

Article

Combining Ad Hoc Text Mining and Descriptive Analytics to Investigate Public EV Charging Prices in the United States

David Trinko ^{1,2,*} , Emily Porter ², Jamie Dunckley ², Thomas Bradley ¹ and Timothy Coburn ¹

¹ Department of Systems Engineering, Colorado State University, Fort Collins, CO 80523, USA; Thomas.Bradley@colostate.edu (T.B.); Tim.Coburn@colostate.edu (T.C.)

² Electric Power Research Institute, Palo Alto, CA 94304, USA; EPorter@epri.com (E.P.); JDunckley@epri.com (J.D.)

* Correspondence: DTrinko@rams.colostate.edu

Abstract: Electric vehicle (EV) charging infrastructure is present all over the United States, but charging prices vary greatly, both in amount and in the methods by which they are assessed. For this paper, we interpret and analyze charging price information from PlugShare, a crowd-sourced EV charging data platform. Because prices in these data exist in a semi-structured textual format, an ad hoc text mining approach is used to extract quantitative price information. Descriptive analytics of the processed dataset demonstrate how the prices of EV charging vary with charging level (Direct Current Fast Charging versus Level 2), geographic location, network provider, and location type. Our research indicates that a great deal of diversity and flexibility exists in structuring the prices of EV charging to enable incentives for shaping charging behaviors, but that it has yet to be widely standardized or utilized. Comparisons with estimates of the levelized cost of EV charging illustrate some of the challenges associated with operating and using these stations.

Keywords: ad hoc text mining; descriptive analytics; data wrangling; EV charging cost; level 2 charging; DC fast charging; spatial variation



Citation: Trinko, D.; Porter, E.; Dunckley, J.; Bradley, T.; Coburn, T. Combining Ad Hoc Text Mining and Descriptive Analytics to Investigate Public EV Charging Prices in the United States. *Energies* **2021**, *14*, 5240. <https://doi.org/10.3390/en14175240>

Academic Editors: Luboš Buzna and Byoung Kuk Lee

Received: 1 June 2021

Accepted: 17 August 2021

Published: 24 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Electricity plays an increasingly important role in powering the U.S. transportation sector with projections of 147–440 TWh of annual consumption by vehicles by 2050 [1,2]. This consumption corresponds to about 4–10% of the current total electricity consumption in the United States [3]. Based on the dataset used in this study, accessed February 2021, there are more than 90,000 charging connectors available at more than 75,000 public charging locations for electric vehicles (EVs). Fueling infrastructure for EVs is unlikely to resemble conventional vehicle fueling infrastructure for a variety of reasons, including the time duration required for fueling, the physical and regulatory differences between electricity and liquid fuels, and the fact that EVs can be charged at home, at the workplace, or in public. Public EV charging infrastructure installed to date has been constructed and operated by a variety of entities under numerous business models. A comprehensive review of public charging prices and price models has not yet been conducted, although this type of summary might be valuable both to sellers and buyers of electricity, as well as to policymakers and other stakeholders. Consumer-facing articles have been published to explain public charging prices to EV drivers [4–6].

Privately operated EV charging infrastructure has been installed and managed by at least 18 companies at public locations in all 50 states, including at grocery stores, hotels, shopping centers, and gas stations. Within and across companies, states, and locations, charging prices can vary greatly. This suggests that companies are pursuing disparate business models. For example, Tesla has installed a centralized network emphasizing long-distance travel that is compatible only with the vehicles Tesla produces, and has intermittently offered free and/or low-cost charging as an incentive for vehicle purchases.

In contrast, other networks, such as ChargePoint, EVgo, and Blink, offer charging at low and high power at a variety of location types, at stations that are operated based on centralized or decentralized models.

Charging prices are assessed at fixed or variable rates during a charging session, as a function of time (seconds, minutes, or hours), energy (kilowatt-hours, kWh), or as a total price per charging session. The majority of charging connectors in the U.S. are Level 2 (L2) chargers, meaning power transfer occurs at an average rate between 6.6 and 19.2 kilowatts (kW) [7]. The majority of the remaining stations have DC fast chargers (DCFC), which provide rates anywhere from 50 to 350 kW. Recognizing that less than 2% of connectors operate at the much slower Level 1, those charging locations are not included in the present analysis. Whereas L2 connectors are largely standardized under the Society of Automotive Engineers' (SAE) J1772 standard, there are three major DCFC connector types that are not mutually compatible: the Tesla Supercharger, the SAE Combined Charging System (CCS), and CHAdeMO (short for "CHArge de MOve"), a standard which is being phased out in favor of CCS for new vehicles.

Although there is no official and comprehensive repository of charging price data for public EV charging stations, PlugShare [8] has obtained price information and other metadata for a substantial portion of the stations in the U.S. via crowd-sourcing through its app and website, and through partnerships with charging station providers. These data ("the dataset"), which largely exist in textual form, are publicly accessible for individual stations via PlugShare's app and website interfaces, but are not publicly accessible in the aggregate form necessary for the application of broad analytics. The authors obtained access to the dataset in aggregate form in order to conduct this study. Due to the many ways that a price signal can be written in textual form, we needed to employ ad hoc text mining and processing methods to reformat a majority of the dataset's price information into quantitative data for analysis.

In Section 2, we present an overview of text mining from the literature. We then describe the dataset in more detail in Section 3 and discuss the text mining and data processing methods employed in our study in Section 4. The results of our analysis are reported in Section 5. The specific contributions of this work are as follows:

- Ad hoc text mining techniques enable quantitative analysis of an otherwise opaque source of EV charging price data;
- Descriptive analytics provide a high-level image of EV charging price variability in the United States; and
- Discussion of trends in observed EV charging prices highlights decision-making implications for EV operators, charging station operators, policymakers, and business innovators.

2. Overview of Text Mining

The concept of text mining (text data mining, text analytics) originated with the ideas of natural language processing in the 1950s. However, it was not until the late 1990s that it began to assume a more prominent role across the analytics landscape. This development occurred in conjunction with a maturing data mining toolkit plus advances in computational power and speed capable of processing large unstructured data sets. More recently, text mining has evolved into a discipline of its own, with numerous applications throughout business, engineering, public health, the physical and social sciences, and other endeavors [9].

The literature on text mining is now quite extensive in both the research and applications domains. Analytical advancements have progressed rapidly with the implementation of newer and faster algorithms and processing capabilities. Materials describing foundational ideas (e.g., [10–12]), as well as advanced methods ([13]) are widely available, and the various tools and techniques have been translated to accommodate a variety of computer languages and platforms (e.g., [14–17]). Madigan [18], Weiss et al. [19], and

Sumathy and Chidambaram [20] provide excellent overviews of the text mining landscape from statistical and data science perspectives.

With its growing importance in the Big Data era, the definition of text mining has become more fluid, expanding to accommodate numerous analytical contexts, ranging from information and content extraction to lexical and sentiment analysis, pattern recognition/categorization, dimensionality reduction, and beyond. Perhaps the most common understanding of text mining in contemporary data analytics revolves around the extraction of word/phrase frequencies and relationships using various clustering and classification techniques [21]. While text mining can logically be thought of as a means for parsing written artifacts for knowledge discovery, it also plays a significant role in the preprocessing and wrangling stages of Big Data analysis, such as reducing semantic, syntactic, and contextual ambiguity [22,23].

Text mining is commonly used to extract, reduce, or regularize information contained in parcels of written material, free-form responses to questions or inquiries, or more conversational communications scraped from social media. It may also be used to effectively analyze text-based transactional records for relevant and recurring content, such as electronic medical reports (transcriptions of physicians' notes pertaining to patient visits, conditions, diagnoses, etc.) [24,25], industrial maintenance files pertaining to failure times and modes [26,27], building maintenance work orders [28], court proceedings (including case files and docket entries) [29], customer service archives [30], and historical exchanges of real estate and mineral leases. The electric vehicle charging records in the dataset represent a similar type of transactional, textual, and numerical data that is amenable to text mining.

In these and other contexts, the approach is more closely aligned with the various aspects of content mining, such as concept extraction, named entity recognition, key word identification, differentiation of implicit or explicit actions and decisions, definition and capture of interesting phrases, and alignment and standardization of abbreviations [31,32]. It is these aspects that are most relevant to our investigation of electric vehicle charging costs. Accomplishing the tasks of text mining, however, often requires a more ad hoc, informal, or even "brute force" approach that involves a combination of human intervention, original scripting, and machine learning [33–35], particularly as the volume of data increases and encompasses more diverse entities. Our analysis of the dataset requires this kind of approach because of its compositional nature and the continuing flow of additional information into the database over time.

3. Data

The data are semi-structured in the sense that they are organized in rows (representing individual charging connectors) and columns (representing variables or attributes pertaining to those charging connectors), although the data entries recorded for several of the attributes exist as words, phrases, or sentences (natural language) that must be refined to extract consistent and usable meanings. Although the database itself is semi-structured, the information associated with some attributes is completely unstructured. The documentation for the application programming interface (API) provides more information about the data organization [8].

We received the data in two separate tranches: 74,237 observations in 2019 and an additional 19,312 observations in 2021, for a total of 93,549 observations. Each observation represents one connector, so a charging station with multiple connectors is represented by multiple observations. A typical station hosts approximately 1.2 connectors on average. Records contain location information (city, state, zip code), charger information (connector type, network; whether charging is free), parking information (location type; whether parking is free), and unstructured price description information. Of these records, 30,756 have interpretable price information. A small sample of data with price descriptions is shown in Figure 1.

City	Connector type	Cost	cost_description	Dual DCFC?	id (connector)
XXXXX	CCS (SAE Combo)	TRUE	\$0.30/min	TRUE	XXXXX
XXXXX	J-1772	FALSE	Free	FALSE	XXXXX
XXXXX	CHAdeMO	FALSE	Currently free!	TRUE	XXXXX
XXXXX	J-1772	TRUE	\$1.00/hour	FALSE	XXXXX
XXXXX	J-1772	TRUE	\$0.19/kWh	FALSE	XXXXX
XXXXX	J-1772	TRUE	First 3 hours \$1 per hour, then \$5 per hour.	FALSE	XXXXX
XXXXX	J-1772	TRUE	18 for valet	FALSE	XXXXX
XXXXX	Tesla Supercharger	TRUE		FALSE	XXXXX

Location Id	Created_at	Network	Open247	Parking Type	POI	State	station_id	Zip
XXXXX	9/1/2020	EVgo	TRUE	Free	Parking Garage/Lot	VA	XXXXX	XXXXX
XXXXX	8/1/2016	ChargePoint	TRUE	Free	Shopping Center	TN	XXXXX	XXXXX
XXXXX	8/1/2020	Non-networked	TRUE	Free	Store	MN	XXXXX	XXXXX
XXXXX	3/1/2012	ChargePoint	FALSE	Pay	Government	VA	XXXXX	XXXXX
XXXXX	4/1/2019	ChargePoint	TRUE	Pay	Parking Garage/Lot	MA	XXXXX	XXXXX
XXXXX	8/1/2013	ChargePoint	TRUE	Free	Government	MI	XXXXX	XXXXX
XXXXX	6/1/2016	Greenlots	TRUE	Pay	Hotel/Lodging	CA	XXXXX	XXXXX
XXXXX	9/1/2017	Supercharger	TRUE	Free	Car Service	MD	XXXXX	XXXXX

Figure 1. Sample from the dataset. Some values are omitted at the request of the data provider. The second half of the columns is shown below the first half of the columns to fit on the page.

