**DATA SCIENCE TOOLBX: PYTHON PROGRAMMING**

**PROJECT REPORT**

(Project Semester January-April 2025)

***ElectroTrend: Insights into Electric Vehicle Adoption Patterns***

Submitted by

Asmit Chaudhary

Registration No: - 12308533

Section: - KM007



Programme -B.Tech. CSE

Course Code: - INT375

Under the Guidance of

**Dr. Mrinalini Rana UID: - 22138**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

****

**DECLARATION**

I, Asmit Chaudhary, student of B.Tech. CSE under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-04-25 Signature

Registration No. 12308533 Asmit Chaudhary

**CERTIFICATE**

This is to certify that Asmit Chaudhary bearing Registration no. 12308533 has completed INT375 project titled, **“****ElectroTrend: Insights into Electric Vehicle Adoption Patterns”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 12-04-2025

**ACKNOWLEDGEMENT**

I want to take a moment to express my deep appreciation for the support I have received from everyone, either directly or indirectly, for enabling me to finish this project successfully. To start, I am grateful to **Dr. Mrinalini Rana** for her guidance, feedback, and steady support during this project.

Her guidance allowed not only academic support but also a wealth of moral support when I needed help staying on track and maintaining my motivation. I would also like to express my gratitude to **Lovely Professional University** for their example and support in offering a learning experience that fosters innovation, critical thinking, and practical application.

The resources and infrastructure they provided were significant factors that enabled me to finish the project. I need to thank my family and close friends for being my backbone throughout the project. Their understanding, optimism, and faith in me provided support, especially as I experienced self-doubt and/or pressure.

Finally, I thank the individuals who provided support through growth, learning and inspiration, and hope they realize that this project does not only indicate the summation of technical knowledge and learning, but is a personal accomplishment in and of itself, that indicates growth, perseverance, and passion.

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Heading** | **Page No.** |
| **I.** | [**INTRODUCTION**](#intro) | **6** |
| **II.** | [**SOURCE OF DATASET**](#source) | **7** |
| **III.** | [**EDA PROCESS**](#EDA) | **8** |
| **IV.** | [**ANALYSIS ON DATASET**](#Analysis) | **9** |
|  | **[Objective 1: EV Growth Over Time](#o1)** | **9** |
|  | [**Objective 2: EV Adoption by County & State**](#o2) | **10** |
|  | [**Objective 3: Correlation Between BEV, PHEV, EV Total**](#o3) | **12** |
|  | [**Objective 4: Boxplot Outlier Visualization for Percent Electric Vehicles**](#o4) | **13** |
|  | [**Objective 5: Counties with 100% EVs**](#o5) | **14** |
|  | [**Objective 6: Scatter Plot - Total Vehicles vs % EVs by Usage Type**](#o6) | **15** |
| **V.** | [**CONCLUSION**](#Conclusion) | **17** |
| **VI.** | [**FUTURE SCOPE**](#Future) | **18** |
| **VII.** | [**REFERENCES**](#refe) | **19** |
|  | IMPORTANT LINKS | **19** |

1. INTRODUTION

Electric vehicles (EVs) are becoming an integral part of the global shift toward cleaner and more sustainable transportation systems. As governments and consumers increasingly prioritize environmental consciousness, EV adoption is steadily rising across urban and rural regions. The deployment of EVs not only promises to reduce greenhouse gas emissions but also reduces dependence on fossil fuels, positioning them as a cornerstone of sustainable urban development.

In recent years, the United States has witnessed significant changes in its electric vehicle landscape, driven by policy incentives, growing environmental awareness, and technological advancements. However, the rate and pattern of adoption are not uniform across regions. Factors such as infrastructure readiness, public perception, demographic variations, and policy support play crucial roles in influencing how and where EVs are being adopted.

As climate change accelerates and air quality deteriorates in densely populated regions, the transition to electric vehicles represents a critical solution for reducing vehicular emissions and promoting cleaner air. Consumers are becoming increasingly aware of their environmental footprint, and this shift in mindset is reflected in the growing preference for EVs across both urban centers and smaller counties. This societal shift underscores the importance of understanding not just how many EVs are on the road, but where and why adoption is happening. By identifying geographic hotspots and usage trends, this project aids in aligning environmental goals with on-ground adoption realities—ultimately supporting smarter, more sustainable transportation planning and public policy formulation.

To support the planning and expansion of EV infrastructure, it is essential to analyze historical trends and adoption behaviors using data-centric approaches. By leveraging publicly available datasets on electric vehicle population statistics by county and state, we can derive insights into growth trends, regional disparities, and behavioral patterns in EV usage. This study, **ElectroTrend: Insights into Electric Vehicle Adoption Patterns**, applies data analysis and visualization techniques to explore these patterns and understand the evolving dynamics of electric mobility in the U.S.

