected areas. The strongest associations appear for highway and amenity related features, indicating the importance of accessibility and urban activity during working days. In contrast, weekend correlations are weaker and, in several cases, negative. Shop, highway, and building related features show reduced or inverse relationships with session counts, while leisure related features display a positive association. This shift indicates a behavioural change between work related and leisure driven mobility patterns, with weekend charging less concentrated in commercial and transport-oriented zones.

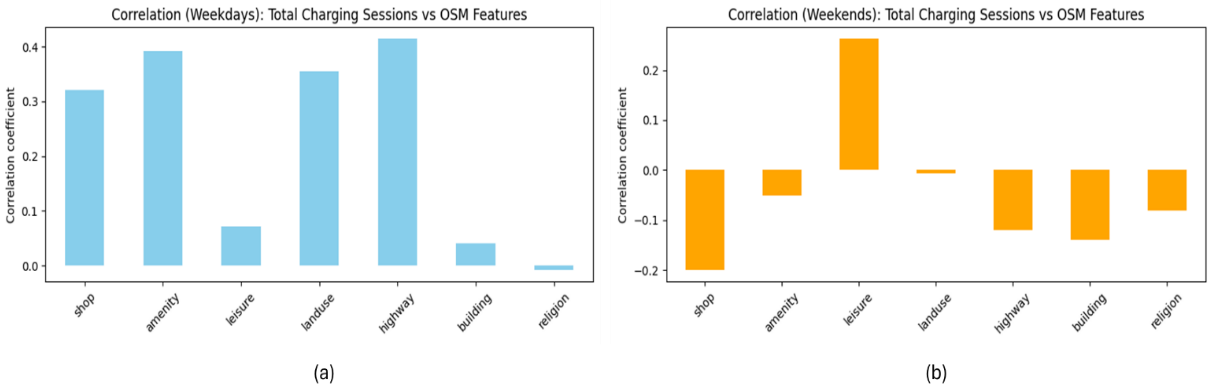


Figure 20 Correlation between total charging sessions and OSM features during (a) weekdays and (b) weekend in Slovenia.

4.3.2. Clustering Results

From the clustering configurations tested, K-Means with Configuration 2 was again the best-scoring combination (Table 9). The four-cluster solution was selected to proceed with the analysis, as it had a better Silhouette score and Davies-Bouldin index than the five-cluster option, and a better Calinski-Harabasz score than the three-cluster option.

Table 10 Cluster evaluation (Slovenia).

Clustering Model	Parameters	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index	Configuration
K-Means	3	0.85	362,602.72	0.43	Configuration 2
K-Means	**4**	**0.85**	**470,443.28**	**0.45**	**Configuration 2**
K-Means	5	0.84	549,178.33	0.47	Configuration 2
K-Means	3	0.56	37,151.25	0.53	Configuration 4
BIRCH	(3, 0.1)	0.54	15,428.85	0.47	Configuration 4
K-Means	4	0.54	42,411.53	0.54	Configuration 4
K-Means	3	0.50	62,831.30	0.67	Configuration 1
K-Means	5	0.48	65,548.71	0.75	Configuration 1

The majority of the sessions (40,389) belong to Cluster 0 – No Flexibility, whereas Cluster 1 - High Flexibility is, once again the smallest group (3,073). Cluster 2 – Minimal Flexibility and Cluster 3 – Average Flexibility are of slightly bigger than the High Flex cluster with 5,328 and 4,294 sessions, respectfully (Figure 21).

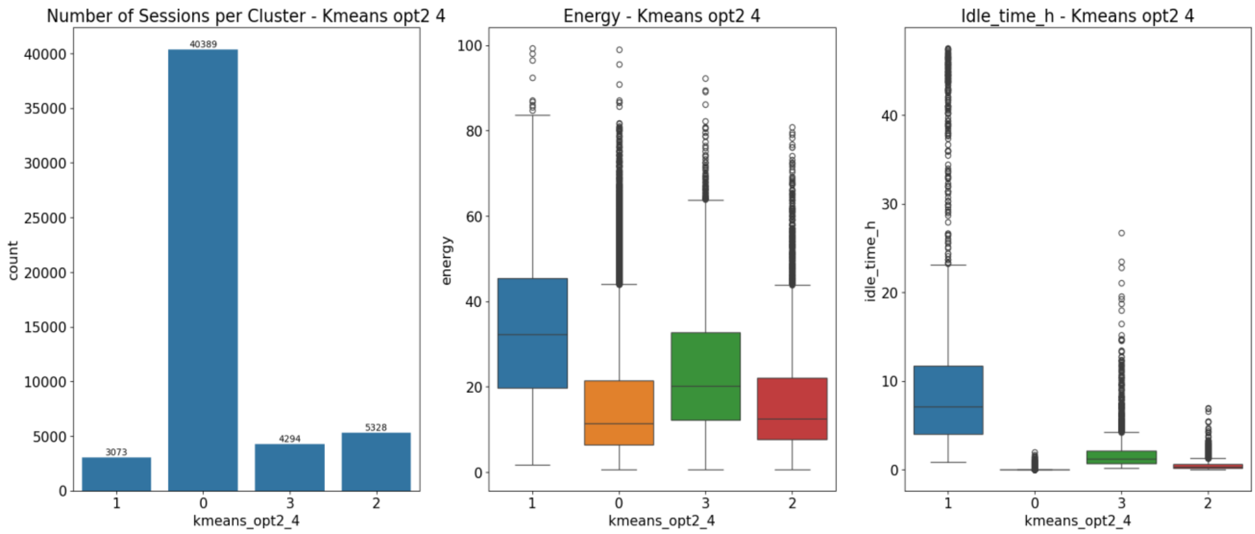


Figure 21 Slovenia: (a) Clusters' size; (b) Energy clusters' boxplots; (c) Idle time clusters' boxplots.

Clusters 0 and 2, which have minimal to no flexibility are the clusters that contain DC EVSEs (48 kW). Cluster 0 sessions are shorter in terms of total parking and charging times. On average they have no idle time and deliver less energy than sessions from other clusters. Cluster 2 – Minimal Flex distinguish itself from the latter cluster because their total parking time. This probably derives from EV drivers that tend to recharge their vehicle when the opportunity arises, not letting their SoC reach very low values. Cluster 1 – High Flex has an average idle time superior to 19h and Cluster 3 – Average Flex of more than 1h 30min. Regarding energy theses clusters almost twice as much energy than Clusters 1 and 0, being the high flexibility, naturally the one with longest idle time and energy delivered per session on average (Table 10).

Table 11 Clusters' profiling (Slovenia).

Cluster	Variable	Min	Max	Mean	Std	Kurtosis	Skewness
0	AVG Charg. Power (kW)	0.09	48.77	9.18	5.84	11.30	2.55
1	AVG Charg. Power (kW)	0.25	21.70	8.69	2.99	1.45	0.20
2	AVG Charg. Power (kW)	0.23	48.11	7.01	3.70	7.43	0.99
3	AVG Charg. Power (kW)	0.22	21.53	8.28	3.68	0.48	0.21
0	Charging time (h)	0.07	16.00	2.06	2.05	12.66	3.12
1	Charging time (h)	0.45	15.59	4.28	2.54	3.17	1.50
2	Charging time (h)	0.08	15.94	2.89	2.69	6.02	2.38
3	Charging time (h)	0.07	15.97	3.63	2.97	2.64	1.71
0	Energy (kWh)	0.60	99.03	15.67	12.84	2.43	1.50
1	Energy (kWh)	1.71	99.28	33.69	16.90	-0.34	0.47
2	Energy (kWh)	0.60	80.86	16.77	13.28	2.34	1.53
3	Energy (kWh)	0.62	92.19	24.23	15.37	0.52	1.02
0	Idle time (h)	0.00	1.96	0.01	0.05	333.41	15.15
1	Idle time (h)	0.81	47.55	9.59	8.91	6.89	2.51

2	Idle time (h)	0.02	6.95	0.50	0.54	20.83	3.47
3	Idle time (h)	0.17	26.76	1.77	1.91	30.87	4.35
0	Total duration (h)	0.17	16.02	2.07	2.04	12.67	3.12
1	Total duration (h)	4.02	48.00	13.87	8.68	6.86	2.52
2	Total duration (h)	0.45	16.44	3.39	2.59	6.75	2.54
3	Total duration (h)	1.66	27.21	5.40	3.11	3.67	1.80

In terms of arrivals throughout the day, both the high- and mid-flexibility clusters show higher concentrations of early-morning to late-evening sessions (Figure 22). In particular, Cluster 1 - High Flex exhibits increasing activity from 5:00, peaking around 20:00, before declining in the evening. Cluster 3 - Mid Flex has the opposite pattern, with arrivals more concentrated in the morning and tapering off around 20:00 (Figure 22). The clusters with no or minimal flexibility have distributions closer to that of the overall sample, with Cluster 2 - Min Flex showing slightly higher activity during the morning (Figure 22).

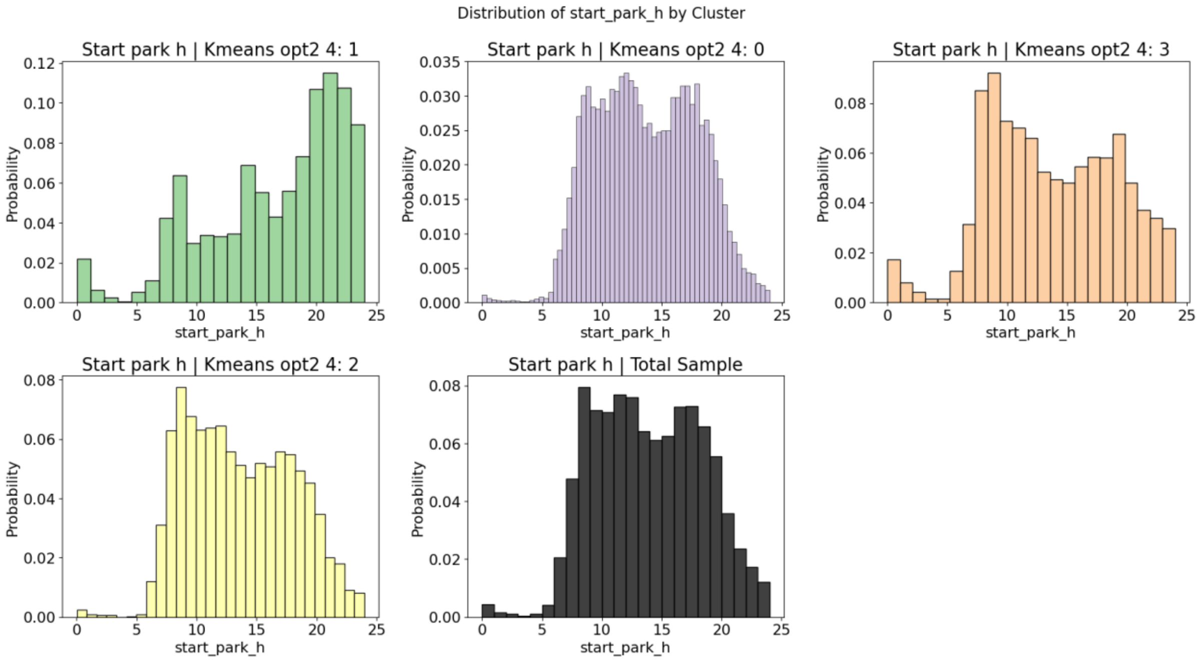


Figure 22 Start parking hour per cluster (Slovenia).

As this dataset has additional geospatial information, the calculated surrounding features (as per Section 3.1.2.) for each EVSE were used to further profile the clusters (Table 11 and Figure 23). Cluster 0 – No Flex represents more approximately 73–78% of sessions across all surrounding features, confirming its dominant role in the system. It shows a relative overrepresentation in shop-related areas (higher session-weighted share and positive z-score), suggesting that non-flexible charging behaviour is particularly common in commercial contexts. However, this cluster is relatively underrepresented in other environmental contexts, which may suggest that non-flexible charging in the kerbside occasion is widespread but not spatially specialised. In contrast, the flexible clusters, although smaller in absolute size, display a higher degree of differentiation. Cluster 1 (High Flex) shows higher relative presence near building-dense and religion-related features, while Clusters 2 and 3 exhibit stronger associations with amenities, highways, leisure areas, and mixed land-use contexts

Table 12 Session-weighted share for each OSM feature (Slovenia).

