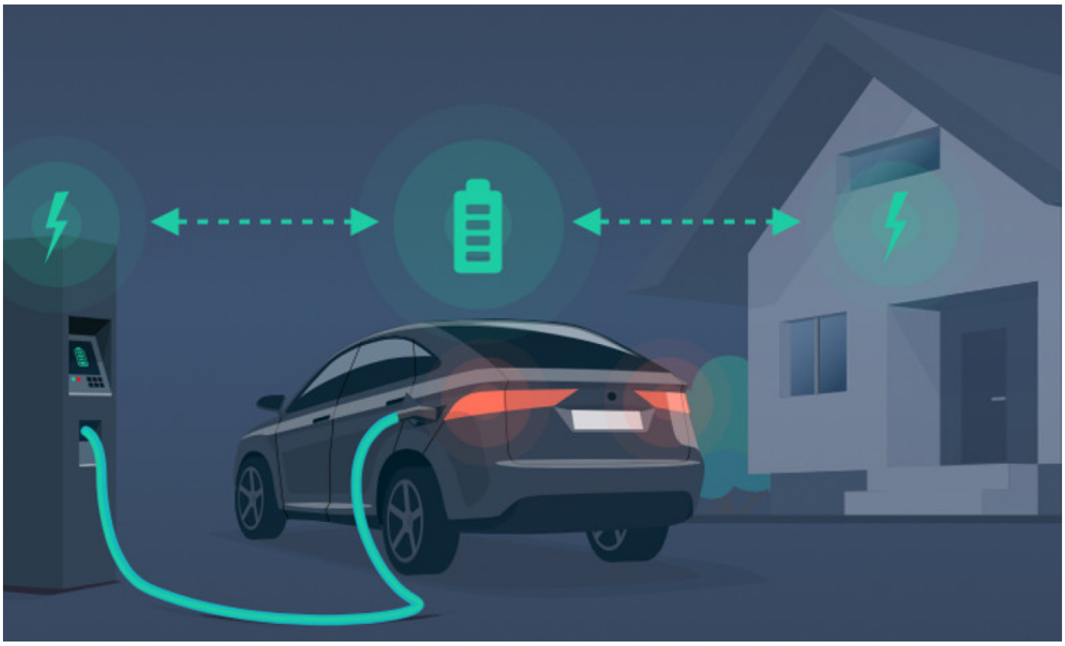

Vehicle to grid "V2G" (OpenSystemsMedia, 2020)

EV Charging Behavior Analysis and Identification for V2G Utilization

Draft Paper

Maksymilian Jasiak
Maximilien Kasteel
John Schafer

Faculty of Civil Engineering
University of California, Berkeley

November 2023

Contents

1 Introduction **1**

2 Literature review **3**

3 Data **5**

3.1 Data Description 5
3.2 Data Set Composition: 6
3.3 Data analysis 6

4 Methods **8**

5 Results **10**

5.1 Start/End Time Analysis 11
5.2 SoC Ratio Analysis 11
5.3 Charge Duration/Session Duration Ratio 13
5.4 Venue Proportion Analysis 13
5.5 Pricing Analysis 14
5.6 Final Results 14

6 Conclusion & Future works **17**

Bibliography **18**

1 Introduction

In order to fully decarbonize, the United States grid must transform entirely in how it is operated and used. In terms of the grid, decarbonization is a multifaceted problem requiring both technological solutions and rethinking how it is operated. Personal electric vehicles (PEVs) and renewable energy generation are two components that will drastically change how the grid functions and is managed. On one hand, increased PEV adoption and use will directly increase the demand of energy. While increased reliance and use of renewable energy resources will undoubtedly lead to more variable power generation. Consequently increased use of PEVs and renewable energy resources require additional technologies for grid balancing and operation, necessitating the need for technologies related to energy storage to help "smooth" the potential gaps between power generations and power demand.

One option that has received a plethora of research and interest recently is the idea of utilizing PEVs not just as a transit resource, but as an energy storage resource as well. As many people can relate to, much of a car's life is stationary, as much as 95% of the time. So why can't an PEV, when not in use, help with grid stability?

Vehicle-to-everything (V2X) is the broad term meant to encapsulate the technology that allows PEVS to release energy back onto the grid when it is needed the most. V2X encapsulates the sub-categories of vehicle-to-grid (V2G), which will be the main focus of this paper, and other applications like vehicle-to-home (V2H).

Currently, most EVs and EV charging stations are not set up with V2X capabilities but the new Ford-F150 lightning and future Chevy Silverado EV will offer capabilities at least for V2H (vehicle to home) capacities. The actual engineering problems associated with V2G will not be discussed and will be left to other research efforts However, before PEVs can effectively participate in grid balancing many questions surrounding the scheduling, location and timing will need to be addressed. As a large multifaceted problem, V2X is a topic of notable academic and industry research.

In theory V2G offers another tool to enable a fully decarbonized grid. However because of the mobility of PEVs and the inherent unpredictability of individual PEV use, it is a challenge to devise schemes that can effectively harness PEVs as an additional battery storage for the grid an appreciable amount. Broadly speaking, in order for PEVs to help load balancing, they need to charge when the grid is underutilized (and electricity prices are low) and then discharge when the grid is at peak load (or even strained). Thus we formulate our objective as: Identifying PEV charging behaviors, and thus what PEV users, in order to inform those users that should be focused on as potential participants in a hypothetical V2G program.

2 Literature review

Within the broader work on V2X, less attention has been paid to classifying the charging behaviors of actual PEV users from real world datasets. However there are still a number notable attempts at classifying PEV charging behavior using many distinct methods, data sources and differing objectives. Some but not all have approached this problem from the motivation to inform V2X efforts.