Price descriptions, in the form of unstructured text, vary widely in format and information content. This has several implications. Due to the nature of crowd-sourced data and the potential for user error, some of the price information may not be accurate or up-to-date. There is no standard way to specify whether price information applies to parking, charging, or both. Price descriptions thus may contain descriptions of prices for both parking and charging, for one or the other, for neither, or for one or multiple different charging levels, without means of resolving the ambiguity. Finally, prices, and their descriptions, do not follow a standard model. Manual interpretation is not feasible for a growing database of more than 30,000 stations with cost descriptions, so an ad hoc algorithmic text interpretation approach is used. Still, for some price descriptions which are inherently ambiguous (examples shown in Table 1), neither algorithmic nor manual text interpretation succeed in extracting meaningful price information.

Table 1. Examples of ambiguous pricing information in the dataset.

Example Entry	Issue
“\$10 for Tesla, \$3 for other vehicles”	No unit of assessment, multiple prices
“varies for non guests”	No price information
“\$3 for 0–4 h of parking, then the price goes up”	Partial price information

Whereas textual price data do not follow a standard format, the dataset does include standard specifications of whether fees exist for (a) parking (“Parking Type” in Figure 1) or (b) charging (“Cost” in Figure 1), or both. Thus, prices for stations with no textual price description but with both free charging and parking can, in theory, be inferred (i.e., the price is \$0). However, since this inference is only possible for free stations, including these data in the general analysis would disproportionately weight free-charging locations. Instead, we assume that the sample of stations with price descriptions, including free stations, constitutes a representative sample of public EV charging stations, and therefore do not infer the price for stations marked as free. Furthermore, there exist records with detailed descriptions of nonzero prices, but that are marked as having both free charging and parking. In such cases, we assume that the price description is accurate.

An overview of how charging connectors are distributed across categories is shown in Figure 2. Among states, California hosts the greatest share by a substantial margin. Among network providers, ChargePoint hosts the greatest share of charging connectors.

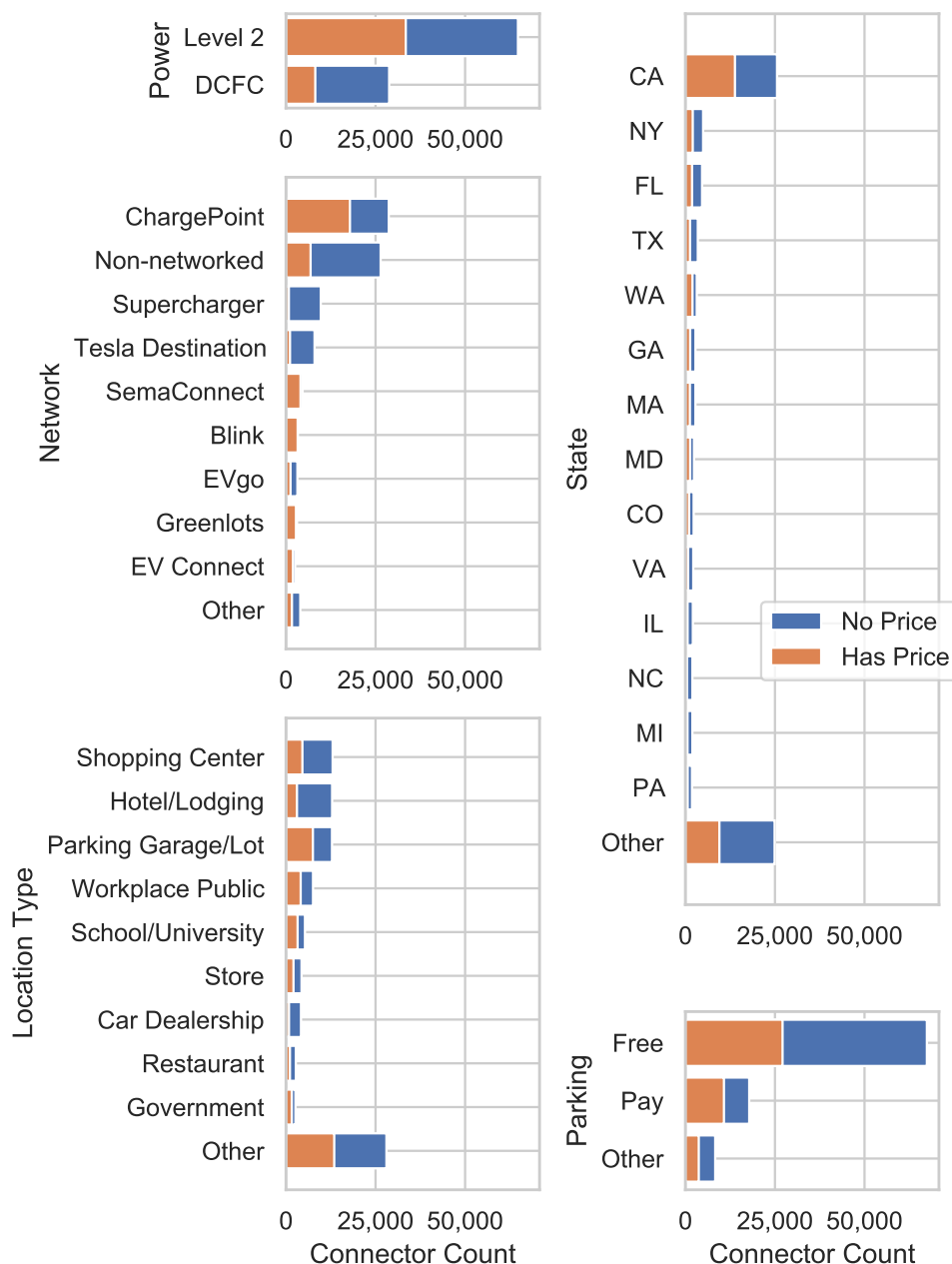


Figure 2. Distribution of electric vehicle (EV) charging connectors across categories of power level, network provider, location type, state, and parking cost. Data shown are for all public U.S. EV charging connectors in the dataset as of February 2021.

4. Text Mining, Processing, and Interpretation

Two challenges must be addressed to enable quantitative analysis of charging prices using this dataset: (1) Prices must be extracted from inconsistently worded price descriptions via reformatting and processing, and (2) fundamental differences in pricing models must be regularized to enable general comparisons. The methods for addressing these challenges are described in Sections 4.1 and 4.2, and details of the overall process are provided in Appendix A.

4.1. Extraction of Charging Price Information

Descriptions of the price of charging were assigned to three basic categories, where costs accrue as a function of (1) units of time charging, or (2) energy consumed, or (3) are assessed as a total price per session, irrespective of session duration. In the first category, costs are typically assessed per hour, minute, or other increment (for example, per 30 s), but sometimes vary during a charging session. For example, the first hour might be free, but each subsequent hour, the price increases by some amount before settling at a final per-hour price. In addition, there might be limits imposed, typically in terms of a minimum or maximum total cost or a charging time limit. Table 2 captures essential elements of the majority of pricing structures in a standardized format.

Table 2. Table headings for populating the details of every interpreted price description from the dataset. An example entry is given for the description “\$0.49 per kilowatt hour (kWh) \$0.50 minimum. First 5 min are free”.

Quantity	Example Entry	
	Value	Unit
Initial price	0	free
Initial price 2	-	-
Initial time window	5	minute
Price next window	0.49	kWh
Next window	-	-
Price next window (2)	-	-
Next window (2)	-	-
Price next window (3)	-	-
Next window (3)	-	-
Minimum	0.5	dollar
Maximum	-	-
Time limit	-	-

To populate Table 2 for every station, price descriptions were first processed to eliminate common language inconsistencies. This involved two steps: (1) vocabulary regularization via string segment replacement, and (2) elimination of extraneous information. The first step involved identifying price-relevant string segments in the data and assembling groups of segments that have an equivalent meaning. For those meanings that can be expressed by multiple different string segments, a consistent and explicit representation of that meaning was chosen, and all equivalent segments were replaced with the consistent representation. This was done via regular expressions in Python [36] (see Appendix A.1 for more details). For example, stations with free charging (for part or all of a session) used terms such as “complimentary”, “no cost”, “\$0.00 per hour”, “free charging”, or “free to charge”; kilowatt-hours could be referred to as “kwhr”, “kilowatt hour”, “kWh”, “kwh”, and sometimes, mistakenly, as a price “per kilowatt” or “per kW/h”. In the second step (removing extraneous information), any non-digit characters that had not been identified as relevant during step 1 were removed. As an example, the description “\$1.25/Hr for first four hours, \$10.00/Hr afterwards” was converted to “\$1.25 IPER HOUR, 4 HOUR, \$10.00 PER HOUR”.

After regularizing vocabulary and removing extraneous characters, descriptions were organized into a format consistent with the headings in Table 2 and separated into expressions that each contain a complete account of the price description. The example description from above has two constituent expressions: “\$1.25 PER HOUR, 4 HOUR”, and “\$10.00 PER HOUR”. An algorithm, detailed in Appendix A.2, was then used to interpret the meaning of each expression and populate the table. The algorithm was developed incrementally. During each iteration of algorithm development, price descriptions that could not be fully interpreted were identified and used to make adjustments to the algorithm to

enable correct interpretation. This process was repeated until interpretation failures could only be attributed to contradictory or otherwise ambiguous pricing structures. In such cases, the algorithm is designed to select the lower of the interpreted prices and label the price as partially interpreted.

4.2. Price Regularization

Pricing structures extracted from price descriptions were regularized by translating from their original units (which include \$/kWh, \$/h, \$/min, and \$/session) into units of \$/kWh. This translation was done by evaluating the effective price, in \$/kWh, that would be assessed in each of a set of charging scenarios (shown in Table 3), assuming constant nominal charging rates. For example, a DCFC station with a price of \$10 per session would be translated, for Scenario 1, to $\frac{\$10}{\text{session}} \div \frac{0.25 \text{ h}}{\text{session}} \div \frac{50 \text{ kWh} \cdot \text{h}}{1 \text{ h}} = \$0.80/\text{kWh}$. In Scenario 3, the same station's effective price would be \$0.20/kWh, because more energy is supplied for the same total cost.

Dynamic prices were similarly regularized as the total cost assessed divided by the total energy supplied. For example, if a DCFC station assesses a session fee (sometimes called “connection fee”) of \$1.00, plus \$0.10/kWh for the first 20 min and \$0.20/kWh thereafter, with a maximum of \$5, computing the effective price requires summing the costs during each applicable time window. Scenario 1, 12.5 kWh in 15 min, falls within the first (20 min) window:

$$\frac{\$1.00/\text{session} + 12.5 \text{ kWh} \times \frac{\$0.10}{\text{kWh}}}{12.5 \text{ kWh}} = \$0.18/\text{kWh}$$

For Scenario 2, 25 kWh in 30 min, two time windows with distinct pricing apply, 0–20 min and 20–30 min:

$$\frac{\$1.00/\text{session} + \left(\frac{20 \text{ min}}{60 \text{ min}} \times 50 \text{ kWh} \times \frac{\$0.10}{\text{kWh}}\right) + \left(\frac{30-20 \text{ min}}{60 \text{ min}} \times 50 \text{ kWh} \times \frac{\$0.20}{\text{kWh}}\right)}{25 \text{ kWh}} = \$0.1733/\text{kWh}$$

For Scenario 3, 50 kWh in 1 h, the maximum price is reached. Thus, the effective price is $\$5.00/50\text{kWh} = \$0.10/\text{kWh}$.

This process was applied to every price description extracted from the dataset. Mean prices per scenario are shown in Figure 3, differentiated by power level (L2 and DCFC) and the original, pre-regularization unit of assessment. The prices presented later in the paper (Figure 4 and on) are the mean of the prices for the three scenarios.

