A Data Driven Analysis of EV Charging Infrastructure

Prashant Baral

School of Computer Science and Engineering
Lovely Professional University Punjab, India
baralprashant09@gmail.com

Abstract—Electric vehicle (EV) adoption has accelerated globally, but the distribution, accessibility, and capacity of charging infrastructure remain uneven and poorly quantified. Existing public datasets contain large volumes of semi-structured information, but lack consolidated analysis that connects charger types, facility categories, access restrictions, and network dominance into a single evaluative framework. This study conducts a structured analysis of EV charging infrastructure using a cleaned and aggregated dataset of approximately 92,000 U.S. charging stations. Power BI was used for preprocessing, categorical normalization, geospatial mapping, and DAX-driven statistical summarization. The analysis focuses on charger levels (Level 1, Level 2, DC Fast), facility classifications, public vs. restricted access, connector types, and network providers, supported by temporal validation through the Date Last Confirmed field. Results reveal strong geographic concentration of stations, dominance of Level-2 chargers, limited deployment of DC Fast charging, and heavy reliance on retail and workplace locations. Public access accounts for most stations, but high-power availability remains low across several regions. The study demonstrates that structured aggregation and visual analytics enable clear identification of infrastructure strengths, accessibility gaps, and deployment patterns relevant for further EV planning.

Keywords—Electric Vehicles, EV Infrastructure, Charging Stations, Data Analysis, Power BI, Data Visualization

I. INTRODUCTION

The development of electric motive power dates to 1827, when Ányos Jedlik constructed one of the earliest functional electric motors, initiating the long trajectory of electric vehicles. Over nearly two centuries, this technology has progressed from experimental prototypes to a strategic component of modern transportation systems. By the 2020s, electric vehicles (EVs) had shifted from niche adoption to mainstream deployment, driven by environmental regulations, sustainability goals, declining battery costs, and national decarbonization commitments. [1]

As EV adoption accelerated, charging infrastructure emerged as the critical operational backbone enabling large-scale electrification. [2] The global transportation infrastructure that used to heavily rely on internal combustion engines and fossil fuels is slowing moving towards electric vehicles which are more sustainable and environment friendly alternative.[1] EV charging infrastructure is typically organized into power-based tiers such as Level 1, Level 2, and various forms of DC fast charging, which differ widely in charging speed, power delivery, and deployment context. This distinct infrastructure involving multiple provider networks, facility categories, connector standards, and access policies—creating persistent challenges in accessibility, interoperability, and regional coverage.[3]

In the United States, the number of charging stations has expanded substantially, but geographic density and charger-type availability remain highly uneven.[4] Level 2 chargers dominate installations in workplaces and retail locations, while DC Fast chargers are concentrated in a limited set of high-traffic corridors and specific network operators.[4][5] The coexistence of multiple technologies, connector standards, provider networks, and access policies introduces substantial heterogeneity into the infrastructure, making some regions better equipped than others. Analyzing these variations across charger types and facility contexts is therefore essential for understanding spatial disparities, identifying infrastructure bottlenecks, and evaluating whether current deployment trends align with the needs of EV users. This motivates the present study's focus on quantifying charger-type prevalence, mapping geographic concentration, and assessing accessibility characteristics within the available dataset.

Despite the availability of national EV-charging datasets, much of the existing information remains under-analyzed in a basic descriptive sense. Most studies focus on policy modeling, forecasting, or mobility simulation, while simple questions about how charger levels, facility types, access categories, and provider networks are distributed across the current infrastructure are often overlooked. A straightforward exploratory analysis can still provide meaningful insights by highlighting what types of chargers dominate, which locations receive the most deployments, and how public accessibility varies. Using a single, cleaned dataset of U.S. charging stations, this study applies direct aggregation and visualization methods to present a clear snapshot of the existing infrastructure, without attempting predictive modeling or complex statistical inference. The objective is to organize the raw attributes into interpretable patterns that can support general understanding of deployment trends across different charger configurations and facility contexts.

A. Data Analysis and Visualization

In the current technology-driven environment, data has become a critical asset that fuels innovation, strategic planning, and evidence-based decision-making across industrial and academic sectors.