Helmus et al. (2020) used a Gaussian Mixture Model (GMM) to develop a typology of charging sessions, which included the descriptions of, Daytime/Overnight, Connection duration, Hours Between Sessions, and Distance Between Sessions for PEVs in the Netherlands. It should be noted that the dataset included the charging session data for individual users. Thus it was possible to build a complete charging history picture for one PEV user. In our case, due to the anonymized nature of the data, we are unable to tie specific charging sessions to specific users. However, their clustering approach using a GMM had success in classifying very distinct clusters.

et al. (2019) summarizes the challenges that increased PEV use poses for the electric grid, informing the motivation to utilize V2X. They present a much broader picture of the different strategies to control PEV charging with respect to the grid. Of particular interest to our work is the discussion on scheduling and clustering PEV charging behavior. Ali Saadon Al-Ogaili et al. presents scheduling as a strategy that “enables grid-to-vehicle and vehicle-to-grid services”. Furthermore they outline the previous work done to cluster charging behavior and note different strategies such as k-means, self-organizing maps and Markov chain methods. Their work highlights the need for preliminary behavior analysis needed to inform more complex projects related to control problems facing the charging of PEVs.

Finally, the study Desai et al. (2018) is probably closest to the type and scope of our work. Ranjit R. Desai et al. examine the charging behaviors of PEVs in the City of Rochester New York and then cluster the behaviors into 5 distinct clusters. The authors considered their unit of analysis as “one daily PEV charging pattern for a single vehicle observed at public charging stations”.We will also have the same unit of analysis. In addition to the charging time of the vehicle, the authors also considered the time the vehicle was parked and not charging. It is also important to note that the largest cluster (Cluster 2) accounted for 46% of their data. Their paper will be a source of great inspiration but we hope to glean additional information purely by the size of our study. Compared to our analysis involving over 300,000 records, Ranjit R. Desai et al. analyzed just under 9,000 records.

3 Data

3.1 Data Description

Plug-In Electric Vehicle (PEV) and Electric Vehicle Supply Equipment (EVSE) Dataset (2019-2022) (Energetics and Cities, 2023)

The increasing adoption of electric vehicles underscores the critical need for comprehensive, current data to comprehend various facets of vehicle electrification. Energetics, in collaboration with diverse partners, has collected and analyzed this dataset. Its primary aim is to assist researchers and policymakers in gaining insights into electric vehicle usage patterns, infrastructure performance, and the overall dynamics of the electric vehicle ecosystem. The dataset spans from Jan 2019 to December 2022, ensuring anonymity to uphold privacy while empowering researchers to conduct analyses that can inform future research and planning in the electric vehicle electrification domain.

This dataset represents a diverse and extensive collection of plug-in electric vehicle (PEV) and electric vehicle supply equipment (EVSE) data. Each charging station is uniquely identified by an `evse_id`. The collaboration involves contributions from various stakeholders in the electric vehicle industry, ensuring a comprehensive representation of charging and driving behaviors.

The dataset covers the entire USA but has been specifically filtered to focus on the Los Angeles metropolitan area. A total of 344 stations (each with a unique `evse_id`) are included, ranging from private home charging plugs to publicly available paid plugs. Due to anonymization, the specific category of each charging station is unknown. However, the dataset indicates whether users are required to pay, indirectly identifying whether the charging station is public or private. From these 344 different stations, a total of 341,291 charging sessions are recorded. This gives plenty of data to analyze the behaviors of users in the LA metropolitan area.

3.2 Data Set Composition:

Initially the data is given in two separate files, one file containing information about the charging station (so 344 rows) and the other describes each charging session (360 605 rows). For ease of use, the two datasets were merged into one containcing all the information. This general dataset comprises several key fields that offer a detailed perspective on electric vehicle usage and charging infrastructure. Each record in the dataset is identified by a unique session ID. The relevant fields include:

`session_id`: A unique identifier for each charging session.
`evse_id`: Identification number for the Electric Vehicle Supply Equipment (EVSE) involved in the charging session.
`connector_id`: Identification number for the connector used in the charging session.
`start_datetime`: The date and time when the charging session commenced.
`end_datetime`: The date and time when the charging session concluded.
`total_duration`: The total duration of the charging session, indicating the time span from start to finish.
`charge_duration`: The duration during which the vehicle was actively charging.
`energy_kwh`: The amount of energy consumed during the charging session, measured in kilowatt-hours.
`start_soc`: The state of charge (SoC) of the vehicle at the beginning of the session.
`end_soc`: The state of charge of the vehicle at the end of the session.
`venue`: The categorization of the charging station location. Classified as either "Business Office", "Corridor", "Multi-Unit Dwelling" or "Undesignated".
`pricing`: The categorization of the charging station payment type. Classifies as either "Free", "Paid" or "Undesignated".

3.3 Data analysis

From the approx. 341,291 charging sessions, the user distribution is made in figure 3.1. It can be seen that two of the most used stations represent approx. 20000 sessions (7% of total). However these stations are classified as "free" so it is probably a private charging location for an office or hotel.