It is important to note that physically delivered charging rates can vary from the nominal rate during a session, particularly with DCFC, which is not accounted for in this analysis. Charging rates are typically less than or equal to the nominal rate and can drop substantially when the battery capacity nears full, especially during DCFC [37]. Thus, converting time-assessed prices to energy-assessed prices using this method results in an underestimate. However, because power delivery curves can vary with the EV model, battery age, ambient temperature, and other factors, the magnitude of underestimation is uncertain. Some regions, with California as an example, have begun to require all new public EV charging stations to assign prices in units of energy, in an effort to ensure price consistency during and between charging sessions and EV models [38].

Table 3. Charging scenarios for Level 2 and DCFC. The regularized price to charge is determined as the mean of the prices for the three scenarios.

Category	DCFC	Level 2
Assumed Charging Power	50 kW	6.6 kW
Scenario 1	12.5 kWh in 15 min	6.6 kWh in 1 h
Scenario 2	25 kWh in 30 min	13.2 kWh in 2 h
Scenario 3	50 kWh in 1 h	19.8 kWh in 3 h

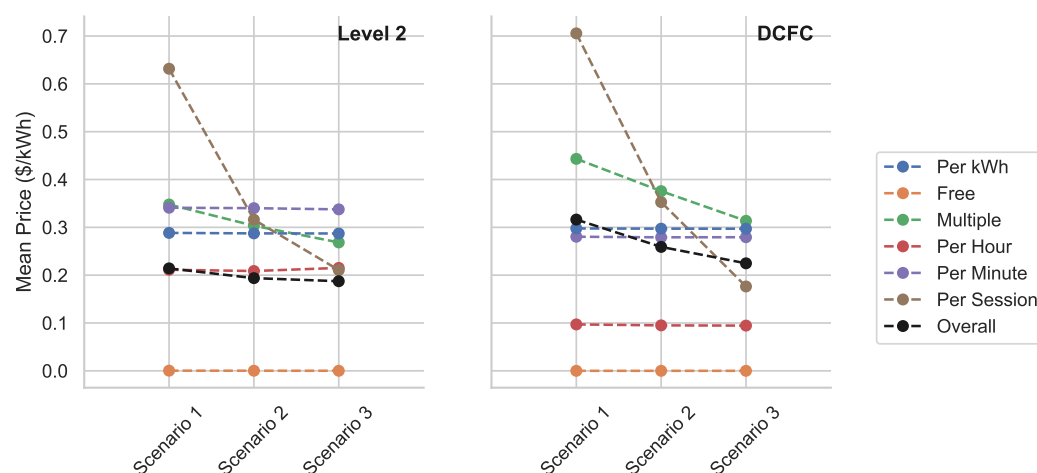


Figure 3. Effective prices, in \$/kWh, for Level 2 and DCFC for each charging scenario, distinguished by the original pricing unit.

Additional complexity representing such mechanisms as membership fees and discounts is present in the business models of some public charging network entities, but the extent to which these are reflected in the dataset is unknown. If a price is available only to subscribers, this fact is not necessarily articulated in the description. By ignoring additional subscription fees, the prices in such cases would appear to be less than they are in reality. However, even if all membership and subscription fees were known, the effect on the regularized charging price is a function of charging behavior, ranging from negligible (costs paid directly for charging are much greater than membership fees) to enormous (membership fee is paid but no charging occurs). Therefore, these pricing mechanisms are considered out of scope for this work.

5. Results

Descriptive analytics, in the form of graphs of the interpreted data, are presented in this section. These analytics are intended to summarize the quantitative data extracted from the dataset, in part to demonstrate the utility and reliability of processing the data using the presented methods. They also provide a high-level overview of public EV charging prices and how they vary within the diverse U.S. public EV charging network. Price variability is present with respect to geography (Figures 4–8), network (Figure 9), location type (Figure 11), and power level (Figure 12).

5.1. Spatial Distribution

Figures 4 and 5 show the total number of L2 stations and the median charging price for all associated connectors, by county, throughout the United States. Similarly, Figures 6 and 7 show the number of DCFC stations and the median charging price for all associated connectors, by county, throughout the United States. Counties without any L2 station records (in the case of Figure 4) or DCFC station records (in the case of Figure 6) are indicated as blank areas. Median prices encompass only those connectors for which unambiguous price information is available. Both L2 and DCFC stations are more highly concentrated on both coasts and in major metropolitan areas in the country's interior. Median charging prices for L2 stations exhibit a somewhat different spatial distribution than do median charging prices for DCFC stations. The median charging price for L2 stations is somewhat more levelized across the country except, perhaps, in the northwest and mid-Atlantic areas, while the median charging price for DCFC stations is distinctly higher in the northwest and northeast regions, and in the upper midwest and northern Texas regions. Note that the disparate sizes of counties from east to west can visually bias perceptions about the spatial distributions, and that adopting more or less granular political jurisdictions can change those perceptions.

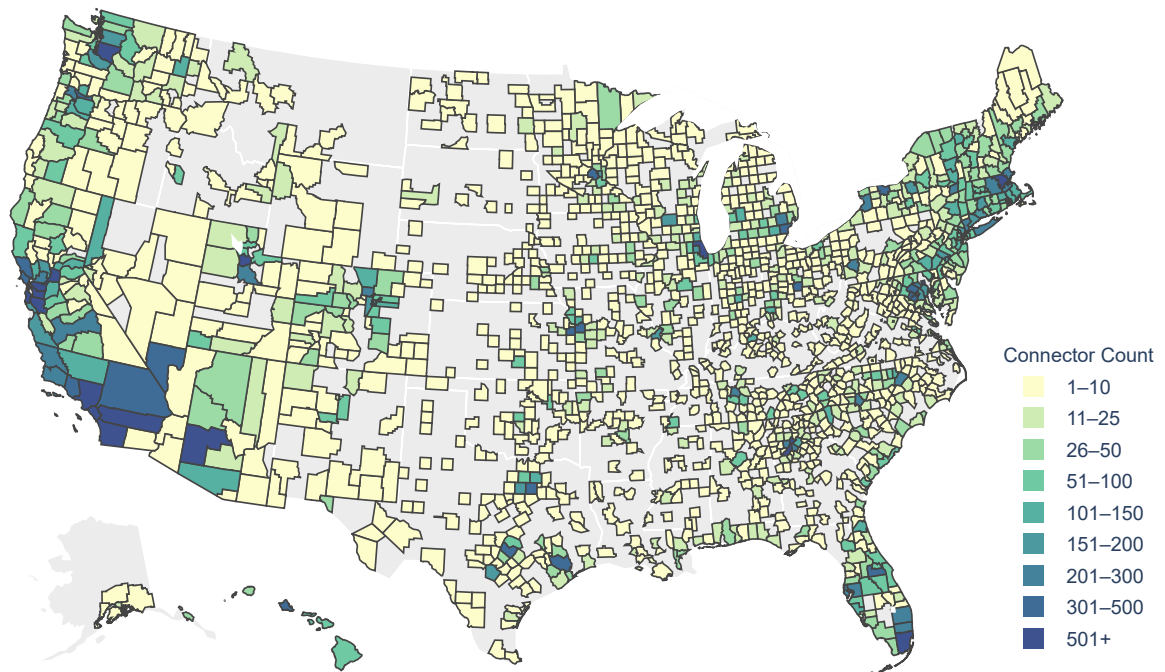


Figure 4. Map illustrating public L2 charging connector counts by county. This includes connectors for which price information could not be extracted.

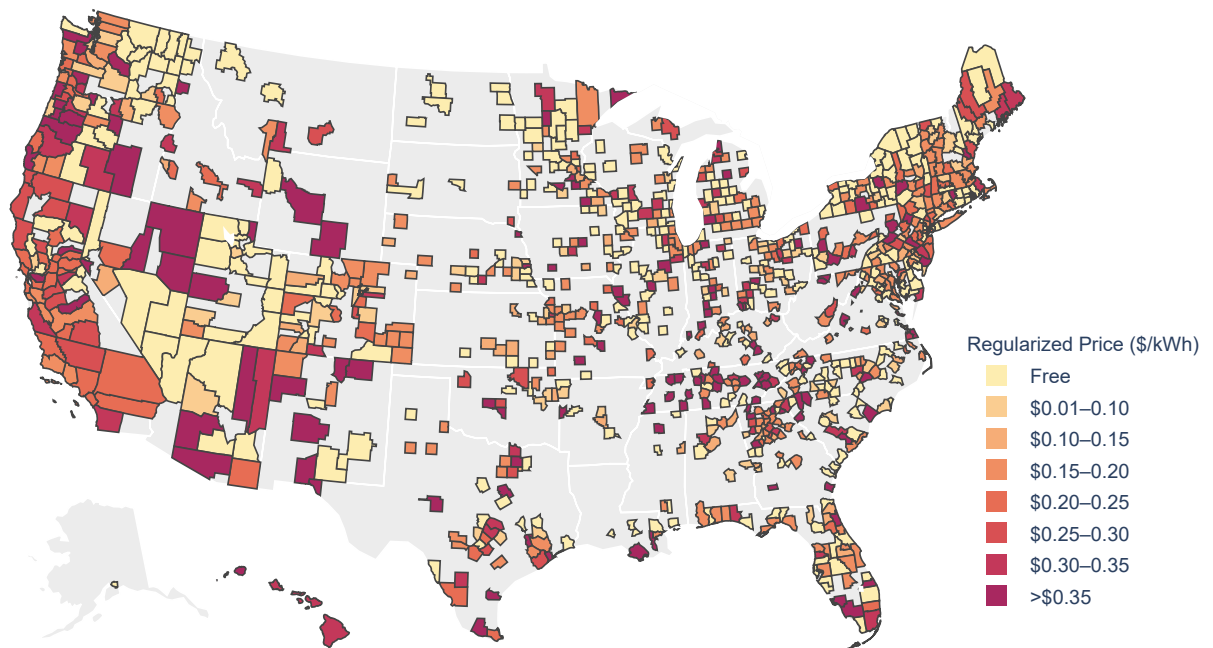


Figure 5. Map illustrating the median L2 charging price for all connectors in each county of the United States, as extracted from the dataset and regularized to \$/kWh. Only connectors with unambiguous textual price information are included.

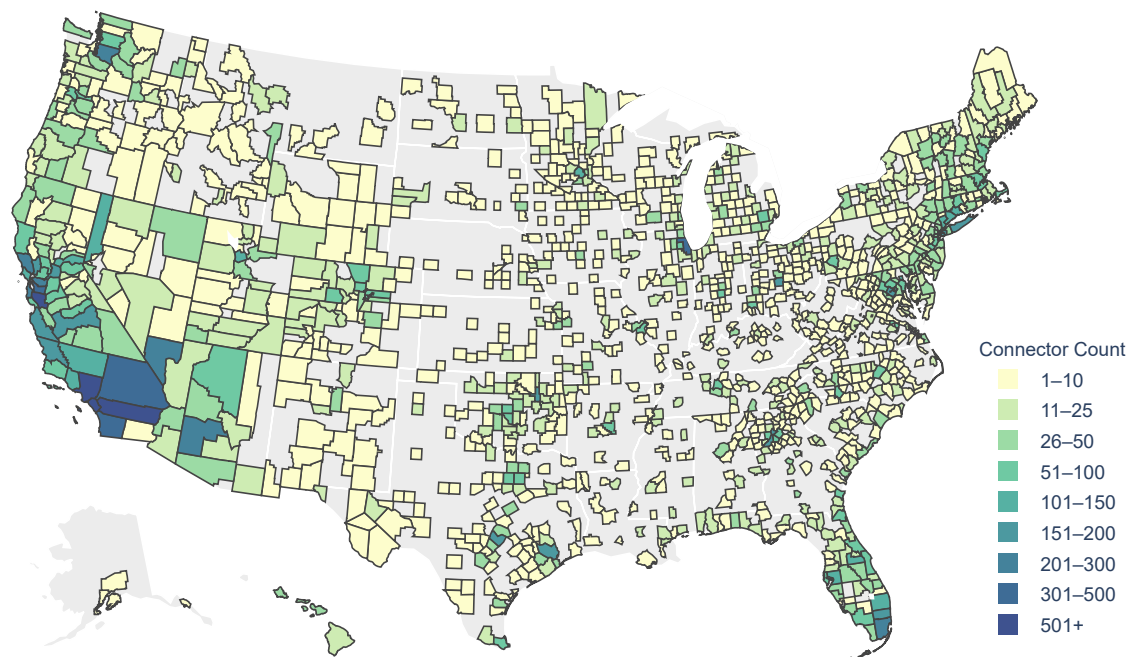


Figure 6. Map illustrating public DCFC charging connector counts by county. This includes connectors for which price information could not be extracted.