The primary motivation for this project is to inform stakeholders—including policymakers, transportation planners, and environmental advocates—about regions that are excelling in or lagging behind EV adoption. A clearer understanding of these trends enables more targeted interventions, infrastructure investments, and awareness campaigns to accelerate the transition toward sustainable transportation.

This project follows a structured pipeline: preprocessing the dataset to handle missing data and standardize variables, performing exploratory data analysis (EDA), detecting outliers, and visualizing trends and correlations using Python libraries such as **Pandas**, **NumPy**, **Matplotlib**, and **Seaborn**. The results offer a data-driven perspective on how EV adoption is progressing over time and across geographies, and how usage patterns (e.g., commercial vs. passenger vehicles) influence electric mobility uptake.

By utilizing empirical data and visual storytelling, this study contributes to a deeper understanding of EV adoption patterns and encourages data-informed strategies for building a greener and more resilient transportation future.

2. SOURCE OF DATASET

The dataset used in this project, titled **"Electric Vehicle Population Size History by County"**, was obtained from the official U.S. open data platform Data.Gov. It provides historical records of electric vehicle registrations categorized by county and state, along with vehicle types, usage classification, and percentage of electric vehicles relative to the total registered vehicles. This rich dataset enabled a comprehensive analysis of EV adoption patterns across different regions in the United States.

[Link](https://catalog.data.gov/dataset/electric-vehicle-population-size-history-by-county) – <https://catalog.data.gov/dataset/electric-vehicle-population-size-history-by-county>

3. EXPLORATORY DATA ANALYSIS (EDA)

The Exploratory Data Analysis (EDA) process for this project involved several systematic steps to understand and prepare the Electric Vehicle Population dataset for analysis.

Initial data inspection was performed using *df.info()* and *df.describe(*) functions to understand the dataset structure, data types, and basic statistics. The *df.nunique()* function was used to identify the number of unique values in each column, helping to understand the categorical variables' cardinality.

The data cleaning process addressed several quality issues:

* **Missing Value Treatment**: Missing values were identified using *df.isnull().sum().* The *dropna()* function removed rows with missing County or State values. The *pd.to\_datetime()* function with errors='coerce' parameter converted Date columns to standard datetime format, and *dropna()* again removed invalid dates. Missing values in "Percent Electric Vehicles" were filled with zeros using *fillna(0).*
* **Data Integrity Verification**: Mathematical consistency was verified using boolean filtering to ensure 1. "EV Total" = "BEV + PHEV" 2. "Total Vehicles" = "EV Total + Non-EV Total". Non-compliant records were excluded using boolean indexing.
* **Standardization**: Text fields were standardized using string methods - *str.upper()* for state names, *str.title()* for county names, and *str.strip()* to remove extraneous whitespace from all text fields.
* **Duplicate Removal**: Duplicate records were identified with *duplicated().sum()* and removed using *drop\_duplicates()* to prevent data redundancy.
* **Outlier Detection**: A custom function implementing the Interquartile Range (IQR) method was used to identify outliers in numeric columns. For each numeric column, Q1 (25th percentile) and Q3 (75th percentile) values were calculated using quantile(), and the IQR computed. Values falling below Q1-1.5*IQR or above Q3+1.5*IQR were flagged as outliers using boolean indexing.

Data transformation for analysis included temporal aggregation using groupby() with datetime periods and geographic aggregation using groupby() with categorical columns.

This EDA process ensured the dataset was clean and properly formatted for the subsequent analytical objectives of the project, providing a solid foundation for reliable insights into electric vehicle adoption patterns.

4. ANALYSIS ON DATASET

Objective 1: EV Growth Over Time

**Introduction**

This analysis examines the temporal evolution of electric vehicle adoption, tracking how the total number of electric vehicles has changed over the dataset's time period. Understanding growth patterns provides insights into adoption acceleration or deceleration and potential seasonal trends.

