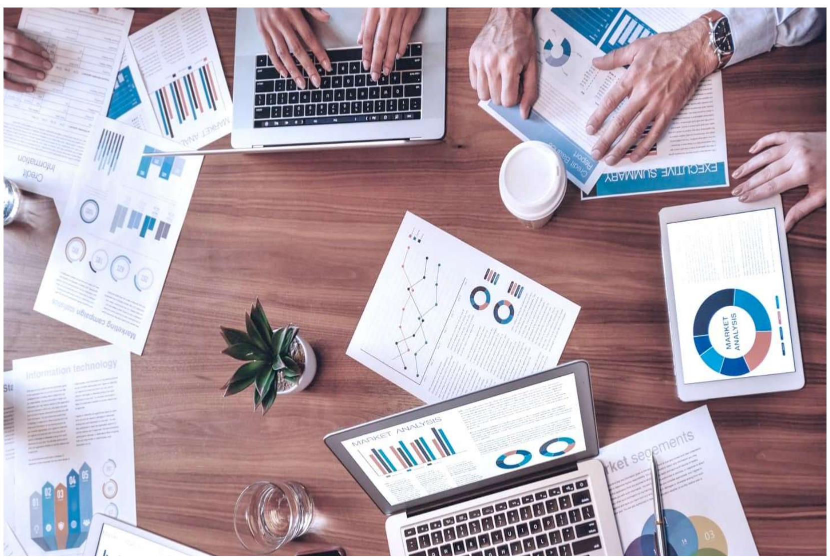
PROJECT -1

ELECTRIC VEHICLES DATA INSIGHT

IE6600 Computation and Visualization for Analytics

Final Report



Group Number 3:

Arya Lokesh Gowda (002249418)
Lokhi Nalam (002649847)
Pooja Arumugam (002872003)
Sathvik Ramappa (002847460)

Part 1: Introduction

With the advent of cutting-edge technology and the previously unheard-of level of data availability, the field of electric vehicle data insight has entered a new era. This change has created a rare chance to investigate and create cutting-edge approaches for detecting the trends in vehicle technology by utilizing the capabilities of deep learning and machine learning techniques. This project explores current trends in the field and clarifies how electric vehicle prediction and prevention are being redefined by this cutting-edge technology.

As a branch of artificial intelligence, machine learning is a potent toolkit of statistical models and algorithms for data analysis and prediction-making. Simultaneously, deep learning, a different branch of machine learning, uses multiple-layer artificial neural networks to simulate intricate correlations between inputs and outputs. The science of the type of vehicles to be used, prediction is about to undergo a paradigm revolution because of the combination of machine learning and deep learning. Algorithms for machine learning have become indispensable resources to predict the better type of electric vehicles to be used. A few algorithms, including support vector machines, random forests, and decision trees, have been used to analyze data in an effort to predict patterns remarkably accurately. These algorithms offer deeper insights into the changing trends and patterns of the activity, beyond just their prediction ability. The allocation of resources and tactics that can successfully combat the trend is made easier by this understanding. Moreover, machine learning algorithms are particularly good at identifying complex relationships between a variety of environmental and demographic factors. Among the variables that are carefully considered are the model, electric vehicle type, and the electric range. The data obtained from this study serves as the basis for the creation of specially designed vehicle types, prediction and prevention plans that are well matched to the needs of communities. In conclusion, this study article takes a tour around the modern terrain of electric vehicle data. It emphasizes how important machine learning and deep learning are to improving our understanding of the trends and strengthening our group's efforts to deter activities more successfully.

Part 2: Summary of Results

- The study highlights the utilization of T-Test to analyze electric vehicle data. Various algorithms have been employed to predict trends accurately, providing insights into changing patterns and aiding in resource allocation and strategy development.
- The dataset includes information on Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs), with a focus on Clean Alternative Fuel Vehicle (CAFV) eligibility. It reveals a dominance of BEVs and provides insights into their distribution and eligibility status.
- Correlation analysis and statistical tests reveal relationships between variables. Negative correlations are observed between electric range and model year, as well as between Base MSRP and Model Year. A t-test indicates no significant difference in electric range between BEVs and PHEVs.

  Various visualizations, including histograms, box plots, bar plots, pie charts, and line plots, help understand the distribution, trends, and patterns in the data. Insights into electric range over model years, normality checks through QQ plots, and comparisons of vehicle types are provided.

Part 3: Data Sources

An all-electric vehicle that stores its electrical energy in one or more batteries—which are charged by connecting the vehicle to an electric power source—is known as a battery electric vehicle, or BEV. An automobile that runs an electric motor using one or more batteries, powers an internal combustion engine or other propulsion source using gasoline or diesel, and recharges by plugging the car into an electric power source is known as a plug-in hybrid electric vehicle, or PHEV.

To qualify for alternative fuel vehicle retail sales and Washington State use tax exemptions, a vehicle must meet the fuel and electric-only range requirements specified in RCW 82.08.809 and RCW 82.12.809. This is known as a Clean Alternative Fuel Vehicle (CAFV) Eligibility. Retail sales of alternative fuel vehicles and Washington State use tax exemptions are contingent upon the sale or lease of these vehicles occurring on or after August 1, 2019, and meeting the purchase price requirements.