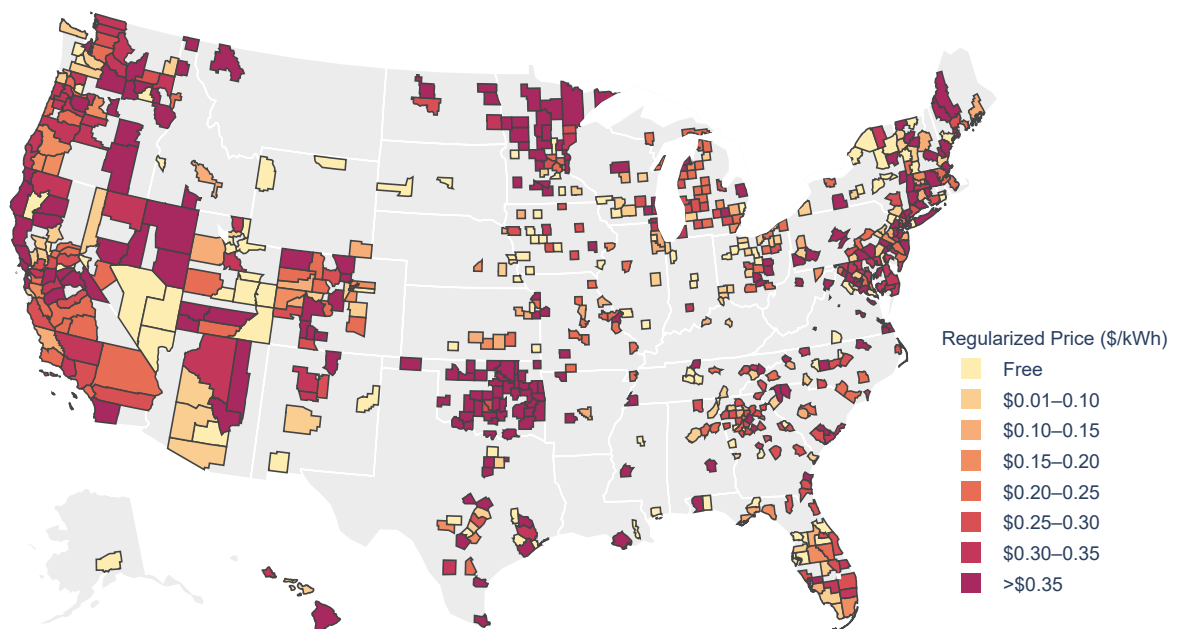


Figure 7. Map illustrating the median DCFC charging price for all connectors in each county of the United States, as extracted from the dataset and regularized to \$/kWh. Only connectors with unambiguous textual price information are included.

Among all L2 stations, the mean effective price to charge across the three cases is 0.277 \$/kWh. Among all DCFC stations, the mean effective price to charge across the three cases is 0.318 \$/kWh. (For reference, the mean cost of residential electricity in the U.S. is 0.133 \$/kWh as of March 2021 [39].) However, effective prices span a wide range. DCFC is consistently more expensive on average than L2, but substantial price variability exists within and between states (Figure 8).

In Figure 8, the states on the horizontal axis are listed in decreasing order of count of records. Although one might expect that states hosting greater numbers of connectors would have lower prices due to increased competition, there is no obvious trend to suggest this is the case. However, it should be reemphasized here that California has many times more records than any other state—more than the total in all 40 states represented by “Other” (see Table A4)—and therefore, that every state’s data are sparse in comparison to California’s.

Additionally, note that in Figure 8 and subsequent similar representations, data distributions are depicted as traditional box-and-whisker plots showing the minimum, maximum, and median values, plus the first and third quartiles. The median, shown as a bold line, may be equal to one or both quartiles if the mode accounts for a sufficiently large fraction of the data.

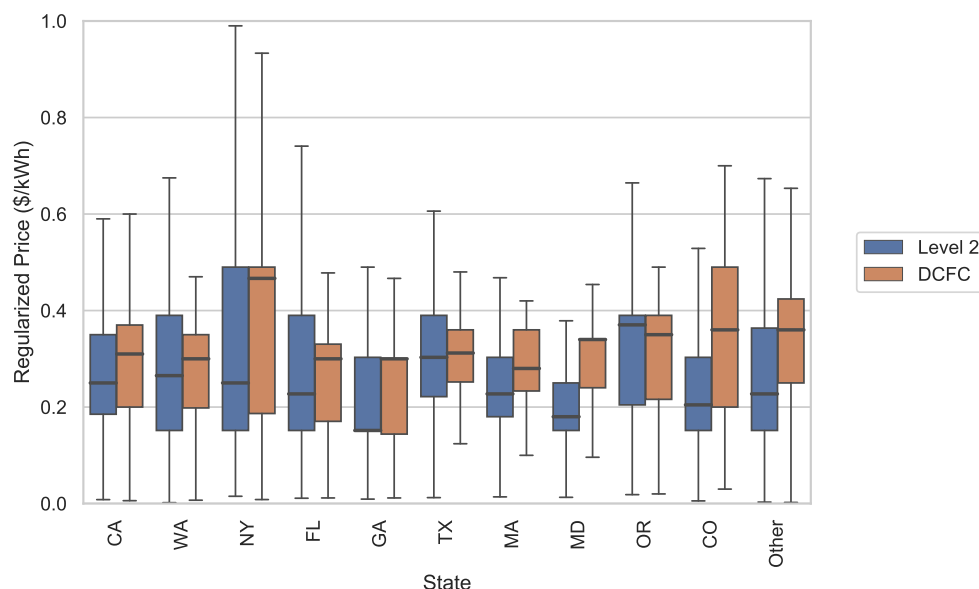


Figure 8. Distributions of L2 and DCFC prices, organized by state. Prices are given for the 10 states with the most charging connectors with price data, and shown in decreasing order of plug counts. The remaining 40 states and Washington D.C. are encompassed by “Other”. The count of records represented by each box is given in Table A4 in the Appendix.

5.2. Networks

Distributions of price by network are shown in Figure 9. Similar to Figure 8, the networks on the horizontal axis are listed in decreasing order based on plug count. If price data are sparse for a network, the price distributions shown may be misleading (see next section). Again referencing Figure 2, connector records are heavily concentrated in the top network, which has even more connector records listed than the state of California.

Still, unlike in the comparison of states in Figure 8, it is clear that some networks have narrower price ranges than others. These differences in price variability may reflect a combination of networks’ spatial span, where widely distributed networks may be subject to a wide variety of utility rates resulting in high price variability, and the extent to which networks impose centralized, network-set pricing, as opposed to station-host pricing.

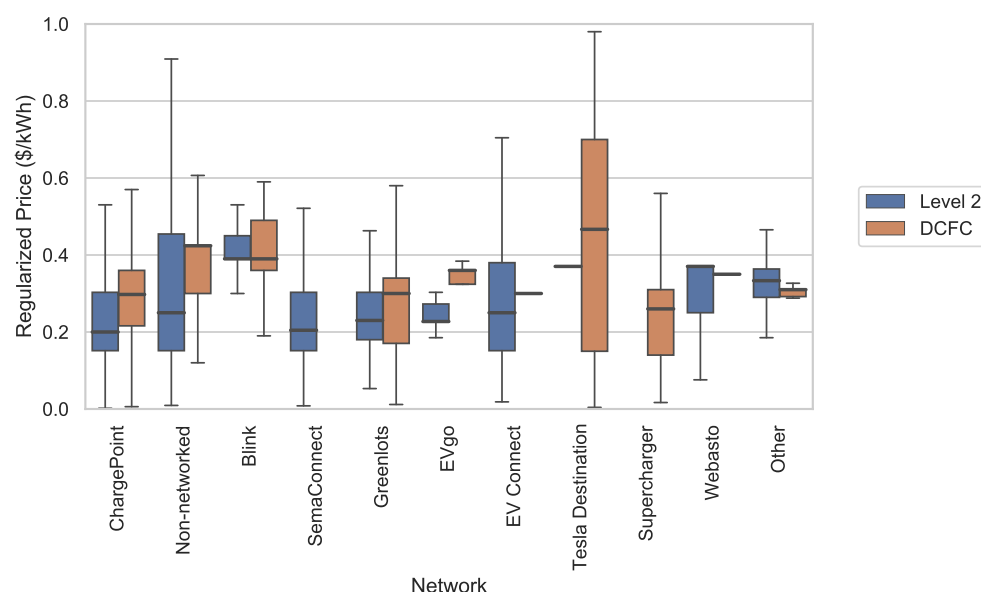


Figure 9. Distributions of L2 and DCFC prices, organized by network operator. Prices are given for the 10 networks with the most charging connectors with price data and shown in decreasing order of plug counts. The remaining 16 networks are encompassed by “Other”. The count of records represented by each box is given in Table A5 in the Appendix.

Missing DCFC Data

When taking into account all levels of charging, Tesla, via its Supercharger and Tesla Destination networks, hosts the second-most stations of any network. However, if considering only DCFC, they account for the overwhelming majority of networked chargers (Figure 10). Since the Tesla network is only available to Tesla drivers through a proprietary app and vehicle interface, Tesla has little incentive to provide accurate pricing information on public-facing third-party apps, such as PlugShare. Accordingly, only a small fraction of their charging connectors have price information in the dataset, and even these prices may be out of date. Our lack of access to most of Tesla’s prices, and those of other DCFC networks, is a major limitation to the DCFC portion of this analysis.

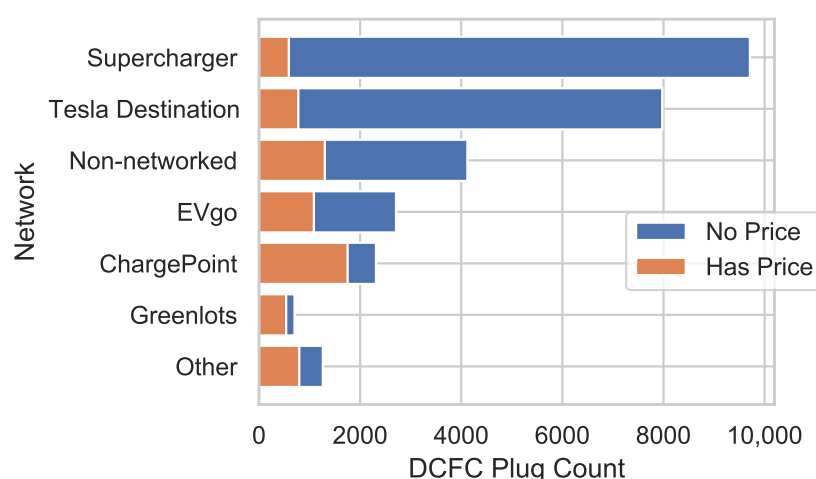


Figure 10. Counts of charging connectors hosted by each network operator, restricted to DCFC.

