© April 2026| IJIRT | Volume 12 Issue 11 | ISSN: 2349-6002

Exploratory Data Analysis on Electric Vehicle Population

Shaik Mufassirunnisa[1], Velugoti Siva Nageswari[2], Gontumukkala Sravani[3], Kuruva Adoni Dharani[4]
[1,2,3,4]*Department of Electrical & Electronics Engineering, G. Pulla Reddy Engineering College (Autonomous), Kurnool, Andhra Pradesh, India*

Abstract- ***The global transition toward clean energy has accelerated Electric Vehicle (EV) adoption, making it critical to analyze population trends for sustainable transportation planning. This paper applies Exploratory Data Analysis (EDA) to a public EV population dataset (261,698 records, 17 features) to study growth patterns, type distributions, manufacturer dominance, regional adoption, and range improvements. Data preprocessing—handling missing values, removing duplicates, correcting data types—was performed using Pandas and NumPy, followed by statistical analysis and multi-method visualization with Matplotlib, Seaborn, and Plotly. Key findings reveal a sharp registration increase post-2021, with Battery Electric Vehicles (BEVs) constituting approximately 79.6% of registrations (208,274 vehicles), Tesla as the dominant manufacturer with 108,771 registered vehicles, and steady electric range improvement in newer model years reaching up to 337 miles. Seattle leads regional adoption with 41,125 registered EVs, while King County accounts for 130,129 registrations. Insights support infrastructure planning, policy formulation, and EV market forecasting.***

Index Terms—Electric Vehicles, Exploratory Data Analysis, BEV, PHEV, Python, Data Visualization, Sustainable Transportation, EV Adoption.

I. INTRODUCTION

Electric vehicles (EVs) are rapidly becoming a cornerstone of modern transportation driven by environmental imperatives and government clean-energy mandates. Unlike internal combustion engine (ICE) vehicles, EVs operate on rechargeable batteries, significantly reducing greenhouse-gas (GHG) emissions and urban air pollution.

The EV ecosystem comprises two primary categories: Battery Electric Vehicles (BEVs), which rely solely on electric power, and Plug-in Hybrid Electric Vehicles (PHEVs), which combine electric motors with conventional fuel. Understanding their distribution and performance trends is critical for researchers, policymakers, and industry stakeholders.

This paper presents a comprehensive EDA on a registered EV population dataset of 261,698 records, uncovering growth trajectories, manufacturer market shares, vehicle-type proportions, and range improvements using Python-based analytical tools.

II. RELATED WORK

The body of literature surrounding EV population analysis spans battery technology, grid integration, charging optimisation, and data-driven methodologies. Peterson et al. [1] examined lithium-ion battery cell degradation under realistic vehicle and vehicle-to-grid (V2G) utilisation profiles, establishing that real-world duty cycles accelerate capacity fade significantly compared to laboratory benchmarks—a finding directly relevant to understanding range limitations observable in older model-year EVs within population datasets.

Khodayar et al. [2] proposed an hourly coordination framework for EV operation alongside volatile wind power generation within a security-constrained unit commitment (SCUC) model. Their work demonstrated that large EV fleets can provide demand-side flexibility, but only if spatial and temporal distribution of vehicles is well understood—motivating the need for accurate EV population data of the kind analysed in this paper.

Hu et al. [3] addressed congestion prevention in distribution grids through coordinated EV charging, showing that uncoordinated charging in high-density urban areas—precisely those identified in this study (Seattle, King County)—creates significant grid stress. Rotering and Ilic [4] further developed optimal charge control strategies for PHEVs in deregulated electricity markets, highlighting the economic incentive structures that influence consumer EV adoption decisions.

McKinney [5] introduced the Pandas library, which provides the core DataFrame structures used throughout EV population preprocessing and aggregation. Waskom [6] developed Seaborn for statistical data visualisation, enabling the heatmaps and distribution plots employed in this EDA, while Hunter [7] created Matplotlib as the foundational 2D graphics environment underpinning all chart outputs. The EV population dataset is sourced from the Kaggle open-data repository [8], and macroeconomic and policy adoption context is drawn from the IEA Global EV Outlook 2023 [9], which documents the global surge in BEV registrations consistent with the post-2021 growth trend observed in this dataset.