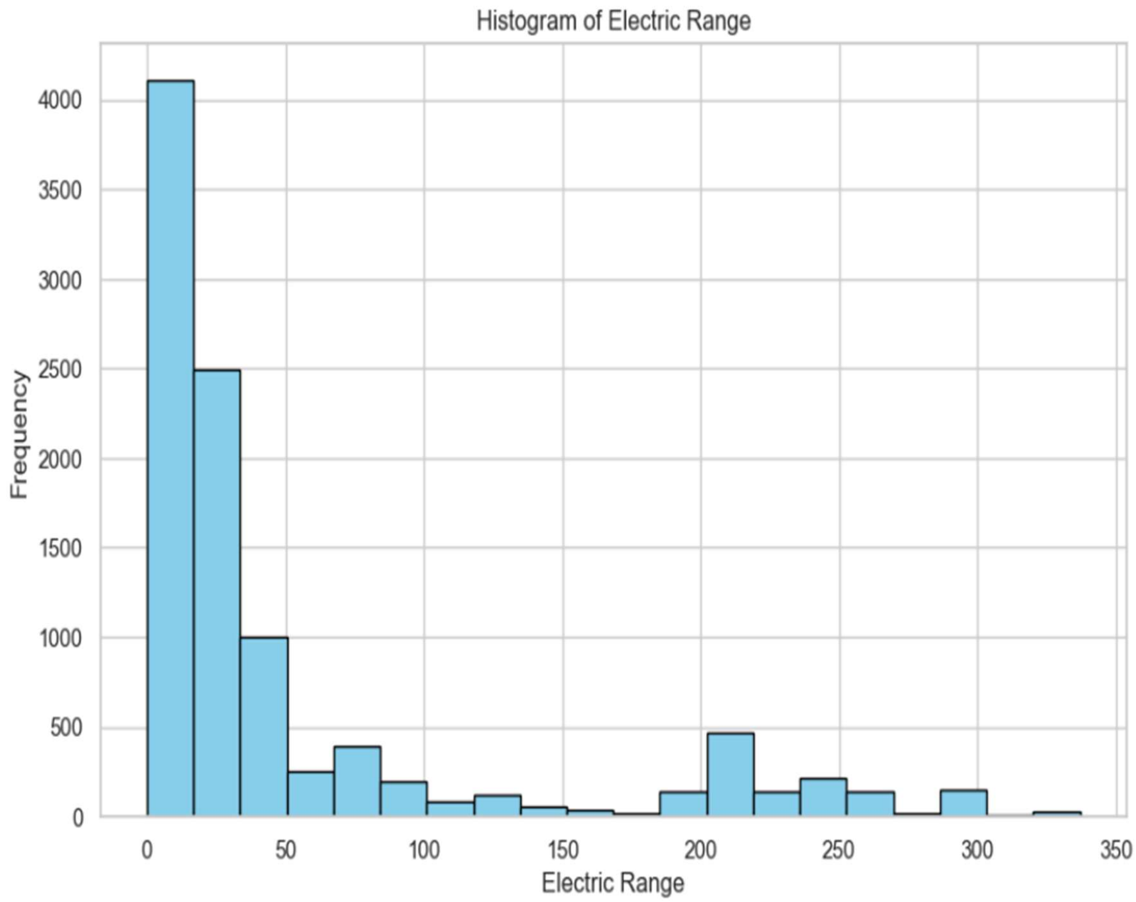
There may be variations in a county's monthly vehicle count between this report and previous ones. At the time of registration, procedures were put in place to assign counties more precisely.

Battery Electric Vehicles (BEVs) are no longer maintained for electric range because the latest models have an electric range of at least thirty miles. Where there is no research on the electric range, zero (0) will be entered.

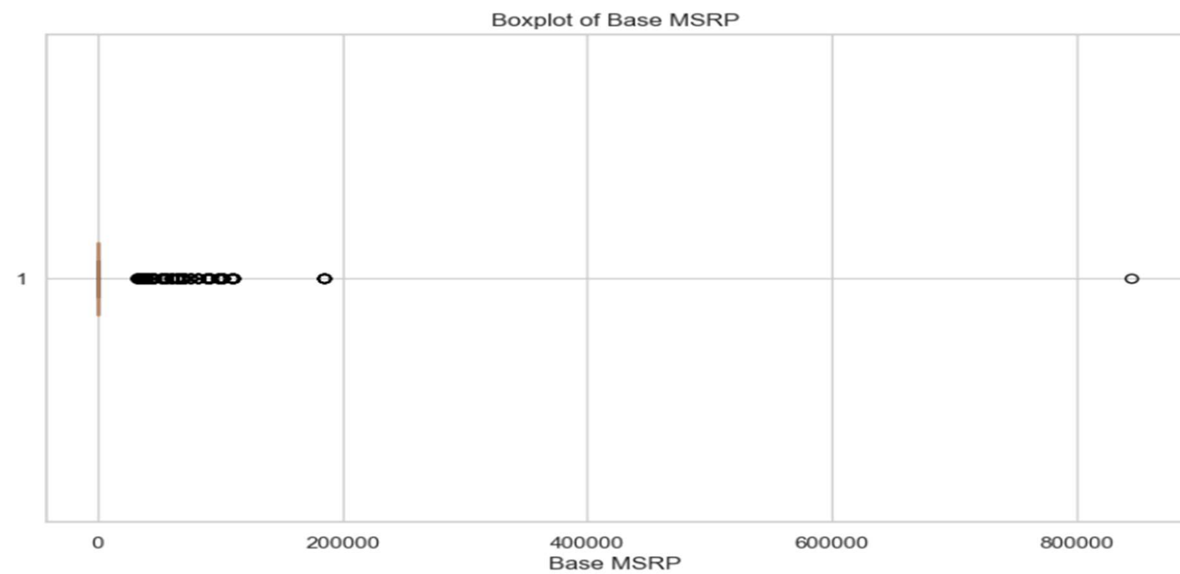
Data consists of Plug-in hybrid electric vehicles and battery electric vehicles that were registered as of December 31, 2023.