5.3. Location Type

In the data, 44 types of charger location, or “places of interest”, are distinguished (Figure 11). While variability between categories appears to be limited relative to that between states or networks, some categories stand out. For example, whereas median

prices at hotels are high, median prices at schools are comparatively modest. This may reflect the role that the necessity of charging plays in setting prices. Visitors to hotels, who are less likely to be near home, presumably have a greater need to charge than do visitors to other location types. Again, sparsity of data should be taken into account (Table A6 in the Appendix B). There are more than five times as many records for parking garages/lots (the most populous category shown) as for restaurants (the least populous category shown).

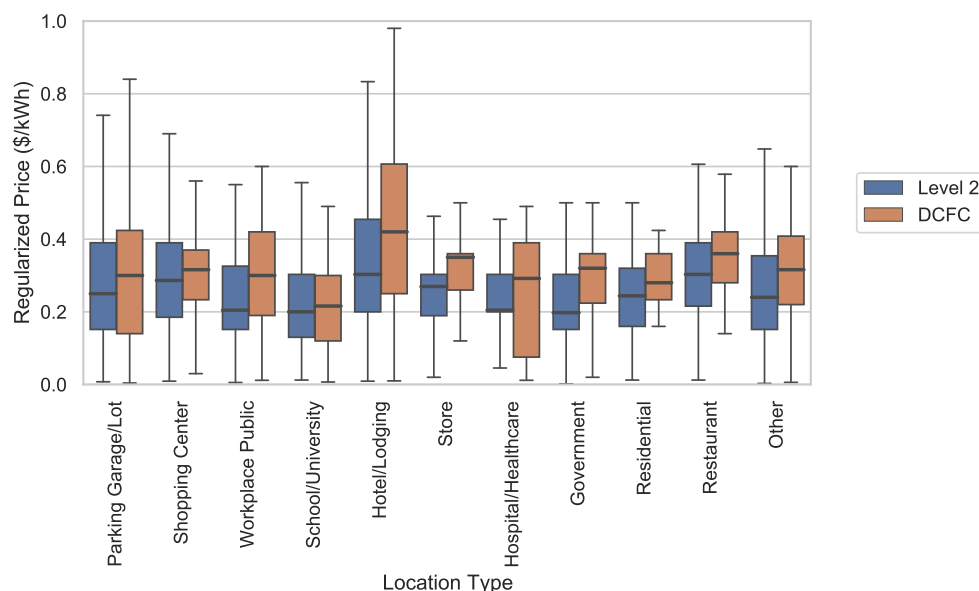


Figure 11. Distributions of L2 and DCFC prices, organized by location type. Prices are given for the 10 location types with the most charging connectors and shown in decreasing order of plug counts. The remaining 34 location types are encompassed by “Other”. The count of records represented by each box is given in Table A6 in the Appendix B.

5.4. Power Level and Units

Variability exists between power levels (DCFC is generally more expensive per kWh than L2) and as a function of the original unit of assessment. As shown in Figure 12, session-based prices vary widely when expressed as regularized prices in \$/kWh. This may be an artifact of the method for regularizing price: since the regularized price is the mean over the three scenarios (Table 3), charging sessions can only range between 1 and 3 h, for L2, and between 15 min and 1 h, for DCFC. It may be rare, for example, that a driver pays an expensive session price to charge for only 15 min, but the price for such a scenario (Scenario 1 for DCFC) is included in the regularized price calculation shown in these results.

Once again, it should be noted that some of the boxes in Figure 12 represent sparse data (see Table A7 in the Appendix B). For example, only 487 of 6834 DCFC stations use a price in units of \$ per hour. The low apparent price for hourly DCFC may thus be an artifact of data sparsity. Alternatively, the sparsity and low apparent prices for hourly DCFC might reflect a psychological aspect of pricing. Relative to L2 prices, DCFC prices expressed as \$/h may appear unusually high to EV operators due to the much higher rate of energy delivery. For example, to deliver energy at an effective price of 0.30 \$/kWh, an L2 station’s hourly price would be 1.98 \$/h, whereas a DCFC station’s hourly price would be 15.00 \$/h. The equivalent price advertised as a price per minute (0.25 \$/min) may be more attractive to EV operators.

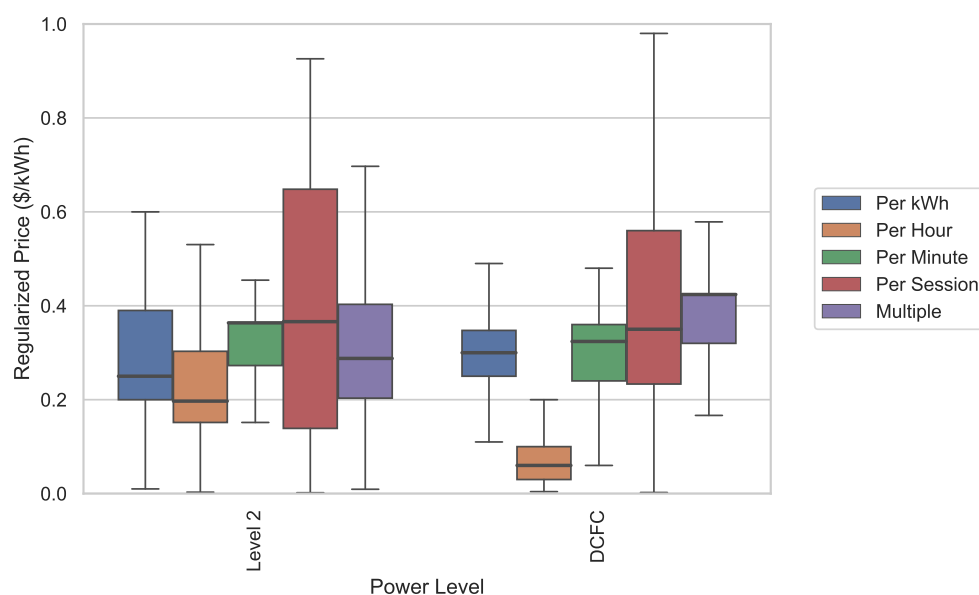


Figure 12. Price distributions for L2 and DCFC, organized by unit of assessment. “Multiple” units of assessment are sometimes applied in dynamic rate structures. The count of records represented by each box is given in Table A7 in the Appendix B.

5.5. Dwell Incentive

Prices can be used as signals to encourage EV operators to extend or shorten the duration of charging sessions. We refer to this as a positive or negative “dwell incentive”. As previously illustrated in Figure 3, for some pricing structures, the effective overall price can change as a function of charging session length. For example, when charging costs are applied as a flat per-session fee, the effective price of energy decreases throughout a charging session. This may serve as an incentive for EV operators to extend charging sessions, potentially to the benefit of nearby retailers. Alternatively, some pricing structures deliberately increase the price of charging during a session, providing an incentive for shorter charging sessions, potentially to the benefit of electricity providers. These are examples of strategies, as highlighted in a 2019 study, to leverage EV operators’ flexibility to adjust the duration and energy consumption of charging sessions [40].

We use a measure of dwell incentive to demonstrate where and how dynamic price structures are implemented. The dwell incentive is calculated by assessing the change in effective price, in \$/kWh delivered, as the session duration increases. If the effective price remains constant irrespective of session duration, the dwell incentive at that station is “neutral”; if the price increases with session duration, the dwell incentive is negative; and if the price decreases with session duration, the dwell incentive is positive.

As shown in Figure 13, the dwell incentive appears to correlate with effective price. On average, stations with a positive dwell incentive charge high effective prices relative to other stations. This suggests a strategy of maximizing revenue per customer (i.e., the drivers who plug in, despite the high price, are incentivized to stay longer), potentially at the expense of fewer customers (some are turned away by the high prices, or because the plug is in use). In contrast, the low average prices in negative dwell incentive structures suggest a strategy of maximizing revenue by increasing plug utilization: the low price encourages drivers to plug in, but the price increases with time to encourage vacating for the next vehicle.

Figure 14 shows that very few stations employ price structures with non-neutral dwell incentives, and in particular, only a few of those employ a negative dwell incentive. It is plausible that pricing mechanisms for influencing dwell behavior, such as idle fees, are assessed more commonly than they appear in price descriptions in the dataset. Still, the typical configuration of EV charging stations, where payment and energy flow are

both managed electronically, provides a unique opportunity to use price signals for load management or utilization improvement purposes.

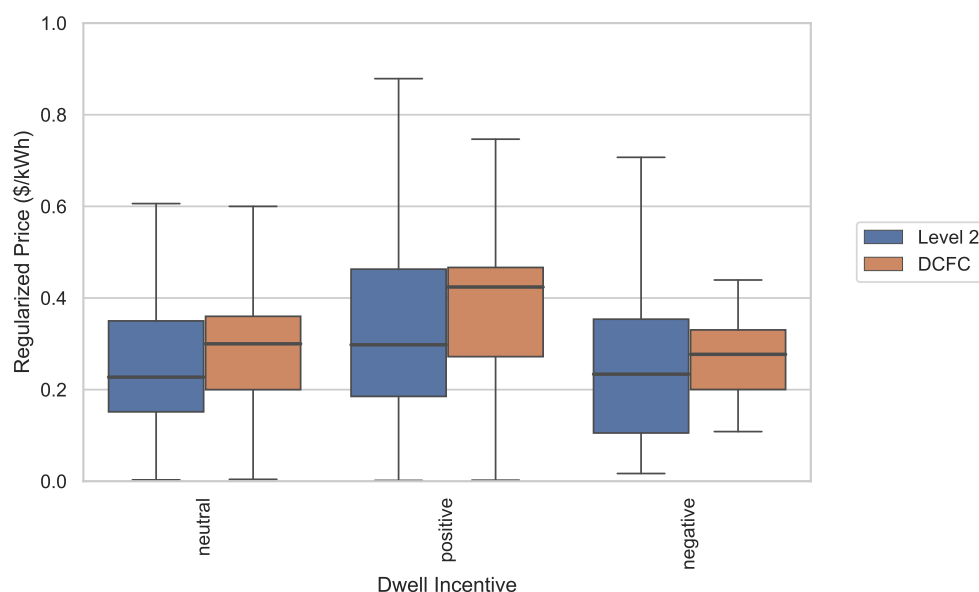


Figure 13. Price distributions for L2 and DCFC, organized by dwell incentive. Dwell incentive refers to whether a price structure encourages (positive) or discourages (negative) extending the duration of charging sessions. The count of records represented by each box is given in Table A8 in the Appendix B.