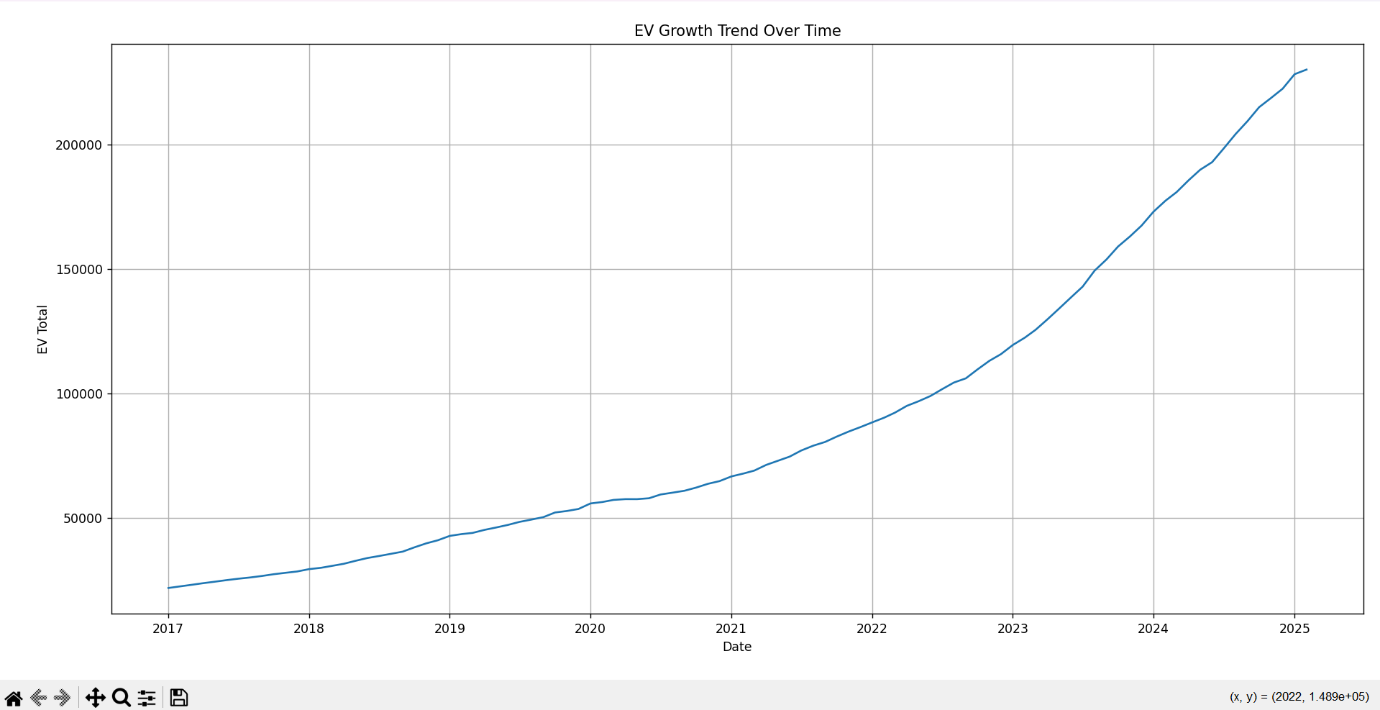
**General Description**

The analysis aggregates electric vehicle counts on a monthly basis to create a time series visualization of EV growth. By summing the "Electric Vehicle (EV) Total" for each month and plotting the results chronologically, this analysis reveals the trajectory of EV adoption over time.

**Specific Requirements, Functions and Formulas**

The implementation utilizes pandas' datetime functionality to convert and group the data by month. Monthly aggregation is performed using *groupby()* with datetime periods and timestamp conversion. Visualization is achieved using Seaborn's *lineplot()* function to create a clear representation of the temporal trend.

**Analysis Results**



**Visualization**

The line graph illustrates the cumulative growth of electric vehicles over time. The visualization reveals the pace of EV adoption and identifies any significant inflection points where growth patterns may have changed. This temporal perspective is crucial for understanding how external factors such as policy changes, technological advancements, or economic conditions might influence EV adoption rates.

Objective 2: EV Adoption by County & State

**Introduction**

This analysis investigates the geographic distribution of electric vehicle adoption, identifying regions with the highest penetration of electric vehicles relative to their total vehicle population.

