Comprehensive Exploratory Data Analysis and Visualization of Electric Vehicle Trends Using Python and Power BI

Mayukh Dutta
Institute of Engineering & Management, Kolkata
University of Engineering and Management
Kolkata, India
mayukh.dutta2002@gmail.com

Ashutosh Saha
Institute of Engineering & Management, Kolkata
University of Engineering and Management
Kolkata, India
asutoshsaha2626@gmail.com

Arpan Biswas
Institute of Engineering & Management, Kolkata
University of Engineering and Management
Kolkata, India
arpanb841@gmail.com

Sudipta Biswas
Institute of Engineering & Management, Kolkata
University of Engineering and Management
Kolkata, India
sudiptabiswas.bwn@gmail.com

Avik Sarkhel
Institute of Engineering & Management, Kolkata
University of Engineering and Management
Kolkata, India
avik305sarkhel@gmail.com

Kaustuv Bhattacharjee
Institute of Engineering & Management, Kolkata
University of Engineering and Management
Kolkata, India
kaustuv.bhattarjee@uem.edu.in

Anirban Das
Institute of Engineering & Management, Kolkata
University of Engineering and Management
Kolkata, India
anirban-das@live.com

I. Abstract

This research investigates trends in electric vehicle (EV) adoption through exploratory data analysis (EDA) and visualization. The dataset, obtained from the Washington State Department of Licensing, originally contained 200,049 records across 17 parameters, which were refined to 182,137 records across 11 key parameters after data preprocessing. Additional features, such as "Urban/Rural" classification and "Electric Utility Type," were engineered to enhance the analysis. The study employs Python-based tools such as Pandas, NumPy, and Matplotlib for analysis, complemented by Power BI for interactive visualization. Key findings include Tesla's market dominance, variations in EV range across counties, and the impact of infrastructure on EV adoption. The integration of Power BI enables real-time insights, facilitating data-driven recommendations for manufacturers and policymakers to enhance EV adoption and infrastructure planning.

A. *Introduction*

Electric vehicles (EVs) are transforming the transportation industry by providing sustainable alternatives to fossil-fuel-powered vehicles. Governments worldwide are investing in EV infrastructure to promote adoption and reduce carbon emissions [1]. As the demand for cleaner energy solutions rises, understanding EV adoption trends is crucial for effective policy-making and infrastructure development [2].

Big Data Analytics (BDA) plays a critical role in examining EV registration data to uncover insights into market distribution, regional adoption patterns, and the impact of infrastructure on EV ownership [3]. By leveraging Python-based analytical methods and Power BI visualization, researchers can extract valuable insights to assist policymakers, manufacturers, and energy providers in making informed decisions [4]. Power BI, with its real-time visualization capabilities, aids decision-making in infrastructure planning and market expansion, ensuring that data-driven strategies align with consumer demands [5].

This study aims to bridge gaps in EV analytics by focusing on structured data processing, predictive modeling, and real-time dashboards to facilitate interactive exploration of EV trends [6]. The integration of machine learning and advanced analytics enables more precise forecasting of market behavior and helps optimize the deployment of charging stations, enhancing the EV ecosystem [7].

B. *Literature Review and Background Study*

Data analytics plays a transformative role across industries by converting unstructured information into actionable insights. Bhatia [8] highlights its importance in sectors such as healthcare and retail, where predictive analytics improves decision-making and operational efficiency. However, challenges like handling unstructured data persist, necessitating advancements in analytical tools. In healthcare, Big Data Analytics (BDA) is revolutionizing hospital management and predictive diagnostics, as Batko & Ślezak [9] discuss, though regulatory constraints and data privacy concerns remain. Similarly, in project management, Mortaji & Shateri [10] demonstrate how data science enhances efficiency through predictive analytics and real-time visualization, transitioning from intuition-based decision-making to data-driven methodologies.