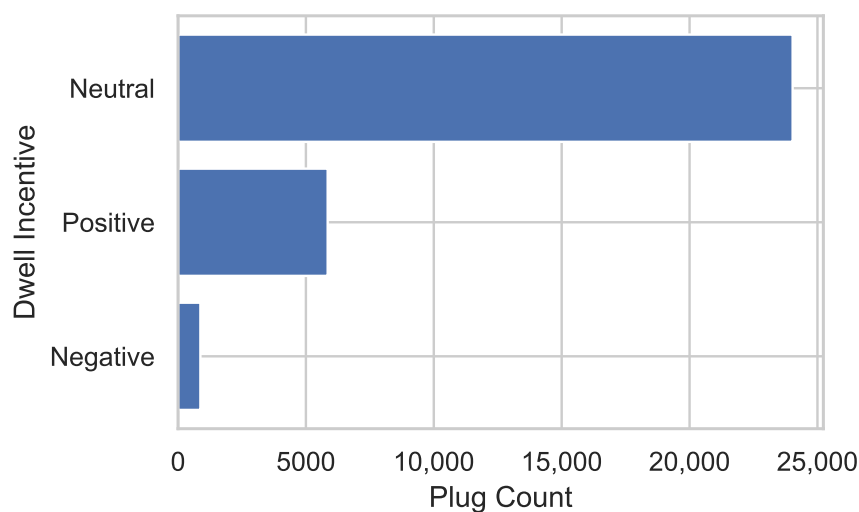


Figure 14. Counts of charging connectors with price structures that provide a positive, negative, or neutral dwell incentive. Dwell incentive refers to whether a price structure encourages (positive) or discourages (negative) extending the duration of charging sessions.

5.6. Comparison with Levelized Cost of Charging

Levelized cost of charging (LCOC) is a metric representing the average cost paid by a station operator to provide charging energy, including initial installation costs and ongoing, time-varying costs throughout the lifetime of the charging equipment. Calculating the difference between LCOC paid by station operators and the average price paid by EV operators is one method for estimating the profit that a station earns.

Median prices obtained from the dataset are higher in every state than the LCOC estimated for station operators. This is illustrated in Figure 15, which compares the prices extracted from the dataset to estimated values of LCOC for different varieties of charging.

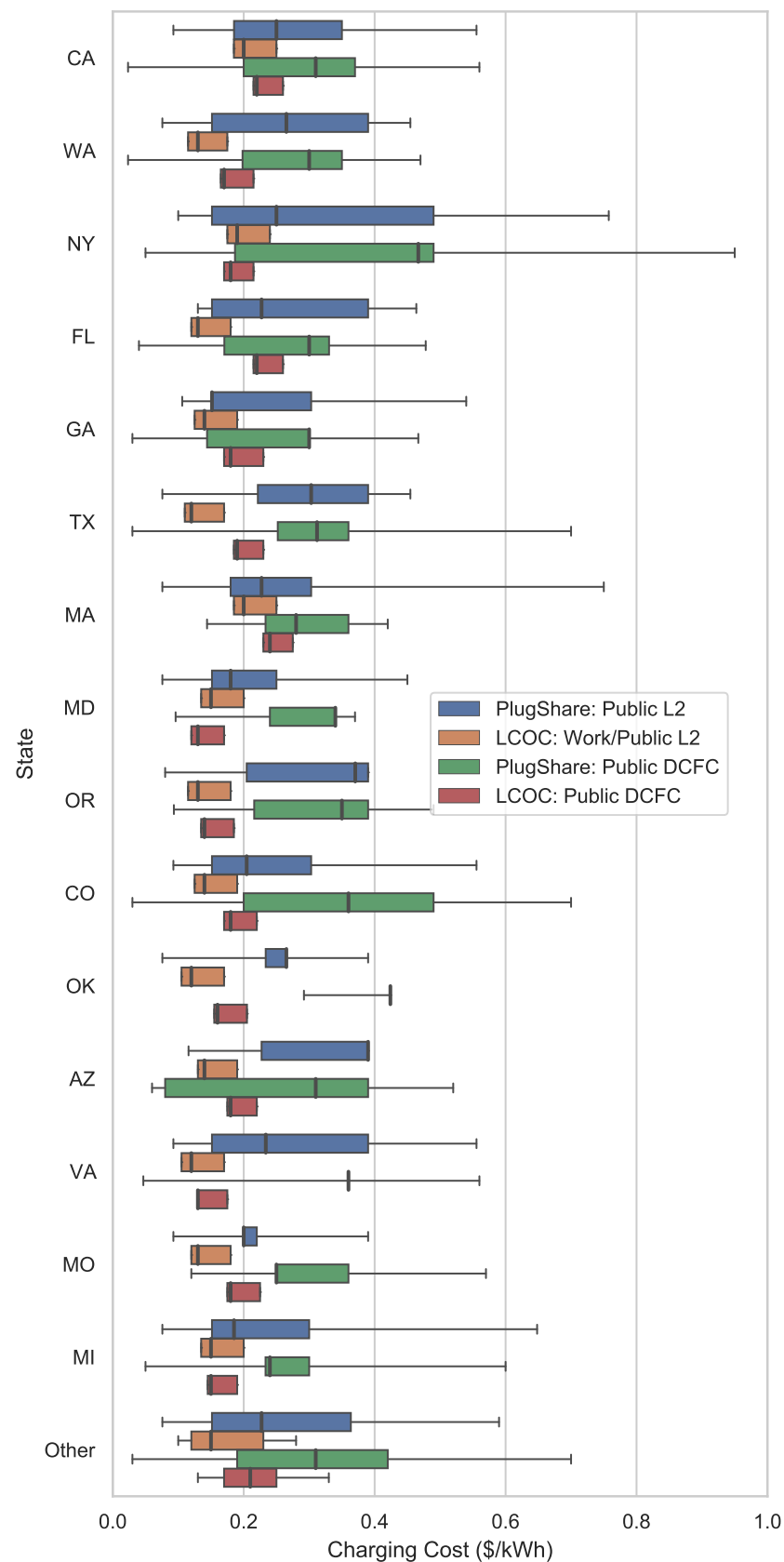


Figure 15. Comparison between prices from the dataset and levelized cost of charging (LCOC), as published in [41]. Prices from the dataset are shown as traditional box-and-whisker plots, where box edges denote quartiles. Data from [41] are shown as a range across lower sensitivity (left edge), baseline (midline), and upper sensitivity (right edge) scenarios.

The LCOC values shown in Figure 15 are taken from a study of 2019 EV charging economics [41]. In this study, researchers detailed the variability of EV charging economics across different charging sites, regions, power levels, and other variables. They estimated LCOC, for an individual charging site, as a function of (a) retail electricity prices, (b) capital and operating costs for the charging equipment, and (c) energy supplied during the lifetime of the equipment. Two sensitivity scenarios (upper and lower) aimed to capture variability in these parameters, leading to higher and lower costs than the baseline scenario.

The comparison in Figure 15 thus serves to emphasize the substantial difference between the estimated LCOC and the actual prices assessed, throughout the U.S., for both L2 and DCFC. One implication of this difference is that the value of energy from a public charging station is substantially higher to a typical EV driver than the cost paid by station operators to provide it. This calls attention to attributes of public EV charging. First, most EV drivers do not have to rely on public infrastructure for the majority of their driving energy, resulting in a different value proposition for drivers at public charging locations relative to home charging or gasoline/diesel refueling. Secondly, utilization may be limited in an early EV market due to the complex means by which infrastructure availability both spurs and reacts to adoption of EVs, representing a restriction to supply that may exert upward pressure on prices. Third, station operators may pay a higher electricity price than nominal retail electricity prices due to pricing mechanisms, such as peak demand tariffs or time-of-use rate schedules, in which case the LCOC would be higher in reality than the estimated values. Each of these attributes is discussed further in the following paragraphs.

5.6.1. Value Proposition for EV Drivers

EV drivers choose from a broader set of refueling locations than do drivers of conventional vehicles, who are confined to refueling at commercial gasoline/diesel stations. This highlights a fundamental difference between the business cases for public charging stations and petroleum refueling stations. Most EV drivers are able to charge at home, and some can charge at the workplace, both of which are likely to be cheaper and more convenient than stopping at a public charging station for either L2 charging or DCFC. Public stations thus serve (a) to enable trips exceeding the EV battery range and/or (b) to provide faster charging than drivers have available at home or work. From the perspective of EV drivers, the value of charging can therefore be considered to be the sum of the direct value of energy and the indirect value of range extension and faster charging (convenience and/or preference), resulting in drivers willing to pay a higher price than the LCOC. An analogous product for which the willingness to pay can be dramatically influenced by differences in convenience and/or preference is water, which usually comes at a significant premium, in bottled form, relative to the price of tap water at home.

5.6.2. Station Utilization in an Early EV Market

Public charging infrastructure and EVs are complexly interrelated in that each increases the value and viability of the other. This is an example of a commonly remarked “chicken-or-egg” problem. If charging is not sufficiently ubiquitous to enable long-distance travel, most people may be unlikely to adopt EVs, but some stations providing widespread charging in an early market will experience low utilization, while EV populations are low. In [41], public L2 connectors were assumed to be utilized 4.5 h per day, whereas DCFC connectors were modeled at varying levels of utilization, from 1–2 charges per day to over 20% utilization. At present, however, these utilization assumptions may yet be overestimates for many stations.

5.6.3. Peak Demand and Time-of-Use Electricity Tariffs

Finally, electricity prices are often designed to discourage high local and aggregate power demands via peak demand and time-of-use tariffs, which can result in high prices for EV charging, especially DCFC. The authors of [41] accounted for the effect of tariff variations on DCFC by testing a total of more than 4000 commercial rates and reporting the

overall average price for each state. Still, they report that the effective price of electricity for DCFC can exceed \$2 per kWh [42]. As utility companies continue to adapt to the emerging demands of EV charging, some charging stations may continue to pay electricity prices according to structures that result in expensive refueling using DCFC infrastructure. Alternative solutions, such as installing means of electricity generation (solar panels) or storage (stationary batteries) to minimize or offset power demands, have been proposed to reduce the cost of electricity and mitigate other challenges with the interactions between the electric grid and EV charging stations [37,43,44].

6. Discussion and Future Directions

Access to a comprehensive source of EV charging price data can facilitate decision-making for EV operators, charging station operators, policymakers, and business innovators. However, such data do not yet exist in an aggregated and accessible format. PlugShare's crowdsourced U.S. dataset is an attractive source of nationwide charging price data, but the unstructured textual format of its price data has hindered its usability. By employing ad hoc text mining to convert the data into a format amenable to direct analysis, this work lays the foundation for studies of a previously underutilized source of data. Descriptive analytics of the converted dataset provide a high-level image of the state of public EV charging across the United States, with emphasis on the wide variability of charging prices in terms of geographic location, network operator, and location type.

EV charging stations operate under a variety of business models and pricing structures that are vastly different from those associated with commercial petroleum fueling stations. The flexibility in price design equips operators with tools to provide incentives for desired charging behaviors, such as ramping prices to discourage long charging sessions. Our analysis suggests that these tools are not yet being used by the majority of EV charging station operators. Further research to understand the effects of potential price designs on customer choices may provide valuable direction for station operators, especially as charging demand increases.

Because it is often an alternative to at-home charging, the business case for public EV charging is distinct from that for conventional fueling. Our research suggests that prices at most stations exceed estimates for the LCOC paid by station operators, resulting in prices well above what consumers would pay at home and highlighting the unique value proposition of public EV charging. This premium in price represents value beyond that of energy, such as convenience, speed, or necessity, but it remains to be seen what prices consumers will accept in a mature EV charging market.

From the perspective of station owners, EV charging infrastructure comes at a high capital cost that must be recouped, whether through a revenue margin on electricity above the LCOC or by other methods, such as increased revenues at an associated business. The wide variety in approaches to public charging suggests that the electric transportation system remains in its developing stages.

Data wrangling and preprocessing can be tedious, time-consuming, and sometimes unproductive pursuits; working with large volumes of unstructured textual information further exacerbates these issues [45]. While text mining provides computational and statistical tools to address the problem, there is still no fully automated way to reduce natural language to numerical data that can be used for quantitative analysis. As illustrated in our work, such circumstances require the use of creative ad hoc approaches to extract useful analytical information. However, we hasten to underscore the imperfections in such approaches and the implications they may ultimately have on modeling results and conclusions. Given the growing interest in EVs and infrastructure to support them, we note the necessity of securing reliable and consistent data on which to construct models for operations and business planning.