Visualizing the kernel density of the EV charging session start and end times in figure 4.3 reveals that the number of charging sessions increases during the morning, and reaches a peak in the afternoon and evening. The observed lag between the increase and decrease

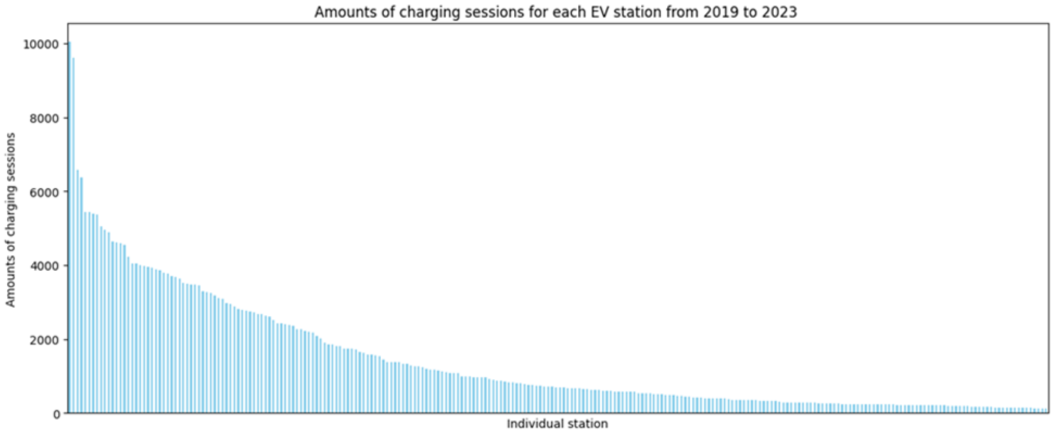

Figure 3.1: Amounts of charging sessions

in density of charging session start times and end times serves as a robust reality check to confirm that logically, there are more start times before end times.

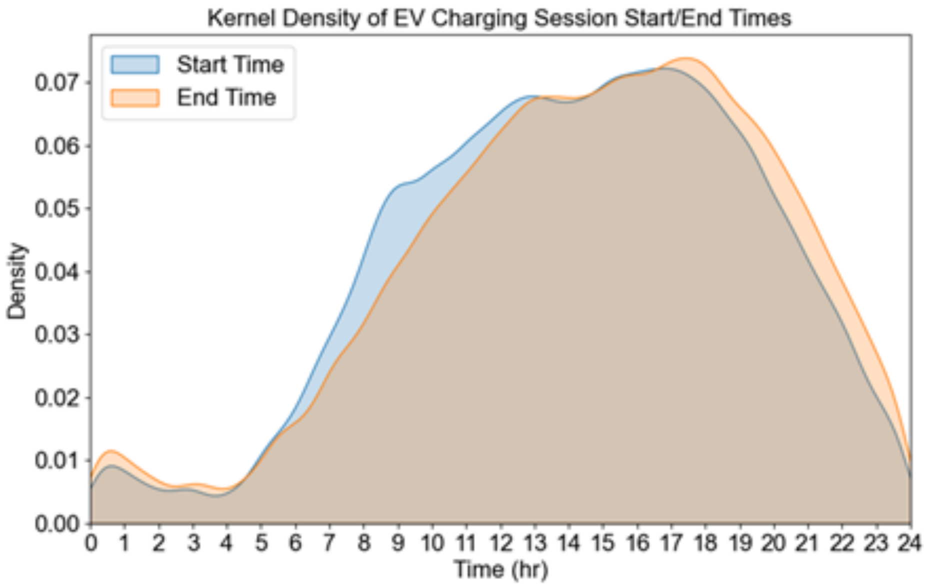

Figure 3.2: Kernel Density of EV Charging session

4 Methods

Next, the start and end times were one-hot encoded into 0.1hr increments across the 24 hours of a day. The proportion of charging sessions at a given time which are actively charging the EV can be expressed as a ratio of the total number of charging sessions captured. A visualization of this proportion of charging sessions with charging ON in figure 4.1 reveals that a majority of charging session occur between approximately 8:00 to 21:00, and reach a peak between 11:00 to 17:00 (7% of all measured charging sessions). Note, a relatively small number of charging sessions occured in the late night and early morning from 23:00 to 6:00.

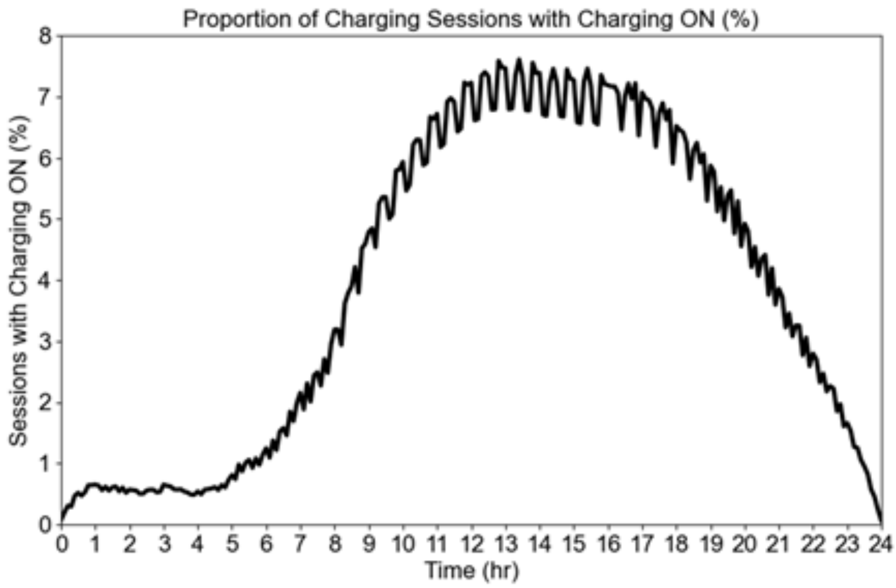

Figure 4.1: Proportion of charging sessions

To further prepare the data for PCA analysis, the categorical features "pricing" and "venue" were also one-hot encoded. Meanwhile, the features "total_duration", "charge_duration", "start_soc", and "end_soc" were standardized. Notably, the "start_soc" and "end_soc" features were of particular interest. As not all charging sessions had this data, the mean of each start and end State of Charge (SoC) was filled. Only 8,521 sessions did not contain SoC data, representing less than 2.5% of our total data.