Part 4: Results and Methods

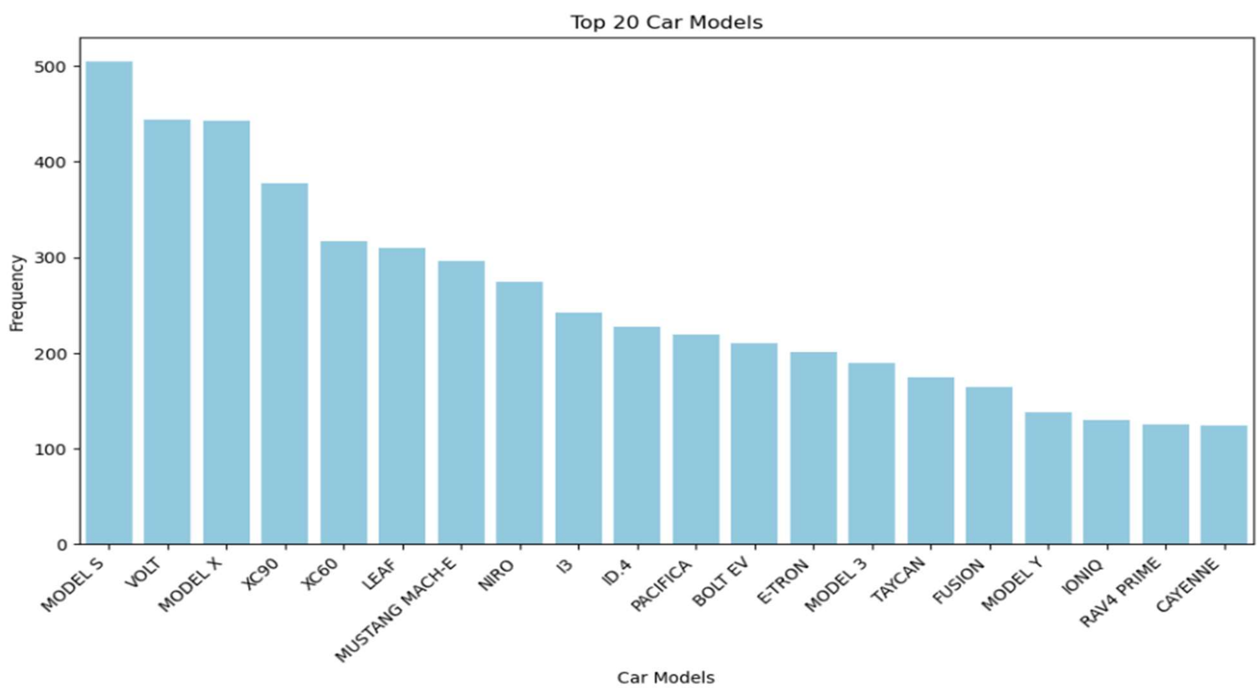
The database was collected from the referenced source, and the built-in pandas and numpy methods were used to explain its analytics. Missing values were addressed by considering a number of parameters throughout the data cleaning process. Duplicate rows were eliminated during the data cleaning procedure.