In this study, we address one limitation to the usability of the dataset, but it suffers other limitations that we are unable to correct. Because the data are not publicly and freely available, the potential for research using the data is limited to those able to pay the

access fees. Furthermore, the restrictions imposed by licensing agreements for non-public datasets inhibit the ability of researchers to provide transparent and reproducible work to the public.

An additional limitation to the usability of the dataset is its method of sourcing. By distributing the labor and costs required to obtain data, “crowd sourcing” can generate large volumes of data that may not be obtainable by other means. However, due to its decentralized sourcing, the value and quality of crowd sourced data can be questioned. Particularly when the data are not made public and open-sourced, the ability of researchers to assess value and quality is limited [46,47]. This study provides an assessment of the value and quality of the dataset in the form of descriptive summaries and analytics.

Even with its limitations, the dataset presently represents one of the best and most current sources of information about charging costs that can be used to inform consumers and operators alike. As described here, the challenge is to reduce the dataset (and similar information sources) into a comprehensible and analytical format that can be effectively employed for decision making. To date, our work has primarily focused on describing the present status of public charging prices in the U.S.; however, we believe continued expansion of the dataset and fine-tuning (training) of our information extraction algorithm will support further investigations that are more predictive and prescriptive in nature. Future modeling work will incorporate the regularized and cleaned data with various operating parameters to help guide the establishment of best practices to promote EV adoption and investment in infrastructure build-out relative to the cost of EV charging.

Author Contributions: Conceptualization, J.D.; methodology, J.D. and D.T.; software, D.T.; validation, D.T.; formal analysis, D.T., E.P. and T.C.; investigation, D.T. and J.D.; resources, J.D. and T.B.; data curation, J.D.; writing—original draft preparation, D.T. and E.P.; writing—review and editing, T.C., J.D. and T.B.; visualization, D.T.; supervision, T.C.; project administration, J.D.; funding acquisition, N/A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of the data used in this analysis, which were obtained from PlugShare (<https://company.plugshare.com/data.html>, accessed on 1 February 2021).

Acknowledgments: The authors gratefully acknowledge PlugShare’s support in navigating and understanding their EV charging dataset.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CHAdEMO	CHArge de MOve (charging standard)
DCFC	Direct current fast charging
EV	Electric vehicle
L2	Level 2 charging
LCOC	Levelized cost of charging
SAE	Society of Automotive Engineers

Appendix A. Details of Text Mining Algorithm

The text mining algorithm for extracting information from the unstructured charging station cost descriptions involves three steps: (1) regularize the vocabulary of the descriptions via string segment replacement (Appendix A.1); (2) identify and interpret groups of regularized terms (Appendix A.2); (3) convert general price structures into terms of \$/kWh (Section 4.2 in the main text).

Appendix A.1. Vocabulary Regularization

The vocabulary regularization step aims to standardize string segments that are different but have equivalent meaning, and to eliminate portions of the text descriptions that are not relevant to the station's pricing structure. The standardized vocabulary is shown in Table A1.

Table A1. Standardized vocabulary substituted into textual price descriptions.

Standard Term	Meaning
DOLLAR	A dollar sign or other indication that a price follows a position in the string
PER	Any indication that the price expression preceding a position is to be assessed in terms of the unit following the position (commonly a forward slash "/")
DECIMAL	A decimal point (as distinguished from a period) that indicates any preceding digit characters should be interpreted as whole numbers, and any following digit characters should be interpreted as digits after the decimal
MAXIMUM	An indication that the preceding or following number (or number/unit combination) expresses an upper bound on charging cost
MINIMUM	An indication that the preceding or following number (or number/unit combination) expresses a lower bound on charging cost; care must be taken to distinguish whether "min" or its variations is meant as "minute" or "minimum"
FIRST	An indication that the following number/unit combination should be understood as the initial applicable price (e.g., "FIRST five minutes are free")
FREE	Expresses that no charge is assessed during the window with which "FREE" is associated
HOUR	Any version of "hour" meant to be interpreted as a unit of time
MINUTE	Any version of "minute" meant to be interpreted as a unit of time
SESSION	An indication that a flat price is assessed irrespective of charging session duration or quantity of energy supplied
KWH	Any version of "kWh" meant to be interpreted as a unit of energy
(digits 0–9)	Any number, whether spelled out in characters or as a digit

Regular expressions (regex) use a standardized syntax to represent textual search patterns, which enables isolating segments of text matching criteria ranging from very broad to very narrow. These, implemented using Python's "re" module, were used to identify instances of the many various equivalents to the Table A1 terms in the unstructured text, and to substitute the standard versions shown in the table. Any characters remaining that are not part of standard terms are removed, leaving only information pertaining directly to price structures.

Context can usually be used to infer the intended meaning of terms that have multiple possible interpretations. For example, "min" is variously used to mean "minute(s)" and "minimum", but in this dataset, when "min" appears immediately after a number, it is interpreted as "minutes", whereas when it appears immediately before or after a time or energy unit (e.g., minutes, hours, kWh), or immediately before a number, it is interpreted as "minimum".

Appendix A.2. Interpretation of Regularized Text

Although regularized price descriptions comprise standardized terms, interpreting meaningful price structures still requires an ad hoc approach. The headings from Table 2, replicated in Table A2, serve as a standard framework for the static and dynamic price structures found in the dataset.

A table with headings from Table A2 is populated, one row per connector in the data, following a process of searching for and replacing key phrases. After a phrase is identified and interpreted, it is removed from the price description. This “search-and-replace” process must therefore proceed in a specific order to avoid capturing fragments of more complete phrases. The regular expressions used in this process are described below.

Table A2. Table headings for populating the details of every interpreted price description, replicated from Table 2.

Initial price
Initial price 2
Initial time window
Price next window
Next window
Price next window (2)
Next window (2)
Price next window (3)
Next window (3)
Minimum
Maximum
Time limit

The sequence of regular expressions in Table A3 are used as inputs, along with the text descriptions, to iterative applications of the “search” function in the Python re module. To avoid interpreting the same phrases multiple times, segments of the text that are successfully captured during a search are removed prior to applying the next expression in the sequence. The expressions are sequenced with the intent of capturing the fullest expressions of price first.

Table A3. Sequence of regular expressions used to interpret vocabulary-regularized price descriptions. Parentheses surround sub-expressions that represent capture groups.

#	Regular Expression
1	(\d*) (DAY HOUR MINUTE KWH) (MAXIMUM MINIMUM)
2	(MAXIMUM MINIMUM) (\d*) (DAY HOUR MINUTE KWH)
3	DOLLAR(\d*) (?:DECIMAL)? (\d*) (MAXIMUM MINIMUM)
4	(MAXIMUM MINIMUM) DOLLAR(\d*) (?:DECIMAL)? (\d*)
5	(\d+) (HOUR MINUTE KWH) FREE
6	(?:DOLLAR) (\d*) (?:DECIMAL)? (\d*) PER (HOUR MINUTE KWH) PER (\d+) (HOUR MINUTE KWH)
7	(?:DOLLAR)? (\d*) (?:DECIMAL)? (\d*) PER (\d*) (?:DECIMAL)? (\d*) (DAY HOUR MINUTE SESSION KWH MONTH SECOND)
8	DOLLAR(\d+) (?:DECIMAL)? (\d*)
9	(\d*) (?:DECIMAL)? (\d*) CENT PER (\d*) (DAY HOUR MINUTE SESSION KWH MONTH)
10	FIRST(\d*) (HOUR MINUTE)
11	(\d*) TO (\d*) (HOUR MINUTE)
12	(?: THEN AFTER) (\d+) (HOUR MINUTE)

For example, the price description “\$0.05/h for 1 h, then \$0.07/h” (after vocabulary regularization: “DOLLAR 0 DECIMAL 05 PER HOUR PER 1 HOUR DOLLAR 0 DECIMAL 07 PER HOUR”, spaces added here between terms for clarity) must be interpreted in a

particular sequence to avoid extracting an incorrect meaning. For example, if the interpretation code were to apply Expression 4 before applying Expression 2, the “\$0.05/h” segment would be removed and interpreted on its own, leaving “for 3 h, then \$0.07/h”—a phrase without a clear meaning—to be interpreted alone.

There are 3 broad categories of information extracted via this method: prices, price windows, and minima/maxima. Prices, here, are single expressions in the form: *some quantity of money per some unit or quantity of units*. Price windows describe the time period during which a price applies, in the form: *some quantity of time or energy*. Minima or maxima are either (1) in the form: maximum/minimum *some quantity of time, energy, or money*; or (2) the form: *some quantity of time, energy, or money* maximum/minimum. Prices are extracted using regular Expressions 4–9; price windows are extracted using regular Expressions 6–7, 9, and 10–12; and minima/maxima are extracted using regular Expressions 1–4.

Price windows are often interpreted simultaneously with prices, and in these cases it is simple to assign each price to a price window, and furthermore to place price windows in the proper sequence. However, sometimes prices and price windows (as interpreted via regular Expressions 10–12) do not appear together. In such cases, prices and price windows are assumed to appear in respective order, i.e., the first price extracted applies during the first price window extracted; the second price extracted applies during the second price window extracted; and so on. There is one key exception, where sometimes a price description ends with the initial price, most often in a form resembling “First X hours are free”. This case is specifically coded for, where the key word “FIRST” triggers a price and the window to which it sequentially corresponds to be assigned as the first price.

These methods for interpreting prices work for the vast majority of price descriptions, but for especially unusually formatted prices they may fail. This risk increases in the case of applying the code to future tranches of data. The interpretation algorithm thus incorporates a suite of methods to recognize when it is likely to have misinterpreted a price, enabling that record to be set aside from analysis. The two most common failure modes are (1) an inconsistency in the extracted information and (2) segments of the text remaining uninterpreted after the full regex sequence. Identifying inconsistent prices involves checking for mismatches in quantity of prices and price windows; multiple incompatible prices without windows; multiple minima or maxima; or improbable numbers, such as prices or time durations exceeding reasonable expectations.

Appendix B. Counts of Records Represented in Figures

The number of records per category shown in Figures 8–9 and 11–13 are given respectively in Tables A4–A8.

Table A4. Record counts represented by each box shown in Figure 8.

Box	# Records
CA—DCFC	2142
CA—Level 2	10,315
WA—DCFC	304
WA—Level 2	1103
NY—DCFC	266
NY—Level 2	1071
FL—DCFC	312
FL—Level 2	944
GA—DCFC	313
GA—Level 2	737
TX—DCFC	179
TX—Level 2	852
MA—DCFC	88
MA—Level 2	911
MD—DCFC	265
MD—Level 2	720
OR—DCFC	185
OR—Level 2	678
CO—DCFC	181
CO—Level 2	667
Other—DCFC	2606
Other—Level 2	5917

Table A5. Record counts represented by each box shown in Figure 9.

Box	# Records
ChargePoint—DCFC	1752
ChargePoint—Level 2	12,386
Non-networked—DCFC	1302
Non-networked—Level 2	2876
Blink—DCFC	197
Blink—Level 2	2916
SemaConnect—DCFC	0
SemaConnect—Level 2	2636
Greenlots—DCFC	537
Greenlots—Level 2	1553
EVgo—DCFC	1085
EVgo—Level 2	154
EV Connect—DCFC	179
EV Connect—Level 2	771
Tesla Destination—DCFC	778
Tesla Destination—Level 2	3
Supercharger—DCFC	590
Supercharger—Level 2	0
Webasto—DCFC	96
Webasto—Level 2	122
Other—DCFC	325
Other—Level 2	498

Table A6. Record counts represented by each box shown in Figure 11.