Location type/Cluster	0	1	2	3
Building	74.5	6.3	10.6	8.6
Shop	77.8	3.9	10.6	7.7
Amenity	74.5	4.9	11.6	9.0
Leisure	74.2	5.9	11.0	8.9
Highway	73.9	5.6	11.4	9.1
Religion	73.5	7.0	10.9	8.5
Land use	72.8	6.1	11.4	9.7

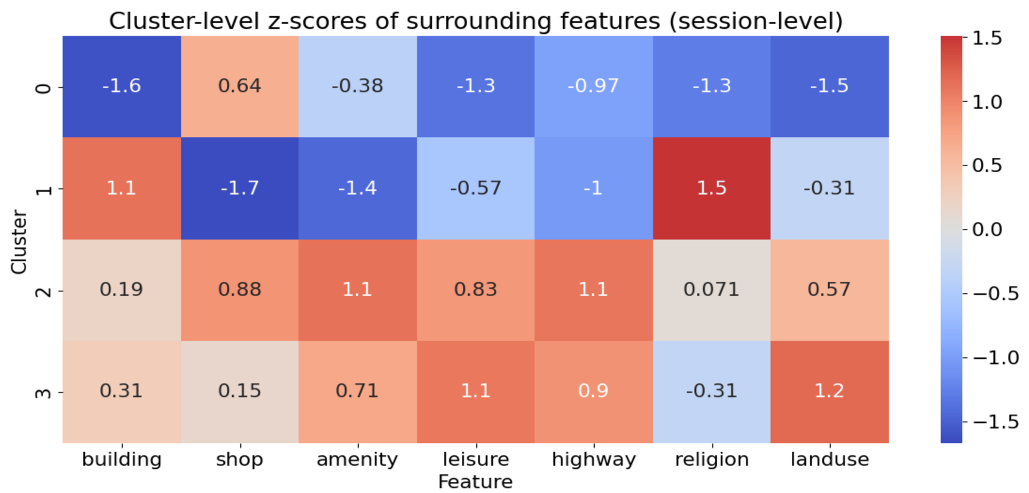


Figure 23 Cluster-level Z-Scores OSM features (Slovenia).

4.3.3. Forecasting Results

Based on the preliminary analyses, the time series of all the clusters were stationary and stable. According to ACF and PACF, the number of lags selected were 1, 2, 3, and 24. The other features remain the same to the ones tested on the other datasets.

Table 13 Forecasting - MAE results (Slovenia).

MAE	Clusters				
	1	2	3	0	AVG
ETS w/encoders	2.81	2.03	2.40	6.44	3.42
ETS w/encoders (log transform)	2.85	2.03	2.47	6.43	3.45
Light GBM – Poisson	3.21	1.87	2.09	5.77	3.23
Light GBM – Tweedie	2.84	1.86	2.11	5.59	3.10

Prophet	3.13	2.00	2.31	5.96	3.35
Prophet (log transform)	3.24	1.98	2.35	5.99	3.39
Random Forest	2.98	1.92	2.17	5.95	3.25
XGBoost – Poisson	3.25	2.03	2.41	5.81	3.38
XGBoost – Tweedie	3.31	2.13	2.54	6.06	3.51

Once again, the MAE results vary across clusters. Cluster 0 – No Flex has the highest errors across the models (Table 12). Because MAE is scale-dependent and Cluster 0 is much larger than the other ones this value is no surprise and does not inherently imply lower predictability. In contrast, Clusters 1, 2, and 3 show lower absolute MAE values, but they are also smaller cluster Among the evaluated models, LightGBM with Tweedie loss (Figure 24) and RF achieved the lowest MAE across, indicating robust performance across heterogeneous flexibility segments. XGBoost – Tweedie was the worst performing model.

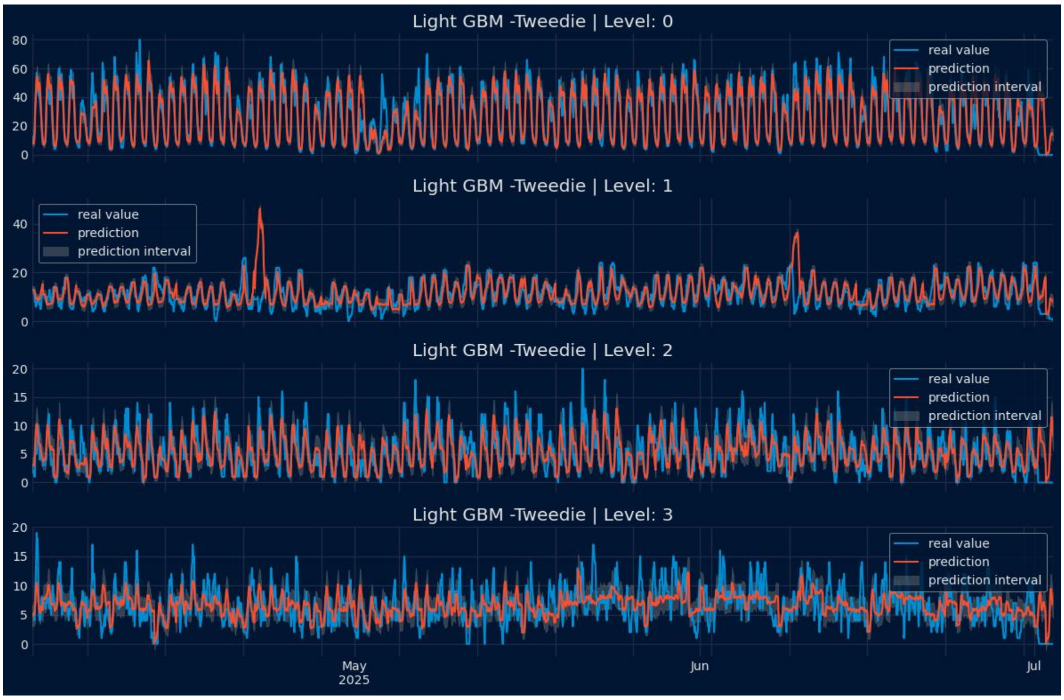


Figure 24 Light GBM forecasting per cluster (Slovenia).

4.4. Madeira Dataset (Light-Duty EVs)

The Madeira Island (Portugal) dataset is a kerbside charging occasion dataset. As mentioned in Sections 2 and 3.1.1. this is a hybrid dataset and because of it, the clustering and forecasting results focuses mostly on metrics evaluating the performance of the algorithms.

4.4.1. Descriptive Analysis

This dataset contains 12,691 sessions spanning three months, in which the number of sessions seems to be steadily increasing (Figure 25). The number of daily sessions varies between day of the week but to a less extent than the Slovenian dataset. Friday is the day registering more sessions and Sunday the least (Figure 25b).

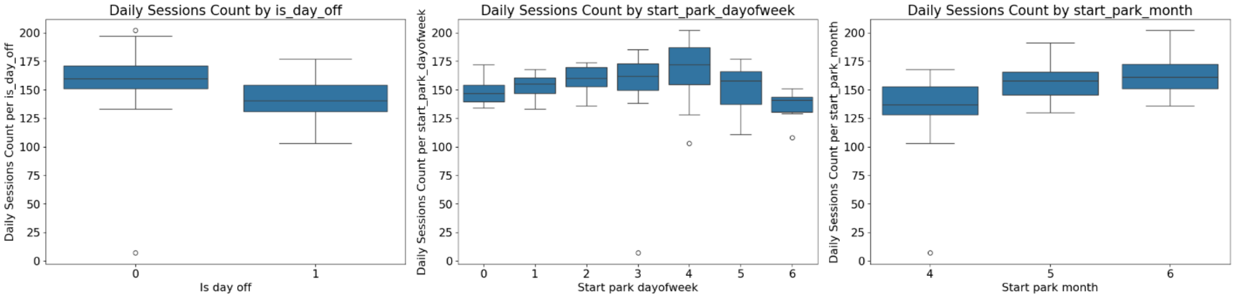


Figure 25 Madeira: (a) Daily sessions boxplot per day off dummy; (b) boxplot per day of the week; (c) boxplot per month.

Like the other kerbside datasets, there is a big overlap between the start and end of the total sojourn time. From 7:00 to 18:00 the number of sessions is roughly the same, and after this period and until the next morning the number of sessions is low. The peak start park occurs at 12:00 whereas the end park time peak happens at 13:00 (Figure 26). 50% of the sessions have a total parking time inferior to 1h 12 min, being the mean value around 2h. The dataset comprises AC and DC chargers.

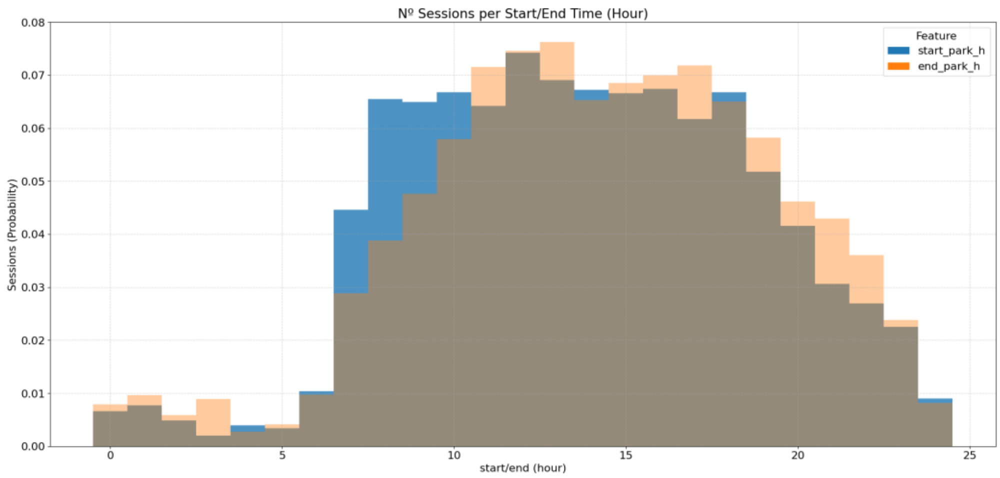


Figure 26 Sessions per hour (Madeira).

4.4.2. Clustering Results

Based on Table 13, K-Means for the second variation showed the best results. Even though the option with 9 clusters shows a better Caliski-Harabasz index with only a small increase on the Davies-Bouldin index, the combination with 4 clusters was the one chosen to proceed the algorithm. This decision was due to the small dimension of the sample.

Table 14 Cluster evaluation (Madeira).

Clustering Method	Parameters	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index	Configuration
K-Means	**4**	**0.90**	**177,804.34**	**0.45**	**Configuration 2**
K-Means	9	0.90	374,850.70	0.48	Configuration 2
K-Means	10	0.90	385,541.51	0.48	Configuration 2
BIRCH	(3, 0.4)	0.59	4,948.27	0.73	Configuration 1
K-Means	3	0.56	7,186.44	0.53	Configuration 4
K-Means	5	0.54	10,131.92	0.52	Configuration 4
K-Means	10	0.54	17,384.50	0.51	Configuration 4
K-Means	3	0.51	17,512.47	0.66	Configuration 1

Four statistically distinct groups were found through clustering. Similarly to the fully empirical dataset, the cluster exhibiting no flexibility was the predominant one, while the cluster with the highest flexibility was the smallest (Figure 27).

