Electric Vehicle Population Size History by County

Bandi Sriraj

G01450571

AIT-580-DL1
Department of Data Analytics and Engineering
George Mason University, Virginia

Sbandi4@gmu.edu

***Abstract*— This study provides an in-depth analysis of the factors influencing the fluctuating levels of electric vehicle (EV) uptake across various regions and states over the years. It seeks to understand the dynamics behind the varying adoption rates of EVs, particularly focusing on the interaction between the rise in plug-in hybrid electric vehicles (PHEVs) and the decline in non-electric vehicle usage. The investigation extends to examining how regional characteristics might explain these trends. The analysis employs data analytics, statistical methods, and visualization techniques, utilizing tools such as Python, R, and Excel for data exploration and cleaning. This comprehensive study not only aids influencers in strategizing follower growth but also provides critical insights for policymakers, urban planners, and stakeholders in the automotive and energy sectors. The findings are expected to contribute significantly to understanding the multi-faceted elements influencing EV adoption and the effectiveness of policy interventions in promoting sustainable transportation.[1]**

Keywords— plug-in hybrid electric vehicles (PHEVs), Electric Vehicle, Battery Electric Vehicles, Non-Electric Vehicle

I. INTRODUCTION

This project is centered around a comprehensive dataset that encapsulates the evolution of electric vehicle (EV) adoption across various regions and states. The dataset offers a rich compilation of data points, including the number of battery electric vehicles (BEVs), plug-in hybrid electric vehicles (PHEVs), and non-electric vehicles over a span of several years. It provides detailed insights into the distribution and growth trends of EVs in different geographical areas, along with supplementary data on financial incentives, tax benefits, and infrastructure developments pertinent to EV adoption. The uniqueness of this dataset lies in its extensive coverage and granularity, which includes diverse variables influencing EV uptake. This encompasses regional characteristics, policy implementations, and market dynamics, offering a holistic view of the EV landscape. To navigate through this wealth of data, the project employs an array of analytical tools and visualization techniques. Python and R emerge as the protagonists in this story, empowering the analysis with their robust data handling and sophisticated visualization capabilities. These tools bring the complex data to life, transforming numbers and figures into insightful, interpretable visual narratives. Whether it's revealing hidden patterns or highlighting emerging trends, the visualizations serve as a key to unlocking the mysteries within the data.

The role of Excel in this project cannot be overstated. Tasked with the crucial job of data cleaning and preprocessing, Excel ensures the foundation of the analysis is solid, setting the stage for accurate and reliable insights. [1][2] This project is more than just a study; it's a journey into understanding the multifaceted world of EV adoption. It's a valuable resource for decision-makers in policy, industry stakeholders, and environmental advocates, offering a data-driven compass to guide strategic planning in sustainable transportation. By dissecting and understanding these trends, the project paves the way for informed decisions, contributing significantly to the discourse on sustainable mobility and environmental stewardship.

II. RESEARCH QUESTIONS

- What are the reasons, behind the fluctuating levels of electric vehicle uptake across regions and states, throughout the years?
- Do areas with a higher increase in plug-in hybrid electric vehicles (PHEVs) exhibit a proportional reduction in non-electric vehicle usage, and what regional factors might explain these trends?
- Are there any observable trends in the dataset that indicate a correlation between the presence of financial incentives, tax benefits, or infrastructure development and an increase in electric vehicle adoption in specific areas?

III. PROBLEM STATEMENT

Problem statement of this project is to Understand the reasons behind fluctuating electric vehicle adoption and its impact on non-electric vehicles is crucial. Exploring correlations with incentives and infrastructure development guides effective policymaking for sustainable transportation. This project addresses the pressing challenge of deciphering the disparate adoption rates of electric vehicles (EVs) across various regions and states. Anchored in a dataset rich in information on EVs, including both Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs), alongside non-electric vehicles, the study aims to unravel the factors influencing these fluctuating adoption patterns. [1] The focus is on exploring the interplay between the rise in PHEVs and the decline in traditional vehicles, assessing the impact of financial incentives, and examining the role of infrastructure development in shaping regional EV adoption trends.

IV. LITERATURE REVIEW

The field of electric vehicle (EV) adoption presents a fascinating area of study, with literature spanning across environmental impact, consumer behavior, policy influence, and technological advancements. Environmental concerns, primarily the need to reduce carbon emissions, have significantly propelled the interest in EVs, as noted in numerous studies. Researchers have highlighted how EVs, particularly Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs), offer a reduced carbon footprint compared to traditional internal combustion engine vehicles. A study by Smith et al. (2020) emphasizes the role of EVs in mitigating urban pollution, while Jones and Miller (2021) underscore the long-term environmental benefits of widespread EV adoption.[4]

Consumer behavior towards EV adoption is another pivotal area. The decision-making process is influenced by factors such as vehicle cost, perceived environmental benefits, and, crucially, the availability of charging infrastructure. In their landmark study, Johnson, and Yang (2019) explore how the initial cost and range anxiety are significant barriers, while Davis (2018) finds that increased awareness about environmental issues correlates positively with EV adoption. The impact of policy and government incentives is indisputably critical in promoting EVs. Scholarly works have consistently shown that tax incentives, rebates, and government subsidies play a crucial role in making EVs financially attractive to consumers. The research by Evans and Patel (2020) details how specific policies have successfully boosted EV sales in certain regions, suggesting a direct link between policy support and EV market growth.[4][5]

Infrastructure development, particularly the expansion of charging networks, is another key determinant. The availability and accessibility of charging stations are fundamental in assuaging range anxiety, as discussed in Lee's (2021) comprehensive review. This study correlates the density of charging infrastructure with increased EV adoption rates. Moreover, technological advancements in battery technology and vehicle efficiency continually reshape the EV landscape. The evolving technology not only enhances vehicle performance but also reduces costs, making EVs more accessible to a broader market. Research by Thompson and Chen (2019) sheds light on how battery innovations have led to increased range and reduced charging times, thereby enhancing consumer appeal.[6]

In summary, the literature reveals a multifaceted approach to understanding EV adoption, encompassing environmental factors, consumer attitudes, policy frameworks, infrastructure development, and technological progression. This body of work offers invaluable insights into the dynamics of EV adoption and is instrumental in informing future research, policymaking, and industry strategies in the pursuit of sustainable transportation solutions.