Data analysis is therefore indispensable in today's data-intensive context, as it transforms unstructured information into actionable knowledge that can improve operational efficiency, support strategic decisions, and guide policy and infrastructure development. Data visualization refers to the graphical representation of data through charts, maps, and other visual formats to expose trends, patterns, and relationships that may not be obvious from raw tables alone. Data analysis processes raw datasets to extract useful knowledge, and visualization helps communicate these findings by presenting them visually in a simple and understandable way.[6]

In contemporary empirical research, data-driven evaluation has become essential for understanding large-scale infrastructure systems. Electric vehicle (EV) charging networks generate large volumes of semi-structured information across charger types, access policies, facility categories, and geographic locations. Quantitative analysis and visualization enable researchers to convert these raw

records into interpretable patterns that reveal distributional gaps, clustering effects, and deployment trends across regions [5].

Fig. 1. Data visualization pipeline [8]

1. Data import stage involves obtaining data from different sources such as databases, sensors, web APIs, or some publicly available repositories. The goal is to gather all the raw information needed for the data analysis and visualization.

2. Data Preparation is the stage where the imported data is cleaned and organized which may include tasks such as correcting errors, removing duplication, converting data types, and normalizing values to ensure the taken dataset is accurate and ready for processing.

3. Data modification is the process of choosing the data that will be visualized and may also involve filtering records, grouping values, combining multiple datasets or selecting important attributes that are crucial for drawing insights.

4. Mapping involves converting the data gathered through the a forementioned procedure to visual elements. Numerical or categorical attributes are mapped to graphical components such as position, shape, color, size, or labels, defining how the data will appear visually.

5. Rendering is the procedure of turning the symmetrical data stated above into a pictorial representation.

The given steps are description form of Fig.1 which represents the data visualization pipeline.[7][8]

B. *Visualization Methodologies for Analysis*

Data analysis of large-scale infrastructure systems requires tools capable of handling heterogeneous, semi-structured, and geographically distributed data. Data virtualization tools are software applications that provide unified, real-time access to data origination from multiple sources without manually consolidating information from multiple sources. These systems remove the need to copy the data manually by providing a virtual layer that lets you query and analyze different data formats in one place. This architecture is particularly relevant for EV-infrastructure research, where datasets contain diverse attributes such as charger-level categories, facility types, network operators, and geographic coordinates. Visualization pipelines using virtualized data make it possible to perform filtering, aggregation, mapping, and statistical summarization with efficiency and reproducibility.[8]

Power BI: Power BI is a widely used business-intelligence and data-visualization platform focused on delivering interactive analytics with strong enterprise integration. It supports automated ingestion of structured and semi-structured files, cleaning and shaping through Power Query, and numerical computation using DAX. Power BI offers a broad range of visualization options, enabling the creation of custom dashboards, reports, and analytical models. It is optimized for speed, scalability, and adaptability across different data environments. From SQL Server and Azure Synapse to AWS S3, Google Big Query, and REST-based data connectors, it supports major data formats and a wide range of cloud and on-premises sources. [9]

Tableau: Tableau is a widely used visualization platform known for its interactive dashboards, strong storytelling features, and broad connectivity to enterprise databases. It offers extensive visual customization, a large library of chart types, and a drag-and-drop interface suitable for exploratory analysis. It proposes volume to produce unique visualizations. It is speedy and adaptable. Unlike traditional BI tools that require extensive technical knowledge, Tableau prioritizes user-friendliness, allowing both technical and non-technical users to create complex visualizations and analyses with ease. Available in both desktop and mobile versions, Tableau ensures that data is accessible anytime, anywhere, fostering a data-driven culture within organizations. [10]

Although both Power BI and Tableau provide advanced visualization features, Power BI is used is used for this study as it is more practical for the study following practical considerations. Power BI is more suitable for EV-infrastructure analysis because it integrates ETL, computation, and mapping in one environment. Power Query handles data cleaning, normalization, and transformation internally, which removes the need for external preprocessing tools. DAX allows direct calculation of charger aggregates, network groups, access-type measures, and time-based metrics, while Tableau requires external scripting for equivalent logic. Power BI's mapping visuals support region filtering, clustering, and density analysis necessary for nationwide EV-station evaluation. It handles medium-scale datasets (~100k rows) with lower system overhead and does not require server deployment. Power BI Desktop is also free, eliminating licensing barriers. These factors make Power BI the more efficient choice for preprocessing and geospatial analysis in this study.[11]

C. *Dashboard*

A dashboard is an analytical interface that consolidates multiple data representations—charts, maps, KPIs, and interactive filters—into a single environment. It enables interpretation of system-level trends without manual computation by translating processed datasets into actionable visual indicators. Core operations such as filtering, cross-highlighting, category-wise breakdown, and comparative analysis allow users to examine multi-dimensional attributes efficiently. Dashboards are widely used in data-driven research because they support repeatable analysis workflows, reduce cognitive load, and maintain metric consistency while enabling anomaly detection and pattern validation.

