Electric Vehicle Population Data

Niharika Bandlamudi: - hbz194

Background:

Electric Vehicle is a vehicle that uses one or more electric motors for propulsion. It can be powered by a collector system, with electricity from extravehicular sources, or it can be powered autonomously by a battery (sometimes charged by solar panels, or by converting fuel to electricity using fuel cells or a generator) EVs include, but are not limited to, road and rail vehicles, surface and underwater vessels, electric aircraft , and electric spacecraft. For road vehicles, together with other emerging automotive technologies such as autonomous driving, connected vehicles, and shared mobility, EVs form a future mobility vision called Connected, Autonomous, Shared, and Electric (CASE) Mobility.

Motivation:

The analysis of electric vehicles in Washington state aims to understand and promote the adoption of alternative fuel vehicles, especially electric ones, due to their significant environmental benefits, such as reduced greenhouse gas emissions and decreased reliance on fossil fuels. By examining the factors influencing Clean Alternative Fuel Vehicle (CAFV) eligibility, policymakers and stakeholders can develop strategies to incentivize electric vehicle adoption and expedite the transition to a more sustainable transportation system.

Furthermore, analyzing this data can provide insights into consumer preferences and market trends related to electric vehicles. Understanding the factors driving CAFV eligibility can help manufacturers tailor their offerings to meet consumer demands, and policymakers can design effective incentive programs to encourage electric vehicle adoption. Overall, the analysis of this data is driven by the goal of advancing sustainable transportation solutions and reducing the environmental impact of traditional fossil fuel vehicles.

The analysis will primarily focus on:

What is the average electric range of vehicles by make and how does it vary among the top manufacturers?

To determine the average electric range of vehicles by make and how it varies among the top manufacturers, I would calculate the mean electric range for each make and then compare these averages among the top manufacturers. This analysis would provide insights into the typical electric range offered by different manufacturers and how they compare to each other in terms of electric vehicle range.

What are the Top 10 counts of cars per City and per County?

To this question, I would use a combination of grouping, summarizing, and sorting functions. This analysis provides us with the top 10 cities and top 10 counties with the highest number of cars, allowing us to identify the cities and counties with the highest car counts.

In which Year Electric Vehicles increased?

To determine the year in which electric vehicle (EV) adoption increased, I will conduct a time series analysis. This analysis will allow us to track the adoption rates of EVs over time and identify the specific year(s) when there was a notable increase in adoption. I will use a line plot or a similar visualization to display the trend in EV adoption rates over the years.

What are the Top 5 vs Bottom 5 Comparison?

To compare the top 5 and bottom 5 EV models based on adoption rates, I will use a combination of frequency analysis and visualization. First, I need to calculate the frequency of each EV model in the dataset to identify the top 5 and bottom 5 models with the highest and lowest adoption rates, then I need to create a bar chart or similar visualization to compare the adoption rates of these models.

Are there demographic patterns that influence EV adoption?

To analyze demographic patterns influencing EV adoption, I will use chi-square tests to determine if there are significant associations between demographic variables and EV adoption. Regression analysis can also help to quantify the impact of demographic factors on EV adoption rates, providing insights into the target demographic for EV marketing and policymaking.

What is the distribution of EV types (eg: batter electric vehicle, plug-in hybrid electric vehicle) in the dataset and how did it evolve overtime?

The distribution of electric vehicle (EV) types can be analyzed by calculating the count or percentage of each EV type for each year. This analysis can provide insights into how the popularity of different EV types has evolved over time. Visualizing this data can further understand the trends and patterns in EV adoption over the years in Washington state.

Is there any relationship between range anxiety and EV adoption?

To analyze the relationship between range anxiety and EV adoption, I will use regression analysis. Specifically, I can regress EV adoption rates on electric range while controlling for other relevant factors such as year, make, and model. This analysis will help us to quantify the impact of electric range on adoption rates, providing insights into how range anxiety influences EV adoption.

Are certain EV models preferred over others by consumers?

To analyze consumer preferences for EV models based on the provided data, I will perform a time series analysis to track the adoption trends of different EV models over time, providing insights into any shifts in consumer preferences.

<u>Description of the Data:</u>

The initial dataset is comprised of 135038 observations with 17 variables describing the data. The data comes from www.kaggle.com and covers cities in Washington state. This dataset spans the time frame of 1997 to 2024-feb, a 27-year period. I am thinking about cutting the start year of my data from 1997 to potentially 2000, however, this is something I will investigate more to determine if this is a well-represented and efficient point to subset at.