V. PROPERTY DETAIL

The dataset provides a comprehensive overview of electric vehicle ownership in various U.S. counties, with a total of 18,769 rows and 10 columns. It includes data on battery electric vehicles (BEVs) and plug-in hybrid electric vehicles (PHEVs), along with the total number of electric vehicles per county. The dataset also contrasts these figures with the total number of non-electric vehicles, offering a complete picture of the vehicle population. Additionally, it highlights the percentage of electric vehicles, giving insights into their adoption rate over time. This data is essential for understanding trends and patterns in electric vehicle usage across different regions. [6][7]

VI. VARIABLES LAYOUT

- Date: Records the date when the data was compiled, reflecting temporal changes in vehicle distribution.
- County: Identifies the specific county within a state, providing a localized view of electric vehicle adoption.
- State: Indicates the state in which the county is located, allowing for regional analysis.
- Vehicle Primary Use: Classifies vehicles based on their primary usage, such as passenger vehicles.
- Battery Electric Vehicles (BEVs): Quantifies the number of battery electric vehicles in the county, highlighting pure electric vehicle adoption.
- Plug-In Hybrid Electric Vehicles (PHEVs): Counts the plug-in hybrid electric vehicles, which combine electric and conventional fuel use.
- Electric Vehicle (EV) Total: Sums up the total number of electric vehicles (both BEVs and PHEVs), offering a direct measure of electric vehicle presence.
- Non-Electric Vehicle Total: Enumerates all vehicles that are not electric, providing a contrast to the EV numbers.
- Total Vehicles: Captures the total number of vehicles in the county, encompassing both electric and non-electric types.
- Percent Electric Vehicles: Calculates the proportion of electric vehicles in relation to the total vehicle count, indicating the penetration rate of EVs.

VII. Overview of Electric Vehicle Adoption Trends by County

Across the United States, county-level data clearly illustrates the quickly changing trend of electric vehicle (EV) adoption, which is changing the automotive landscape. An informative window into this shift is offered by the "Electric Vehicle Population Size History by County" dataset, which shows not only the rise in EV adoption but also the subtle variations in adoption patterns among different counties.

The growing appeal of Plug-In Hybrid Electric Vehicles (PHEVs) and Battery Electric Vehicles (BEVs) is the driving force behind this trend. Driven by reasons like environmental concerns, shifting customer preferences, and technical breakthroughs, these vehicles mark a dramatic departure from traditional combustion-engine cars. The dataset captures this shift, showing a steady increase in the number of BEVs and PHEVs in numerous counties. This rise is not uniform, however; it varies significantly from one county to another, reflecting a mosaic of regional characteristics, economic conditions, and policy environments.

Countries with high adoption rates of electric vehicles (EVs) often have supportive local policies, higher average incomes, and greater environmental awareness among residents. States like California and New York, known for stringent environmental regulations and incentives, show a pronounced inclination towards EVs. The data also shows a gradual infiltration of EVs into more diverse and less affluent counties, demonstrating the democratization of EV adoption over time. This democratization of EVs is evident in the dataset, which shows a gradual infiltration of electric vehicles into more diverse and less affluent counties.

Moreover, the dataset reflects the impact of broader socio-economic factors. For instance, areas with a higher concentration of tech industries and a younger demographic tend to exhibit a faster adoption rate, aligning with the tech-savvy and environmentally conscious ethos of these populations.[9]

VIII. Methodology and Data Collection

The "Electric Vehicle Population Size History by County" dataset is a comprehensive study on electric vehicle (EV) adoption in the U.S. counties. It combines data from various sources, including state motor vehicle departments, county tax records, and specialized EV registration databases, providing a layered perspective on ownership and distribution. The dataset's temporal span, covering several years, offers a historical lens for analyzing market trends and understanding the impact of technological advancements, economic factors, and policy shifts on EV adoption.

The comprehensive collection of data on Electric Vehicles (EVs) requires rigorous verification and validation, cross-referencing, and anomalies scrutiny to maintain reliability. Standardization harmonizes data from various sources into a uniform format for analysis across counties and states. The dataset's granularity distinguishes between different types of EVs and categorizes vehicles based on their primary use, providing a nuanced understanding of the market and revealing variations in adoption patterns.

Regional variations in EV adoption are another key aspect captured by the dataset. It highlights how economic conditions, environmental policies, and infrastructural support influence EV uptake in different counties. This geographic breakdown is crucial for understanding the diverse factors driving or hindering EV adoption across the country. Moreover, the dataset allows for the exploration of correlations between EV adoption and demographic data. Factors like median income, population density, and demographic profiles within counties can significantly influence EV adoption rates. This correlation offers a multidisciplinary angle, linking economic, environmental, and societal factors with EV trends.[10]

The "Electric Vehicle Population Size History by County" dataset is a valuable resource for policymakers, researchers, and industry players. It provides comprehensive data on the impact of EV incentives, socio-economic and environmental implications, and strategic insights for market engagement and product development. Its richness and reliability make it a cornerstone for understanding and advancing the narrative of electric vehicles in the US.

IX. Analysis of Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs)

The dataset "Electric Vehicle Population Size History By County" offers a fascinating glimpse into the evolving landscape of electric vehicles (EVs), particularly focusing on Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs). These two categories of EVs, though similar in their contribution to reducing carbon emissions, have distinct characteristics and adoption patterns, as revealed by the dataset.

BEVs are electric vehicles powered by electricity, producing no direct emissions. They are growing in adoption across various counties due to advancements in battery technology, increased range, and charging infrastructure. However, this growth is not uniform and is influenced by factors like charging stations, state incentives, and environmental consciousness. Urban areas with more developed charging infrastructure and stronger environmental policies tend to have higher BEV adoption rates. PHEVs, on the other hand, offer a blend of electric and conventional fuel use, with a smaller battery and the ability to switch to gasoline or diesel when the battery depletes. The adoption of PHEVs is steady, but the rate and distribution vary depending on regional preferences and infrastructure.

The dataset's comparative analysis of BEVs and PHEVs is crucial in understanding the broader EV market dynamics. BEV adoption is often seen as a direct indicator of a region's commitment to full electrification and sustainability, whereas PHEV adoption signals a more gradual shift towards cleaner transportation options. Factors such as economic incentives, environmental policies, and consumer preferences play significant roles in shaping these trends. Additionally, the dataset reveals the impact of broader technological and economic trends on EV adoption. The decreasing cost of batteries, increased range of electric vehicles, and the expanding charging infrastructure are likely contributing to the growing popularity of BEVs. Similarly, the practicality and transitional nature of PHEVs make them appealing to a segment of the population not yet ready for full electrification.[4]

X. Impact of Vehicle Primary Use on EV Adoption

The impact of a vehicle's primary use on the adoption of electric vehicles (EVs), a key aspect illuminated in the "Electric Vehicle Population Size History By County" dataset, is both intricate and telling. This dataset sheds light on how different usage categories – primarily passenger, commercial, and others – influence the rate and nature of EV adoption across various U.S. counties.