As a result, the observation matrix used for analysis has 261 features. Columns 1 through 250 represented the encoding of the start and end charging times, while columns 251 to 261 included the `charge_duration`, `start_soc`, `end_soc`, and the encoding of the `venue` and `pricing` scheme.

The observation matrix attains its ultimate configuration with dimensions (341219, 261). Subsequent to this, a principal component analysis (PCA) was conducted. Next, the cumulative variance graph was computed in figure 4.3, revealing that 80% of the total variance could be elucidated through the utilization of 21 principal components. Following this determination, the K-means elbow method was employed.

INCLUDE K-Means elbow (too computationally expensive to compute until the final results calculations). However based upon initial analysis, K values between 10-15 seem appropriate.

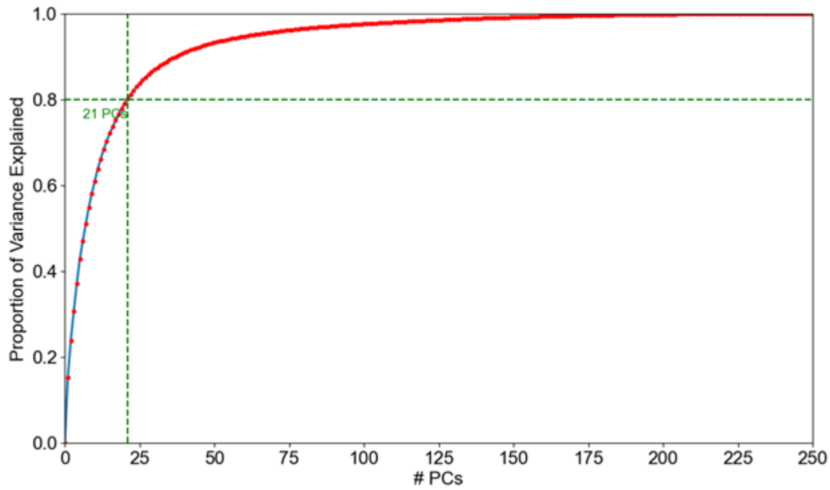

Figure 4.2: Proportion of variance

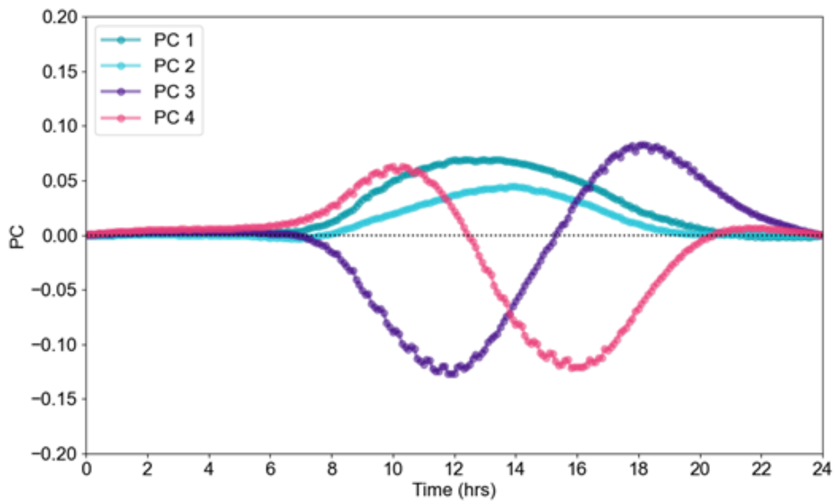

Figure 4.3: First four PCs

5 Results

After using PCA to reduce the complexity of the data and applying K-means to group the data, we obtained 12 distinct clusters. Our results focus on the individual characteristics of the clusters, their meanings, and, in some cases, the inferences we can draw based on these characteristics. The percentage of the total data represented by each cluster is shown below.

- **Cluster 1**: 73089.1098 sessions (21.42%)
- **Cluster 2**: 58792.0337 sessions (17.23%)
- **Cluster 3**: 90252.4255 sessions (26.45%)
- **Cluster 4**: 14638.2951 sessions (4.29%)
- **Cluster 5**: 20063.6772 sessions (5.88%)
- **Cluster 6**: 409.4628 sessions (0.12%)
- **Cluster 7**: 7029.1114 sessions (2.06%)
- **Cluster 8**: 18732.9231 sessions (5.49%)
- **Cluster 9**: 5220.6507 sessions (1.53%)
- **Cluster 10**: 19176.5078 sessions (5.62%)
- **Cluster 11**: 14535.9294 sessions (4.26%)
- **Cluster 12**: 19244.7516 sessions (5.64%)