Commercial tools such as Tableau and Microsoft Power BI provide structured environments for constructing these interfaces, combining data modeling, visualization components, and interaction controls. Power BI dashboards integrate data ingestion, transformation, and visualization within one pipeline, supporting reproducible modeling and the generation of dynamic views such as geospatial maps, categorical summaries, and KPI indicators. Features like slicers, drill-downs, and cross-highlighting enable users to explore patterns across hierarchical and regional dimensions.

In this study, the dashboard serves as the primary visualization layer for the EV-infrastructure dataset. It combines cleaned and modeled attributes—including charger density, state-wise station counts, network-operator distribution, and access-type segmentation—to provide a direct, interpretable assessment of spatial and categorical disparities across North America.[8]

D. *Research Questions:*

This study is structured to address the following Research

Questions (RQs):

RQ 1: How are EV charging stations distributed across states?

RQ 2: Which cities have the highest concentration of charging stations?

RQ 3: What facility types host the most EV stations?

RQ 4: How do different charge types distributed across different Facility type?

RQ 5: What is the distribution of restricted vs public access stations?

RQ 6: How has the EV charging infrastructure grown over time?

RQ 7: How are stations distributed over connector types?

II. METHODOLOGY

A. Data Source:

The study uses the Alternate Fuels Stations (Electric) dataset from the Alternate fuels Data Center Platform of U.S. Department of Energy. The dataset contains information on approximately 92,000 EV charging stations, including attributes such as: Charger Level (Level 1, Level 2, DC Fast) , Facility Type (Retail, Workplace, Public, Fleet, Government, etc.), Connector Types, Network Providers, Access Restrictions (Public, Restricted/Private), Geolocation (Latitude/Longitude) and Date Last Confirmed. This dataset provides a comprehensive view of EV infrastructure across the United States and Canada, enabling geographic and categorical analyses.[9][12]

B. Data Cleaning:

Prior to analysis, The original dataset contained approximately 92,000 records of EV charging stations obtained from the AFDC repository. Preliminary preprocessing was conducted in Power BI's Power Query Editor to ensure structural consistency and analytical suitability. Only attributes relevant to geospatial and infrastructural assessment were retained—station identifiers, location details (city, state, ZIP, latitude, longitude), charger counts (Level 1, Level 2, DC Fast), network information, and connector types. All non-essential metadata fields were excluded to reduce dimensionality and improve model efficiency.

Each column was assigned an appropriate data type to prevent computational errors during aggregation: location and network attributes were treated as text, charger counts as whole numbers, and geographic coordinates as decimal values. Missing numerical values were replaced with zero to prevent null propagation in quantitative measures, while categorical nulls (e.g., EV Network and Connector Types) were standardized to "Unknown." Duplicate entries, which frequently occur in the AFDC dataset, were removed to eliminate redundancy.

Categorical normalization was performed for fields that contained heterogeneous representations. Access-type labels were consolidated into three classes (Public, Private, Unknown). Similarly, infrastructure location types—which appeared in more than forty raw formats—were mapped into seven consistent facility categories: Parking, Retail, Workplace, Public/Government, Hospitality, Recreation/Education, and Other. These transformations established a consistent categorical structure necessary for comparative analysis.

C. Data Transformation:

To improve computational efficiency, an aggregated analytical table was constructed by grouping records at the State–City level. For each group, total counts of Level 1, Level 2, and DC Fast chargers, along with the total number of stations, were computed. This reduced the dataset to approximately 1,500 consolidated records while preserving all information required for state-wise and city-wise analysis.