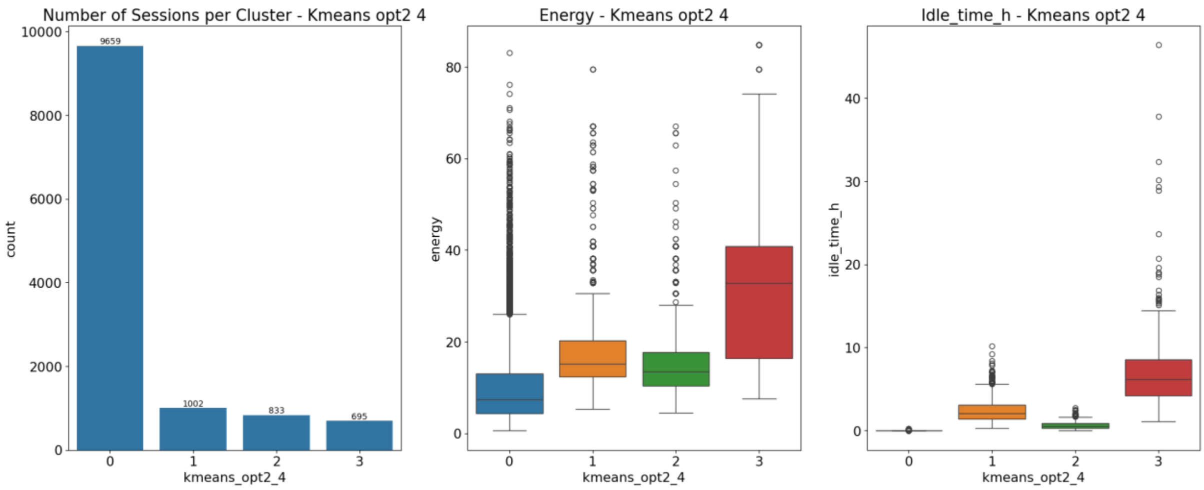


Figure 27 Madeira: (a) Clusters' size; (b) Energy clusters' boxplots; (c) Idle time clusters' boxplots.

For this dataset, parking start time is one of the empirical features, so a brief analysis of arrivals per cluster was performed. Cluster 3 - High Flex is most likely to start after 17:00 and continues until around 1:00 AM, while the Mid Flex cluster (Cluster 1) peaks in the morning before 10:00 and gradually decreases throughout the day (Figure 28). Minimal and No Flex clusters show distributions similar to the overall sample (Figure 28). Although this is a hybrid dataset, arrival patterns resemble those seen in the Slovenian dataset, providing support for the validity of both the clustering methodology and the simulation approach.

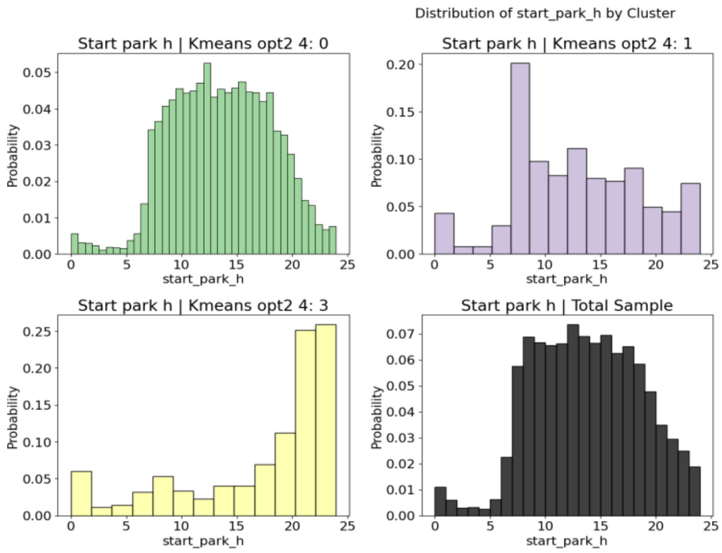


Figure 28 Start parking hour per cluster (Madeira).

4.4.3. Forecasting Results

As per the preliminary analysis, all the time series were stationary. Except for Cluster 2 – Min Flex, the time series were also stable. Window features and forecasting horizon were the same of other datasets, but the lags were set to 1, 2, 3, and 24.

XGBoost was the worst performing method, whereas ETS with encoders and Prophet were the best. ETS was marginally better than Prophet. Cluster 0 – No Flex that the highest errors across the models, while Cluster 2 – Min Flex was the cluster with the smallest MAE (Table 14).

Table 15 Forecasting - MAE results (Madeira).

MAE	Clusters				
	1	2	3	0	AVG
ETS w/encoders	1.48	1.07	1.67	2.48	1.68
ETS w/encoders (log transform)	1.49	1.05	1.76	2.57	1.72
Light GBM – Poisson	1.56	1.24	2.21	2.70	1.93
Light GBM – Tweedie	1.63	1.16	2.15	2.63	1.89
Prophet	1.49	1.08	1.64	2.54	1.69
Prophet (log transform)	1.46	1.05	1.71	2.64	1.71
Random Forest	1.67	1.14	1.91	2.73	1.86
XGBoost – Poisson	1.75	1.18	2.82	2.91	2.17
XGBoost – Tweedie	2.01	1.29	2.97	2.91	2.29

4.5. Danish Dataset

4.5.1. Descriptive Analysis

This is a 1-year length synthetic dataset based on real-world residential charging patterns. This is the only dataset where it is visible that some of the flexibility potential of the charging sessions is already being used (Figure 29). Comparing to a kerbside or an office occasion, residential charging is longer and happens mostly during evening and nighttime (Figure 29). The start parking time peaks at 17:00 and the end of the sojourn time peaks at 7:00. Interestingly, while the initial parking time distribution has flatter peak (lower kurtosis), the end parking time is narrower (higher kurtosis). On average, parking times lasted more than 15h and the maximum charging rate is 11.55 kW. This corresponds to Danish residential grid connections typically being three-phase, with most chargers rated at 11 kW three-phase.

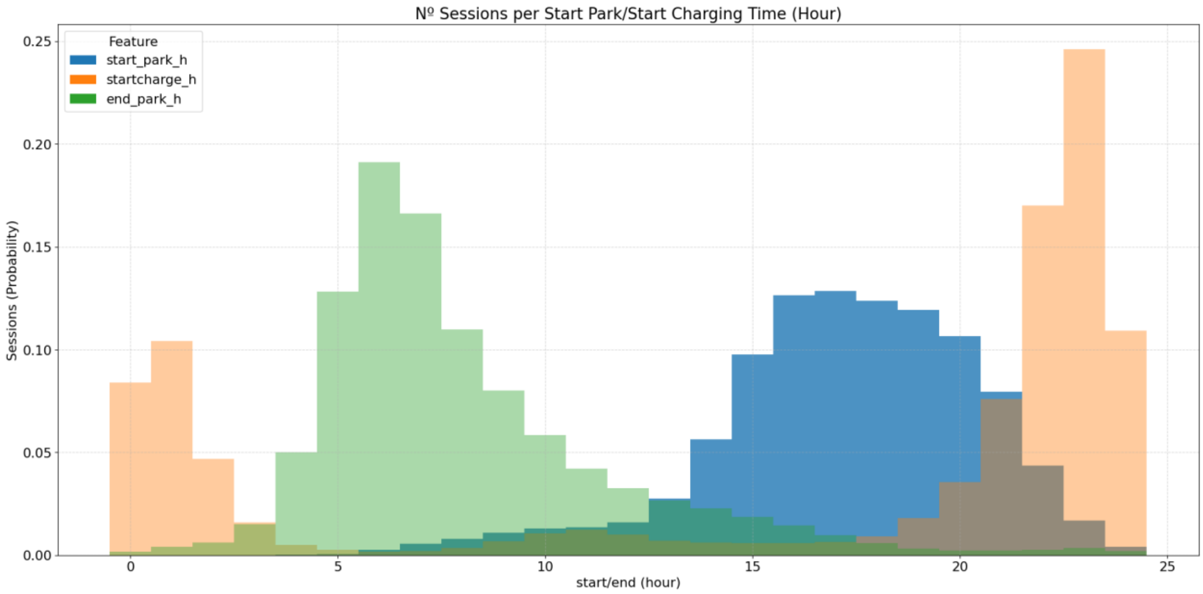


Figure 29 Sessions per hour (Denmark).

4.5.2. Clustering Results

This dataset did not have sessions without idle time, thus clustering configurations excluding sessions with zero idle time were not run (Table 15). The analysis proceeded with 10 clusters, as preliminary tests with 11 and 12 clusters showed that 10 clusters produced better-defined cluster structures.

Table 16 Cluster evaluation (Denmark).

Clustering Model	Parameters	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index	Configuration
K-Means	11	0.52	1,580,122.09	0.52	Configuration 2
K-Means	**10**	**0.52**	**1,486,818.48**	**0.52**	**Configuration 2**
K-Means	12	0.52	1,679,955.45	0.52	Configuration 2

K-Means	6	0.34	215,175.96	0.89	Configuration 1
K-Means	5	0.33	213,685.64	0.95	Configuration 1
K-Means	7	0.33	215,900.58	0.90	Configuration 1

From the 10 clusters found, all showed flexibility potential (Figure 30). Despite the idle time of the clusters being statistically distinct (except for the Cluster 4 and Cluster 6), the average parking time of many groups is between 11h and 17h. Cluster 8 has the highest flexibility with long charging sessions and high loads, and Cluster 7 is the one with shorter sessions and less charged load. Concerning energy delivered, the ten clusters can be grouped into three categories: clusters with small loads (1, 3, 6, 7), clusters with average of around 20 kWh (2 and 9), and clusters with average close to 40 kWh of energy delivered (0, 4, and 8). This likely reflects different user charging behaviours, where some users charge frequently with small loads while others charge less often but for longer sessions. This is consistent with the average daily driving distance in Denmark of around 40 km, corresponding to roughly 8 kWh of consumption assuming an efficiency of 5 km/kWh. As a result, users may not charge every day but rather every three to five days. Contrary to previous datasets, the cluster with the least flexibility in not the largest (in number of sessions) but it is instead the smallest-sized group.

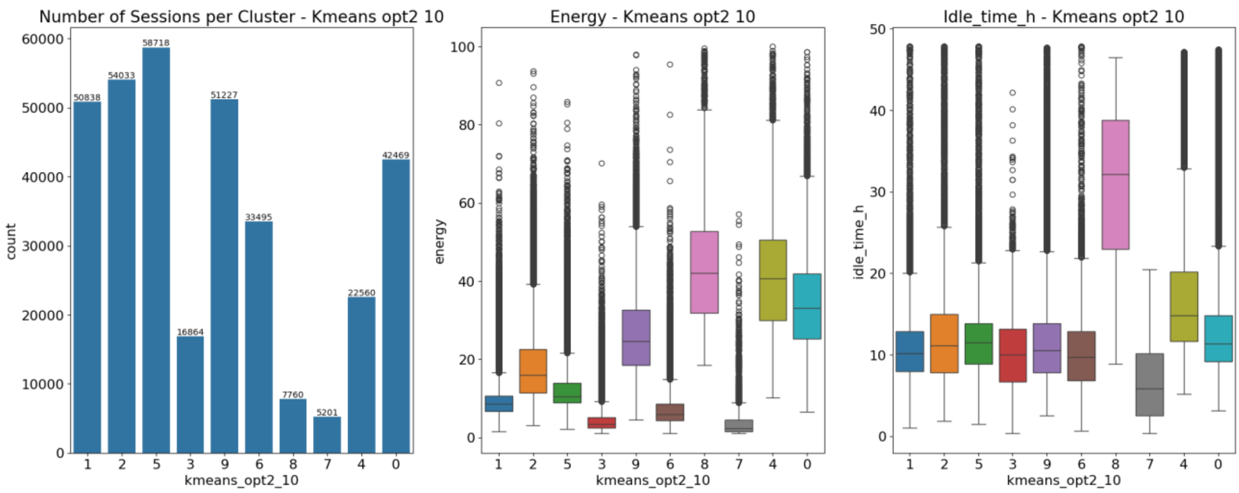


Figure 30 Denmark: (a) Clusters' size; (b) Energy clusters' boxplots; (c) Idle time clusters' boxplots.

Figure 31 shows the parking start time distribution for three of the ten clusters, as these were the ones that differed most from the overall sample distribution (also in the figure). Cluster 7, which has the highest proportion of day-off sessions relative to weekday sessions, seem to reflect a weekend arrival pattern, peaking between 10:00 and 11:00 AM. Cluster 8, which has the highest flexibility potential, and Cluster 3 also show higher concentrations of morning sessions than other clusters. These patterns may represent users who do not drive every day, or in the case of Cluster 3, people working night shifts.