Looking ahead, Challa [11] explores how emerging technologies like AI, IoT, and cloud computing are reshaping data analytics applications, with direct implications for sustainability programs and resource optimization. In the EV sector, data analytics is crucial for infrastructure planning and vehicle performance optimization. Chen [12] emphasizes how EV-generated data, including charging patterns and battery health, contributes to sustainability improvements. Additionally, Hussain et al. [13] highlight the role of edge computing and IoT sensors in enhancing real-time traffic analysis and power consumption efficiency in smart cities. Data synthesis techniques, such as those introduced by Li et al. [14], leverage Gibbs sampling to generate real-world charging data for infrastructure planning.

In the automotive sector, Wang [15] discusses AI-driven analytics, including deep learning and cognitive automation, which are shaping the future of vehicle manufacturing. Power BI has also emerged as a critical tool in vehicle data analysis, with studies by Patil et al. [16] and Tirupati et al. [17] showcasing its effectiveness in monitoring vehicle performance and enabling predictive analytics. The role of Power BI in business intelligence extends to enhancing strategic decision-making across industries, as explored by Metre et al. [11] and Dhake et al. [12].

Furthermore, Gonçalves et al. [13] emphasize the significance of data analytics in optimizing EV fleet charging infrastructure, while Amara-Ouali et al. [14] review open data sources for EV charging to improve load forecasting. Additional studies by Demirci et al. [15] and Li et al. [16] provide insights into vehicle-to-grid (V2G) integration and large-scale charging behavior analysis. Finally, Revathi [17] underscores how data science enhances EV design, battery management, and predictive maintenance, addressing key sustainability challenges in the electric vehicle ecosystem.

II. Methodology

A. Data Collection and Preprocessing

The dataset initially contained $N_{\text{raw}} = 200,049$ records across $P_{\text{raw}} = 17$ parameters. Data cleaning involved handling missing values, correcting misclassifications, and standardizing attributes.

Missing Values Handling: Let M_i be missing values in column i, then the cleaned column C_i is:

$$C_i = X_i \setminus M_i \tag{1}$$

where X_i is the original data in column i.

Final dataset size:

$$N_{\text{clean}} = 182,137, \quad P_{\text{clean}} = 11 \tag{2}$$

Feature Engineering:

Electric Utility Type Classification: A categorical variable representing the ownership type of electric utilities is classified as:

$$E = \begin{Bmatrix} \text{Federal, Investor-Owned, Municipal,} \\ \text{Political Subdivision, Cooperative} \end{Bmatrix} \tag{3}$$

This classification enables a comparative analysis of EV adoption across different utility ownership models.

Urban/Rural Classification: A new column, **Urban/Rural**, categorizes each county based on its classification:

$$U = \begin{cases} \text{Urban}, & \text{if county is classified as urban} \\ \text{Rural}, & \text{if county is classified as rural} \end{cases} \tag{4}$$

This binary classification allows for an assessment of how EV adoption varies between urban and rural regions.

Filtering Specific Data Subsets: To analyze specific subsets of electric vehicles (EVs), a filtering method was applied based on different categorical attributes such as vehicle type, manufacturer, and other relevant criteria:

$$\text{Subset} = EV_{\text{filtered}}\left[EV_{\text{filtered}}\left[\text{ColumnName}\right] == \text{DesiredValue}\right] \tag{5}$$

where:

- EV_{filtered} represents the pre-processed dataset.
- `ColumnName` is the attribute used for filtering (e.g., “Make” for manufacturers, “Electric Vehicle Type” for BEV/PHEV).
- `DesiredValue` specifies the target category (e.g., “TESLA” for Tesla vehicles, “Battery Electric Vehicle (BEV)” for fully electric cars).

B. Analytical Approach

Overall EV Trends: The percentage share of EVs in a specific region i is calculated as:

$$S_i = \left(\frac{EV_i}{EV_T}\right) \times 100 \tag{6}$$

where:

- EV_T = Total number of EVs in the dataset
- EV_i = Number of EVs in region i

This metric helps assess the concentration of EVs across different geographical areas.

Market Share of EV Manufacturers: The market share of an EV manufacturer j is determined using:

$$\text{Market Share}_j = \left(\frac{M_j}{EV_T}\right) \times 100 \tag{7}$$

where:

- M_j = Number of EVs manufactured by company j
- EV_T = Total number of EVs

This calculation evaluates manufacturer dominance in the EV market.