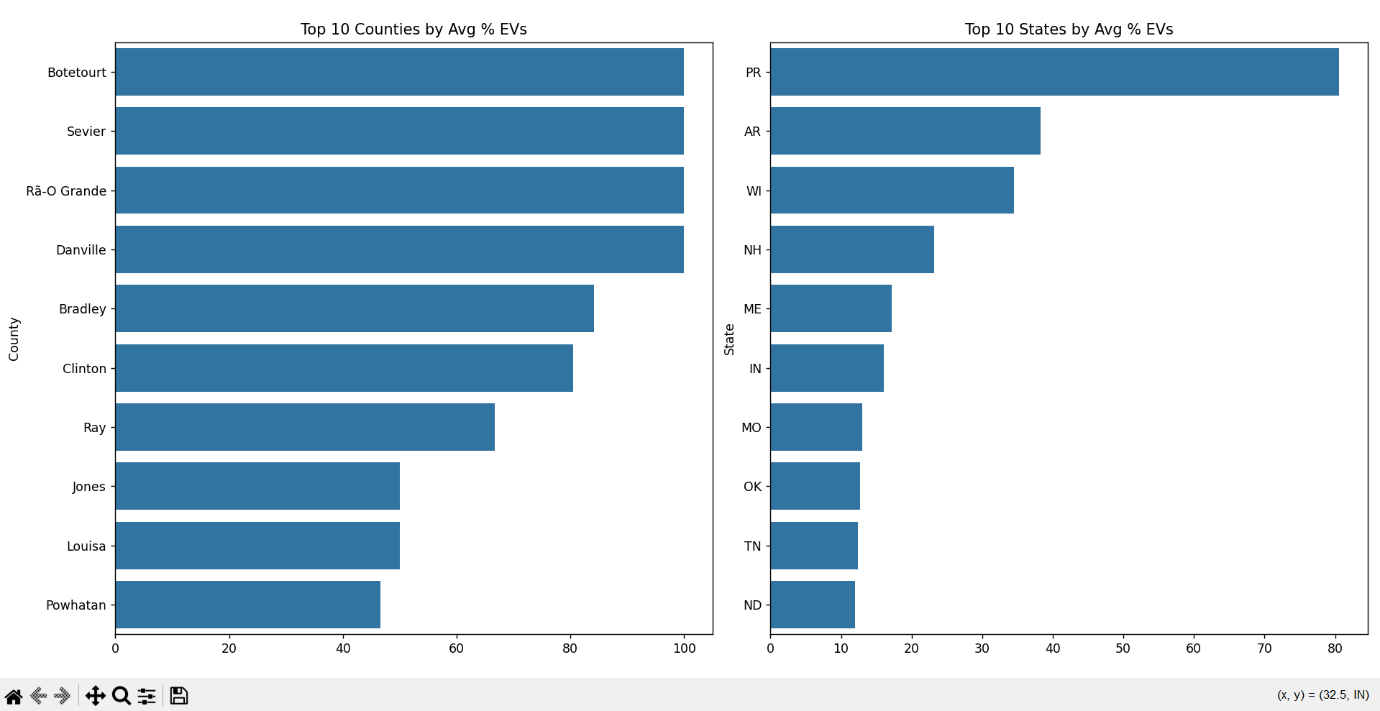
**General Description**

By aggregating and comparing the "Percent Electric Vehicles" metric across different counties and states, this analysis highlights geographic disparities in EV adoption. The focus on percentage rather than absolute numbers provides a normalized view that accounts for differences in population and vehicle density.

**Specific Requirements, Functions and Formulas**

The analysis uses pandas *groupby()* operations to calculate mean EV percentages by region, with *sort\_values()* and *head()* functions to identify the top regions. Visualization employs Seaborn's *barplot()* function to create side-by-side bar charts facilitating comparison between different geographic entities.

**Analysis Results**



**Visualization**

The dual bar charts display the top 10 counties and states with the highest average percentage of electric vehicles. This visualization reveals which regions are leading in EV adoption relative to their total vehicle population, providing insights into potential geographic factors, policies, or demographic characteristics that might influence EV adoption rates.

Objective 3: Correlation Between BEV, PHEV, EV Total

**Introduction**

This analysis explores the relationships between different types of electric vehicles: Battery Electric Vehicles (BEVs), Plug-In Hybrid Electric Vehicles (PHEVs), and the total electric vehicle count.