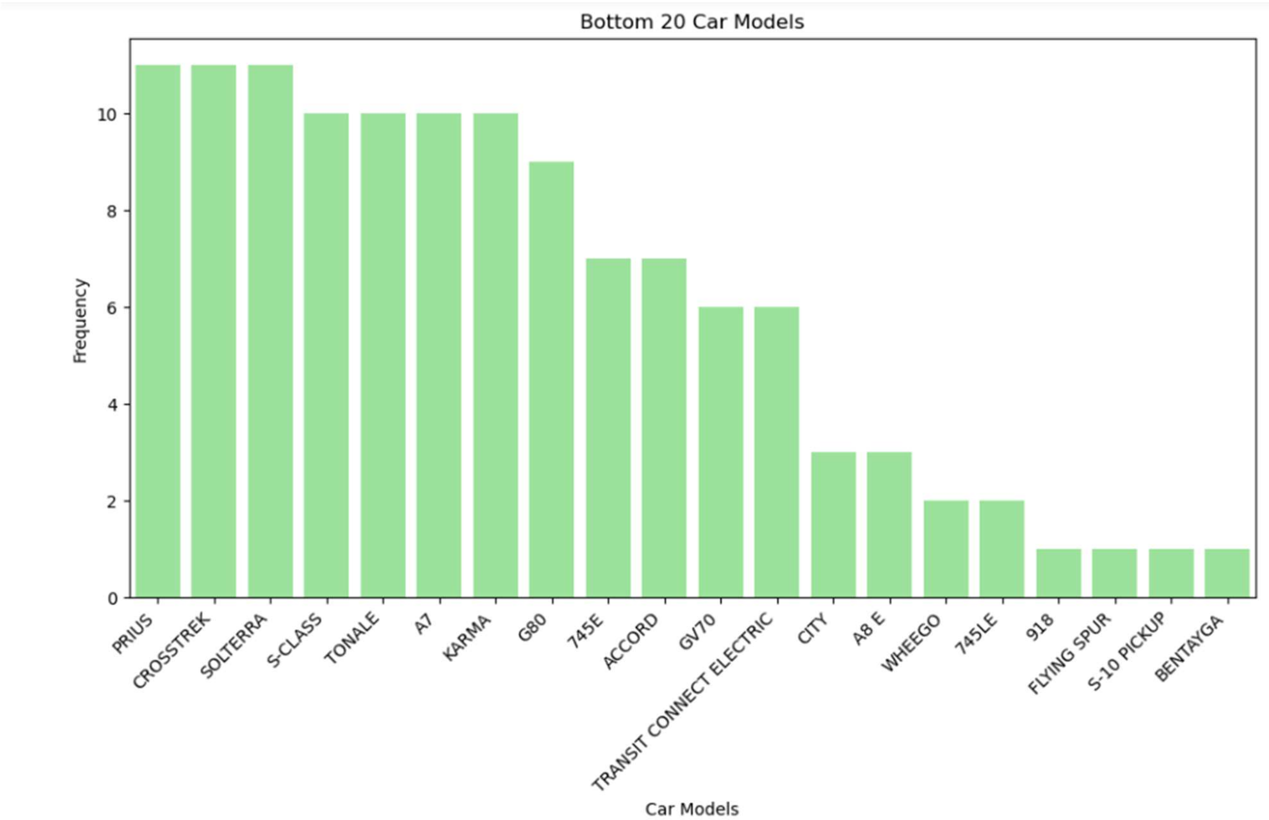


The dataset's electric range value distribution is represented by this histogram. The range of electric values is shown on the x-axis, which is separated into 20 bins. Within each bin, the frequency or count of data points is shown on the y-axis. The distribution of electric range values throughout the dataset is made easier to understand by this visualization.

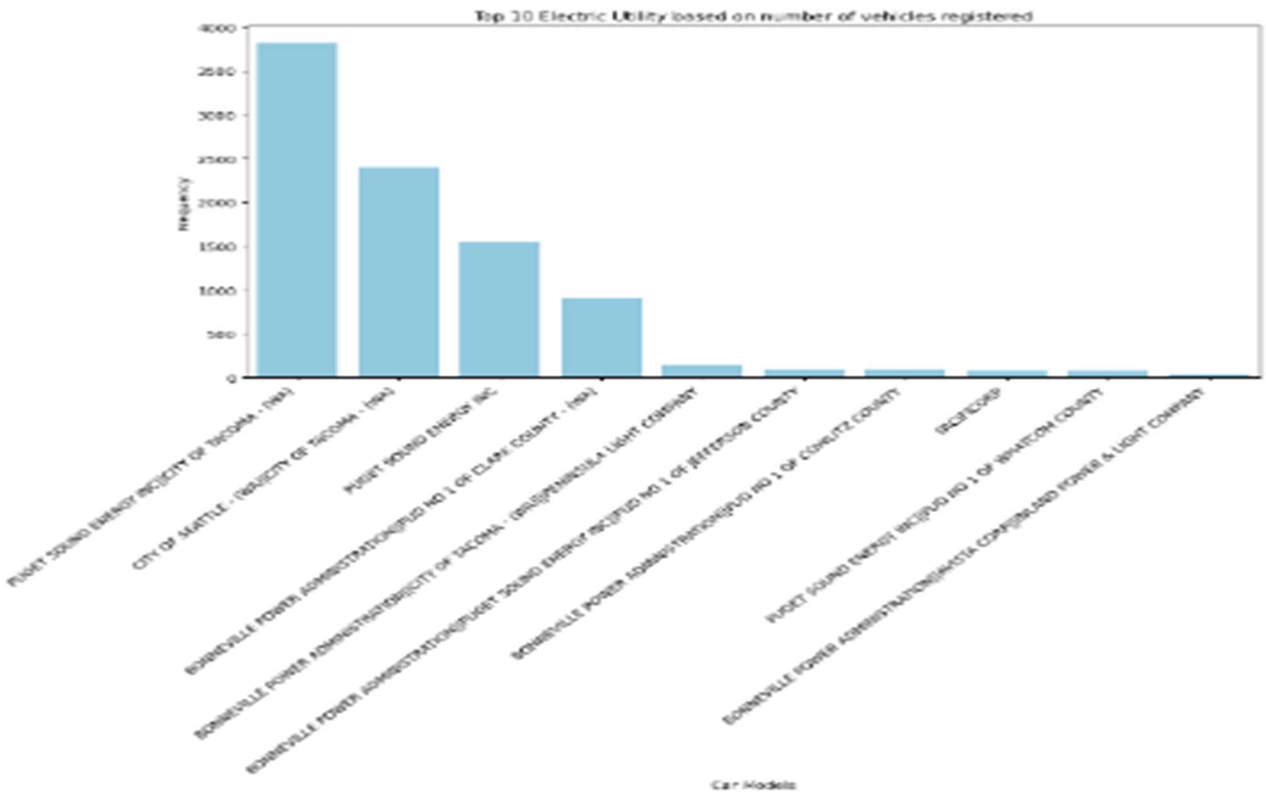


The pattern of distribution of the "Electric Range" and "Base MSRP" columns in the dataset is revealed by these visualizations. The boxplot displays the distribution of base MSRP values, including the median, quartiles, and any possible outliers, while the histogram displays the frequency distribution of electric range values.





- There are over 120 unique car models registered in Washington State Department of Licensing.
- MODEL S has the highest population of 505, followed by VOLT and MODEL X with 444 and XC90 with 378.
- There are over 25 car models that have a population lower than 20 with 918, FLYING SPUR, S-10 PICKUP and BENTAYGA with the population of 1.



- There are 54 Electric Power Retail Service territories.
- The PUGET SOUND ENERGY INC dominates the population with more than 5000 out of 9500 vehicles registered under it.