III. DATASET DESCRIPTION

A. Overview

The dataset contains 261,698 registered EV records with 17 feature columns sourced from the Kaggle public open-data repository [8]. It covers both BEVs and PHEVs with attributes spanning identification, location, vehicle specifications, and policy eligibility. The dataset spans model years 2000 to 2026 and covers 854 unique cities across 52 states, with Washington State comprising the dominant share of records. Global EV adoption context is drawn from the IEA Global EV Outlook 2023 [9].

B. Feature Categories

- Identification: VIN (1–10) with 15,753 unique values; DOL Vehicle ID (261,698 unique).
- Location: County (236 unique), City (854 unique), State (52 unique), Postal Code (1,053 unique), Electric Utility (76 unique).
- Vehicle Details: Make (46 manufacturers), Model (181 models), Model Year (21 unique years, 2000–2026), EV Type (BEV/PHEV).
- Technical: Electric Range (0–337 miles), Base MSRP, Legislative District.
- Policy/Mapping: CAFV Eligibility (3 categories), Vehicle Location (GPS), 2020 Census Tract.

IV. METHODOLOGY

A. Data Preprocessing

The raw dataset was loaded via pd.read_csv() and inspected with df.info(), revealing 261,698 entries across 17 columns. The dataset exhibited minimal missing values: Legislative District had the highest missing rate (~0.24%), while most other columns had rates below 0.007%. Preprocessing steps included: (i) dropping rows with null or zero values in County, City, and Postal Code; (ii) converting Postal Code to integer type; (iii) retaining rows with missing Electric Range (3 nulls) and Base MSRP (3 nulls) to preserve data volume; and (iv) extracting latitude/longitude from Vehicle Location for geospatial analysis. The cleaned dataset retained 261,688 entries with ≤3 null values per analytical column.

B. EDA Techniques

Statistical summaries (value_counts(), describe(), groupby()) were computed across four analytical dimensions: temporal growth (model year), vehicle-type proportions (BEV vs. PHEV), manufacturer market share, and electric range distribution. Unique value analysis was performed across all 17 columns to understand cardinality. Correlation analysis via heatmap was applied to numerical features.

C. Visualization Methods

Six visualization techniques were employed: (1) Pie charts—proportional category distributions; (2) Bar charts—cross-category frequency comparison; (3) Count plots—direct frequency of categorical records; (4) Histograms—numerical feature distributions using Matplotlib [7] and Plotly; (5) Scatter plots—bivariate correlations for Electric Range vs. Model Year; (6) Heatmaps—correlation matrix visualization using Seaborn [6]. Data manipulation throughout used the Pandas library [5]. A Folium interactive map was additionally used for geospatial vehicle distribution.

V. RESULTS AND DISCUSSION

A. *EV Distribution by Model Year*

Fig. 1 shows EV registrations per model year from 2000–2026. Registrations remained negligible before 2011 (fewer than 50 vehicles) and modest between 2011–2018 (631 to 14,522 vehicles/year). Growth accelerated sharply after 2021, reaching 20,934 in 2021, 29,645 in 2022, and peaking at 60,214 for the 2023 model year—the highest single-year count in the dataset. The 2024 model year recorded 49,869 vehicles, 2025 shows 29,495 registrations, and 2026 (still in early registration) accounts for 5,448 vehicles. This trajectory aligns with global EV market trends reported by the IEA [9], driven by declining battery costs, expanded model availability, and strengthened government incentive programmes.