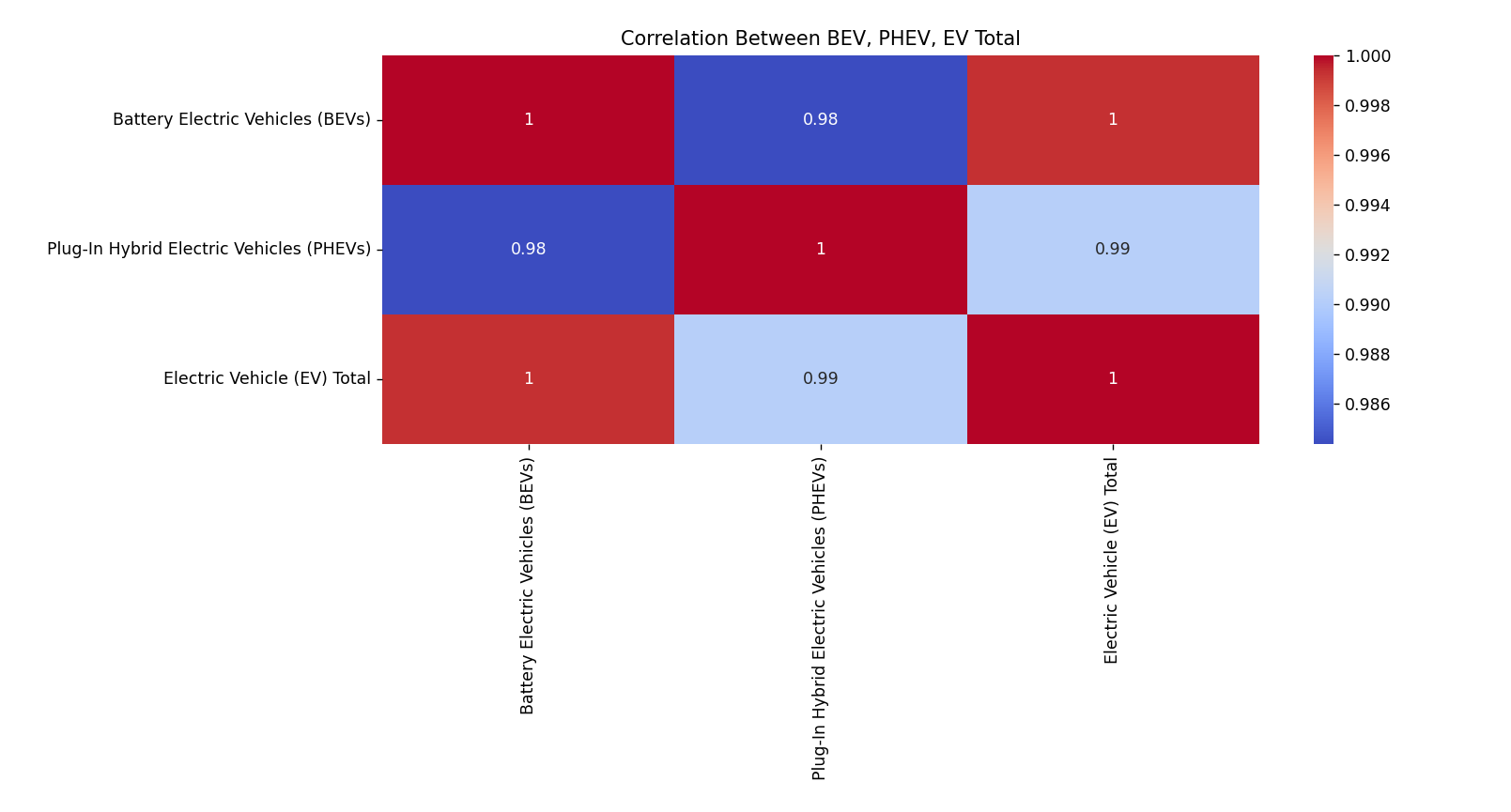
**General Description**

By calculating correlation coefficients between these three variables, this analysis reveals how strongly the adoption of BEVs relates to PHEVs and how each contributes to the overall electric vehicle population.

**Specific Requirements, Functions and Formulas**

The analysis uses pandas' *corr()* function to compute Pearson correlation coefficients between the three vehicle count variables. A heatmap visualization with annotated values using Seaborn's *heatmap()* function provides a clear representation of the relationship strengths.

**Analysis Results**



**Visualization**

The correlation heatmap illustrates the strength of relationships between different types of electric vehicles. The visualization reveals whether regions with high BEV adoption also tend to have high PHEV adoption, or if there are inverse or independent relationships between these vehicle types. Understanding these correlations provides insights into consumer preferences and potential substitution effects between different electric vehicle technologies.

Objective 4: Boxplot Outlier Visualization for Percent Electric Vehicles

**Introduction**

This analysis focuses on identifying and visualizing outliers in the distribution of electric vehicle adoption percentages across different regions.