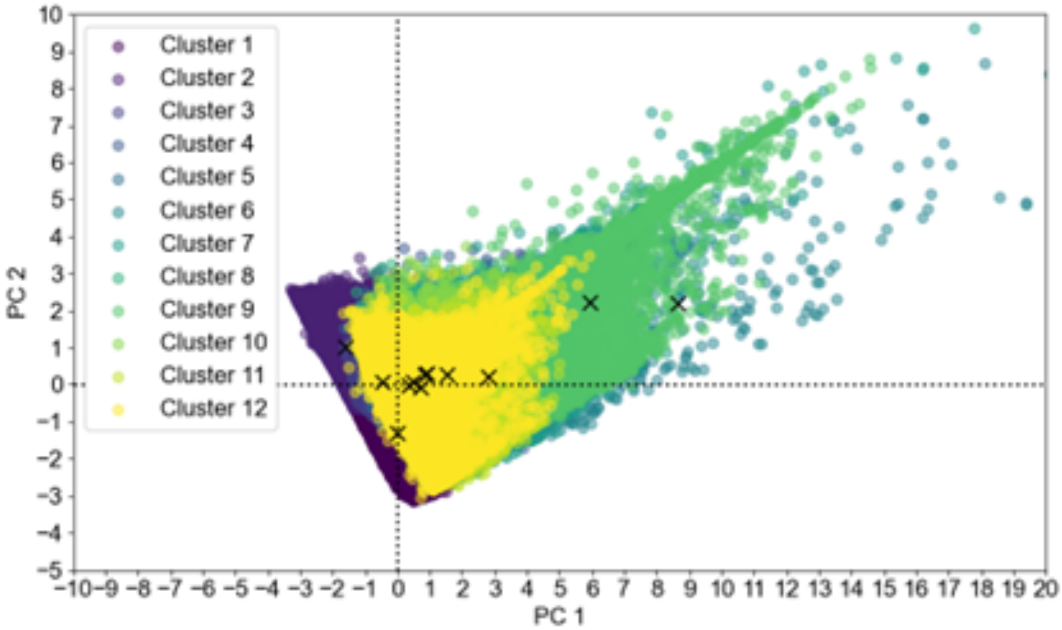

Figure 5.1: PC1 vs PC2

5.1 Start/End Time Analysis

Of the most importance to our analysis, is the start and end times of charging across the different clusters shown in Figure 5.2. Across the twelve clusters, it is obvious the diversity in charging patterns displayed. Clusters 1,2,3 (which represent most of the data) have the most irregular charging patterns of all the data. Not much can be gleaned from these visualizations other than the realization that the 65% of the data these clusters represent, charge, seemingly in a uniform random manner throughout the day. From our interpretation, these clusters capture the charging sessions, where there is no inherent pattern. Although disappointing at first, we believe this result shows that many charging sessions are too random to predict, yet this implies that the rest of the data potentially does have inherent patterns to analyze.

The rest of the clusters showcase neat charging patterns, with some showcasing very little variability in the start and end times. Of particular note are clusters 6 & 7 which are the only clusters to have start times after their end times, implying the charging session was overnight.

5.2 SoC Ratio Analysis

Figure 5.3 shows the ratio between the Start and end State of Charge (SoC) for each cluster. A lower ratio would imply more actual charging, while a lower ration would capture