C. *Statistical Analysis*

Mean EV Range: The average range of EVs in the dataset is given by:

$$\mu_R = \frac{1}{N}\sum_{i=1}^{N} R_i \tag{8}$$

where:

- R_i = Range of the i-th EV
- N = Total number of EVs

This provides insights into the average electric range of vehicles, which is crucial for understanding technological advancements and consumer preferences.

D. *Tools and Technologies*

1. Pandas: Used for data cleaning, transformation, and exploratory data analysis.
2. Matplotlib: Created visualizations such as pie charts, bar graphs, and line plots.
3. NumPy: Assisted in performing statistical calculations.
4. Python: The primary programming language for data handling, analysis, and visualization.
5. Google Colab Notebook: Enabled cloud-based analysis and real-time collaboration.
6. MS Excel: Used for initial preprocessing and data filtering.

III. Problem Statement and Proposed Solution

A. *Problem Statement*

1. Definition

The project focuses on analyzing a large dataset of Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) registered through the Washington State Department of Licensing. The dataset presents challenges such as missing data, misclassifications, and complexities in analyzing attributes like electric utility types and vehicle variants. The objective of the project is to clean, process, and analyze the data to extract meaningful insights that can assist organizations, policymakers, and EV manufacturers in improving their strategies and decision-making.

2. Relevance

This problem is highly significant as it deals with understanding the adoption and performance of electric vehicles in Washington State—a key factor in the transition to sustainable transportation. By addressing these issues, the project aims to provide a reliable and detailed overview of the EV market. The insights obtained will be beneficial to manufacturers, government agencies, and consumers, enabling data-driven decisions regarding EV production, infrastructure planning, and policy formulation.

3. Scope of the Problem

The scope of this project includes analyzing an EV dataset containing various parameters such as electric range, MSRP, make, and county. The primary focus is on data accuracy and meaningful interpretation, which includes:

a. Identifying and correcting data inconsistencies (e.g., misclassified vehicle types).

b. Performing a company-wise breakdown of top EV manufacturers.

c. Conducting a county-wise analysis of EV adoption trends.

d. Addressing missing data issues while ensuring data integrity.

e. This study is limited to Washington State's EV dataset and does not cover global EV trends or factors beyond the dataset's attributes.

B. *Proposed Solution*

1. Approach

a. Dataset Preprocessing: Handled null values, removed inconsistencies, and standardized data for analysis.

b. Comprehensive Analysis: Conducted whole-dataset and manufacturer-wise breakdowns to derive meaningful insights.

2. Design & Architecture

a. Collaborative Platform: Google Colab was used for data analysis, ensuring an interactive and scalable workflow.

b. Feature Engineering: Additional columns like "Urban/Rural" and "Electric Utility Type" were introduced to categorize data effectively.

c. Data Completion: Missing values in "Base MSRP" and "Electric Range (MILE)" were researched and corrected using reliable sources.

d. Eligibility Refinement: The CAFV eligibility column was updated after manual verification of vehicle models and their ranges.

e. Classification Corrections: Misclassified Plug-in Hybrid Electric Vehicles (PHEVs) were reassigned to Battery Electric Vehicles (BEVs) where necessary.

f. Data Filtering: Entries with unknown utility types were removed to ensure a cleaner dataset.

g. Trend Identification: Clustered data based on model year, vehicle type, and electric range to track adoption patterns and technological growth.

3. Workflow

a. Data Cleaning: Removed inconsistencies and missing values using Python and Excel.

b. Exploratory Analysis: Conducted structured analysis at both dataset-wide and manufacturer levels.

c. Visualization: Presented key findings using pie charts and bar graphs for better interpretability.

IV. Results

1. Market Trends and Manufacturer Insights

a. Tesla leads the EV market with 44% of total registrations, followed by Chevrolet and Nissan at 7% each.

b. Battery Electric Vehicles (BEVs) dominate with 79% of total registrations, while Plug-in Hybrid Electric Vehi- cles (PHEVs) account for 21%.

c. Geographical analysis shows that King County has the highest EV adoption rate (56%), followed by Snohomish (13%) and Pierce (8%).