Box	# Records
Parking Garage/Lot—DCFC	901
Parking Garage/Lot—Level 2	4495
Shopping Center—DCFC	1692
Shopping Center—Level 2	1897
Workplace Public—DCFC	341
Workplace Public—Level 2	2825
School/University—DCFC	103
School/University—Level 2	2229
Hotel/Lodging—DCFC	718
Hotel/Lodging—Level 2	1338
Store—DCFC	629
Store—Level 2	1109
Hospital/Healthcare—DCFC	38
Hospital/Healthcare—Level 2	1315
Government—DCFC	234
Government—Level 2	1013
Residential—DCFC	28
Residential—Level 2	1027
Restaurant—DCFC	416
Restaurant—Level 2	517
Other—DCFC	1741
Other—Level 2	6150

Table A7. Record counts represented by each box shown in Figure 12.

Box	# Records
DCFC—Multiple	1693
DCFC—Per Hour	487
DCFC—Per Minute	1883
DCFC—Per Session	973
DCFC—Per kWh	1798
Level 2—Multiple	2853
Level 2—Per Hour	9626
Level 2—Per Minute	794
Level 2—Per Session	1546
Level 2—Per kWh	9046

Table A8. Record counts represented by each box shown in Figure 13.

Box	# Records
neutral—DCFC	4543
neutral—Level 2	19,487
positive—DCFC	2256
positive—Level 2	3595
negative—DCFC	42
negative—Level 2	833

References

1. Annual Energy Outlook. U.S. Energy Information Administration. 2020. Available online: <https://www.eia.gov/outlooks/aeo/pdf/AEO2020%20Transportation.pdf> (accessed on 18 January 2021).

2. Electric Vehicle Outlook. Bloomberg New Energy Finance. 2019. Available online: <https://about.bnef.com/electric-vehicle-outlook/> (accessed on 3 January 2020).
3. U.S. Energy Information Administration. What is U.S. Electricity Generation by Energy Source? 2020. Available online: <https://www.eia.gov/tools/faqs/faq.php?id=427&t=3> (accessed on 18 January 2021).
4. EPRI Program 18: Electric Transportation. Consumer Guide to Electric Charging, 2019. Available online: <https://www.epri.com/research/products/3002009442> (accessed on 5 January 2020).
5. Berman, B. The Real Price of EV Public Charging. 2019. Available online: <https://www.plugincars.com/guide-to-public-charging-costs.html> (accessed on 5 January 2020).
6. Gorzelany, J. What It Costs to Charge An Electric Vehicle. 2019. Available online: <https://www.myev.com/research/ev-101/what-it-costs-to-charge-an-electric-vehicle> (accessed on 5 January 2020).
7. EV Charging Statistics. EVAdoption. 2019. Available online: <https://evadoption.com/ev-charging-stations-statistics/> (accessed on 5 January 2020).
8. PlugShare API. Available online: <https://developer.plugshare.com/docs/#introduction> (accessed on 24 February 2021).
9. Anandarajan, M.; Hill, C.; Nolan, T. Practical Text Analytics: Maximizing the Value of Text Data; In *Advances in Analytics and Data Science*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; Volume 2, doi:10.1007/978-3-319-95663-3.
10. Feldman, R.; Sanger, J. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*; Cambridge University Press: Cambridge, UK, 2007.
11. Miner, G. *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*; Academic Press: Cambridge, MA, USA, 2016.
12. Zhai, C.; Massung, S. *Text Data Management and Analysis: A practical Introduction to Information Retrieval and Text Mining*; ACM Association for Computing Machinery: New York, NY, USA, 2016.
13. Aggarwal, C.C. *Machine Learning for Text*; Springer: Berlin/Heidelberg, Germany, 2019.
14. Kwartler, T. *Text Mining in Practice with R*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2017.
15. Silge, J.; Robinson, D. *Text Mining with R: A Tidy Approach*; O'Reilly Media: Sebastopol, CA, USA, 2017.
16. Bengfort, B.; Bilbro, R.; Ojeda, T. *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2018.
17. Chakraborty, G.; Pagolu, M.; Garla, S. *Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS*; SAS Institute: Cary, NC, USA, 2018.
18. Madigan, D. Text Mining: An Overview. Columbia University. Available online: <http://www.stat.columbia.edu/~madigan/W2025/notes/IntroTextMining.pdf> (accessed on 7 May 2021).
19. Weiss, S.M.; Indurkha, N.; Zhang, T. *Fundamentals of Predictive Text Mining*; Texts in computer science; Springer: London, UK; New York, NY, USA, 2010; OCLC: ocn639164988.
20. Sumathy, K.L.; Chidambaram, M. Text Mining: Concepts, Applications, Tools and Issues An Overview. *Int. J. Comput. Appl.* **2013**, *80*, 29–32. doi:10.5120/13851-1685.
21. Allahyari, M.; Pouriyeh, S.; Assefi, M.; Safaei, S.; Trippe, E.D.; Gutierrez, J.B.; Kochut, K. A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *arXiv* **2017**, arXiv:1707.02919.
22. Martinez, A.R. 4—Data Mining of Text Files. In *Data Mining and Data Visualization*; Rao, C., Wegman, E., Solka, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2005; Volume 24, pp. 109–131. doi:10.1016/S0169-7161(04)24004-4.
23. Stavrianou, A.; Andritsos, P.; Nicoloyannis, N. Overview and semantic issues of text mining. *ACM Sigmod Rec.* **2007**, *36*, 23–34. doi:10.1145/1324185.1324190.
24. Kocbek, S.; Cavedon, L.; Martinez, D.; Bain, C.; Manus, C.M.; Haffari, G.; Zukerman, I.; Verspoor, K. Text mining electronic hospital records to automatically classify admissions against disease: Measuring the impact of linking data sources. *J. Biomed. Informatics* **2016**, *64*, 158–167. doi:10.1016/j.jbi.2016.10.008.
25. Sun, W.; Cai, Z.; Li, Y.; Liu, F.; Fang, S.; Wang, G. Data Processing and Text Mining Technologies on Electronic Medical Records: A Review. *J. Healthc. Eng.* **2018**, *2018*, 4302425. doi:10.1155/2018/4302425.
26. Christopher Pereira, P. Text-Mining Maintenance Records to Automate the Identification and Grouping of Failure Modes. In Proceedings of the Offshore Technology Conference, Houston, TX, USA, 4–7 May 2020; doi:10.4043/30737-MS.
27. Wang, H.; Liu, Z.; Xu, Y.; Wei, X.; Wang, L. Short text mining framework with specific design for operation and maintenance of power equipment. *CSEE J. Power Energy Syst.* **2020**, doi:10.17775/CSEEJPES.2019.01120.
28. Gunay, H.B.; Shen, W.; Yang, C. Text-mining building maintenance work orders for component fault frequency. *Build. Res. Inf.* **2019**, *47*, 518–533, doi:10.1080/09613218.2018.1459004.
29. Scott, E.S.; Vafaie, H.; Gungordu, Z.; Horowitz, C.; Brown, B. Text Mining for Quality Control of Court Records. Proceedings of the 2014 ACM Symposium on Document Engineering, New York, NY, USA, 16–19 September 2014, 2014.
30. Muller, O.; Junglas, I.; Debortoli, S.; Brocke, J.v. Using Text Analytics to Derive Customer Service Management Benefits from Unstructured Data. *MIS Q. Exec.* **2016**, *15*, 243–258.
31. Park, Y.; Byrd, R.J. Hybrid Text Mining for Finding Abbreviations and their Definitions. In Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing, Pittsburgh, PA, USA, 3–4 June 2001.
32. Bedathur, S.; Berberich, K.; Dittrich, J.; Mamoulis, N.; Weikum, G. Interesting-phrase mining for ad-hoc text analytics. *Proc. Vldb Endow.* **2010**, *3*, 1348–1357, doi:10.14778/1920841.1921007.

33. Wachsmuth, H. Text Analysis Pipelines. In *Lecture Notes in Computer Science*; Springer International Publishing: Cham, Switzerland, 2015; Volume 9383, doi:10.1007/978-3-319-25741-9.
34. Xhymshiti, D. 10 Pure Python Functions for Ad Hoc Text Analysis, 2020. Available online: <https://towardsdatascience.com/10-pure-python-functions-for-ad-hoc-text-analysis-e23dd4b1508a> (accessed on 9 May 2021).
35. Iarrobino, M. The Evolution of Text Mining—Trends We’re Seeing Across R&D Organizations, 2021. Available online: <http://www.copyright.com/blog/trends-evolution-text-mining/> (accessed on 23 May 2021).
36. Regular Expression Operations in Python. Python 3.9.5 Documentation. Available online: <https://docs.python.org/3/library/re.html> (accessed on 1 May 2021).
37. Wang, S.; Lu, L.; Han, X.; Ouyang, M.; Feng, X. Virtual-battery based droop control and energy storage system size optimization of a DC microgrid for electric vehicle fast charging station. *Appl. Energy* **2020**, *259*, 114–146, doi:10.1016/j.apenergy.2019.114146.
38. California Department of Food and Agriculture Division of Measurement Standards. Zero Emission Vehicle Projects. Available online: <https://www.cdffa.ca.gov/dms/programs/zevfuels/> (accessed on 24 April 2021).
39. U.S. Energy Information Administration. Electric Power Monthly: Average Price of Electricity to Ultimate Customers by End-Use Sector. 2021. Available online: https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_6_a (accessed on 30 May 2021).
40. Limmer, S. Dynamic Pricing for Electric Vehicle Charging—A Literature Review. *Energies* **2019**, *12*, 3574, doi:10.3390/en12183574.
41. Borlaug, B.; Salisbury, S.; Gerdes, M.; Muratori, M. Levelized Cost of Charging Electric Vehicles in the United States. *Joule* **2020**, *4*, 1470–1485, doi:10.1016/j.joule.2020.05.013.
42. Muratori, M.; Kontou, E.; Eichman, J. Electricity rates for electric vehicle direct current fast charging in the United States. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109235, doi:10.1016/j.rser.2019.06.042.
43. Bhatti, A.R.; Salam, Z.; Aziz, M.J.B.A.; Yee, K.P.; Ashique, R.H. Electric vehicles charging using photovoltaic: Status and technological review. *Renew. Sustain. Energy Rev.* **2016**, *54*, 34–47, doi:10.1016/j.rser.2015.09.091.
44. Ashique, R.H.; Salam, Z.; Bin Abdul Aziz, M.J.; Bhatti, A.R. Integrated photovoltaic-grid dc fast charging system for electric vehicle: A review of the architecture and control. *Renew. Sustain. Energy Rev.* **2017**, *69*, 1243–1257, doi:10.1016/j.rser.2016.11.245.
45. Rattenbury, T.; Hellerstein, J.M.; Heer, J.M.; Kandel, S.; Carreras, C. *Principles of Data Wrangling: Practical Techniques for Data Preparation*, 1st ed.; O’Reilly: Sebastopol, CA, USA 2017.
46. Chai, C.; Fan, J.; Li, G.; Wang, J.; Zheng, Y. Crowdsourcing Database Systems: Overview and Challenges. In Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE), Macau, China, 8–12 April 2019; pp. 2052–2055, doi:10.1109/ICDE.2019.00237.
47. Niu, H.; Silva, E.A. Crowdsourced Data Mining for Urban Activity: Review of Data Sources, Applications, and Methods. *J. Urban Plan. Dev.* **2020**, *146*, 04020007, doi:10.1061/(ASCE)UP.1943-5444.0000566.