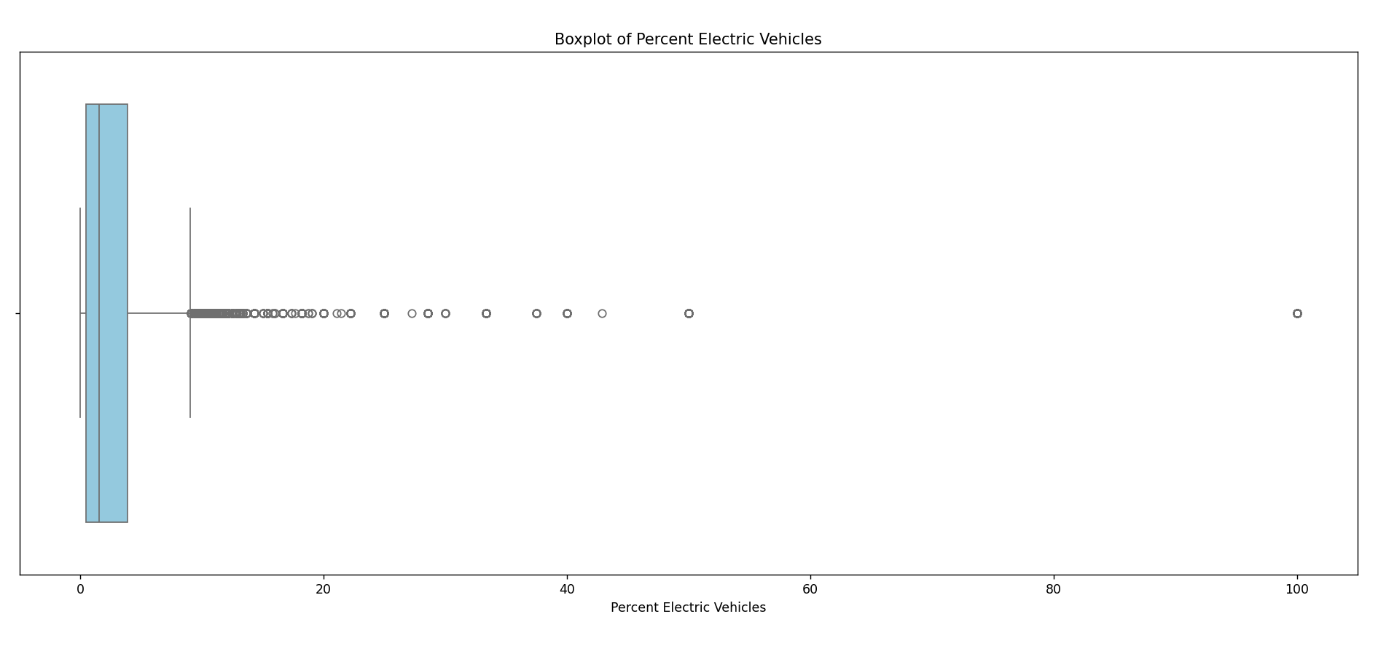
**General Description**

Using a boxplot representation of the "Percent Electric Vehicles" metric, this analysis highlights the central tendency, dispersion, and presence of outliers in electric vehicle adoption rates. This visualization provides insights into typical EV penetration rates and regions with exceptionally high or low adoption.

**Specific Requirements, Functions and Formulas**

The analysis utilizes Seaborn's *boxplot()* function to create a statistical visualization of the data distribution. This approach leverages the boxplot's ability to simultaneously display median, quartiles, and outlier values.

**Analysis Results**



**Visualization**

The boxplot visualizes the distribution characteristics of electric vehicle percentages across the dataset. This visualization reveals the median EV adoption rate, the typical range of adoption rates (interquartile range), and identifies regions with unusually high or low EV penetration. The presence and extent of outliers provides insights into the uniformity or disparity of EV adoption across different regions.

Objective 5: Counties with 100% EVs

**Introduction**

This analysis identifies and examines counties that have achieved complete electric vehicle adoption, where all registered vehicles are electric.

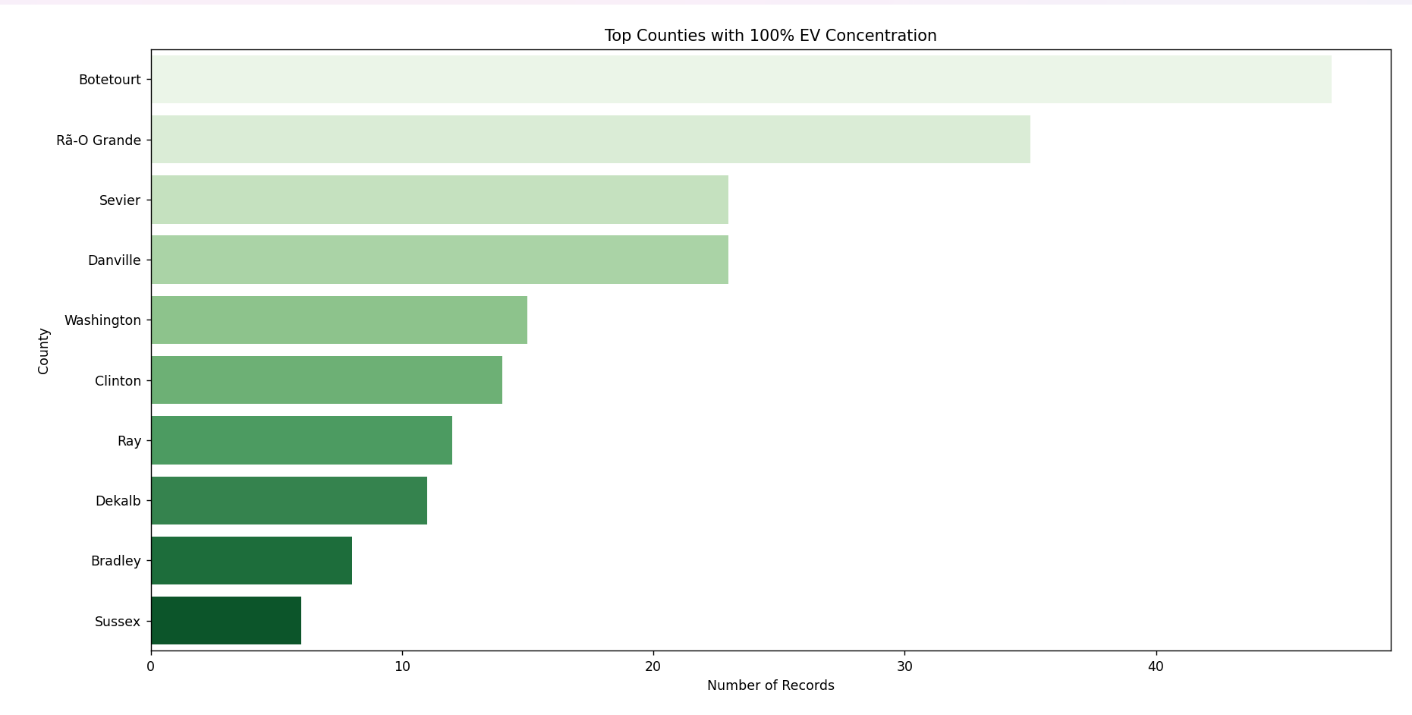
**General Description**

By filtering the dataset for records with "Percent Electric Vehicles" equal to 100, this analysis highlights regions that have reached full electrification of their vehicle fleet. The frequency of such records by county indicates which areas consistently maintain complete EV adoption.