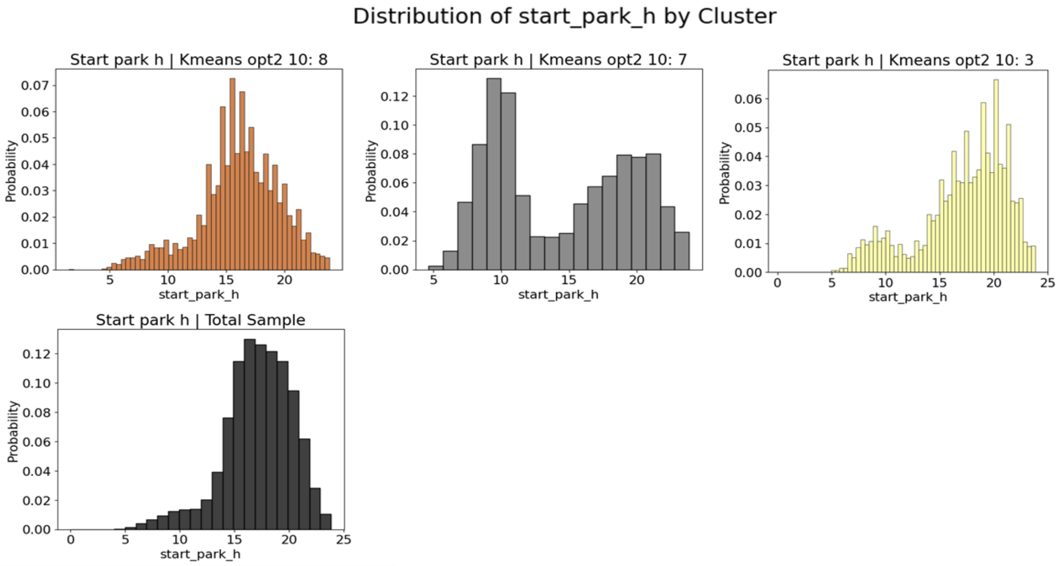


Figure 31 Start parking hour of Cluster 3, 7, 8, and total sample (Denmark).

4.5.3. Forecasting Results

All the time series were stable and stationary. Seven lags were set for the forecasting (1-6, and 24). Exponential smoothing with encoders was the best performing method, whereas Light GBM and XGBoost with Poisson were the worst performing models. Cluster 7 has the smallest errors whereas Clusters 9 and 5 the biggest ones (Table 16).

Table 17 Forecasting - MAE results (Denmark).

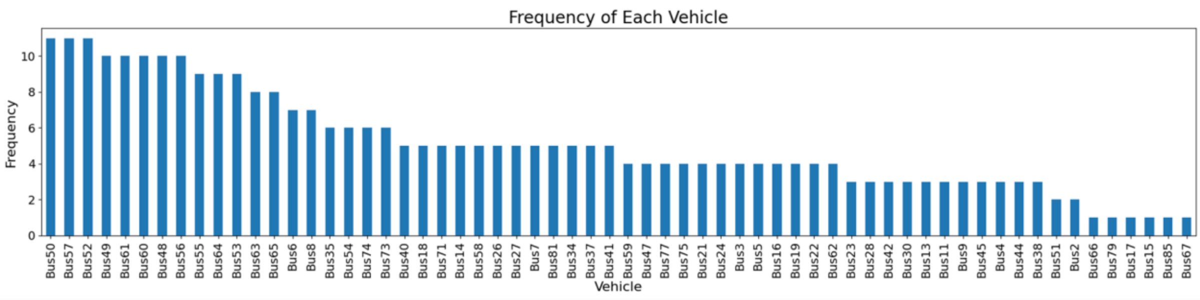
Forecasting Model / MAE	Clusters										
	1	2	3	4	5	6	7	8	9	0	AVG
ETS w/encoders	7.78	8.34	4.27	6.24	9.23	6.01	2.18	4.20	8.18	7.57	6.40
ETS w/encoders (log transform)	8.33	8.47	4.38	6.18	9.17	6.51	2.15	4.19	8.32	7.58	6.53
Light GBM - Poisson	7.69	8.61	5.28	7.50	10.0	6.91	3.17	5.55	7.85	8.48	7.10
Light GBM – Tweedie	7.38	8.13	4.62	7.50	8.83	6.66	2.53	5.39	7.63	8.37	6.70
Prophet	7.69	8.31	3.93	6.68	8.71	5.75	2.02	5.01	8.38	7.90	6.44
Prophet (log transform)	8.59	8.79	4.48	6.83	9.29	6.68	2.03	5.11	8.82	8.33	6.89
Random Forest	7.67	7.84	4.64	6.97	8.71	6.82	2.65	4.83	7.51	7.90	6.55
XGBoost - Poisson	7.62	8.21	4.69	7.49	8.66	6.49	2.83	5.64	7.54	7.83	6.70
XGBoost – Tweedie	7.75	8.22	4.79	7.46	8.53	6.55	3.03	5.74	7.63	7.87	6.76

4.6. Milanese e-Bus Dataset Data Imputation

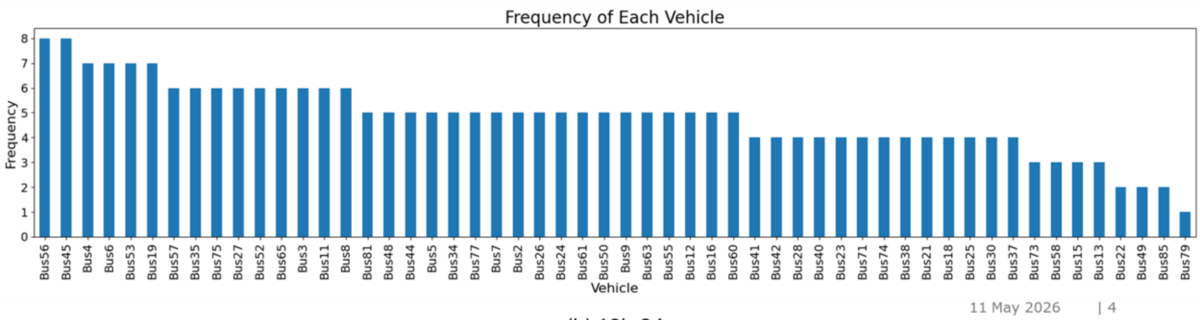
As previously mentioned, the proposed methodology was not fully applied to this dataset. Nevertheless, a clustering approach was implemented to address the large amount of missing information, and its results are presented in this section.

Figure 32 highlights the strong imbalance in the frequency of charging sessions across buses for the 14lu24 and 18lu24 data collection periods. While some buses are well represented, others have only a limited number of recorded events, indicating incomplete operational coverage across vehicles and routes. In addition, there is also a low availability of usable data for flexibility analysis (e.g., idle time). For instance, only 16.22% (42 samples in 14lu24) and 7.30% (23 samples in 18lu24) of the data contain sufficient information to estimate flexibility envelopes. Overall, just 15.51% of the dataset (284 observations) meets the minimum requirements for flexibility assessment.

Furthermore, even within the more complete data entries, most observations only allow the calculation of one idle time (temporal flexibility) component, either before or after charging, since both are rarely recorded simultaneously. This limitation, combined with sparse data for certain routes and vehicles, restricts the direct estimation of total idle time and other flexibility metrics. As a result, data pre-processing and missing data handling are critical to ensure that the analysis is not biased toward a small subset of well-documented cases and can better reflect the overall operational behaviour of the fleet.



(a) 14lu24



(b) 18lu24

Figure 32 Milan: Frequency of charging sessions by vehicle for the observation periods (a) 14lu24 and (b) 18lu24.

4.6.1. Feature Selection

Because a large portion of the dataset contains missing or incomplete operational information, it was necessary to identify the most relevant variables that could explain charging flexibility. To achieve this, a correlation analysis was conducted between the available operational variables and the calculated temporal flexibility indicators (*beforeFlexiblity* and *afterFlexibility*). Where Before (Temporal) Flexibility (equation 3) is defined as the time interval between a bus returning to the depot and the start of its charging process.

$$\textbf{\textit{Before Flexibility}} = \textbf{\textit{Start of Charging Time}} - \textbf{\textit{Return Time}} \quad \textbf{(3)}$$

And After (Temporal) Flexibility (equation 4) is defined as the time interval between the completion of charging and the scheduled departure of an electric bus from the depot.

$$\textbf{\textit{After Flexibility}} = \textbf{\textit{Departure Time}} - \textbf{\textit{End of Charging Time}} \quad \textbf{(4)}$$

Two correlation heatmaps were generated to examine the relationships among the variables (Figure 33). The first heatmap (Figure 33a) evaluates correlations between battery and charging characteristics, including *SOC_start*, *SOC_end*, *Energy_kWh*, *MaxPower_kW*, and the idle time/temporal flexibility indicators (*beforeFlexibility* and *afterFlexibility*). The second heatmap analyses time-related operational variables (Figure 33b), such as *StartHour*, *EndHour*, *DayOfWeek*, *DayOfMonth*, and *Month*, together with the idle time metrics.

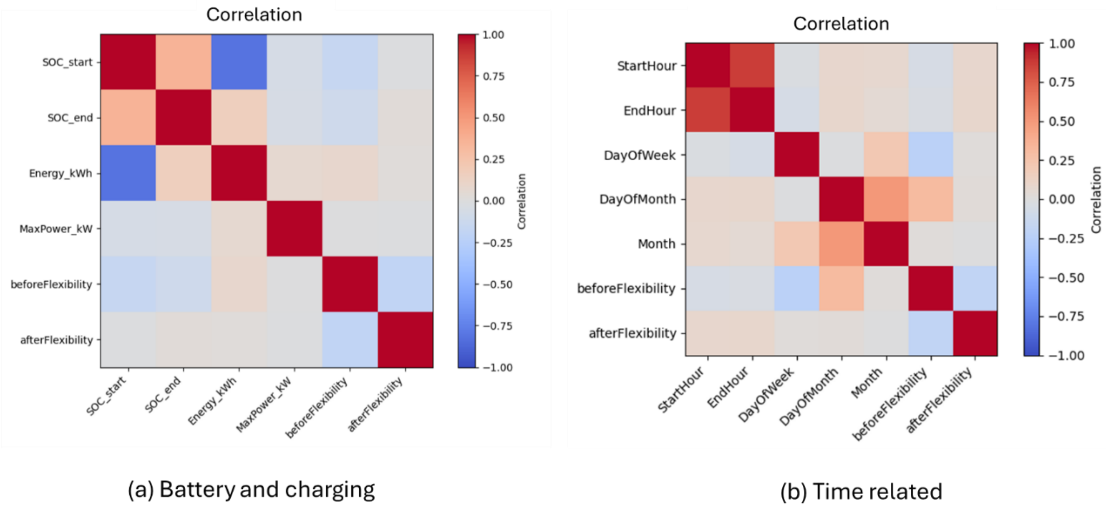


Figure 33 Milan: (a) correlation heatmap between battery and charging variables; (b) correlation heatmap between time-related operational variables idle time (temporal flexibility) metrics.

The correlation analysis revealed several meaningful relationships. The SoC at the beginning and end of charging (*SOC_start* and *SOC_end*) show a clear relationship with the amount of energy delivered during the charging session. For example, *SOC_start* has a strong negative correlation with *Energy_kWh* (-0.817), which is expected since buses that arrive with a lower battery level require more energy to recharge. Therefore, these variables provide relevant information about the charging state and indirectly about the available idle time (temporal flexibility).

Charging start and end times also show a strong correlation (≈ 0.87), indicating that charging sessions follow relatively consistent operational schedules. Although their direct correlation with flexibility temporal metrics is moderate, these variables capture the temporal structure of bus operations, making them important predictors.

Calendar-based variables such as *DayOfWeek* and *DayOfMonth* were also considered. These variables help represent operational patterns in the public transport system, since bus schedules and charging behaviour can vary depending on weekday service intensity, weekend operations, or seasonal variations.

Based on this analysis, a subset of features was selected for the idle time estimation models:

- Before Idle Time (temporal *beforeFlexibility*): The selected input variables are *StartHour*, *EndHour*, *SOC_start*, *SOC_end*, *DayOfWeek*, *DayOfMonth*.
- After Idle time (temporal *afterFlexibility*): The selected variables are *StartHour*, *EndHour*, *DayOfWeek*, *DayOfMonth*, *SOC_start*, *SOC_end*, *before_real_imputed_median*.