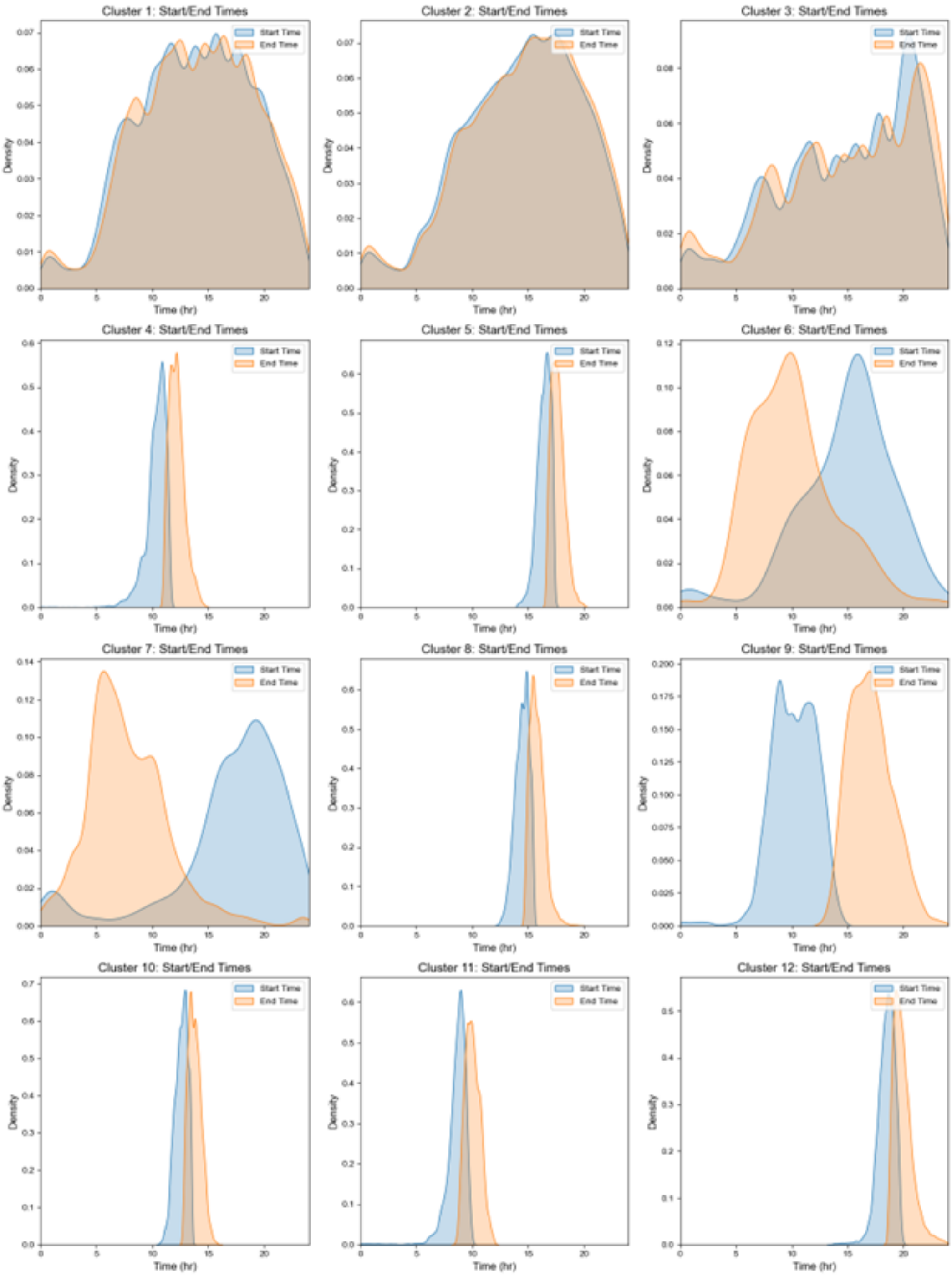

Figure 5.2: Cluster analysis

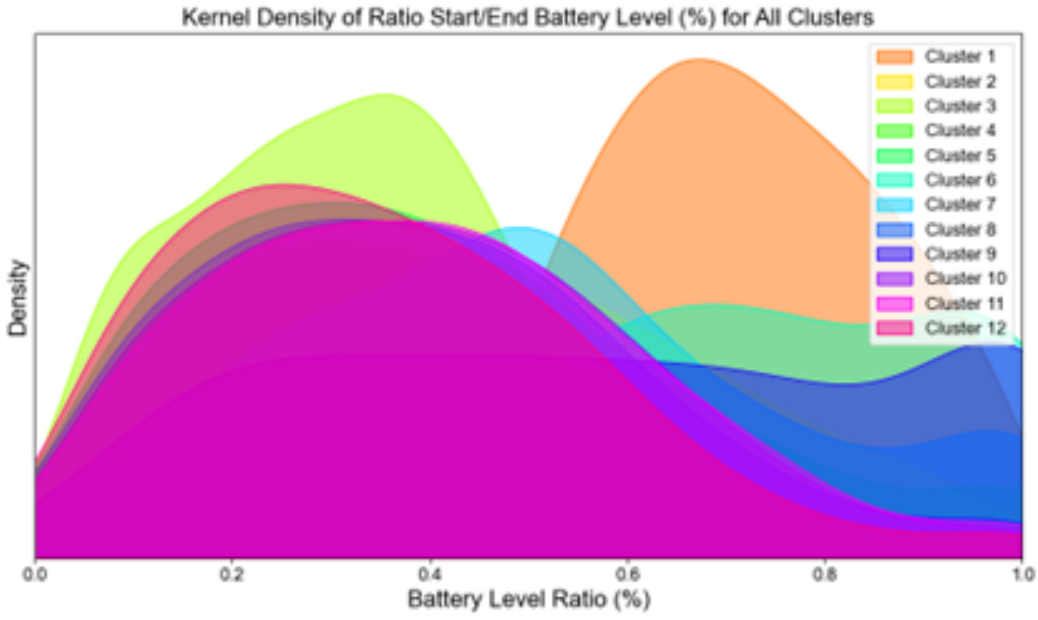

Figure 5.3: Kernel density of Battery level

those sessions where relatively little energy was transferred to the battery. Of particular note are clusters 8 & 6, which have the highest ratio, indicating those sessions routinely deliver very little energy. Besides adding further evidence that cluster 6 encapsulates outliers, this shows that cluster 8 tends to either charge for a very short amount of time, or does not actually need that much charging.

5.3 Charge Duration/Session Duration Ratio

The ratio between the actual charge duration and session duration is a measure of how much time was actually spent charging the vehicle as compared to the vehicle simply connected to the charger. As one would expect, most clusters are close to 1, indicating very little time was spent not actually charging the vehicle. However, clusters 6, 7 and 8 are distinct from the others. As mentioned cluster 6 is a group of outliers (some sessions in this group were connected to chargers for weeks at a time).

5.4 Venue Proportion Analysis

From figure 5.5, we can see that cluster 9 represents a distinct group that charges at a Business Office much more frequently than other groups.