STATISTICAL ANALYSIS:

Important details regarding the connections between the various numerical variables in the dataset are provided by the correlation matrix's output.

In the matrix, the correlation coefficient between two variables is represented by each cell.

The linear relationship between two variables' strength and direction are measured by the correlation coefficient. It has a range of -1 to 1, where:

- When one variable rises, the other rises proportionately, indicating a perfect positive correlation of 1.
- When one variable rises, the other falls proportionately, indicating a perfect negative correlation,
  of -1.

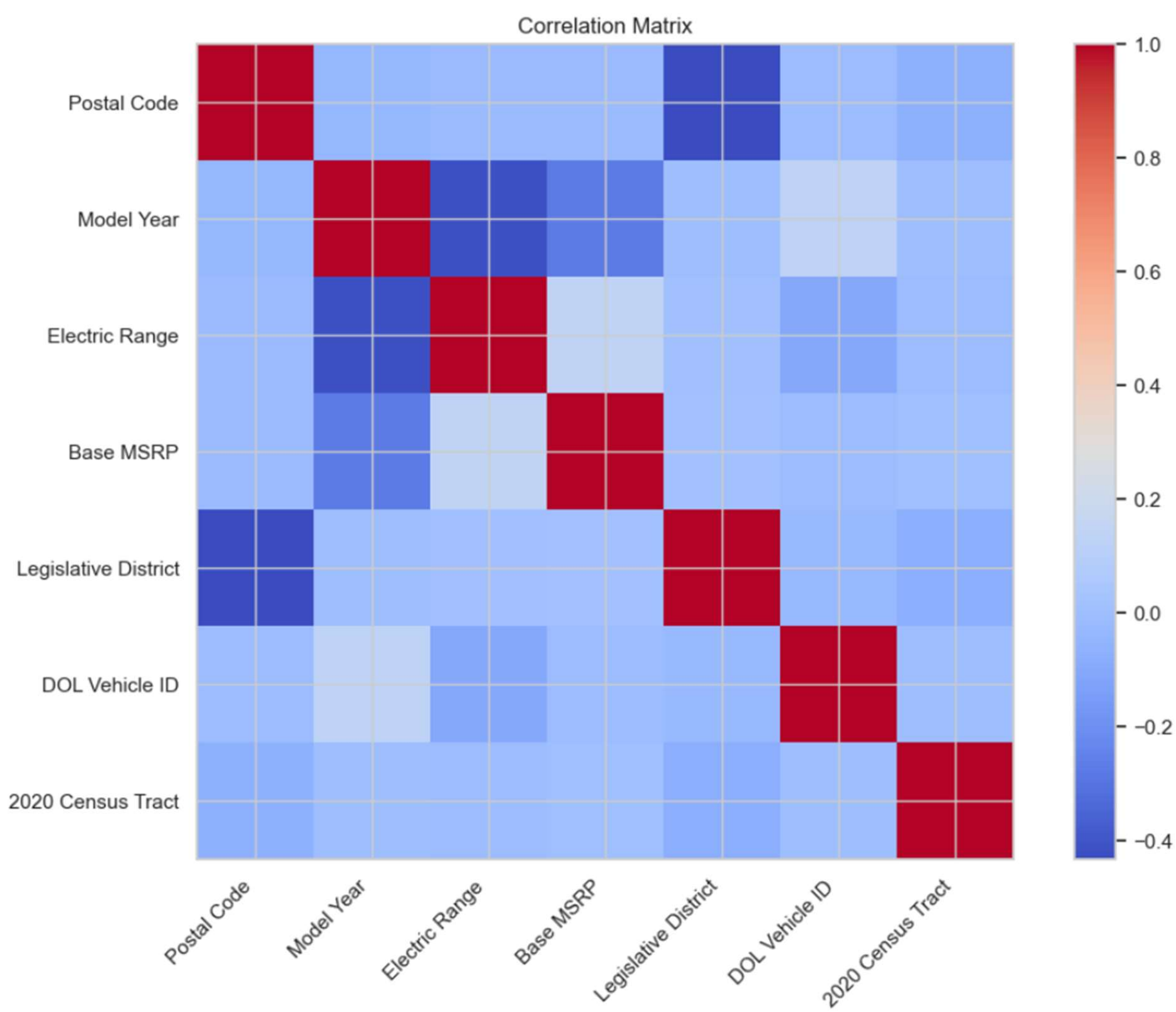
When the correlation coefficient is 0, it means that there is no systematic relationship between the variables.

The correlation coefficients of each variable with itself are found on the diagonal of the matrix, and they are always 1 (perfect correlation).

Interpreting the heatmap:

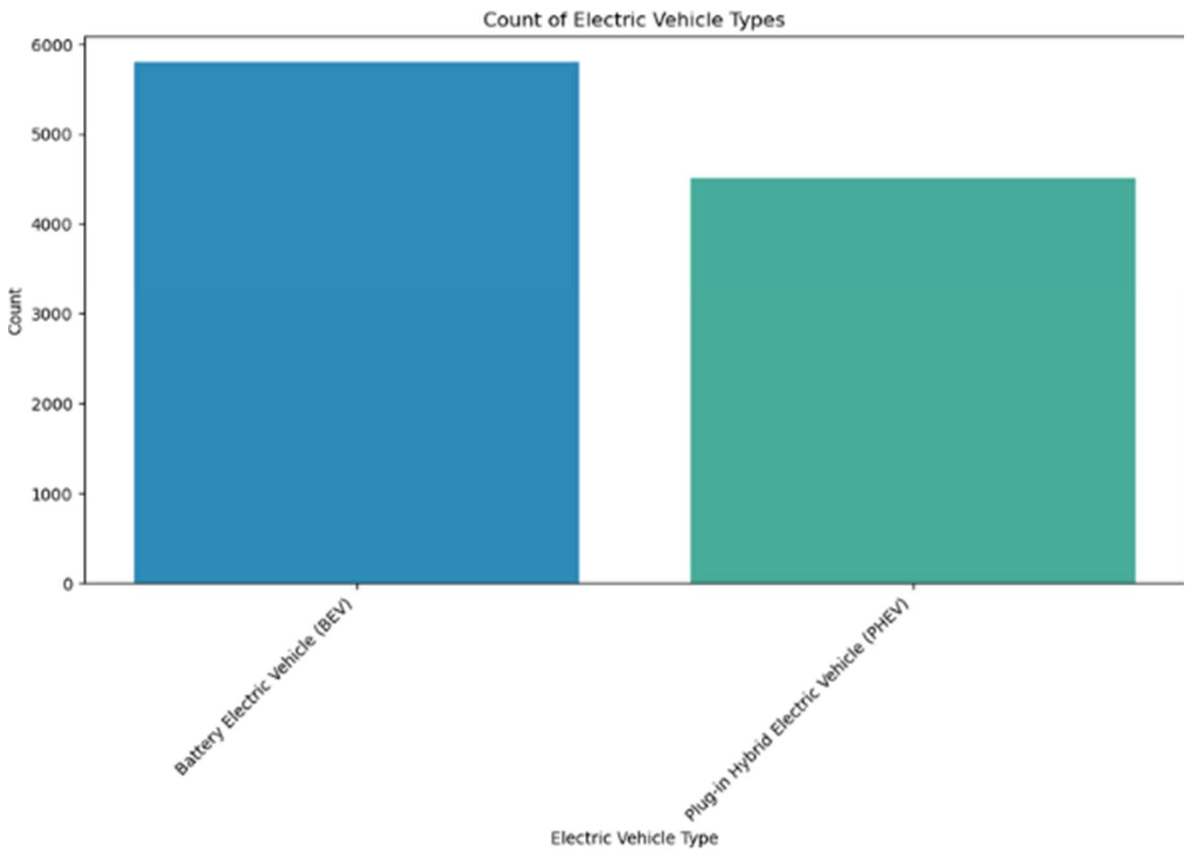
Reds and other warm colors denote positive correlations, in which the variables tend to rise together. Blues and other cool hues show negative correlations, in which one variable tends to rise while the other falls. Stronger correlations are represented by darker shades, and weaker correlations by lighter shades.

- Negative correlation between electric range and model year can be observed
- Negative correlation can be observed between Base MSRP and Model Year



Bar plot - Count of Electric Vehicle Types

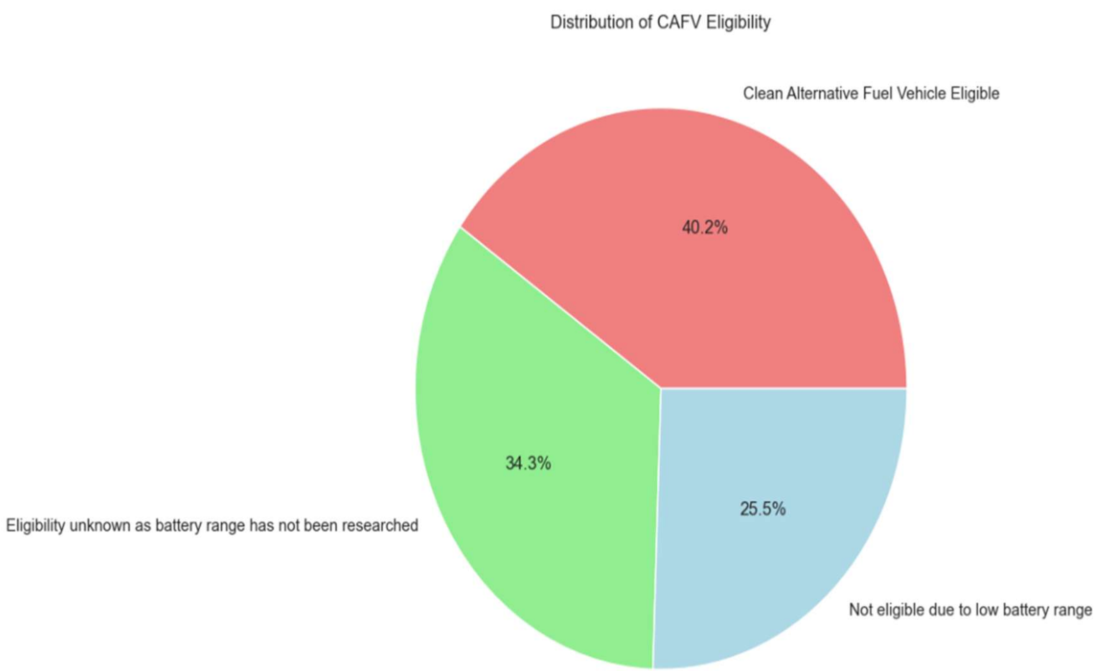
- Each bar corresponds to a specific electric vehicle type, and the height of the bar indicates how many instances of that type are present in the dataset
- As seen in the adjacent figure, the count of battery electric vehicles are higher than that of plug-in hybrid electric vehicle
- However, the difference between the two is a very low margin.