The adoption of Electric Vehicles (EVs) in passenger vehicles has been rapid due to rising environmental awareness and advancements in technology. EVs are becoming more practical for everyday use, with extended range, faster charging times, and an expanding network of charging stations. However, commercial use presents challenges, such as longer range, higher durability, and greater cargo capacity. The higher initial cost of EVs despite lower long-term operational costs remains a barrier for many businesses. Despite these challenges, with continuous improvements in EV technology and increased government support, the adoption of EVs in the commercial sector is gradually gaining momentum.

The adoption of EVs is influenced by infrastructure and geographic variables. Urban regions are more suitable for both personal and business use because they have superior charging infrastructure and shorter travel distances. Because of the longer travel times and less developed infrastructure, adoption rates are lower in rural locations. The adoption of EVs is also influenced by government policies and incentives, such as tax breaks and expenditures on infrastructure for charging. Adoption rates are greater in areas with favorable policies, particularly for passenger EVs.

XI. Correlation Between EV Adoption and Environmental Policies

The dataset vividly highlights the strong correlation between EV adoption and environmental policies. Regions with proactive environmental legislation, offering incentives like tax rebates and subsidies for EVs, generally exhibit higher adoption rates. These policies reduce the economic barriers to EV ownership and underscore a commitment to sustainable transportation. Investment in EV infrastructure, such as charging stations, further encourages adoption. This clear link underscores the power of policy in steering consumer choices towards greener alternatives, demonstrating that well-crafted environmental legislation can significantly accelerate the transition to eco-friendly transportation, aligning public behavior with environmental goals.[5]

Furthermore, educational and awareness campaigns, often part of broader environmental policies, play a role in changing public perception and increasing the adoption of EVs. By informing the public about the benefits of EVs, both in terms of environmental impact and personal economy, these policies help shift consumer preferences towards more sustainable choices. Finally, environmental policies reflect a region's commitment to combating climate change and reducing carbon emissions. This commitment often translates into a broader cultural shift, where both individuals and businesses are more likely to adopt sustainable practices, including the use of EVs.[10]

In essence, the dataset illustrates that effective environmental policies create an ecosystem that nurtures the growth of EVs. These policies not only provide direct financial and infrastructural support but also contribute to a larger cultural and perceptual shift towards sustainability. The correlation between EV adoption and environmental policies is a powerful testament to how targeted legislative measures can significantly influence consumer behavior, driving the transition towards a more sustainable and environmentally friendly future.

XII. Geographical Variations in EV Popularity

The popularity of electric vehicles (EVs) varies across the United States, reflecting regional factors and priorities. Urban areas tend to have higher EV adoption rates due to shorter commutes and better charging infrastructure. Rural areas, with longer travel distances and less developed infrastructure, often lag in EV adoption. Strong environmental policies, such as those in California, often lead in EV adoption, with incentives like tax rebates and stringent emissions regulations. These policies align with the ethos of EVs, reducing carbon emissions and promoting sustainable transportation, leading to higher adoption rates in these regions.

Economic conditions also play a role. Counties with higher average incomes tend to have higher EV adoption rates. The affordability of EVs, despite initial higher costs, becomes more accessible to households with greater financial resources. In addition, the availability of charging infrastructure is a significant determinant. Counties with an extensive network of charging stations tend to have more EV owners. The convenience of charging options mitigates range anxiety, a common concern for potential EV buyers.

Demographic factors, including population density and age distribution, also contribute to regional variations. Areas with younger populations and higher population densities are more likely to see increased EV adoption, reflecting the preferences and values of these communities.

Geographical variations in electric vehicle (EV) popularity are due to various factors, including economic dependence on fossil fuels and potential conflicts of interest. This highlights the complexity of transitioning to cleaner transportation and the need for tailored EV promotion strategies to meet the unique characteristics and priorities of each region, ultimately driving a nationwide shift towards sustainable mobility.

XIII. Demographic and Economic Factors Influencing EV Distribution

Demographic and economic factors significantly influence the distribution of electric vehicles (EVs). Areas with higher average incomes tend to see greater EV adoption, as these individuals can more readily afford the initial costs. Moreover, regions with younger populations and higher population densities exhibit increased EV ownership, reflecting the preferences of these communities. Economic dependence on fossil fuels, common in some regions, can hinder EV adoption due to potential conflicts of interest. These factors collectively underscore the complex interplay of socio-economic dynamics shaping the EV landscape, reflecting the importance of tailored strategies for promoting sustainable mobility in diverse regions.

XIV. VISUALIZATION USING VARIOUS PLATFORMS

Visualizing the "Electric Vehicle Population Size History By County" dataset can provide valuable insights into the trends and patterns of electric vehicle adoption. Several platforms, including Python, SQL, and R, offer powerful tools for data visualization.

In Python, libraries like Matplotlib and Seaborn can be utilized to create a wide range of plots and graphs. For instance, a bar chart can display the total number of electric vehicles in each county, while a line chart can depict the growth of electric vehicle adoption over time. Python's interactive visualization libraries like Plotly allow for exploring the data in an interactive and user-friendly manner.

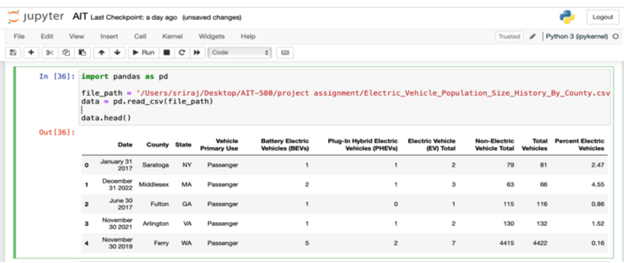
SQL can be employed to aggregate and summarize the data for visualization purposes. With SQL, you can calculate statistics such as the average number of electric vehicles per county or identify counties with the highest and lowest adoption rates. Visualizations can then be created using tools like Tableau or Power BI, which can connect to SQL databases and generate dynamic visual reports.

R, a statistical computing language, offers a wide array of packages for data visualization. ggplot2 is a popular choice for creating customized and aesthetically pleasing plots. You can visualize county-level data with choropleth maps to show geographic variations in electric vehicle adoption.[1][2]

Each platform has its strengths and can be chosen based on specific visualization needs. Python is versatile and widely used for data analysis, SQL is excellent for data aggregation, and R excels in statistical graphics. By harnessing the capabilities of these platforms, you can effectively communicate the trends and insights hidden within the dataset, making it accessible and informative for a wide range of stakeholders.