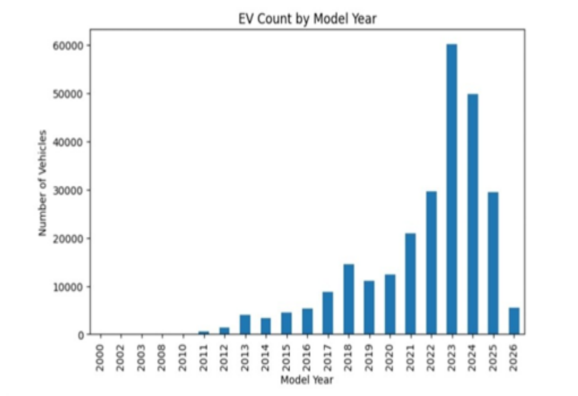

Fig. 1. EV registrations by model year. Sharp growth is evident post-2021

B. *Vehicle Type Distribution*

As shown in Fig. 2, the dataset contains 208,274 Battery Electric Vehicles (BEVs) and 53,414 Plug-in Hybrid Electric Vehicles (PHEVs), yielding a BEV share of approximately 79.6% and a PHEV share of 20.4%. The dominant BEV share indicates consumer preference for fully electric solutions, attributable to zero-emission operation, lower running costs, and improving range. The declining PHEV share suggests that range anxiety is diminishing as BEV technology and charging infrastructure mature.

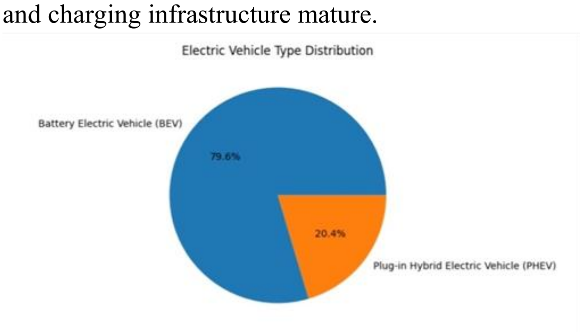

Fig. 2. BEV vs. PHEV distribution. BEVs dominate at 79.6%.

C. *Manufacturer-wise Distribution*

Fig. 3 presents the top-10 EV manufacturers by registration volume. Tesla leads by a substantial margin with 108,771 units—approximately 41.5% of total registrations—followed by Chevrolet (18,908), Nissan (16,224), Ford (13,987), KIA (12,849), Toyota (10,845), BMW (10,808), Hyundai (8,936), Rivian (7,932), and Volvo (6,796). The top EV model is the Tesla Model Y with 54,717 registrations, followed by Model 3 (37,772), Nissan Leaf (13,852), Model S (7,945), and Chevrolet Bolt EV (7,873). Tesla's dominance reflects early market entry, a proprietary Supercharger network, and continuous model innovation.

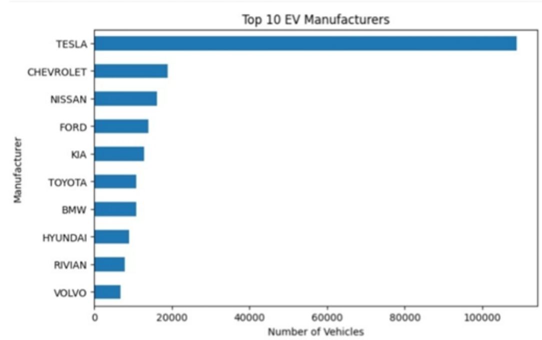

Fig. 3. Top 10 manufacturers by EV registration volume.

D. *Electric Range Analysis*

Fig. 4 illustrates the electric range distribution by vehicle type across all vehicles (count = 261,685; mean = 42.6 miles; std = 81.2 miles; min = 0; max = 337 miles). The highly skewed distribution—with the 50th percentile at 0 miles—reflects the large proportion of records with unreported range, predominantly PHEVs. When filtered by vehicle type, two distinct clusters emerge: a low-range cluster (10–60 miles) for PHEVs and a high-range cluster (200–337 miles) for BEVs, with BEV outliers reaching up to 337 miles. This confirms that advances in battery energy density, thermal management, and powertrain efficiency in recent model years have directly addressed the range anxiety barrier throughout this research.