D. Visualization Preparation:

Following data cleaning and transformation, the dataset was modeled in Microsoft Power BI with validated relationships to ensure consistent filtering. Geographic, categorical, and numerical fields were mapped to their corresponding visualization attributes. Geospatial layers used standardized latitude–longitude coordinates with state and city identifiers to represent charger density and regional variation. Level 1, Level 2, and DC Fast counts were implemented as measures to maintain consistent aggregation. Categorical fields— including facility type, access classification, vehicle class, and network operator—were encoded for controlled filtering. Core KPIs (total chargers, station counts, network share, connector-type ratios) were computed using DAX. The visualization model consisted of four components: geospatial distribution, categorical composition, charger-type summaries, and network/access segmentation. This structure supports a reproducible, multi-dimensional evaluation of North American EV-charging infrastructure

III. LITERATURE REVIEW

The literature review/survey is tabulated in the form of a table I where it compares the work of previous authors on the topic and presents their work, which dataset they used, conclusions they have drawn, and mentioned limitations of their paper.

TABLE I: Literature review on Techniques used for EV
Infrastructure Data Analysis

Year	Technique and Dataset	Proposed Work	Limitations
2025 [5]	Used national-scale U.S. EV charging station datasets and applied spatial coverage gap analysis.	The study systematically mapped the distribution of EV charging stations across the United States and identified regions where infrastructure is insufficient or completely missing. It highlights the geographic mismatch between EV adoption trends and charging availability.	The focus remained largely on spatial coverage only; the research did not examine charger-level details such as Level 1/2/DC Fast availability, network providers, access restrictions, or facility types, which limits its ability to explain *why* gaps exist.
2024 [3]	Reliability and resilience data from the U.S. National Renewable Energy Laboratory.	This work analyzed how charging station reliability, downtime, and resilience influence user adoption of electric vehicles. It demonstrates that poor uptime and poorly located stations negatively affect charging confidence.	The study centers heavily on reliability but does not combine these metrics with geospatial distribution, charging power levels, connector standards, or public–private access patterns, leading to a partial view of infrastructure gaps.
2024 [13]	EV battery datasets combined with an Extended Kalman Filter + Deep Learning model.	The paper developed a predictive platform for estimating electric vehicle battery health using a hybrid EKF and deep learning architecture, improving long-term battery performance forecasting.	The study is focused on vehicle-side analytics rather than charging infrastructure. It does not explore station density, charger types, or accessibility factors relevant to infrastructure planning.
2024 [14]	Technical performance data of fast-charging systems across different deployment scenarios.	The research examined strategies for deploying sustainable and scalable fast-charging stations, emphasizing power requirements, grid constraints, and environmental aspects.	The analysis stays technical and does not include user-side accessibility, geographic coverage, or real-world public charging dataset evaluation.
2024 [1]	Review of EV integration studies focusing on city-level legislation, infrastructure, and grid impact	This work explores how cities integrate EVs into transportation and energy systems, covering policy frameworks, infrastructure placement, and overall technology evolution.	The study focuses on high-level urban integration and does not provide quantitative assessment using large national charging station datasets.
2024 [2]	Comprehensive literature review of global EV fast-charging technology advancements.	The authors surveyed recent progress in fast-charging standards, charger hardware, thermal technologies, and charging protocols, providing a consolidated view of the technological landscape.	The study is purely descriptive and lacks real-world data analysis. It does not evaluate how fast chargers are distributed geographically or how users access them.
2024 [4]	National EV charger dataset combined with demographic and socio-economic factors.	The paper investigated how socio-economic inequality influences access to EV charging stations, identifying disparities linked to income, housing, and environmental factors.	The work does not include charger reliability, power levels, or connector type availability, resulting in an incomplete assessment of infrastructure quality.

IV. Results and Discussions

This section shows solutions for the Research Questions raised in the data analysis.

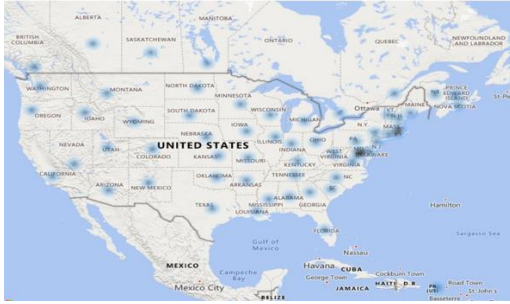RQ 1: How are EV charging stations distributed across states?