**Specific Requirements, Functions and Formulas**

The analysis uses boolean filtering and *value\_counts()* function to identify counties with complete EV adoption. Visualization is accomplished using Seaborn's *barplot()* function to display the frequency of 100% EV records for each county.

**Analysis Results**



**Visualization**

The bar chart displays counties with records showing 100% electric vehicle adoption, ranked by the frequency of such records. This visualization identifies regional leaders in complete vehicle electrification, highlighting areas that may have implemented particularly effective policies or possess unique characteristics conducive to full EV adoption.

Objective 6: Scatter Plot - Total Vehicles vs % EVs by Usage Type

**Introduction**

This analysis investigates the relationship between vehicle fleet size and electric vehicle adoption rates across different vehicle usage categories.

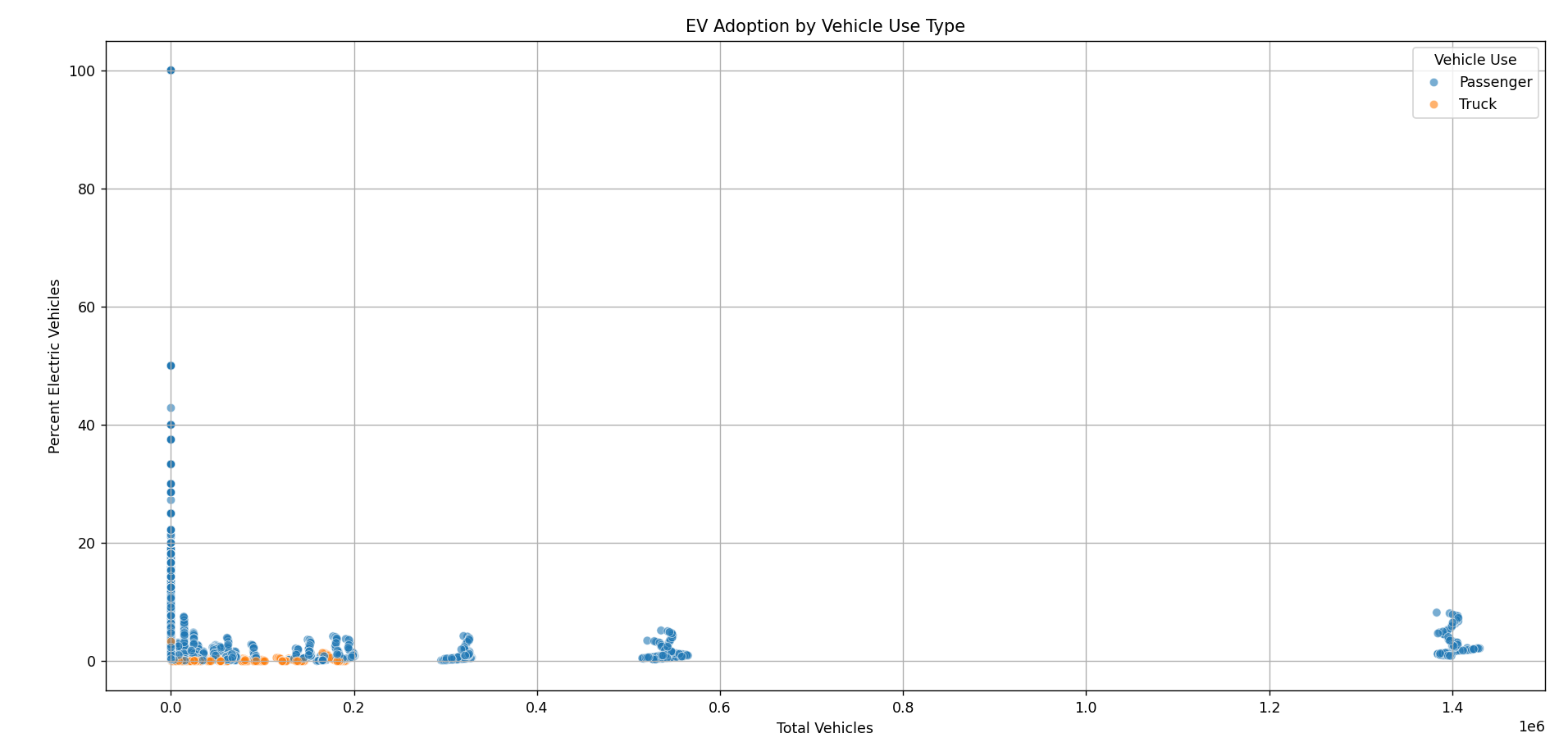
**General Description**

By plotting "Total Vehicles" against "Percent Electric Vehicles" and differentiating by "Vehicle Primary Use," this analysis reveals how EV adoption varies with fleet size and vehicle purpose. This multi-dimensional view provides insights into how different sectors or usage types contribute to overall EV adoption.