2. Performance and Range Analysis

a. Snohomish County has the highest average EV range at 240 miles, suggesting a preference for long-range electric vehicles.

b. King County follows with an average range of 231 miles, highlighting strong infrastructure support and consumer preference for high-range EVs.

c. EVs priced below $40,000 show higher adoption rates, indicating a cost-sensitive consumer base.

3. Policy and Infrastructure Influence

a. Regions with better charging infrastructure experience a significantly higher adoption rate of EVs.

b. Utility providers play a key role, with Puget Sound Energy Inc. and the City of Tacoma serving 40% of EV users.

c. Government policies and incentives have been major drivers in EV adoption, with CAFV-eligible vehicles comprising 90% of total registrations.

Table I: EV Market Trends and Analysis

Category	Key Findings
Market Trends	Tesla: 44% Chevy, Nissan: 7% each BEVs: 79%, PHEVs: 21% King: 56%, Snohomish: 13% Pierce: 8%
Performance	Snohomish: 240 mi King: 231 mi EVs ¡ $40K: More adoption
Policy	Better charging → More EVs Puget Sound & Tacoma: 40% CAFV Eligible: 90%

V. Discussion

1. EV Market Growth and Manufacturer Strategy

a. Tesla's market dominance suggests a strong brand preference, with innovative models and extended range influencing consumer choices.

b. Chevrolet and Nissan's moderate market share highlights their role in the affordable EV segment, appealing to budget-conscious buyers.

c. The high percentage of BEVs (79%) over PHEVs indicates a growing consumer trust in fully electric vehicles rather than hybrids.

2. The Role of Infrastructure in EV Adoption

a. Counties with a higher density of charging stations have witnessed increased adoption, confirming the direct impact of infrastructure on consumer decisions.

b. Range anxiety remains a key factor, as reflected in the strong preference for longer-range vehicles in high adoption regions.

3. Future Implications for EV Adoption

a Affordability remains a critical aspect, and manufacturers need to focus on cost-effective solutions to increase adoption rates.

b. Improvements in charging infrastructure and battery technology will be key in driving future market expansion. Additional research on EV lifecycle impact can further enhance sustainability by addressing concerns related to battery production and recycling.

VI. Conclusion

1. Summary of Findings

This study analyzed EV adoption trends in Washington State, identifying key insights on manufacturer dominance, vehicle preferences, and regional adoption rates. The dataset was cleaned and preprocessed to improve accuracy, and relationships between electric range, MSRP, and policy incentives were explored.

2. Significance of EV Market Growth

The dominance of Tesla, Chevrolet, and Nissan highlights the evolving consumer preference for BEVs over PHEVs. The findings suggest that affordable pricing and extended range will play a key role in future EV adoption.

3. Market Share and Consumer Preferences

Tesla leads the EV market with 44% of total registrations, while Chevrolet and Nissan follow at 7% each. Additionally, Battery Electric Vehicles (BEVs) dominate the market, comprising 79% of total registrations, indicating a growing consumer preference for fully electric vehicles over hybrids.

4. Regional EV Trends and Electric Range

Insights Counties with stronger infrastructure and policy support saw higher EV adoption rates. King County led the adoption at 56%, followed by Snohomish (13%) and Pierce (8%). Meanwhile, Snohomish County recorded the highest average EV range at 240 miles, showcasing consumer preference for longer-range EVs.

5. The Role of Utility Providers in EV Expansion

An analysis of utility providers showed that Puget Sound Energy Inc. and the City of Tacoma supplied electricity to 40% of EV owners, highlighting the importance of electricity providers in supporting EV adoption

VII. Future Work

1. Expansion of Dataset Parameters

Future studies should incorporate additional parameters such as recharge time, number of airbags, safety ratings, and vehicle classifications (sedan, SUV, hatchback, etc.) to provide deeper insights into consumer preferences and vehicle performance.