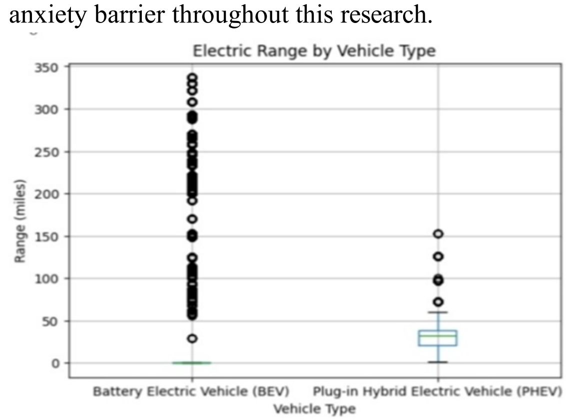

Fig. 4. Electric range by vehicle type showing BEV and PHEV clusters.

VI. KEY INSIGHTS

- Rapid Growth: EV registrations grew from negligible levels pre-2011 to 60,214 in 2023, confirming the market's transition from early-adopter to mainstream-consumer phase.

- BEV Dominance: BEVs account for 79.6% of registrations (208,274 vehicles), reflecting a decisive shift toward fully electric mobility.
- Market Concentration: Tesla leads with 108,771 vehicles (~41.5% share); the Tesla Model Y is the single most registered model (54,717 units).
- Range Progress: Electric range shows a strong positive correlation with model year, with the maximum range reaching 337 miles in the latest model years.

VII. RECOMMENDATIONS

- Expand public charging infrastructure in high-EV-density cities (Seattle, Bellevue, Tacoma) near transit hubs, malls, and workplaces.
- Develop inter-city fast-charging corridors (Seattle–Tacoma–Olympia; Seattle–Spokane) to eliminate highway range anxiety.
- Integrate solar-powered charging at government buildings and universities to align EV rollout with renewable energy goals.
- Extend CAFV incentive thresholds to include mid-range PHEVs as a transitional measure toward full electrification.
- Establish dedicated EV service centres in high-density counties (King, Snohomish, Pierce) to address maintenance barriers as the installed base grows.

VIII. CONCLUSION

This paper presented a systematic EDA on a 261,698-record EV population dataset using Python (Pandas, NumPy, Matplotlib, Seaborn, Plotly, Folium). The study confirmed exponential adoption growth post-2021 (peaking at 60,214 registrations in 2023), decisive BEV market dominance (79.6%, 208,274 vehicles), Tesla's market leadership (108,771 vehicles, ~41.5% share), the Tesla Model Y as the most registered model (54,717 units), and a strong positive correlation between electric range and model year (max range = 337 miles).

Future work will incorporate machine-learning forecasting models for EV adoption prediction, real-time data integration, interactive dashboards for policymakers, and cross-country comparative analysis to enrich understanding of global EV dynamics.

ACKNOWLEDGMENT

The authors would like to thank Dr. C. Harinatha Reddy, M.Tech., Ph.D., SMIEEE, Associate Professor, Department of Electrical & Electronics Engineering, G. Pulla Reddy Engineering College (Autonomous), Kurnool, for his invaluable guidance and support

REFERENCE

[1] S. B. Peterson, J. Apt, and J. F. Whitacre, "Lithium-ion battery cell degradation resulting from realistic vehicle and vehicle-to-grid utilization," *Journal of Power Sources*, vol. 195, no. 8, pp. 2385–2392, 2010.

[2] M. E. Khodayar, L. Wu, and M. Shahidehpour, "Hourly coordination of electric vehicle operation and volatile wind power generation in SCUC," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1271–1279, 2012.

[3] J. Hu, S. You, M. Lind, and J. Østergaard, "Coordinated charging of electric vehicles for congestion prevention in the distribution grid," *IEEE Transactions on Smart Grid*, 2014.

[4] N. Rotering and M. Ilic, "Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets," *IEEE Transactions on Power Systems*, 2011.

[5] W. McKinney, "Data Structures for Statistical Computing in Python," *Proceedings of the 9th Python in Science Conference*, 2010.

[6] M. Waskom, "Seaborn: Statistical Data Visualization," *Journal of Open-Source Software*, 2021.

[7] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, 2007.

[8] Kaggle Dataset, "Electric Vehicle Population Data," Available: https://www.kaggle.com

[9] International Energy Agency (IEA), "Global EV Outlook 2023," Available: https://www.iea.org