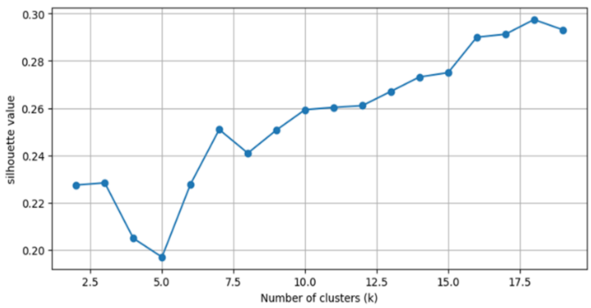
This feature selection process reduces the dimensionality of the dataset while preserving the variables that are most closely related to operational behaviour and charging conditions.

4.6.2. Optimal Number of KNN Clusters

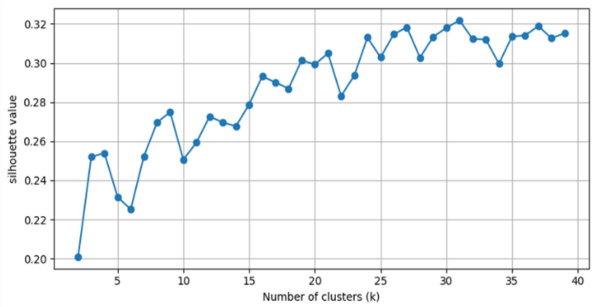
After selecting the relevant features, a KNN clustering was applied to then estimate the missing idle times values in the dataset. To determine the appropriate number of clusters, only the silhouette method was used.

For the before idle time (*beforeFlexibility*) variable the silhouette scores calculated for different numbers of clusters (k) are shown in Figure 34a. The results indicate that the silhouette value gradually increases as the number of clusters grows. The highest score is observed around k = 18, suggesting that, from a purely theoretical perspective, 18 clusters would provide the best separation among the observations. Nonetheless, using a large number of clusters can reduce the number of data points within each group, which becomes problematic when the dataset already contains many missing values.

For this reason, a practical compromise was adopted. Instead of using the theoretical optimum of 18 clusters, the analysis uses 9 clusters. This number still preserves meaningful differences between operational patterns while ensuring that each cluster contains a sufficient number of observations. As a result, the clustering structure can be applied to the entire dataset and used to estimate missing before idle time (*beforeFlexibility*) values more reliably.



(a) Before Flexibility



(b) After Flexibility

Figure 34 (a) shows the silhouette scores for different cluster numbers (k) in the before-flexibility case, while (b) shows the results for the after-flexibility configuration.

A similar clustering approach was applied to estimate missing idle time after charging (temporal *afterFlexibility*) values. Although the highest score was obtained at $k = 31$ Figure 34b, this would, once again, result in many small clusters with insufficient data points for reliable estimation.

To address this, a practical solution of 8 clusters was adopted. This configuration captures the main operational patterns while ensuring sufficient data within each cluster, providing a balanced and reliable basis for imputing missing after flexibility values.

4.6.3. Flexibility Data Imputation Results

After identifying the relevant features and applying the KNN-based estimation approach, the missing temporal flexibility values were addressed. The clustering groups allowed observations with similar operational characteristics to be used for estimating the missing flexibility metrics. Once these idle times were defined flexibility envelopes (equation 2) could be calculated as well.

For idle time before charging (temporal *beforeFlexibility*), a total of 1,547 values were missing in the original dataset. Using the clustering approach with 9 clusters and the selected features (*StartHour*, *EndHour*, *SOC_start*, *SOC_end*, *DayOfWeek*, *DayOfMonth*), all missing observations were successfully estimated. As a result, 1,547 values were filled, and no estimation failures occurred.

For idle time after charging (temporal *afterFlexibility*), the dataset also contained 1,547 missing values. Using 8 clusters and the selected features (*StartHour*, *EndHour*, *DayOfWeek*, *DayOfMonth*, *SOC_start*, *SOC_end*, *before_real_imputed_median*), most missing observations were estimated. In this case, 1,364 values were successfully filled, while 183 observations could not be estimated, mainly due to insufficient similarity with other observations or limited available information within some clusters.

Overall, this process significantly improved the completeness of the dataset. By reconstructing the missing flexibility values, the dataset now provides a much broader representation of charging operations across vehicles, time periods, and operating conditions. This improvement enables more reliable analysis of charging flexibility and allows the dataset to be used for further modelling and operational studies related to electric bus charging management.

4.7. Milanese e-Bus Preliminary Clustering Analysis

Applying another clustering technique after a KNN-based imputation would not be methodologically sound, as the imputation process itself introduces artificial similarity into the data. As a result, any subsequent clustering may reflect these imposed relationships rather than genuine underlying patterns, leading to biased or circular results.

For this reason, the clustering step of the methodology was performed on the original dataset entries that had enough information to calculate one of the idle time components (after or before temporal flexibility) to preserve its inherent structure and avoid bias.

Moreover, given the characteristics of the dataset, only the clustering results for the flexibility envelopes calculated using the after-charging idle time (temporal *afterFlexibility*) are presented. A similar analysis could be conducted using before-charging idle time, or ideally, clustering should be performed on the total temporal flexibility, combining both before- and after-charging idle periods.

Table 18 Clustering evaluation (Milan).

Clustering Model	Parameters	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index	Configuration
K-Means	2	0.63	724.40	0.49	Configuration 2
K-Means	3	0.59	908.58	0.50	Configuration 2
K-Means	9	0.55	1,778.03	0.51	Configuration 2
K-Means	4	0.52	878.25	0.55	Configuration 2
K-Means	4	0.45	229.47	0.73	Configuration 1
K-Means	8	0.44	266.25	0.71	Configuration 1
K-Means	9	0.41	266.51	0.75	Configuration 1
K-Means	3	0.41	173.50	0.88	Configuration 1

Based on the clustering evaluation metrics k=2 (Configuration 2) offers the best Silhouette score, whereas the best Caliski-Harabasz indexes are from k=3 and k=9 (Configuration 2). All these three options have similar Davies-Bouldin index values (Table 17). Given that k=9 would be a value very big for the current sample size, the analysis proceeded with k=3.

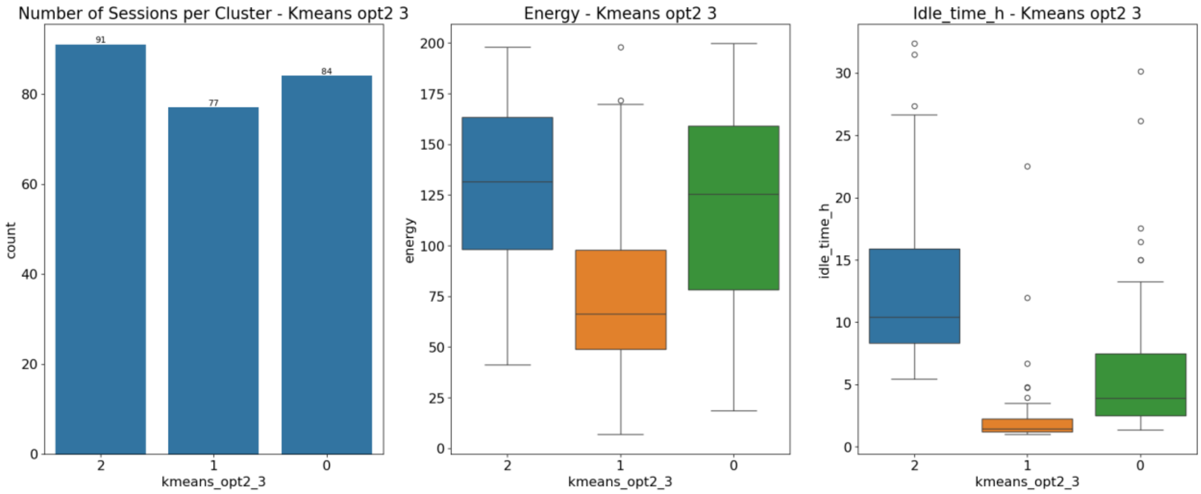


Figure 35 Clustering boxplots (Milan).

The clustering results identify three clusters of similar size (Figure 35). Although the groups were formed based on flexibility features, energy and after idle time in this case, their distribution across the time of day is also clearly distinct. These clusters can be broadly interpreted as three operational patterns: afternoon-to-evening charging with high flexibility (Cluster 2), mid-day charging with the lowest flexibility (Cluster 1), and late-night charging with moderate flexibility (Cluster 3) (Figure 36). Similar to the residential charging occasion analysed in this deliverable, there was no No-Flexibility cluster.

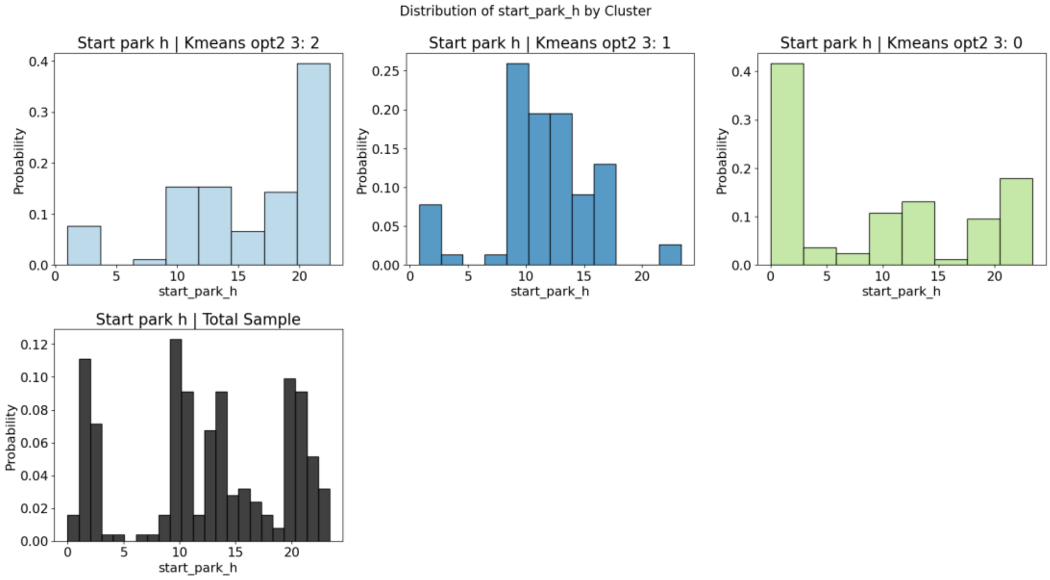


Figure 36 Start parking hour per cluster (Milan).

Overall, the three-cluster solution captures key operational patterns in a simple and interpretable way, distinguishing the flexibility potential of e-bus charging sessions throughout the day. Further analysis with a more complete dataset or using total idle time (total temporal flexibility), may justify a higher number of clusters to better capture variability across bus routes. This could provide more granular insights to support route optimisation and smart charging strategies.

4.8. Madeira Dataset (Port Preliminary Clustering Analysis)

The Madeira port vessel charging dataset comprises one week of simulated data for 10 vessels, totalling 122 sessions. The charging happens more often during weekdays, particularly on Mondays, which also the start day of the simulation (Figure 37).

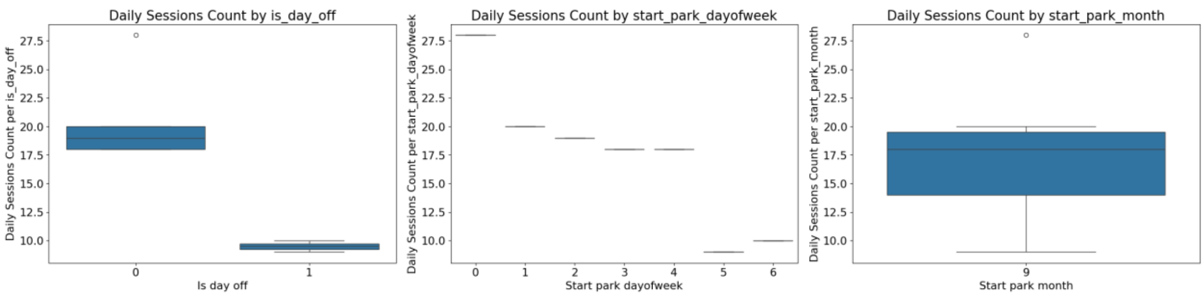


Figure 37 General descriptive data (Madeira Port).

Given the small sample size, the BIRCH algorithm was not tested, and the clustering configurations for selecting the optimal number of clusters also included the possibility of only two groups. Table 18 show the top results for the Silhouette, Calinski-Harabasz, and Davies-Bouldin metrics.