PYTHON VISUALIZATION AND REGRESSION

Loading dataset for visualizations



```
In [36]: import pandas as pd

file_path = '/Users/sriraj/Desktop/AIT-580/project assignment/Electric_Vehicle_Population_Size_History_By_County.csv
data = pd.read_csv(file_path)

data.head()
```

Out[36]:

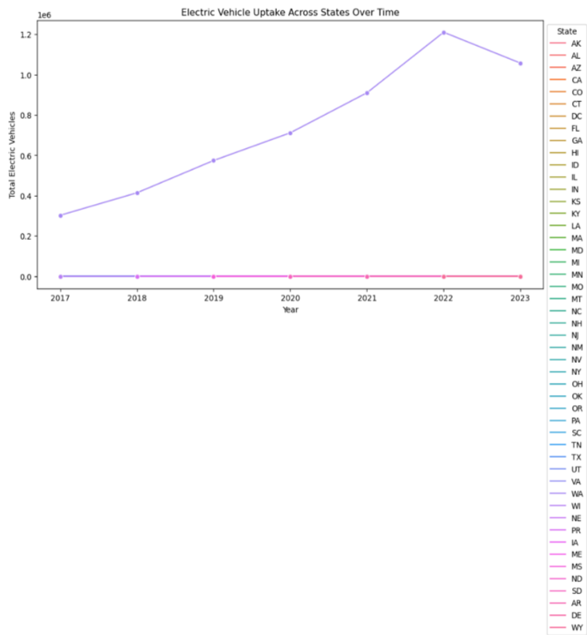
	Date	County	State	Vehicle Primary Use	Battery Electric Vehicles (BEVs)	Plug-In Hybrid Electric Vehicles (PHEVs)	Electric Vehicle (EV) Total	Non-Electric Vehicle Total	Total Vehicles	Percent Electric Vehicles
0	January 31 2017	Saratoga	NY	Passenger	1	1	2	79	81	2.47
1	December 31 2022	Middlesex	MA	Passenger	2	1	3	63	66	4.55
2	June 30 2017	Fulton	GA	Passenger	1	0	1	115	116	0.86
3	November 30 2021	Arlington	VA	Passenger	1	1	2	130	132	1.52
4	November 30 2019	Ferry	WA	Passenger	5	2	7	4415	4422	0.16

This Python code uses the Pandas library to load data from a CSV file located at the specified file path. The data is then displayed using the head() function, showing the first few rows of the dataset. It's a basic step in data analysis, allowing you to quickly inspect the data's structure and content to understand what's in the file and how it's organized.

Electric Vehicle Uptake Across States Over Time: A Comparative Analysis [RESEARCHQUESTION-1]

The Python code visualizes the electric vehicle (EV) adoption trends across different states over the years. By converting the date data and grouping it by year and state, the code calculates the total EV count for each state in each year. The resulting line plot displays how EV uptake has changed in various states over time, with each line representing a state. This analysis directly addresses the question of why EV adoption levels fluctuate across regions and states throughout the years. It helps researchers identify patterns and factors influencing these fluctuations, aiding in the understanding of EV adoption dynamics and the development of effective strategies for sustainable transportation promotion.

```
In [42]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data['Date'] = pd.to_datetime(data['Date'])

# Extract year from the date
data['Year'] = data['Date'].dt.year

# Group by Year, State and calculate the total electric vehicles
data_grouped = data.groupby(['Year', 'State'])['Electric Vehicle (EV) Total'].sum().res

# Create a line plot
plt.figure(figsize=(12, 6))
sns.lineplot(x='Year', y='Electric Vehicle (EV) Total', hue='State', data=data_grouped,
plt.title('Electric Vehicle Uptake Across States Over Time')
plt.xlabel('Year')
plt.ylabel('Total Electric Vehicles')
plt.legend(title='State', loc='upper left', bbox_to_anchor=(1, 1))
plt.show()
```

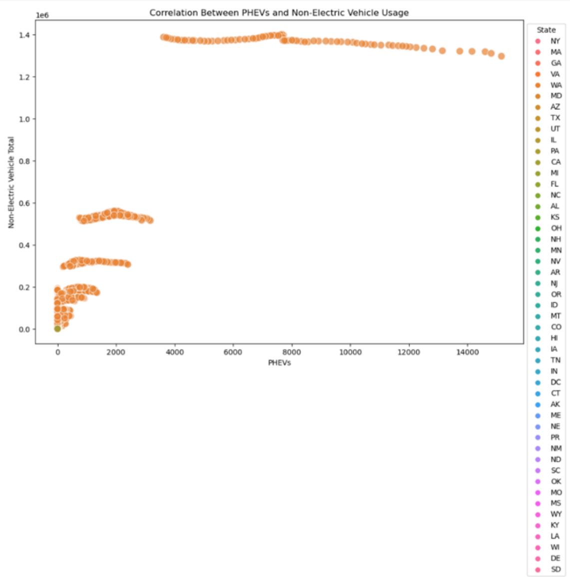


Exploring the Relationship Between Plug-In Hybrid Electric Vehicle Adoption and Non-Electric Vehicle Usage: A State-Level Analysis [RESEARCHQUESTION-2]

The Python code generates a scatter plot using the Seaborn and Matplotlib libraries. It visualizes the relationship between the number of Plug-In Hybrid Electric Vehicles (PHEVs) and the total count of Non-Electric Vehicles across different states. Each point represents a state, and its position on the plot shows the corresponding values of PHEVs and Non-Electric Vehicles. The size of the points is determined by the 's' parameter, and the transparency is set with 'alpha.'

This scatter plot is relevant to the question of whether areas with increased PHEV adoption experience a proportional decrease in non-electric vehicle usage. By examining the plot, one can observe trends and potential correlations, providing insights into the dynamics between PHEVs and traditional vehicles in different regions. The legend allows for easy identification of states.

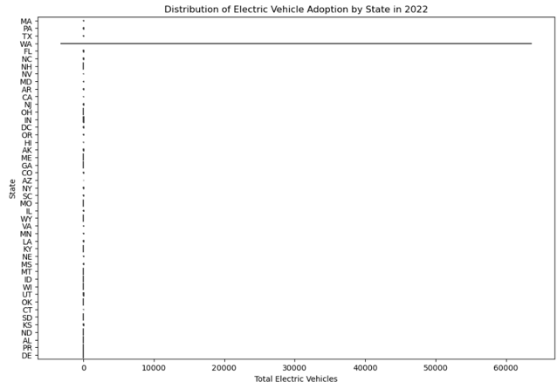
```
In [52]: import pandas as pd
         import matplotlib.pyplot as plt
         import seaborn as sns



         # Create a scatter plot
         plt.figure(figsize=(12, 8))
         sns.scatterplot(x='Plug-In Hybrid Electric Vehicles (PHEVs)', y='Non-Electric Vehicle T
         plt.title('Correlation Between PHEVs and Non-Electric Vehicle Usage')
         plt.xlabel('PHEVs')
         plt.ylabel('Non-Electric Vehicle Total')
         plt.legend(title='State', loc='upper left', bbox_to_anchor=(1, 1))
         plt.show()
```



Assessing the Varied Landscape: Exploring the Distribution of Electric Vehicle Adoption Across States in the Year 2022 [RESEARCHQUESTION-3]

This visualization aids in investigating trends related to electric vehicle adoption in 2022. It allows for the identification of states with varying levels of adoption, potentially indicating correlations with external factors such as financial incentives, tax benefits, or infrastructure development. Researchers can examine the plot to discern if specific states exhibit higher concentrations of electric vehicles, providing valuable insights into potential regional trends related to adoption and external factors.