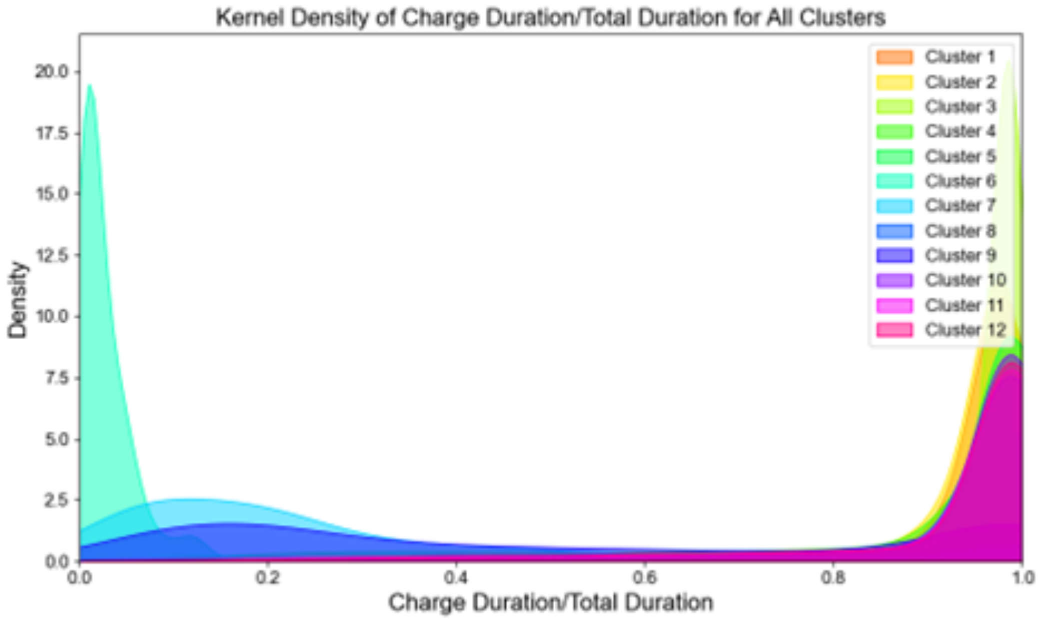

Figure 5.4: Kernel density total duration

5.5 Pricing Analysis

Quite distinct in this graph are clusters 6, 7 and 9 which have the highest proportion of free charging sessions. While clusters 1 & 2 have the highest proportion of paid sessions.

5.6 Final Results

Aggregating these analyses, we describe each cluster as follows:

- **Cluster 1 (21.42%)**: Sessions with uniform and random charging behavior.
- **Cluster 2 (17.23%)**: Sessions with uniform and random charging behavior.
- **Cluster 3 (26.45%)**: Sessions with uniform and random charging behavior.
- **Cluster 4 (4.29%)**: Charges between 11 am and 1 pm.
- **Cluster 5 (5.88%)**: Charges between 4 pm and 6 pm.
- **Cluster 6 (0.12%)**: Charges between 9 am and 4 pm. Outliers with long connection times, little to no charging.
- **Cluster 7 (2.06%)**: Charges between 8 pm and 6 am. Connected for a long time with no charging. Tends to be a free charging session, implying this is the residential charging group.

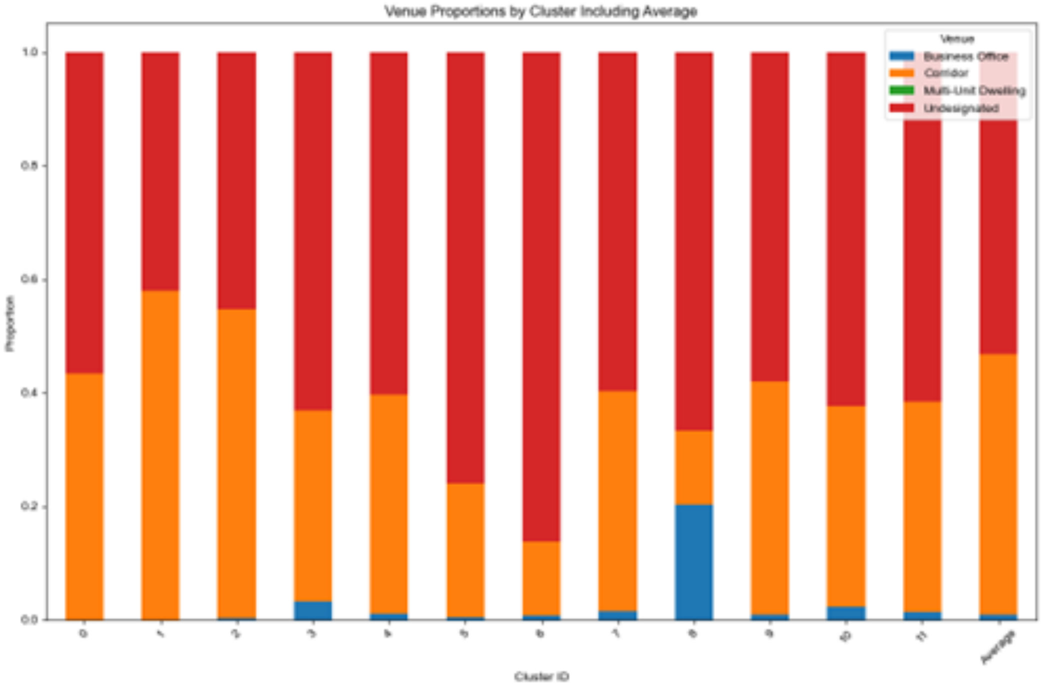

Figure 5.5: Venue Proportions by clusters

- **Cluster 8 (5.49%)**: Charges between 2 pm and 4 pm. Charges relatively little. Connected to the charger for longer.
- **Cluster 9 (1.53%)**: Charges between 10 am and 6 pm. Tends to charge at business locations. Sessions also tend to be free.
- **Cluster 10 (5.62%)**: Charges between 12 pm and 2 pm.
- **Cluster 11 (4.26%)**: Charges between 9 am and 10 am.
- **Cluster 12 (5.64%)**: Charges between 7 pm and 9 pm.

As is easily seen, there are a number of groups that tend to charge at quite specific times. Yet given the features we have, there does not seem to be much else that makes them distinct from other sessions.

However, clusters 7, 8 and 9 stand out from the rest given the other features beside session start and end time. Cluster 7 is almost undoubtedly a residential-esque charging behavior where the session takes place overnight and is free. Furthermore Cluster 9 is almost the opposite. These sessions tend to take place all day, are also free but tend to be at a business setting. It is inferred these sessions correspond to some sort of working charging behavior where the users are able to charge while they are at the office, potentially in an employer sponsored parking lot.