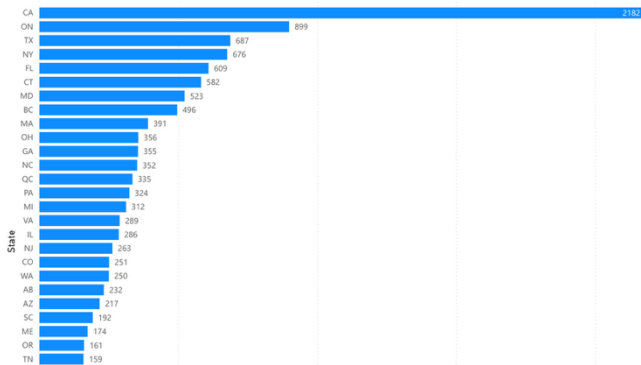Fig. 2. Map represents EV stations across different states.

Fig. 3. Chart represents states with max stations

Conclusion: Fig.2. and Fig.3 show that EV station distribution is highly uneven across states. California holds the largest concentration, followed by Orlando, Texas and New York. Many central and mountain-region states show fewer installations. The visualization confirms a coastal-dominated deployment pattern, indicating infrastructure growth is tied to population density and early EV adoption markets.

RQ 2: Which cities have the highest concentration of charging stations?
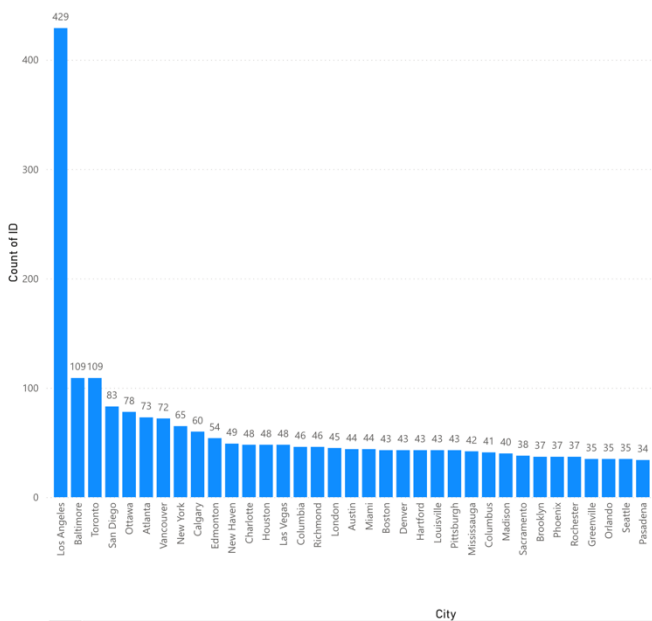

Fig. 4. City wise total station count

Conclusion: The Fig .4. identifies a small group of cities contributing disproportionately to national EV infrastructure. Cities such as Los Angeles, Baritone, Toronto, San Diego, and Ottawa appear at the top, each hosting a substantially higher number of stations than the median city. Most cities show low station counts, revealing a strong urban clustering effect.

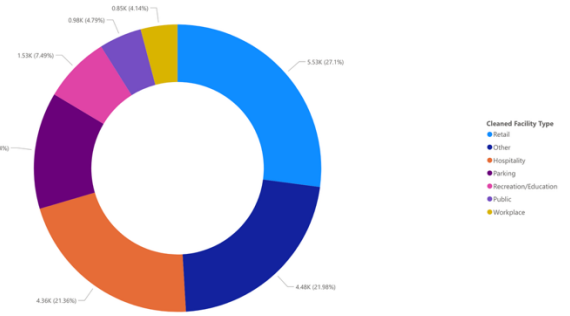RQ 3: What facility types host the most EV stations?

Fig. 3 Pie chart represents facility type with max stations

Conclusion: The pie chart shows that retail locations and hospitalities host most charging stations. Residential and government locations form a smaller share. This indicates that national infrastructure is primarily positioned for public and commercial use, not home-based deployment.

RQ 4: How are the station types are distributed across different Facility type?
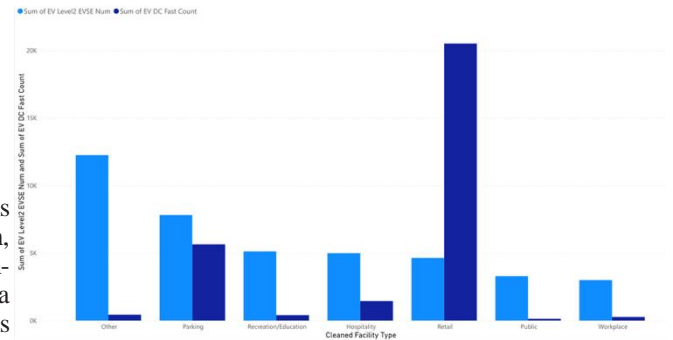

Fig. 5. Facility-type wise Level 2 and EV DC fast count

Conclusion: The chart shows that Level 2 chargers dominate across almost all facility categories while the retail is completely dominated by DC Fast chargers.