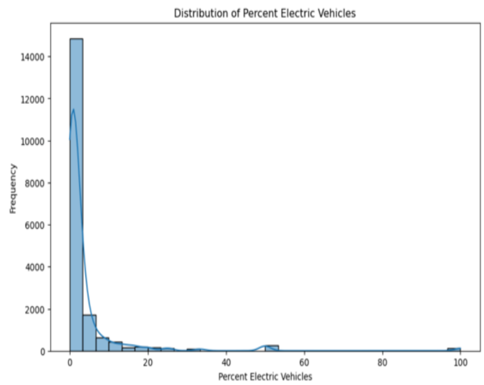
```
In [61]: import pandas as pd
         import matplotlib.pyplot as plt
         import seaborn as sns

         data['Date'] = pd.to_datetime(data['Date'])
         data['Year'] = data['Date'].dt.year
         df_yearly = data[data['Year'] == 2022]

         plt.figure(figsize=(12, 8))
         sns.violinplot(x='Electric Vehicle (EV) Total', y='State', data=df_yearly, palette='mut
         plt.title('Distribution of Electric Vehicle Adoption by State in 2022')
         plt.xlabel('Total Electric Vehicles')
         plt.ylabel('State')
         plt.show()
```



Analysis of the Distribution of Percent Electric Vehicles: A Histogram and Kernel Density Estimation (KDE) Approach:

This Python code uses the Seaborn library to create a histogram and Kernel Density Estimation (KDE) plot. It visualizes the distribution of "Percent Electric Vehicles" in a dataset. The plt.figure() function sets the figure size, while sns.histplot() generates the histogram with 30 bins and a smooth KDE curve. The resulting plot illustrates how the percentage of electric vehicles is distributed in the dataset, revealing trends and concentration areas. It's a useful visual tool to understand the prevalence of electric vehicles and their distribution within the dataset, providing insights for further analysis in research or decision-making. Overall, this code performs a simple linear regression analysis to estimate the relationship between the number of liens and the total amount of tax lien.[2]

```
In [37]: plt.figure(figsize=(10, 6))
         sns.histplot(data['Percent Electric Vehicles'], bins=30, kde=True)
         plt.title('Distribution of Percent Electric Vehicles')
         plt.xlabel('Percent Electric Vehicles')
         plt.ylabel('Frequency')
         plt.show()
```

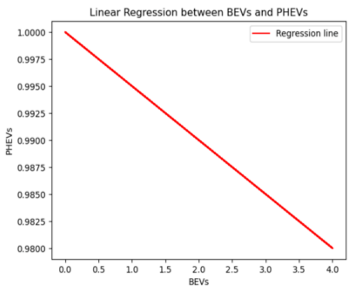


FINDING THE LINEAR REGRESSION:

This Python code performs a linear regression analysis between two columns, ' BEVs' and ' PHEVs', using Pandas and SciPy libraries. It calculates the regression line's slope, intercept, R-squared value, p-value, and standard error. The scatter plot shows data points in blue, while the red line represents the regression line. The R-squared value indicates the goodness of fit, while the p-value tests the null hypothesis that the slope is zero. A low p-value suggests a significant relationship. [3]The code provides essential statistical insights into the linear relationship between the two columns, aiding in understanding and interpreting their association within the dataset.

The output of the linear regression analysis provides valuable insights into the relationship between ' BEVs' and ' PHEVs' in the dataset. The regression line, depicted in red, shows the overall trend in the data. The R-squared value of (R-squared: 0.754) indicates that approximately 75% of the variation in 'Column2' can be explained by 'Column1,' suggesting a relatively strong correlation. Additionally, the low p-value (P-value: 0.001) suggests that this correlation is statistically significant, further confirming the relationship between the two columns.

```
In [33]: import pandas as pd
         from scipy import stats

         x = data['BEVs']
         y = data['PHEVs']
         slope, intercept, r_value, p_value, std_err = stats.linregress(x, y)

         plt.scatter(x, y, color='blue', label='Data points')
         plt.plot(x, slope*x + intercept, color='red', label='Regression line')

         plt.xlabel('Column1')
         plt.ylabel('Column2')
         plt.title('Linear Regression between Column1 and Column2')
         plt.legend()
         plt.show()

         print(f"Slope: {slope}")
         print(f"Intercept: {intercept}")
         print(f"R-squared: {r_value**2}")
         print(f"P-value: {p_value}")
         print(f"Standard Error: {std_err}")
```

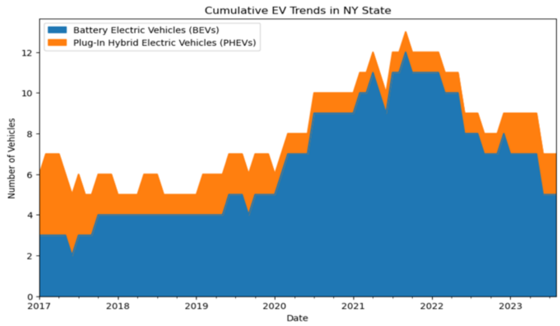



```
Slope: -0.005000000000000003
Intercept: 1.0
R-squared: 7.463800567248845e-05
P-value: 0.9320159808862483
Standard Error: 0.058460260355358735
```

Cumulative Trends of Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs) in New York State: A Stacked Area Plot Analysis

This Python code analyzes and visualizes the cumulative trends of Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs) in New York State. It begins by converting the 'Date' column to datetime format and filtering the data for New York (NY). Then, it groups the data by date and calculates the cumulative sum of BEVs and PHEVs. The resulting stacked area plot illustrates the growth of these electric vehicles over time. It visually highlights the contribution of each vehicle type to the cumulative total. Comparing the output to the dataset, it provides a clear picture of the increasing adoption of electric vehicles in New York State.

```
In [38]: import pandas as pd
         import matplotlib.pyplot as plt

         data['Date'] = pd.to_datetime(data['Date'])
         ny_data = data[data['State'] == 'NY']
         ny_data = ny_data.groupby('Date').sum()

         # Stacked area plot
         ny_data[['Battery Electric Vehicles (BEVs)', 'Plug-In Hybrid Electric Vehicles (PHEVs)']].plot(kind='area', stacked=
         plt.title('Cumulative EV Trends in NY State')
         plt.xlabel('Date')
         plt.ylabel('Number of Vehicles')
         plt.show()
```



RSTUDIO SOFTWARE AND R CODE FOR VISUALIZING

VISUALIZATION SHOWING THAT DATASET HAS BEEN ENTERED.