Table 19 Clustering evaluation (Madeira Port).

Clustering Model	Parameters	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index	Configuration
K-Means	**3**	**0.93**	**4,792.49**	**0.07**	**Configuration 2**
K-Means	2	0.93	4,663.67	0.11	Configuration 2
K-Means	2	0.89	433.66	0.08	Configuration 4
K-Means	2	0.68	306.14	0.42	Configuration 1
K-Means	3	0.66	924.46	0.35	Configuration 4
K-Means	4	0.52	148.06	0.51	Configuration 3

Overall, the evaluation metrics (Table 18) indicate that k = 3 (Configuration 2) provides the best clustering solution, followed by k = 2 (Configuration 2). Given the small sample size, a low number of clusters is expected. In the case of k = 2, the clustering essentially separates sessions into two broad groups, the flexible sessions and low-to-no flexibility ones (Figure 38). The second cluster includes instances with some idle time, although in most cases this remains below 20 minutes.

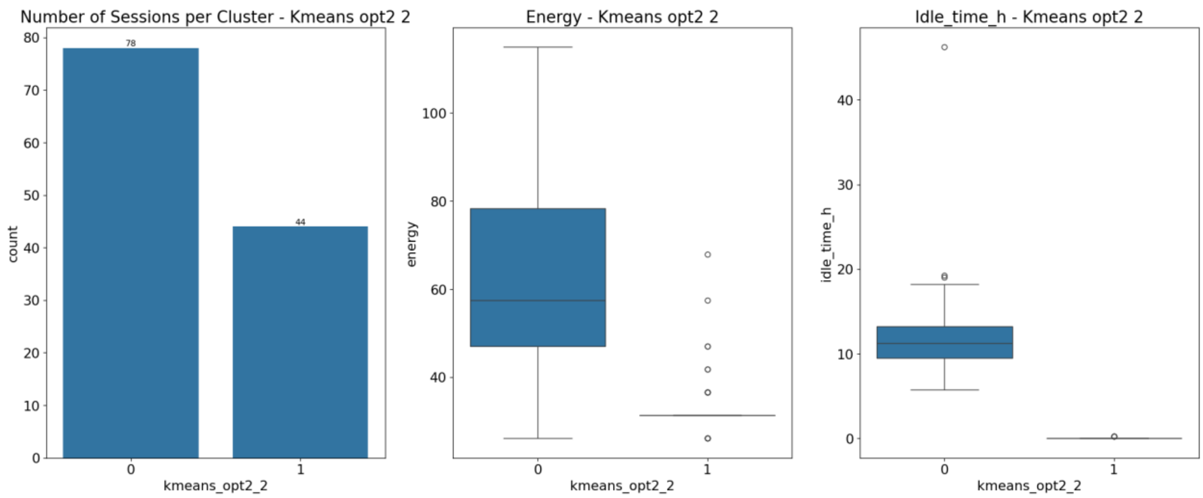


Figure 38 Clustering boxplots (Madeira Port, k=2).

When looking at the cluster sizes in the k=3 solution (Figure 39), the low-to-no-flexibility group from k=2 is split into two, with the No-Flexibility cluster being very small and containing only four observations. Further analysis showed that these entries correspond to different vessels with similar charging patterns, suggesting that they are not errors but rather represent a specific, but uncommon, operational behaviour. With a larger and more representative dataset, a higher number of clusters could provide greater granularity in identifying different levels of flexibility. And, while k = 3 performs best for the current sample, further analysis with more comprehensive data might reveal that k = 2 or k = 4 better capture the flexibility needs of port charging operations.

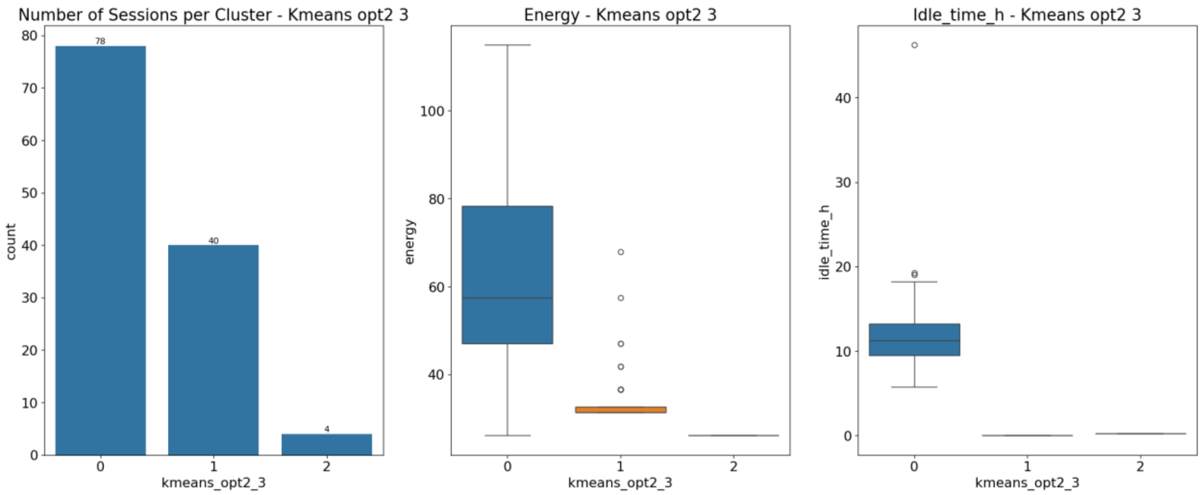


Figure 39 Clustering boxplots (Madeira Port, k=3).

4.9. Discussion

Flexibility in the electric mobility sector has a broad definition and has been quantified and operationalised in several ways across different level of the grid, and by different operators. The present deliverable analysed the flexibility potential of charging sessions after vehicle connection. Meaning that this study does not account for potential pre-connection flexibility incentives such as economic, social, and psychological factor, which may influence the decision of when, where, and for how long to charge a vehicle.

4.9.1. Datasets comparison

Across the explored datasets, the flexibility potential of kerbside charging sessions in Boulder, Funchal, and Slovenia proved to be relatively limited, especially when compared to residential charging profiles. Office and campus charging showed higher flexibility potential than kerbside but below residential availability levels.

Even though the cluster centroids differed, the ideal number of clusters for the kerbside charging datasets was the same (four clusters). For Slovenia and Boulder, five clusters also appeared plausible. However, given the sizes of the datasets and the need for sufficiently large sub-samples for the subsequent forecasting step, this setting was not adopted. The proportion of entries per cluster was also similar across datasets, with most data belonging to the no-flexibility cluster and the fewest to the high-flexibility cluster.

In terms of arrival time for kerbside locations, high-flexibility clusters tended to start later in the day and stay connected overnight, mid-flexibility clusters were concentrated in the morning, minimal-flexibility clusters showed a bimodal distribution with a stronger morning peak, and the no-flexibility cluster closely resembled the overall sample distribution. This shows that common temporal and flexibility patterns exist across datasets representing the same charging occasion, despite differences in geographic location (Southern Europe, Western Europe, and the US) and scale (city versus country).

As far as office, campus and residential charging sessions are concerned, the cluster-specific patterns were also aligned with expected EV-driver behaviour for this charging occasions. For example, the weekend-oriented cluster peaked later in the morning, and clusters with higher flexibility potential due to very long idle times and high energy delivered seem to reflect users who charge less frequently but for longer. For the campus/office dataset the morning arrival peak was predominant with the departure peak matching the expected schedule of a nine to five office.

These patterns should be considered when activating these clusters for specific flexibility services. For instance, office and campus cluster groups, weekend residential sessions, and kerbside mid-flexibility clusters are more appropriate for PV-load absorption, whereas night-time clusters may be relevant for peak shaving or electricity bill optimisation strategies.

Naturally the location where these clusters occur and where flexibility is needed also plays a role in their activation. The brief geo-location profiling done for the Slovenian case adds depth and actionable insights to this type of analysis. In the analysed dataset, spatial information was integrated into the profiling, and the findings suggest that flexibility-related charging behaviour is somewhat spatially structured, whereas non-flexible sessions follow a more general baseline pattern, occurring across diverse environments. Similarly, the patterns observed in the Danish residential dataset and the ACN office location were aligned with people’s daily activities and working schedules. These observations support the idea that charging behaviour is closely linked to human routines and mobility patterns [55].

Regarding charging infrastructure, the Slovenian and ACN datasets contained both AC and DC charging, whereas the Danish residential, Madeira electrified vessels, and Boulder charger were AC only (<22 kW). In contrast, the Milanese e-Bus depot dataset consisted exclusively of DC chargers, which reflects the high energy demand and operational requirements of e-buses. The Madeira EV dataset included infrastructure capable of supporting DC charging, although within the dataset it operated only at AC charging power levels.

These differences are particularly relevant from a grid integration and flexibility perspective. AC chargers, especially residential ones, are typically connected to LV grids, where flexibility services are mainly associated with local load management, peak shaving, or voltage support. DC fast chargers, on the other hand, are commonly connected to MV grids due to their higher power demands, enabling different forms of flexibility provision such as higher-capacity balancing services.

This distinction is also reflected in the clustering results. In the Slovenian dataset, low-flexibility clusters tended to contain a higher concentration of DC chargers. Similarly, in the ACN dataset, the no-flexibility cluster was associated with the highest registered charging powers, although not on average. This suggests that high-power charging sessions may be linked to more constrained charging needs, such as highway charging, where charging speed and immediacy are prioritised. This makes direct flexibility activation less feasible, with the exception of e-buses depot DC charging. Even though the presented results are preliminary only, the clusters showed overall moderate to high flexibility potential and may therefore be suitable for flexibility services. Their less stochastic charging windows further supports this potential.

Nevertheless, in low-flexibility clusters, price-based incentives might play an important role. Rather than activating flexibility directly during charging sessions, dynamic tariffs or differentiated pricing schemes can indirectly influence user behaviour. This could encourage charging at different times, the selection of alternative charging locations, or a reduced reliance on high-power charging when it is not strictly necessary. Thus, economic signals may too contribute to shaping charging demand in ways that support grid operation and congestion management.

4.9.2. Theoretical Implications and Future Research

Despite using a “big-data” clustering technique in which different algorithm parameters such as the threshold parameter h, which controls the maximum diameter of sub-cluster, were tested, BIRCH was consistently outperformed by K-Means. As datasets continue to grow with increasing data availability and more consumers shifting to electric transport modes, future research could explore alternative large-scale clustering techniques or extensions of the classical BIRCH model.

The profiling step of the methodology combines multiple techniques, including visualisations (histograms, boxplots, and bar charts), descriptive statistics (minimum, maximum, mean, etc.), geo-location analysis, non-parametric significance tests, and a KDE step to estimate cluster-based probabilities of features relevant for planning and flexibility deployment. While KDE is valuable, it is one of the most time- and memory-intensive steps, especially for the Danish dataset with over 300,000 sessions. This load can be reduced by limiting the number of features analysed or reducing the number of parameters tested on KDE cross-validation as currently the code tests Gaussian and Exponential functions. The other two most resourceintensive steps are the calculation and evaluation of the optimal clustering and the forecasting procedure (post preliminary analysis).

One way to further reduce the running time of the algorithm is to decrease the number of tested configurations. For instance, this can be achieved by limiting the number of parameter combinations tested or by running only one clustering technique (e.g., BIRCH or K-Means). Given the exploratory nature of this deliverable, testing a wide range of configurations was necessary. However, future analyses could narrow the set of options and start from the most promising settings of this deliverable. Additionally, the Silhouette score is not recommended for very large datasets because its computation scales with N^2, making it computationally expensive.

The geo-location cluster profiling presented in this analysis sets the groundwork for further research into the spatial nuances of charging flexibility. While existing literature addresses aspects of this topic, flexibility is often analysed primarily through power, energy, or time-based metrics. In this study, a session-level approach was applied to directly link charging flexibility patterns to the environments in which they take place. Future research could extend this by focusing on the EVSE infrastructure itself and explore the spatial typology of the charging stations. Tree-based algorithms or other machine learning models could also be applied to enable a more detailed behavioural analysis. However, tree-based algorithms should be used carefully with unbalanced groups as they can be biased towards larger sub-samples.