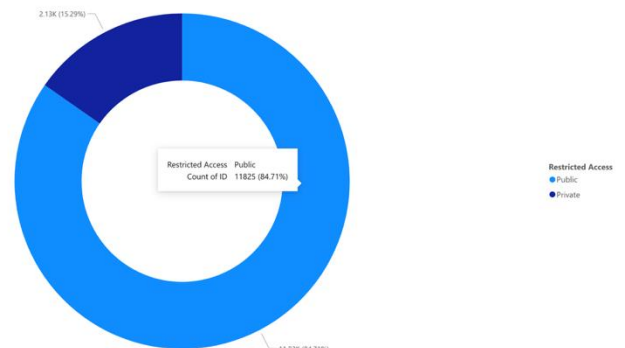RQ 5: What is the distribution of restricted vs public access stations?

Fig. 6. Private vs public access stations

Conclusion: Public access accounts for most charging stations. Restricted or private-access stations (workplaces, fleet locations, gated properties) form a smaller fraction. This indicates a reasonable level of national accessibility but also highlights that a percentage of infrastructure remains unavailable to general users.

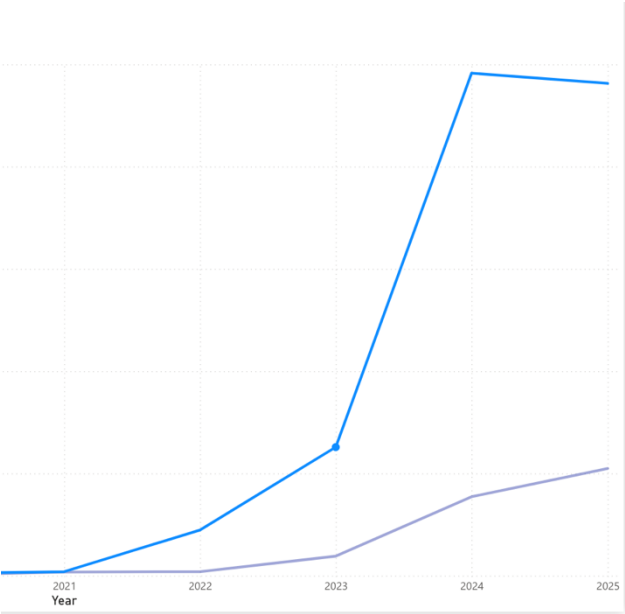RQ 6: How has the EV charging infrastructure grown over time?

Fig.7. EV growth over year, blue represents public access while purple represents private access.

Conclusion: The line chart shows a consistent and steep upward trend in confirmed station installations, with the sharpest growth occurring in recent years.
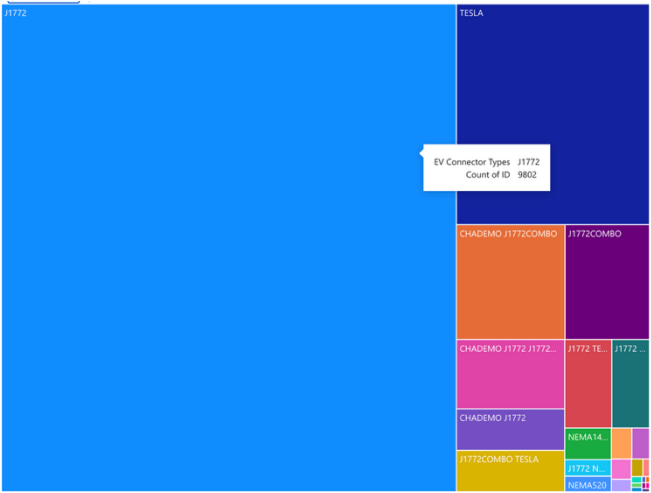
RQ 7: How are stations distributed over connector types?

Fig .8. Distribution of stations across connector types

Conclusion: The tree map shows J1772 and Tesla as the dominant connector standards. CHAMEDO also has large presence through its different types, while the CHAMEDO J1772COMBO and J1772COMBO remain major player. This distribution confirms that the market is consolidating around a small set of standardized connectors.