These variables include both categorical (e.g., county, city, state, make, model, electric vehicle type, eligibility status, electric utility) and numerical (e.g., postal code, model year, electric range, MSRP, legislative district, DOL vehicle ID, census tract) data, providing a comprehensive overview.

As for the cleanliness of the data, the data is fairly clean and does not have too many missing values. The dataset aims to capture the growing adoption of electric vehicles and includes various details about these vehicles. This data set demonstrates the Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) that are currently registered through Washington State Department of Licensing (DOL). The attributes of every completely electric passenger car are listed in this dataset.

The wide temporal scope (1997 to 2024) could pose challenges in analyzing trends over time, considering the evolving nature of electric vehicle technology and regulations. Additionally, the dataset's focus on cities in Washington state may limit its generalizability to other regions. Lastly, the inclusion of the electric utility provider variable may require careful consideration depending on the analysis goals.

<u>Proposed Analysis:</u>

The analysis of the data will comprise of logistic and linear regressions, Anova, time series analysis, two sample t-tests and visual representations of the data analysis such as histograms to help interpret the information. Using logistic regression, I will try to explore the relationship between several factors and the likelihood of a vehicle being eligible for Clean Alternative Fuel Vehicle (CAFV) incentives. The use of this type of model can help us understand the factors that influence the eligibility of electric vehicles for Clean Alternative Fuel Vehicle (CAFV) incentives in Washington state. By building a model that predicts CAFV eligibility based on variables such as electric range, manufacturer, and model year, I can identify the key factors that drive eligibility. This analysis can provide insights into the characteristics of electric vehicles that are more likely to be eligible for incentives, which can inform policy decisions aimed at promoting the adoption of electric vehicles. Additionally, the logistic regression model can help evaluate the effectiveness of current incentive programs and guide the development of future incentive strategies.

The first step in the analysis is data preprocessing, which includes checking for missing values and encoding categorical variables into numerical format. Exploratory data analysis (EDA) is then conducted to understand the distribution of variables and identify patterns and relationships between them. Next, relevant features that could influence CAFV eligibility are selected based on domain knowledge and EDA results. A logistic regression model is built using these features to predict CAFV eligibility. The model is evaluated using metrics such as accuracy, precision, recall, and F1-score to assess its performance.

The coefficients of the logistic regression model are interpreted to understand the impact of each feature on CAFV eligibility and identify the most key features driving eligibility. The model is validated using cross-validation to ensure its generalizability, and regularization techniques are applied to optimize the model and prevent overfitting.

Finally, the trained logistic regression model is used to predict the CAFV eligibility of new electric vehicles. The results are analyzed to gain insights into the factors influencing CAFV eligibility and provide recommendations for policymakers and stakeholders to promote the adoption of electric vehicles and optimize CAFV incentive programs.

References:

Electric Vehicle population data: Electric Vehicle Population Data (kaggle.com)

Appendix – Variable Descriptions:

Raw data: Electric Vehicle Population Data | Data.WA | State of Washington

Includes observations from 1997 to February 2024.

VIN (Vehicle Identification Number): A unique identifier for each vehicle.

County: The county where the vehicle is registered.

City: The city where the vehicle is registered.

State: The state where the vehicle is registered (in this case, WA for Washington).

Postal Code: The postal code of the vehicle's registration location.

Model Year: The year the vehicle was manufactured.

Make: The manufacturer of the vehicle (e.g., BMW, TESLA, CHEVROLET).

Model: The model name of the vehicle (e.g., X5, MODEL 3, VOLT, PACIFICA).

Electric Vehicle Type: The type of electric vehicle (e.g., Plug-in Hybrid Electric Vehicle (PHEV), Battery Electric Vehicle (BEV)).

Clean Alternative Fuel Vehicle (CAFV) Eligibility: The eligibility status for Clean Alternative Fuel Vehicle incentives.

Electric Range: The electric range of the vehicle in miles.

Base MSRP: The Manufacturer's Suggested Retail Price of the vehicle.

Legislative District: The legislative district associated with the vehicle's registration.

DOL Vehicle ID: The Department of Licensing Vehicle ID for the vehicle.

Vehicle Location: The geographical location of the vehicle in latitude and longitude coordinates.

Electric Utility: The electric utility provider for the vehicle.

2020 Census Tract: The 2020 Census Tract associated with the vehicle's registration location.

These attributes provide detailed information about each Electric Vehicle in the dataset, including its characteristics, location, and registration details.