In terms of forecasting, statistical models from the ARIMA family were not explored, as they tend to be computationally slow for very large datasets. Among the models tested, Exponential Smoothing, Prophet, and Random Forest were the most time-consuming. For the defined sliding window, with retraining every 24 hours, the Random Forest algorithm required approximately one hour to run across all groups in the Danish dataset (over 300,000 sessions), with Exponential Smoothing (ETS) and Prophet taking even longer. In contrast, the other machine learning techniques were significantly faster, requiring less than ten minutes in total.

These running times can be reduced, for instance, by decreasing the retraining frequency, number of trees and/or shortening the window length. The use of the SKforecast and Darts packages enabled intuitive multi-series modelling and avoids data leakage. Future research should consider such auxiliary packages, or equivalent, when developing machine learning forecasting frameworks.

Additionally, further studies may explore alternative clustering and forecasting techniques as well as different sets of variables. For instance, the inclusion of weather metrics could improve the forecast. Combining multiple techniques (e.g., ensembles) might also lead to smaller prediction errors. Regarding lag and window features, this deliverable used sine/cosine transforms of the hour and week to capture seasonality within these periods. In addition, it included not only lag features but also rolling attributes, such as weekly, 24-hour, and 12-hour averages, the weekly coefficient of variation, and a dummy lag-1 variable (1 if t-1 had sessions, 0 otherwise). While this may increase endogeneity, it can also be an effective way to predict values for clusters with many zeros, as it was the case for kerbside high-flexibility clusters. Testing models with different timestamp transforms (e.g., wavelets), numbers of lags and window features through a fine-tuning step could further improve forecasting results.

MAPE was not used as a forecasting evaluation metric because the dependent variable contained instances of the value 0. Instead, MAE and MSE were used. An important goal of the methodology was to identify which forecasting model performs better overall. Nevertheless, using MAE and MSE limits the analysis to comparing methods rather than evaluating whether certain clusters are predicted more accurately than others. This is particularly relevant when cluster sizes are unbalanced like in this deliverable. Future algorithms can address this limitation by using an evaluation metric that allows a direct comparison of forecasting accuracy across clusters, rather than only across methods.

The deliverable code generates several outputs, including Excel files with identified outliers, clustering evaluation metrics, key profiling attributes such as descriptive statistics and significance test results, and probability distribution functions for each group. The clustering models themselves are stored as pickle objects that can be retrieved when needed. In addition to reducing running time, this supports reproducibility. Forecasting preliminary results, MAE values and run times, are also stored in an Excel file. These outputs are complemented by visualisations, including plots and charger location maps, when such information is available.

A limitation of the algorithm is that flexibility potential is assessed in a pre-activation context, meaning that the flexibility envelopes used always reflect the full theoretical flexibility of a session. As a result, flexibility is not dynamically updated once a charging session has started, and any reduction in available flexibility during the session is not captured. Given that the primary goal of the current methodology is planning and profiling, this limitation is not an issue for this analysis. However, it should be considered when moving to scheduling or real-time operations, where updated and adjusted flexibility estimates are required.

Regarding e-buses, the full methodology proposed in this deliverable was not applied due to limited dataset size, missing data, and interrupted time spans. A preliminary clustering-based flexibility analysis using the proposed approach, although the forecasting stage was not implemented. The same happens for the port vessel-charging case. Even if there were larger datasets for these cases, it is important to note that depot and port operations from an aggregated perspective follow predefined schedules and fixed routes where users know what bus routes exist and their schedule, often involving a limited known number of vehicles. As a result, the forecasting component would likely provide less additional insight compared to more stochastic EV charging occasions (e.g., kerbside charging). For these more closed systems, simpler probabilistic forecasting approaches might be sufficient.

Despite these limitations, the high proportion of missing information about key features for flexibility calculation such as idle time, made this dataset suitable for exploring an alternative application of clustering. In this context, a KNN algorithm was used as part of the data pre-processing stage, alongside filtering incomplete records and addressing missing values. This method enabled a more reliable imputation of missing temporal (idle time) values while preserving meaningful operational patterns. Overall, this process significantly improved dataset completeness, allowing for a more representative analysis of charging behaviour across buses, time periods, and operating conditions, and supporting further modelling and operational studies related to e-bus charging management such as smart charging scheduling.

5. CONCLUSION

Task 3.3 developed an AI-based methodology to support the integration of EVs as flexibility resources across multiple grid levels. The model is designed for integration into planning, simulation, and operational tools developed in subsequent work packages.

The proposed methodology enables the identification and prediction of EV charging flexibility using AI-based clustering and forecasting techniques. Clusters are defined based on flexibility potential, characterised by energy and idle time envelopes, and are then utilised to predict flexibility availability throughout the day.

These clusters can be used to simulate flexibility profiles in empirical datasets, while the forecasting results can support scheduling optimisation, pricing models, and bidding strategies. From an operational perspective, clustering-based flexibility activation can also reduce the risk of violating charging constraints by prioritising highly flexible sessions.

A key of finding deliverable is that EV flexibility shows a strong and consistent temporal structure across contexts, indicating well-defined flexibility windows. The highest flexibility potential is associated with long-duration charging sessions, particularly overnight kerbside charging and extended residential stays. These conditions enable sustained flexibility provision due to high idle times. Residential charging during weekends and late mornings also represents highly flexible periods with prolonged connection durations. Office and campus charging provides moderate flexibility, mainly during midday idle periods between arrival and departure. In contrast, kerbside charging is the most constrained charging occasion with flexibility mainly concentrated in evening-to-overnight periods.

These results highlight that EV flexibility is strongly time-dependent and closely aligned with human behavioural routines, with statistically consistent but inherently stochastic activation periods emerging across different charging environments.

From a methodological perspective, the framework is transferable across diverse datasets and provides a scalable basis for integrating EV flexibility into future energy system planning and operation.

5.1. Summary

Seven datasets were analysed, including kerbside, office/campus, residential charging occasions, as well as preliminary e-bus and vessel cases. Residential charging had the highest levels of flexibility, followed by office or campus charging. On the other hand, Kerbside charging exhibited the lowest flexibility levels, with the no-flexibility cluster being predominant. The preliminary analysis of e-bus and vessel charging also revealed distinct operational charging patterns and flexibility groups, with a higher proportion of sessions classified in flexible clusters compared to low- or no-flexibility groups.

Despite geographical and scale differences, kerbside datasets showed consistent structural patterns, with similar cluster distributions and temporal behaviour. Most sessions from these datasets were concentrated in low-flexibility clusters. Office, campus, and residential datasets closely align with expected behavioural patterns for these charging occasions.

Additionally, geo-location profiling (done for the Slovenian case) suggests that flexible charging behaviour is somewhat location-dependent, while non-flexible behaviour is more uniform. This further supports the link between charging behaviour and human routines and mobility patterns.

These results highlight the importance of combining temporal and spatial context in flexibility activation strategies. For instance, office and campus cluster groups, weekend residential sessions, and kerbside mid-day clusters are more appropriate for PV-load absorption, whereas night-time clusters may be more appropriate for peak shaving or electricity bill optimisation strategies.

5.2. Progress Achieved and Lessons Learned

The proposed methodology proved robust and transferable across datasets of varying size, structure, and origin, including empirical, hybrid, and synthetic datasets.

A key achievement is the development of a lightweight clustering framework based solely on charging session attributes (energy and idle time), enabling flexibility estimation without requiring detailed user or vehicle-level data. Temporal information is introduced only after clustering for profiling and forecasting. This approach keeps the analysis focused and ensures that meaningful insights can be drawn from both the clustering and demand prediction stages.

Nevertheless, the absence of SoC, vehicle, and user-level data limits deeper behavioural and technical analysis. For instance, without SoC information, full Vehicle-to-everything (V2X) solutions cannot be analysed. Similarly, while price models based on charging supply and demand can established based on this methodology, the impact of economic incentives or the participation behaviour cannot be fully captured without user related data.

In terms of performance, K-Means provided better cluster separation than BIRCH across the analysed datasets. For the forecasting, although the best performing model varied depending on the dataset, Random Forest consistently ranked among the best-performing models. Overall, the results demonstrate the applicability of the methodology across different EV charging contexts.

The methodology is designed for pre-activation analysis, meaning flexibility is static and not dynamically updated during charging sessions. While suitable for planning and profiling, future work should extend this to real-time operational settings.

5.3. Upcoming Steps and Deliverables

This deliverable provides a robust profiling approach based on flexibility envelopes, where among other techniques, KDE is used to characterise the distribution of charging sessions within each cluster. These profiles can then be used to support simulation. Moreover, the resulting clusters are then used as the basis for time series forecasting. This enables the prediction of the number of active/new sessions per hour in a day-ahead scenario, which can be used as a data input for the *T3.4 – Flexibility simulation tools* and support the prequalification recommendations in *T3.5 – Flexibility contracting and pre-qualification*.

Given the broad scope of techniques tested in this deliverable, relevant parts of the algorithm can be later integrated into the operational tools developed in *WP4 - Grid and spatial mapping models* and *WP6 - AI tool for planning of grids and charging infrastructure*.

Finally, due to its profiling component, the developed methodology can potentially contribute to the analysis of the results of the demo WPs (*WP7 – Nordic Europe demo*, *WP8 – Southern Europe Demo*, *WP9 – Western Europe Demo*). In this deliverable, the e-bus and vessel datasets were only briefly explored. As previously mentioned, the main reason is that, currently, only small datasets are available, which are not yet fully representative of these charging scenarios. As the demo work packages progress and more information becomes available, a full application of the D3.3 methodology will be possible and could be incorporated into the results discussion of these WPs.

6. REFERENCES

1. Olivella-Rosell, P.; Lloret-Gallego, P.; Munné-Collado, Í.; Villafafila-Robles, R.; Sumper, A.; Ottessen, S.Ø.; Rajasekharan, J.; Bremdal, B.A. Local Flexibility Market Design for Aggregators Providing Multiple Flexibility Services at Distribution Network Level. *Energies (Basel).* **2018**, *11*, doi:10.3390/en11040822.

2. Plaum, F.; Ahmadiahangar, R.; Rosin, A.; Kilter, J. Aggregated Demand-Side Energy Flexibility: A Comprehensive Review on Characterization, Forecasting and Market Prospects. *Energy Reports* 2022, *8*, 9344–9362, doi: 10.1016/j.egyr.2022.07.038.

3. Gschwendtner, C.; Knoeri, C.; Stephan, A. Mind the Goal: Trade-Offs between Flexibility Goals for Controlled Electric Vehicle Charging Strategies. *iScience* **2023**, *26*, doi:10.1016/j.isci.2023.105937.

4. Nazari, M.; Hussain, A.; Musilek, P. Applications of Clustering Methods for Different Aspects of Electric Vehicles. *Electronics (Switzerland)* **2023**, *12*, doi:10.3390/electronics12040790.

5. Märtz, A.; Langenmayr, U.; Ried, S.; Seddig, K.; Jochem, P. Charging Behavior of Electric Vehicles: Temporal Clustering Based on Real-World Data. *Energies (Basel).* **2022**, *15*, doi:10.3390/en15186575.

6. Wagner, L.P.; Reinpold, L.M.; Kilthau, M.; Fay, A. A Systematic Review of Modeling Approaches for Flexible Energy Resources. *Renewable and Sustainable Energy Reviews* **2023**, *184*, 113541, doi:10.1016/j.rser.2023.113541.

7. Lee, Z.J.; Li, T.; Low, S.H. ACN-Data: Analysis and Applications of an Open EV Charging Dataset. In Proceedings of the e-Energy 2019 - Proceedings of the 10th ACM International Conference on Future Energy Systems; Association for Computing Machinery, Inc, June 15 2019; pp. 139–149, doi: 10.1145/3307772.3328313.

8. City of Boulder Electric Vehicle Charging Station Data 2023. https://open-data.bouldercolorado.gov/maps/95992b3938be4622b07f0b05eba95d4c

9. Knezović, K.; Marinelli, M.; Zecchino, A.; Andersen, P.B.; Traeholt, C. Supporting Involvement of Electric Vehicles in Distribution Grids: Lowering the Barriers for a Proactive Integration. *Energy* **2017**, *134*, 458–468, doi:10.1016/j.energy.2017.06.075.