2. Integration of Machine Learning Models

Applying predictive analytics and machine learning techniques can help identify future EV adoption trends. Models like regression analysis, decision trees, and clustering algorithms can improve trend prediction accuracy.

References

[1] M. K. Bhatia, “Data Analysis and its Importance,” International Research Journal of Advanced Engineering and Science, vol. 2, no. 1, pp. 166–168, 2017

[2] K. Batko and A. Ślezak, “The Use of Big Data Analytics in Healthcare,” Journal of Big Data, vol. 9, no. 3, Jan. 2022

[3] S. T. H. Mortaji and S. Shateri, “The Role of Data Science in Enhancing Project Management Practices: A Case Study in the Pharmaceutical Industry,” Journal of Data Analytics, vol. 3, no. 1, pp. 1–12, Jun. 2024

[4] N. Challa, “DATA ANALYTICS AND ITS IMPACT ON FUTURE,” Corrosion and Protection, vol. 51, no. 1, p. 1, Jan. 2023

[5] X. Chen, “The analysis of electric vehicle integration based on big data technology,” Applied and Computational Engineering, vol. 22, pp. 8–13, 2023.

[6] M. M. Hussain, M. M. S. Beg, M. S. Alam, and S. H. Laskar, “Big Data Analytics Platforms for Electric Vehicle Integration in Transport Oriented Smart Cities,” International Journal of Digital Crime and Forensics, vol. 11, no. 3, pp. 23–42, Jul. 2019

[7] Z. Li, Z. Bian, Z. Chen, K. Ozbay, and M. Zhong, “Synthesis of electric vehicle charging data: A real-world data-driven approach,” Communications in Transportation Research, vol. 4, pp. 100128–100128, Dec. 2024

[8] Z. Wang, ”Application of data analytics in automobile manufacturing—Evidence from existing literature,” Advances in Engineering Technology Research EEMAI, vol. 5, pp. 429–431, 2023.

[9] S. Patil, A. Tambitkar, R. Phadatare, and V. Pol, “Vehicle data analyzation using Power BI,” J. Emerg. Technol. Innov. Res. (JETIR), vol. 11, no. 7, 2024.

[10] K. Kishor, N. A. Joshi, P. Singh, A. Chhapola, N. S. Jain, and A. Gupta, “Leveraging Power BI for Enhanced Data Visualization and Business Intelligence,” Universal Research Reports, vol. 10, no. 2, pp. 676–711, Sep. 2024

[11] K. V. Metre et al., “An Introduction to Power BI for Data Analysis,” International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 1s, pp. 142–147, 2024

[12] “Impact of Power Bi on Business,” International Research Journal Of Modernization In Engineering Technology And Science, Mar. 2024

[13] F. Gonçalves, L. de Abreu Borges, and R. Batista, “Electric Vehicle Charging Data Analytics of Corporate Fleets,” World Electric Vehicle Journal, vol. 13, no. 12, p. 237, Dec. 2022

[14] Y. Amara-Ouali, Y. Goude, P. Massart, J.-M. Poggi, and H. Yan, “A Review of Electric Vehicle Load Open Data and Models,” Energies, vol. 14, no. 8, p. 2233, Apr. 2021

[15] A. Demirci, S. M. Tercan, U. Cali, and I. Nakir, “A Comprehensive Data Analysis of Electric Vehicle User Behaviors Toward Unlocking Vehicle-to-Grid Potential,” IEEE Access, vol. 11, pp. 9149–9165, 2023

[16] Z. Li, Z. Xu, Z. Chen, C. Xie, G. Chen, and M. Zhong, “An empirical analysis of electric vehicles ”charging patterns,” Transportation Research Part D: Transport and Environment, vol. 117, p. 103651, Apr. 2023

[17] R. S, “The Role Of Data Science In Advancing Electric Vehicles: A Comprehensive Analysis,” International Journal of Engineering Research & Technology, vol. 12, no. 12, Dec. 2023

[18] A. K. Sharma et al., ”Classification of Indian Classical Music with Time-Series Matching Deep Learning Approach,” IEEE Access, vol. 9, pp. 102041–102052, 2021