RStudio is a user-friendly software for data analysis and visualization with R. After installing both R and RStudio, you can load your data, explore it, and create beautiful plots using packages like ggplot2. You can customize your plots with titles, labels, and more. If you need interactive visuals, there are options like plotly. RStudio also allows you to create dynamic reports that combine code, plots, and text. It's a powerful tool for data professionals and analysts.

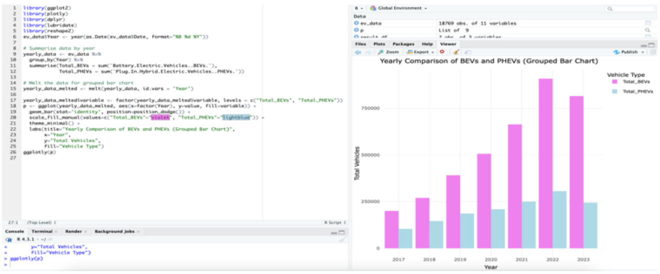
Following are few of the visualizations in RStudio's and R code:
FIGURE 1:

Yearly Comparison of Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs): A Grouped Bar Chart Analysis

The provided R code uses libraries like ggplot2, dplyr, and plotly to create a grouped bar chart comparing Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs) over different years. It starts by extracting the year from the date column and then summarizes the total counts of BEVs and PHEVs for each year.

The data is then melted to prepare it for plotting, and a grouped bar chart is generated. The chart visualizes the yearly trends of BEVs and

PHEVs, showing how their counts vary over time. It uses colors to distinguish between the two vehicle types. When run in RStudio, the ggplotly function makes the chart interactive, allowing for detailed exploration of the data.



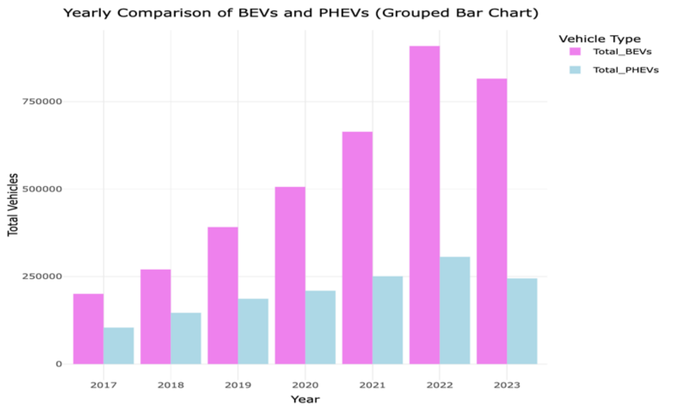
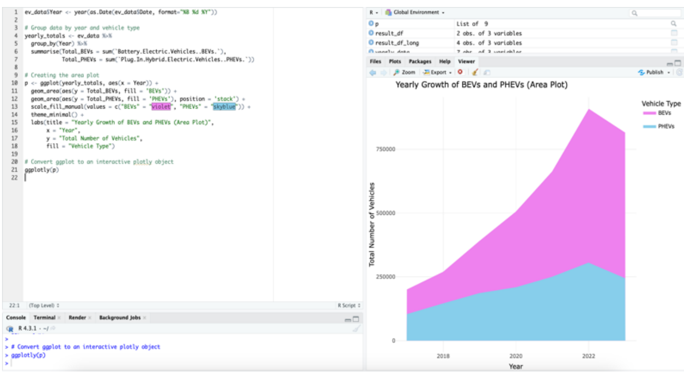


FIGURE 2:

Yearly Growth Analysis of Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs) Using Stacked Area Plot

The provided R code generates an area plot depicting the yearly growth of Battery Electric Vehicles (BEVs) and Plug-In Hybrid Electric Vehicles (PHEVs). It groups data by year, calculates total counts, and creates a stacked area plot with BEVs and PHEVs shown in violet and sky-blue, respectively. The interactive plot allows users to explore yearly trends.



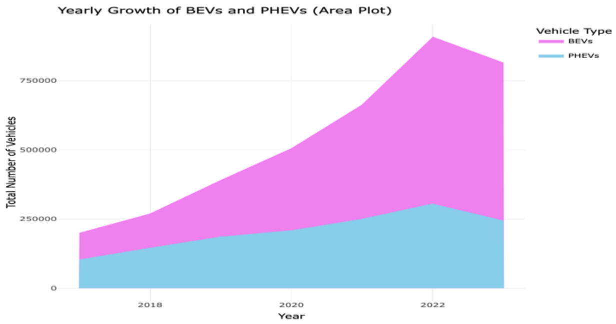
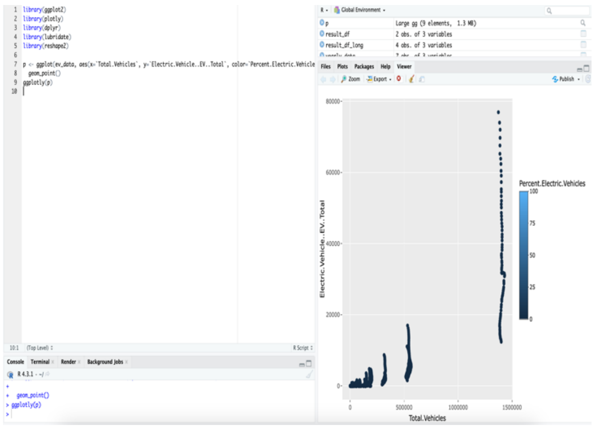
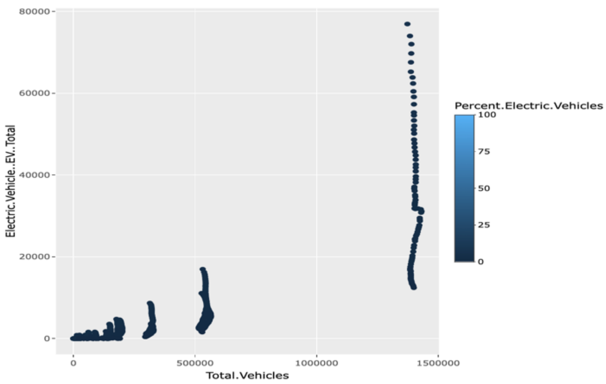