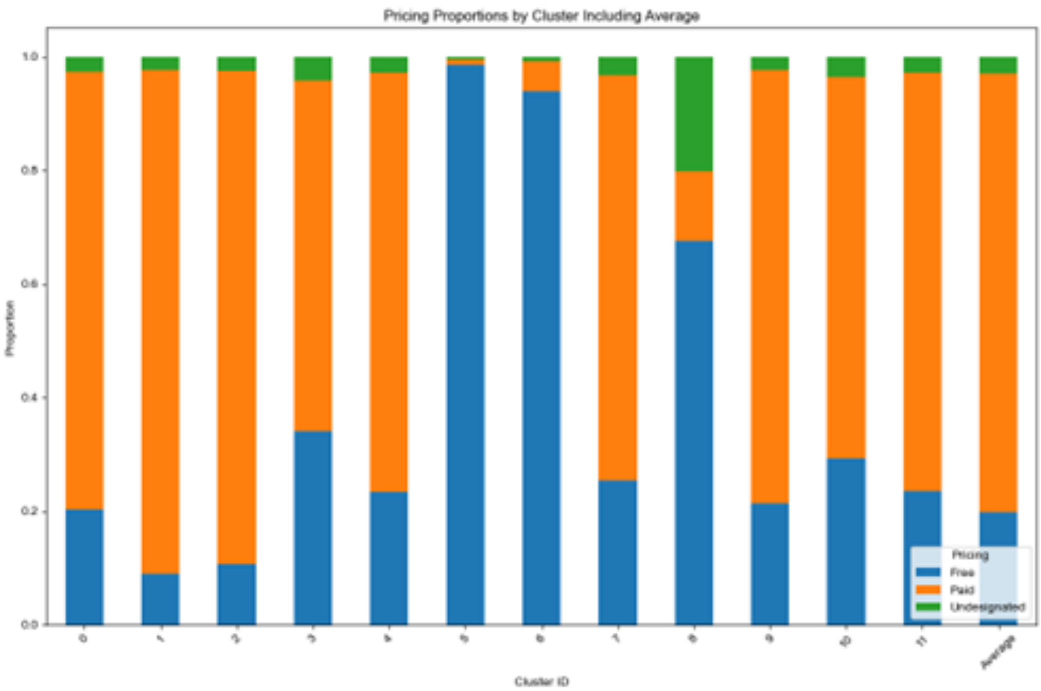

Figure 5.6: Pricing Proportions by clusters

Cluster 8 is somewhat peculiar given the tight window of charging and the fact that this charging does not seem to deliver much energy (the charge duration to total duration ratio is very low). Potentially this cluster encapsulates those who charge after work during errands (at the grocery store for example) on a very regular basis as each session does not deliver much total charge. This group could be seen as having "charge anxiety".

In terms of V2G, there are many options in which to see our analysis. For one, cluster 9 seems to be ripe for V2G applications. They are connected to charges throughout the day when energy is cheap and could theoretically plug back into the grid when they return home to deliver energy back to the grid, just when the duck curve is reaching its biggest inflexion point.

To that point any of the clusters that have daytime charging behaviors could be tapped to help stabilize the grid later on in the day.

And finally, cluster 7 which is the residential overnight charger, could be tapped in order to change their charging behavior to better align with grid stabilization.

The most important part is that our analysis can help those looking to identify users that can immediately be utilized for V2G efforts and those users whose behavior is most "problematic" charging right at 6pm in order to steer them to more helpful charging patterns.

6 Conclusion & Future works

V2X in theory, is a compelling solution to grid management under the certain conditions of increased power needs of a decarbonizing economy. However, many challenges abound in the actual implementation of V2X programs considering the inherently stochastic nature of PEV usage and charging patterns.

However, our analysis of charging data in the LA metro region offers a particular method of identifying users based on certain charging behaviors. At first the anonymized nature of the data presented seemingly presented an obstacle to identifying particular behaviors. But as we have showcased, our method successfully is able to discriminate between distinct groups of users such as the residential overnight chargers, and the daytime office worker chargers. This type of analysis would be quite valuable to efforts to identify certain groups of PEV users, especially in data constrained, or anonymized situations.

With that said, future analysis would benefit greatly with more fine grain data sources. In particular, geographic location data for the charges would enable further network based analysis. Furthermore, if individual users had unique identification, a specific profile of a particular user (as opposed to our aggregate methods) could be built.

Bibliography

Desai, R. R., Chen, R. B., and Armington, W. (2018). A pattern analysis of daily electric vehicle charging profiles: Operational efficiency and environmental impacts. *Journal of Advanced Transportation*, 2018:Article ID 6930932, 15 pages.

Energetics and Cities, C. (2023). Electric vehicle widescale analysis for tomorrow's transportation solutions (ev watts) project. Partnership between Energetics, Clean Cities, ChargePoint, Sawatch Labs, and U.S. Department of Energy. Led by Bryan Roy (EV WATTS Program Manager), Ewan Pritchard (Analysis Team Lead), Yash Pavuluri (Data Analyst), and U.S. Department of Energy.

et al., A.-O. (2019). Review on scheduling, clustering, and forecasting strategies for controlling electric vehicle charging: Challenges and recommendations. *IEEE Access*, 7:128353–128371.

Helmus, J. R., Spoelstra, J., Refa, N., Lees, M., and van den Hoed, R. (2020). A data-driven typology of electric vehicle user types and charging sessions. *Transportation Research Part C: Emerging Technologies*, 115:102637.