V. CONCLUSION

The analysis of the U.S. EV charging infrastructure using the DOE Alternative Fuel Stations dataset shows that national deployment is expanding rapidly but remains structurally uneven. Charging stations are heavily concentrated in a small set of states and urban centers, while large portions of the Midwest and rural regions continue to show minimal coverage. Level 2 chargers dominate the national network, confirming that the majority of installations support slower, destination-based charging rather than long-distance travel. DC Fast chargers remain limited and are clustered around

retial and commercial corridors, exposing clear gaps in high-power accessibility.

Infrastructure placement is driven primarily by retail, and parking facilities, indicating that the charging network is built around public and commercial use rather than residential support. Public access represents most stations, but a non-trivial share of chargers remain restricted to employees, fleets, or private organizations. Temporal analysis confirms steady year-on-year growth in station installations, aligned with policy incentives and rising vehicle adoption. Connector-type distribution also shows consolidation around J1772 and TESLA, with legacy standards losing relevance.

Overall, the results highlight three systemic issues: geographic imbalance, limited fast-charging density, and dependency on commercial facility placement. These findings underscore the need for targeted regional build-outs, strategic fast-charger deployment, and more standardized infrastructure planning if the U.S. and Canada aims to support large-scale EV adoption in the coming decade.

REFERENCES

[1] Alonso-Cepeda, Antonio, et al. "A review on electric vehicles for holistic robust integration in cities: history, legislation, meta-analysis of technology and grid impact." *Applied Sciences* 14.16 (2024): 7147.

[2] Zentani, Ahmed, Ali Almaktoof, and Mohamed T. Kahn. "A comprehensive review of developments in electric vehicles fast charging technology." *Applied Sciences* 14.11 (2024): 4728.

[3] Powell, Bonnie, and Caley Johnson. *Impact of electric vehicle charging station reliability, resilience, and location on electric vehicle adoption*. No. NREL/TP-5R00-89896. National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2024.

[4] Ermagun, Alireza, and Joshua Tian. "Charging into inequality: A national study of social, economic, and environment correlates of electric vehicle charging stations." *Energy Research & Social Science* 115 (2024): 103622.

[5] Hanig, Lily, et al. "Finding gaps in the national electric vehicle charging station coverage of the United States." *Nature Communications* 16.1 (2025): 561.

[6] Olowe, Kehinde Josephine, et al. "Conceptual review on the importance of data visualization tools for effective research communication." *International Journal of Engineering Research and Development* 20.11 (2024): 1259-1268.

[7] Qin, Xuedi, et al. "Making data visualization more efficient and effective: a survey." *The VLDB Journal* 29.1 (2020): 93-117.

[8] Chowdary, Pullela Harish, et al. "From Athens to Rio: A Comprehensive Data Analysis and Visualization of 120 Years of Olympic History." *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. IEEE, 2024.

[9] Mohammed, A., Saif, O., Abo-Adma, M. *et al.* Strategies and sustainability in fast charging station deployment for electric vehicles. *Sci Rep* **14**, 283 (2024). https://doi.org/10.1038/s41598-023-50825-7

[10] Parhusip, H. A., Trihandaru, S., Heriadi, A. H., Santosa, P. P., & Puspasari, M. D. (2022). Data Exploration Using Tableau and Principal Component Analysis. *JOIV: International Journal on Informatics Visualization*, 6(4), 911-920.

[11] Sahaya, Alfiansyah Putra Nur, et al. "Powering Sales Insights: A Comparative Analysis of Data Visualization Tools, Microsoft Power BI vs Tableau." *2024 9th International Conference on Business and Industrial Research (ICBIR)*. IEEE, 2024.

[12] U.S. Department of Energy, "Alternative Fuel Stations Dataset," Alternative Fuels Data Center (AFDC), 2025. [Online]. Available: https://afdc.energy.gov/data_download

[13] Li DC, Felix JR, Chin Y-L, Jusuf LV, Susanto LJ. Integrated Extended Kalman Filter and Deep Learning Platform for Electric Vehicle Battery Health Prediction. *Applied Sciences*. 2024; 14(11):4354. https://doi.org/10.3390/app14114354

[14] Bhargava, Mandava Geetha, K. T. P. S. Kiran, and Duvvada Rajeswara Rao. "Analysis and design of visualization of educational institution database using power bi tool." *Global Journal of Computer Science and Technology*