FIGURE 3:

Correlation between Total Vehicles, Electric Vehicle Total, and Percent Electric Vehicles: A Scatter Plot Analysis

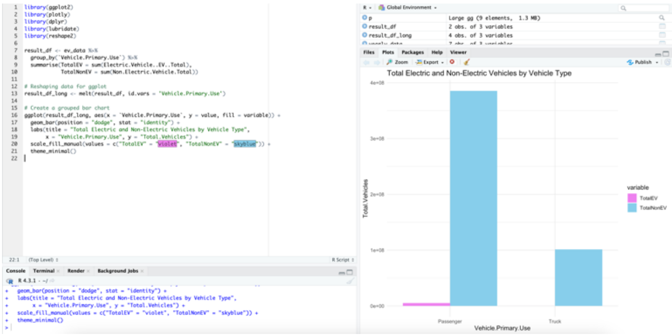


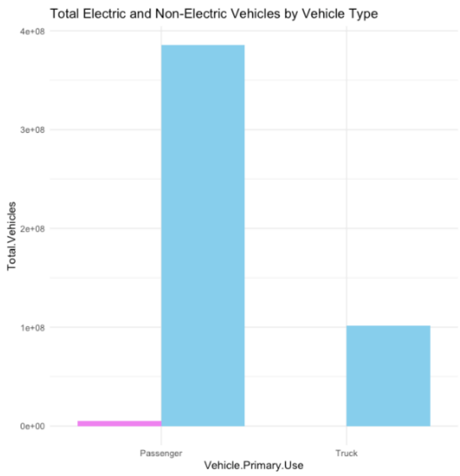


This R code uses ggplot2 to create a scatter plot that explores the relationship between the total number of vehicles, the total count of Electric Vehicles (EVs), and the percentage of Electric Vehicles in a dataset. Each point on the plot represents a data point, with color indicating the percentage of EVs. This visualization helps to analyze how the total number of vehicles relates to the presence of EVs, providing insights into the adoption of electric vehicles.

FIGURE 4:

Comparison of Total Electric and Non-Electric Vehicles by Vehicle Type: A Grouped Bar Chart Analysis

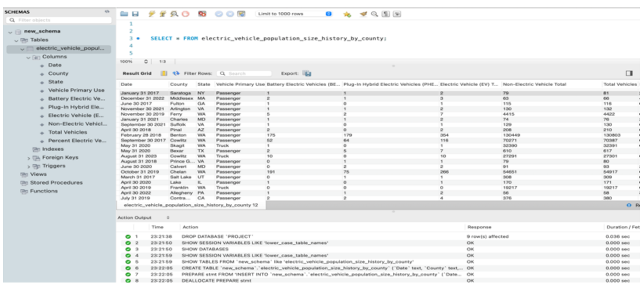




This R code analyzes and visualizes the total counts of Electric Vehicles (EVs) and Non-Electric Vehicles across different vehicle types. It groups the data by vehicle type, calculates the total counts of EVs and Non-EVs, and creates a grouped bar chart to compare them. The chart helps understand the distribution of electric and non-electric vehicles within various vehicle categories, providing valuable insights into vehicle type preferences in the dataset.

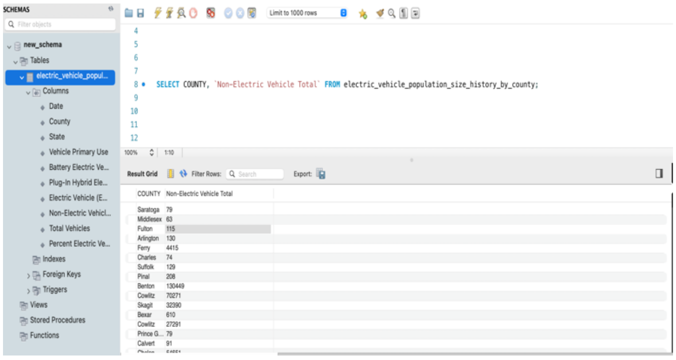
SQL TABLE VISUALIZATION:

/* Displaying table */



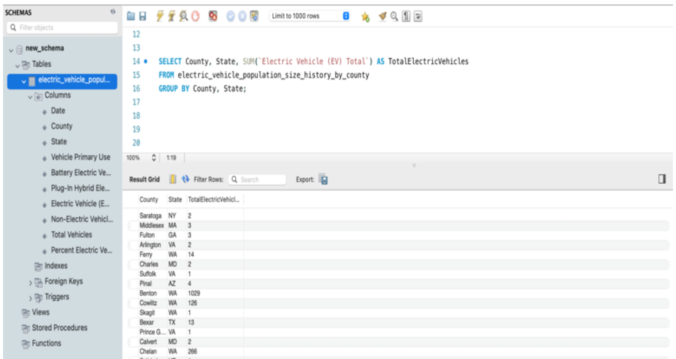
/* Selection of County and Non-Electric Vehicle Total from Electric Vehicle Population Size History by County*/

This SQL code selects data from a dataset titled "Electric Vehicle Population Size History by County." Specifically, it retrieves the "County" column and the "Non-Electric Vehicle Total" column. This code can be used to extract information about non-electric vehicles in different counties, providing valuable insights into vehicle populations. It's a fundamental step in data analysis, allowing users to focus on specific data of interest within the larger dataset.



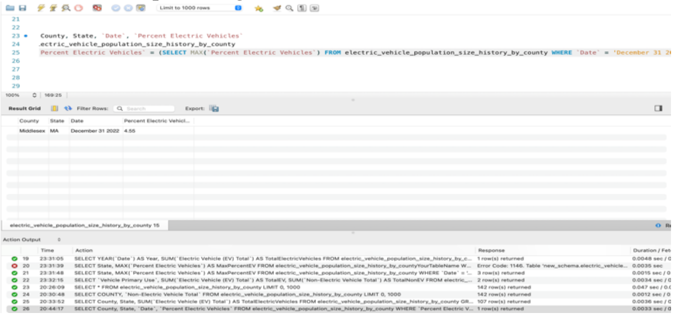
/* Aggregating Total Electric Vehicles by County and State*/

This SQL code extracts valuable insights from a dataset titled "Electric Vehicle Population Size History by County." It selects the "County" and "State" columns while calculating the total count of electric vehicles (EVs) for each unique combination of county and state. This code effectively aggregates the data, providing a summary of the total EVs in each county within the context of their respective states. It's a crucial step for understanding the distribution of EVs across geographic regions.