It is simple to compare the frequency of various electric vehicle types with this visualization. We can obtain insights into the distribution of electric vehicle types and rapidly determine which types are more common in the dataset.

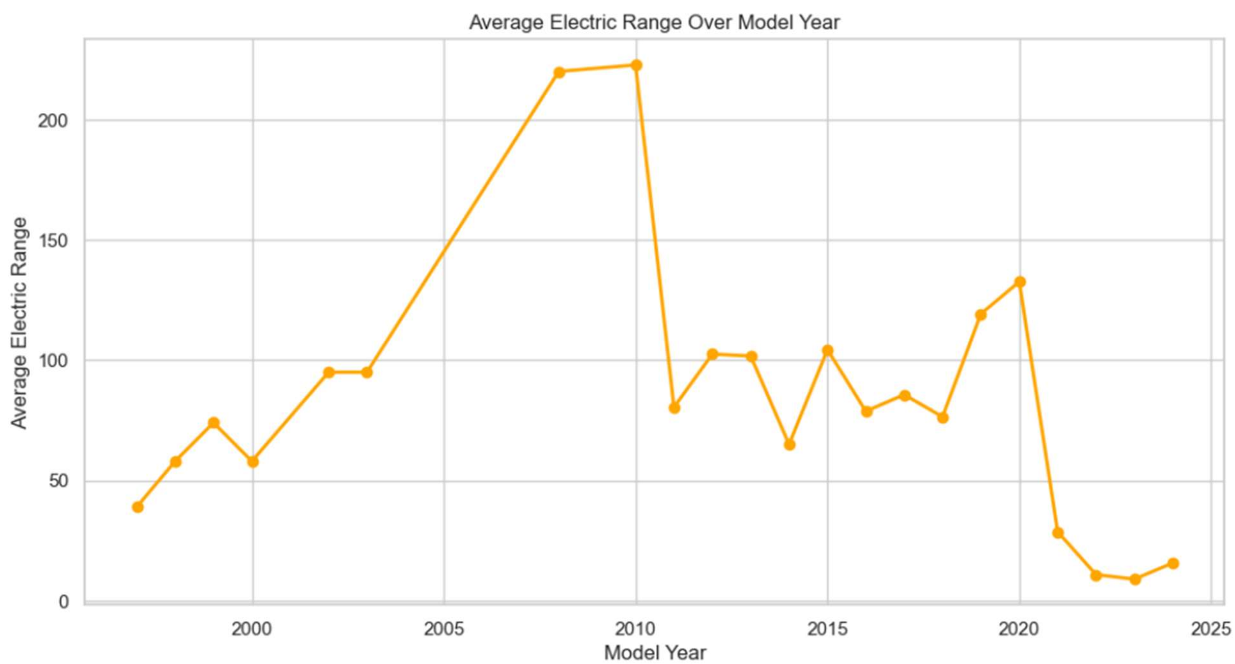
Distribution of Clean Alternative Fuel Vehicle (CAFV) Eligibility

- The distribution of Clean Alternative Fuel Vehicle (CAFV) eligibility statuses in the dataset is visualized in the pie chart's output. This is how you understand it:
- Pie Slices: Every pie slice denotes a different CAFV eligibility status.
- Slice Size: Compared to the total number of vehicles in the dataset, the size of each slice represents the percentage of vehicles with a given eligibility status.
- The chart displays the distribution of CAFV eligibility categories.
- Labels on the chart show the percentage distribution of each category.
- Majority of the CAFV eligibility is occupied by the clean alternative fuel vehicle.
- 25.5% of vehicles are not eligible due to low battery range.
- 34.3% of eligibility is unknown since the battery range has not been researched.