10. Hamwi, M.; Lizarralde, I.; Legardeur, J. Demand Response Business Model Canvas: A Tool for Flexibility Creation in the Electricity Markets. *J. Clean. Prod.* **2020**, *282*, 124539, doi:doi.org/10.1016/j.jclepro.2020.124539.

11. Degefa, M.Z.; Sperstad, I.B.; Sæle, H. Comprehensive Classifications and Characterizations of Power System Flexibility Resources. *Electric Power Systems Research* 2021, *194*, doi: 10.1016/j.epsr.2021.107022.

12. Chemudupaty, R.; Hornek, T.; Pavić, I.; Potenciano Menci, S. Optimizing Trading of Electric Vehicle Charging Flexibility in the Continuous Intraday Market under User and Market Uncertainties. *Appl. Energy* **2025**, *381*, doi:10.1016/j.apenergy.2024.125103.

13. Massaoudi, M.; Davis, K.R.; Akramul Haque, K. Analysis and Quantification of Demand Flexibility for Resilient Distribution Networks: A Systematic Review. *IEEE Access* **2025**, *13*, 42650–42668, doi:10.1109/ACCESS.2025.3548526.

14. Gonzalez Venegas, F.; Petit, M.; Perez, Y. Plug-in Behavior of Electric Vehicles Users: Insights from a Large-Scale Trial and Impacts for Grid Integration Studies. *eTransportation* **2021**, *10*, doi:10.1016/j.etran.2021.100131.

15. Hu, J.; Zhou, H.; Zhou, Y.; Zhang, H.; Nordströmd, L.; Yang, G. Flexibility Prediction of Aggregated Electric Vehicles and Domestic Hot Water Systems in Smart Grids. *Engineering* **2021**, *7*, 1101–1114, doi:10.1016/j.eng.2021.06.008.

16. Zhang, Y.; Guo, Z.; Jiao, F.; Li, H.; Zhang, X.; Ukil, A. Joint EV Eco-Charging Strategy Considering Spatiotemporal Flexibility in Transportation- Distribution Networks. *IEEE Transactions on Transportation Electrification* **2025**, *11*, 11606–11618, doi:10.1109/TTE.2025.3579812.

17. Golab, A.; Loschan, C.; Zwickl-Bernhard, S.; Auer, H. The Value of Flexibility of Commercial Electric Vehicle Fleets in the Redispatch of Congested Transmission Grids. *Energy* **2025**, *316*, doi:10.1016/j.energy.2025.134385.

18. Lai, S.; Dong, Z.Y.; Yang, H.; Tao, Y.; Yip, C.; Qiu, J.; Zhao, J. Renewable Energy Following Strategy via Quantified Aggregated Heterogeneous Electric Vehicle Flexibility. *IEEE Trans. Smart Grid* **2025**, *16*, 3873–3886, doi:10.1109/TSG.2025.3580386.

19. Calearo, L.; Marinelli, M.; Ziras, C. A Review of Data Sources for Electric Vehicle Integration Studies. *Renewable and Sustainable Energy Reviews* 2021, *151*, doi: 10.1016/j.rser.2021.111518.

20. Marzbnai, F.; Osman, A.H.; Hassan, M.S. Electric Vehicle Energy Demand Prediction Techniques: An In-Depth and Critical Systematic Review. *IEEE Access* 2023, *11*, 96242–96255, doi: 10.1109/ACCESS.2023.3308928.

21. Xin, F.; Yang, X.; Beibei, W.; Ruilin, X.; Fei, M.; Jianyong, Z. Research on Electric Vehicle Charging Load Prediction Method Based on Spectral Clustering and Deep Learning Network. *Front. Energy Res.* **2024**, *12*, doi:10.3389/fenrg.2024.1294453.

22. Forte, M.; Morais, H.; Lascano, C. Clustering Applications in Electric Mobility, University of Lisbon, 2023. MSc thesis.

23. Yang, A.; Zhang, G.; Tian, C.; Peng, W.; Liu, Y. Charging Behavior Portrait of Electric Vehicle Users Based on Fuzzy C-Means Clustering Algorithm. *Energies (Basel).* **2024**, *17*, doi:10.3390/en17071651.

24. Genov, E.; Cauwer, C. De; Kriekinge, G. Van; Coosemans, T.; Messagie, M. Forecasting Flexibility of Charging of Electric Vehicles: Tree and Cluster-Based Methods. *Appl. Energy* **2024**, *353*, doi:10.1016/j.apenergy.2023.121969.

25. Guzman, C.P.; Gomes, E.; Pereira, L.; Morais, H. A Computational Implementation for Creating Electric Vehicles Profiles. In Proceedings of the 2023 IEEE PES Innovative Smart Grid Technologies Europe; IEEE, 2023, doi: 10.1109/ISGTEUROPE56780.2023.10407312.

26. Carlton, G.J.; Sultana, S. Electric Vehicle Charging Station Accessibility and Land Use Clustering: A Case Study of the Chicago Region. *Journal of Urban Mobility* **2022**, *2*, doi:10.1016/j.urbmob.2022.100019.

27. Abdullahi, M.R.; Lu, Q.C.; Hussain, A.; Tripura, S.; Xu, P.C.; Wang, S.X. Location Optimization of EV Charging Stations: A Custom K-Means Cluster Algorithm Approach. *Energy Reports* **2024**, *12*, 5367–5382, doi:10.1016/j.egyr.2024.09.075.

28. Kim, K.; Kim, G.; Yoo, J.; Heo, J.; Cho, J.; Ryu, S.; Kim, J. Data-Driven Clustering Analysis for Representative Electric Vehicle Charging Profile in South Korea. *Sensors* **2024**, *24*, doi:10.3390/s24216800.

29. Forte, M.; Guzman, C.P.; Pereira, L.; Morais, H. A Comprehensive Review of Clustering Methods Applications in Electric Mobility. *Transportation Engineering* **2025**, *21*, doi:10.1016/j.treng.2025.100360.

30. Forte, M.; Guzman, C.P.; Lekidis, A.; Morais, H. Clustering Methodologies for Flexibility Characterization of Electric Vehicles Supply Equipment. *Green Energy and Intelligent Transportation* **2025**, 100304, doi:10.1016/j.geits.2025.100304.

31. Shirkhorshidi, A.S.; Aghabozorgi, S.; Wah, T.Y.; Herawan, T. Big Data Clustering: A Review. In Proceedings of the Lecture Notes in Computer Science (including subseries

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer Verlag, 2014; Vol. 8583 LNCS, pp. 707–720, doi: 10.1007/978-3-319-09156-3_49.

32. Dave, M.; Gianey, H. Different Clustering Algorithms for Big Data Analytics: A Review. In Proceedings of the Proceedings of the 5th International Conference on System Modeling & Advancement in Research Trends : SMART-2016; Prof. Rakesh Kumar Dwivedi : Principal College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, 2016; p. 350.

33. Hanghang, T.; Kang, U. *Big Data Clustering*. In *Data Clustering*; 2018; pp. 259–276.

34. Lang, A.; Schubert, E. BETULA: Fast Clustering of Large Data with Improved BIRCH CF-Trees. *Inf. Syst.* **2022**, *108*, doi:10.1016/j.is.2021.101918.

35. Halkidi, M.; Batistakis, Y.; Vazirgiannis, M. On Clustering Validation Techniques. *J. Intell. Inf. Syst.* **2001**, *17*, 107–145, doi:10.1023/A:1012801612483.

36. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts, 2018; https://otexts.com/fpp3/.

37. Kolassa, S.; Rostami-Tabar; Siemsen Enno *Demand Forecasting for Executives and Professionals*; Chapman and Hall/CRC, 2023;

38. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: 100,000 Time Series and 61 Forecasting Methods. *Int. J. Forecast.* **2020**, *36*, 54–74, doi:10.1016/j.ijforecast.2019.04.014.

39. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. M5 Accuracy Competition: Results, Findings, and Conclusions. *Int. J. Forecast.* **2022**, *38*, 1346–1364, doi:10.1016/j.ijforecast.2021.11.013.

40. Amezquita, H.; Guzman, C.P.; Morais, H. *A Computational Implementation to Forecast Electric Vehicles Usage in the Power System*; **2024** IEEE 22nd Mediterranean Electrotechnical Conference (MELECON), 1375–1379, doi: 10.1109/MELECON56669.2024.10608512.

41. Buratto, W.G.; Muniz, R.N.; Nied, A.; Barros, C.F. de O.; Cardoso, R.; Gonzalez, G.V. Wavelet CNN-LSTM Time Series Forecasting of Electricity Power Generation Considering Biomass Thermal Systems. *IET Generation, Transmission and Distribution* **2024**, *18*, 3437–3451, doi:10.1049/gtd2.13292.

42. Yaghoubi, E.; Yaghoubi, E.; Khamees, A.; Razmi, D.; Lu, T. A Systematic Review and Meta-Analysis of Machine Learning, Deep Learning, and Ensemble Learning Approaches in Predicting EV Charging Behavior. *Eng. Appl. Artif. Intell.* **2024**, *135*, doi:10.1016/j.engappai.2024.108789.

43. Bellizio, F.; Dijkstra, B.; Fertig, A.; Van Dijk, J.; Heer, P. Machine Learning Approaches for the Prediction of Public EV Charge Point Flexibility. *Sustainable Energy, Grids and Networks* **2025**, *42*, doi:10.1016/j.segan.2025.101657.

44. Taylor, S.J.; Letham, B. *Forecasting at Scale*, The American Statistician, **2018**, *72(1)*, doi: 10.1080/00031305.2017.1380080.

45. Seong, B.; Lee, K. Intervention Analysis Based on Exponential Smoothing Methods: Applications to 9/11 and COVID-19 Effects. *Econ. Model.* **2021**, *98*, 290–301, doi:10.1016/j.econmod.2020.11.014.

46. Osman, H.; Azab, A.; AlGhazi, A.; Duffuaa, S.; Baki, F. Forecasting Electric Vehicle Charging Loads Using Random Forest and Gene Expression Programming Ensemble Models. *Results in Engineering* **2025**, *28*, doi:10.1016/j.rineng.2025.108369.

47. Yapıcı, A.T.; Abut, N.; Erfidan, T. Comparing the Effectiveness of Deep Learning Approaches for Charging Time Prediction in Electric Vehicles: Kocaeli Example. *Energies (Basel).* **2025**, *18*, doi:10.3390/en18081961.

48. Amezquita, H.; Guzman, C.P.; Morais, H. Forecasting Electric Vehicles' Charging Behavior at Charging Stations: A Data Science-Based Approach. *Energies (Basel).* **2024**, *17*, doi:10.3390/en17143396.

49. Bayarri, M.J.; Berger, J.O.; Datta, G.S. Objective Bayes Testing of Poisson versus Inflated Poisson Models. In *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*; Institute of Mathematical Statistics, 2008; pp. 105–121, doi: 10.1214/074921708000000093.

50. Hainaut, D.; Trufin, J.; Denuit, M. Response versus Gradient Boosting Trees, GLMs and Neural Networks under Tweedie Loss and Log-Link. *Scand. Actuar. J.* **2022**, *2022*, 841–866, doi:10.1080/03461238.2022.2037016.

51. Saha, D.; Alluri, P.; Dumbaugh, E.; Gan, A. Application of the Poisson-Tweedie Distribution in Analyzing Crash Frequency Data. *Accid. Anal. Prev.* **2020**, *137*, doi:10.1016/j.aap.2020.105456.

52. Herzen, J.; Lässig, F.; Piazzetta, S.G.; Neuer, T.; Tafti, L.; Raille, G.; Van Pottelbergh, T.; Pasieka, M.; Skrodzki, A.; Huguenin, N.; et al. Darts: User-Friendly Modern Machine Learning for Time Series. *Journal of Machine Learning Research* **2022**, *23*, 1–6.

53. Amat Rodrigo, J.; Escobar Ortiz, J. (2026). *skforecast (v0.22.0)*. Zenodo. doi:10.5281/zenodo.8382787.

54. IST-ID AHEAD-EU/Clustering-Forecasting-T3.3 Available online: https://github.com/AHEAD-EU/Clustering-Forecasting-T3.3 (accessed on 14 May 2026).

55. Blue, S.; Shove, E.; Forman, P. Conceptualising Flexibility: Challenging Representations of Time and Society in the Energy Sector. *Time Soc.* **2020**, *29*, 637–658, doi:10.1177/0961463X19842474.