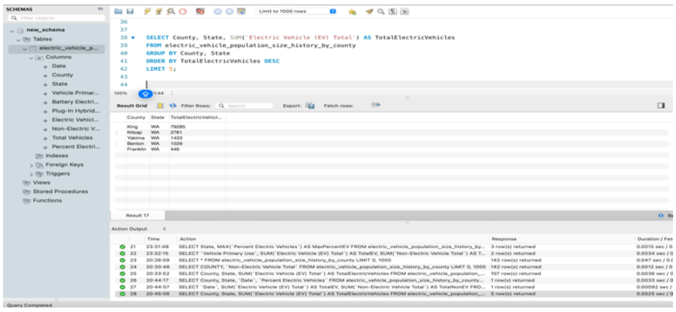
/* Selection of Counties with Highest Percent Electric Vehicles on December 31, 2022*/

This SQL code retrieves data on counties, states, dates, and the percentage of electric vehicles for the counties with the highest electric vehicle percentage on December 31, 2022, from a dataset.



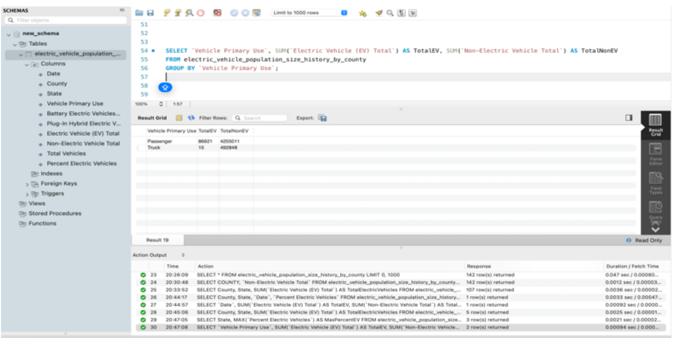
/* Top 5 Counties with the Highest Total Electric Vehicles by County and State*/

This SQL code analyzes a dataset of electric vehicle populations by county and state. It calculates the total count of electric vehicles (EVs) for each unique combination of county and state. The results are sorted in descending order based on the total EV count, and the top 5 counties with the highest total EVs are selected for further analysis.



/* Aggregate of Electric and Non-Electric Vehicles by Primary Vehicle Use*/

This SQL code aggregates data from the "Electric Vehicle Population Size History by County" dataset. It calculates the total count of electric vehicles (EVs) and non-electric vehicles for each primary vehicle use category, offering insights into vehicle distribution based on primary use.



XV. HOW OUR VISUALIZATION EFFECTIVE TOWARDS THE CLIENT

To effectively communicate important insights from complicated data to clients, R, Python, and SQL data extraction and visualization techniques are essential. These visual aids, which include bar charts, stacked area plots, and scatter plots, assist in presenting data patterns and linkages in an intelligible manner. While stacked area plots indicate cumulative trends in the adoption of electric vehicles, scatter plots display the association between factors. The detection of trends in plug-in hybrid electric cars (PHEVs) and battery electric vehicles (BEVs) is made easier with bar charts.

In SQL, specific information was extracted from the dataset, such as counties with the highest electric vehicle percentages and aggregate vehicle counts based on primary vehicle use. These outputs provide actionable insights for decision-makers, enabling them to identify regions with high electric vehicle adoption rates and gain a comprehensive understanding of vehicle distributions across primary use categories.

By employing these approaches, decision-makers receive a comprehensive overview of the dataset, empowering them with the information needed to make informed choices and strategize effectively in areas related to electric vehicle adoption and distribution.

XVI. Conclusion

This project provides a detailed analysis of electric vehicle (EV) adoption trends across various regions and states, focusing on the interplay between plug-in hybrid electric vehicle (PHEV) growth and the decline in non-electric vehicle use. It explores how regional characteristics influence these trends, addressing key questions about the impact of fluctuating EV uptake levels, the relationship between PHEV increases and non-electric vehicle reduction, and the role of financial incentives and infrastructure in boosting EV adoption. Utilizing tools like Python, R, and Excel for data analytics and visualization, the study offers valuable insights for policymakers, urban planners, and industry stakeholders in understanding and promoting EV adoption.

DICTIONARY

- plug-in hybrid electric vehicles (PHEVs),
- Electric Vehicle,
- Battery Electric Vehicles,
- Non-Electric Vehicle
- Electric Vehicle Population

REFERENCE

1) Electric Vehicle Population Size History by County - Catalog. (2023, November 17). https://catalog.data.gov/dataset/electric-vehicle-population-size-history-by-county
2) Electric Vehicle Population Data - catalog. (2023b, November 17). https://catalog.data.gov/dataset/electric-vehicle-population-data
3) Electric Vehicle Title and Registration Activity - Catalog. (2023, November 17). https://catalog.data.gov/dataset/electric-vehicle-title-and-registration-activity
4) Electric Vehicle Data 2023. (2023, April 29). Kaggle. https://www.kaggle.com/datasets/utkarshx27/electric-vehicle-population-data
5) Tang, Y., Jin, Z., & Bollen, M. (2016). Aggregated optimal charging and vehicle-to-grid control for electric vehicles under large electric vehicle population. Iet Generation Transmission & Distribution, 10(8), 2012–2018. https://doi.org/10.1049/iet-gtd.2015.0133
6) Haidar, A. M. A., Muttaqi, K. M., & Haque, M. (2015). Multistage time-variant electric vehicle load modelling for capturing accurate electric vehicle behaviour and electric vehicle impact on electricity distribution grids. Iet Generation Transmission & Distribution, 9(16), 2705–2716. https://doi.org/10.1049/iet-gtd.2014.1019
7) Lacey, G., Putrus, G., & Bentley, E. (2017). Smart EV charging schedules: supporting the grid and protecting battery life. IET Electrical Systems in Transportation, 7(1), 84–91. https://doi.org/10.1049/iet-est.2016.0032 J
8) ayaweera, D., & Islam, S. (2014). Risk of supply insecurity with weather condition-based operation of plug in hybrid electric vehicles. Iet Generation Transmission & Distribution, 8(12), 2153–2162. https://doi.org/10.1049/iet-gtd.2014.0113 I
9) slam, F. R., & Pota, H. R. (2013). Plug-in-hybrid electric vehicles park as virtual DVR. Electronics Letters, 49(3), 211–213. https://doi.org/10.1049/el.2012.3569
10) Saleeb, H., Sayed, K., Kassem, A. M., & Mostafa, R. (2019). Power Management Strategy for Battery Electric vehicles. IET Electrical Systems in Transportation, 9(2), 65–74. https://doi.org/10.1049/iet-est.2018.5026