Electric Range over Model Year

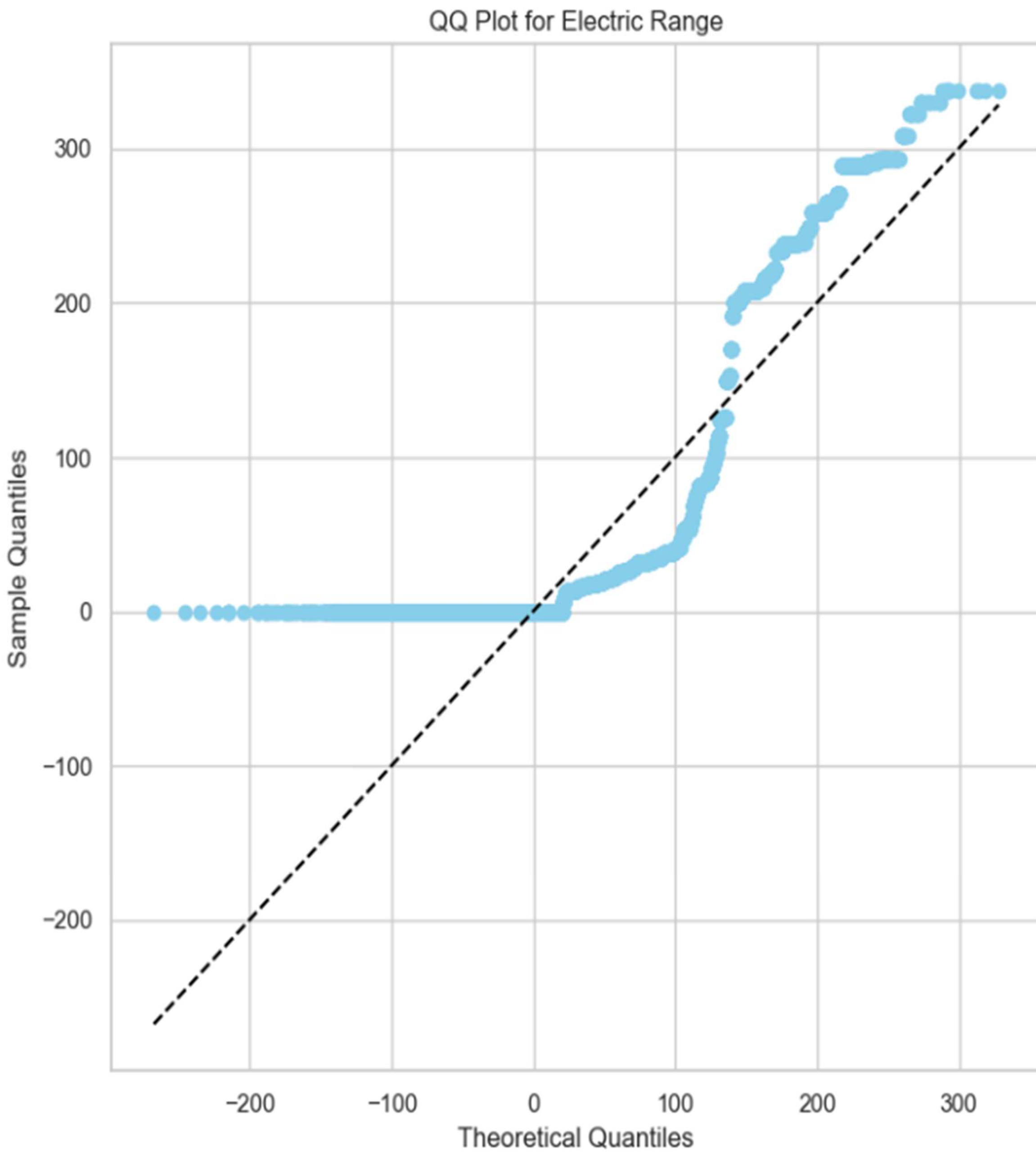
This visualization allows for the exploration of how the average electric range has changed over different model years. It provides insights into potential trends or patterns in the electric range of vehicles over time.



- The average electric range for a given model year is represented by each point on the plot.
- The average electric range trend or pattern over various model years is represented visually by the line joining these points.
- Markers: Each data point has an orange circle (a circular marker) to draw attention to particular observations and make it simpler to distinguish between different values.
- The plot title "Average Electric Range Over Model Year" sets the scene for the visualization.
- "Model Year" is the label on the x-axis that indicates which model year is being represented.
- "Average Electric Range" is the label on the y-axis that indicates the mean electric range.
- Gridlines: Added to the plot to serve as a reference for the values and help in interpretation.

QQ Plot to check for normality in Electric Range

It is possible to determine visually whether the dataset's electric range values follow a normal distribution. The data may be roughly normally distributed if the points in the QQ plot closely match the reference line. In all other cases, departures from normalcy are indicated by deviations from the reference line.



- The plot is designed to check for normality in the distribution of the 'Electric Range' variable.
- The x-axis represents the theoretical quantiles expected from a normal distribution.
- The y-axis represents the actual quantiles observed in the sample data ('Electric Range').
- The diagonal dashed line represents the line of perfect normality. This line helps in visually assessing how close the observed data points are to the expected quantiles.
- Blue markers represent scatter plot where the x-axis represents the expected quantiles, and the y-axis represents the sorted observed 'Electric Range' data.
- Points on the plot are scattered in a pattern around the diagonal line.
- The points closely following the diagonal line suggests that the distribution of 'Electric Range' is approximately normal.
- Deviations from the line indicates departures from normality.

T – Test

- T-statistic = 27.10126236775644, p-value = 1.9947448841779136
- The p-value exceeds the significance level (0.05).
- There is no significant difference in Electric Range between Battery Electric Vehicles (BEV) and Plug-in Hybrid Electric Vehicles (PHEV).

This T-test analysis suggests that the Electric Range between Battery Electric Vehicles (BEV) and Plug-in Hybrid Electric Vehicles (PHEV) is not statistically significant, indicating comparable performance in terms of range.

Part 5: Limitations and Future Work

Limitations

- The dataset is limited to electric vehicles registered in Washington State, which may not represent global or national trends. Therefore, generalizing findings beyond this scope should be done cautiously.
- While efforts were made to address missing values and clean the data, there might still be gaps or inaccuracies in the dataset, which could affect the analysis and conclusions drawn.
- The study mentions over 120 unique car models but does not delve into specific model characteristics or features, which could provide deeper insights into electric vehicle trends and preferences.
- The analysis primarily focuses on internal variables like electric range and model year, but external factors such as policy changes, market dynamics, and technological advancements could also influence electric vehicle trends.

Future Work

- Including data from other regions or countries could provide a broader perspective on electric vehicle trends and enhance the generalizability of findings.
- Further exploration of specific vehicle models, features, and customer preferences could offer more nuanced insights into electric vehicle adoption and usage patterns.
- Utilizing advanced predictive modeling techniques could forecast future electric vehicle trends, aiding policymakers, manufacturers, and stakeholders in decision-making and planning.
- Incorporating external variables such as government policies, infrastructure development, and consumer behavior could enrich the analysis and provide a more comprehensive understanding of electric vehicle dynamics.

Overall, while the study offers valuable insights into electric vehicle trends and analysis techniques, there is ample room for future research to address limitations and further explore this evolving field.