**Specific Requirements, Functions and Formulas**

The analysis employs Seaborn's *scatterplot()* function to create a multi-variable visualization. The use of the *hue* parameter for color differentiation by vehicle usage type adds an additional analytical dimension.

**Analysis Results**



**Visualization**

The scatter plot illustrates the relationship between total vehicle population and electric vehicle percentage across different vehicle usage types. This visualization reveals whether larger fleets tend to have higher or lower EV adoption rates and how this relationship varies by vehicle purpose. The pattern of points across the plot provides insights into sector-specific EV adoption trends and potential barriers or enablers for electrification in different vehicle use categories.

5. CONCLUSION

The Electric Vehicle Population Size History Analysis project has revealed several significant insights into the patterns and trends of electric vehicle adoption across different regions and time periods. The temporal analysis demonstrates a consistent upward trajectory in electric vehicle adoption, indicating growing acceptance and integration of electric mobility solutions. However, the growth rate varies significantly across different geographic regions, with certain counties and states emerging as clear leaders in EV penetration.

The correlation analysis between BEVs and PHEVs revealed interesting relationships between different electric vehicle technologies, suggesting potential market preferences and adoption patterns. The identification of outliers in EV adoption percentages highlighted the disparities in electrification progress, with some regions achieving exceptional adoption rates while others lag significantly behind.

Particularly noteworthy is the identification of counties with 100% electric vehicle records, representing pockets of complete electrification that could serve as case studies for broader implementation. The relationship between fleet size and EV adoption across different vehicle usage types uncovered sector-specific patterns that provide valuable insights for targeted electrification strategies.

These findings collectively contribute to our understanding of the electric vehicle transition landscape, revealing both progress made and challenges that remain in the shift toward sustainable transportation. The geographic and temporal disparities identified suggest that contextual factors, potentially including policy environments, infrastructure availability, and socioeconomic conditions, play significant roles in determining electric vehicle adoption rates.

6. FUTURE SCOPE

The analysis conducted in this project opens several avenues for future research and expansion. First, incorporating additional variables such as charging infrastructure density, local incentive programs, and socioeconomic indicators could provide deeper insights into the factors driving regional disparities in EV adoption. This multi-factorial analysis would help identify the most influential determinants of electric vehicle penetration.

Second, expanding the temporal scope of the analysis could enable more sophisticated time series modeling, including forecasting future EV adoption trends based on historical patterns. Predictive modeling could help stakeholders anticipate future EV growth and plan accordingly for infrastructure and policy needs.

Third, granular analysis of vehicle usage types could be enhanced by investigating specific sectors or industries that show particularly high or low EV adoption rates. Understanding the unique challenges and opportunities in different sectors could inform targeted strategies for accelerating electrification across diverse vehicle fleets.

Fourth, comparative analysis with international EV adoption data could provide a global context for the findings, highlighting how the observed patterns compare to trends in other countries with different policy environments and market conditions.

Finally, integrating environmental impact metrics could translate the observed EV adoption patterns into quantifiable environmental benefits, such as carbon emission reductions and air quality improvements. This would connect the technical analysis to broader sustainability goals and provide a more comprehensive view of the significance of electric vehicle adoption.

7. REFERENCES

[1] Pandas Documentation, "pandas.DataFrame.groupby," 2023. [Online]. Available: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html>

[2] Seaborn Documentation, "seaborn.lineplot," 2023. [Online]. Available: <https://seaborn.pydata.org/generated/seaborn.lineplot.html>

[3] Matplotlib Documentation, "matplotlib.pyplot," 2023. [Online]. Available: <https://matplotlib.org/stable/api/pyplot_api.html>

[4] NumPy Documentation, "NumPy v1.24 Manual," 2023. [Online]. Available: <https://numpy.org/doc/stable/>

IMPORTANT LINK

LinkedIn Link – “<https://www.linkedin.com/posts/asmit03_electrotrend-electricvehicles-sustainability-activity-7316756637661179905-ncAN?utm_source=share&utm_medium=member_desktop&rcm=ACoAAEX-MVYBnLKCfjIe2EBhNJuIyO-XnETJjks”>

GitHub Link – “<https://github.com/Asmit03/Python-Project---ElectroTrend”>

Google Drive Link – [“https://drive.google.com/drive/u/1/folders/1p5ETJec440\_bhBz035S2Jz5w\_VDz2TnK”](“https:/drive.google.com/drive/u/1/folders/1p5ETJec440_bhBz035S2Jz5w_VDz2TnK”)