Electric Vehicles Data Insight

Data Acquisition and Inspection

Dataset Overview

Dataset: **Electric Vehicle Population Data**

This dataset shows the Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) that are currently registered through Washington State Department of Licensing (DOL).

This dataset is crucial for understanding the distribution, characteristics, and legislative aspects of electric vehicles in Washington.

Sample Data

	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Vehicle Location	Electric Utility	2020 Census Tract
0	5YJYGDEF5L	Thurston	Lacey	WA	98516.0	2020	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	291	0	22.0	124535071	POINT (-122.7474291 47.0821119)	PUGET SOUND ENERGY INC	5.306701e+10
1	1N4BZ1CP1K	King	Sammamish	WA	98074.0	2019	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	150	0	45.0	102359449	POINT (-122.0313266 47.6285782)	PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)	5.303303e+10
2	5YJXCDE28G	King	Kent	WA	98031.0	2016	TESLA	MODEL X	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	200	0	33.0	228682037	POINT (-122.2012521 47.3931814)	PUGET SOUND ENERGY INC||CITY OF TACOMA - (WA)	5.303303e+10

Data Cleaning and Preparation

Data Preprocessing Steps

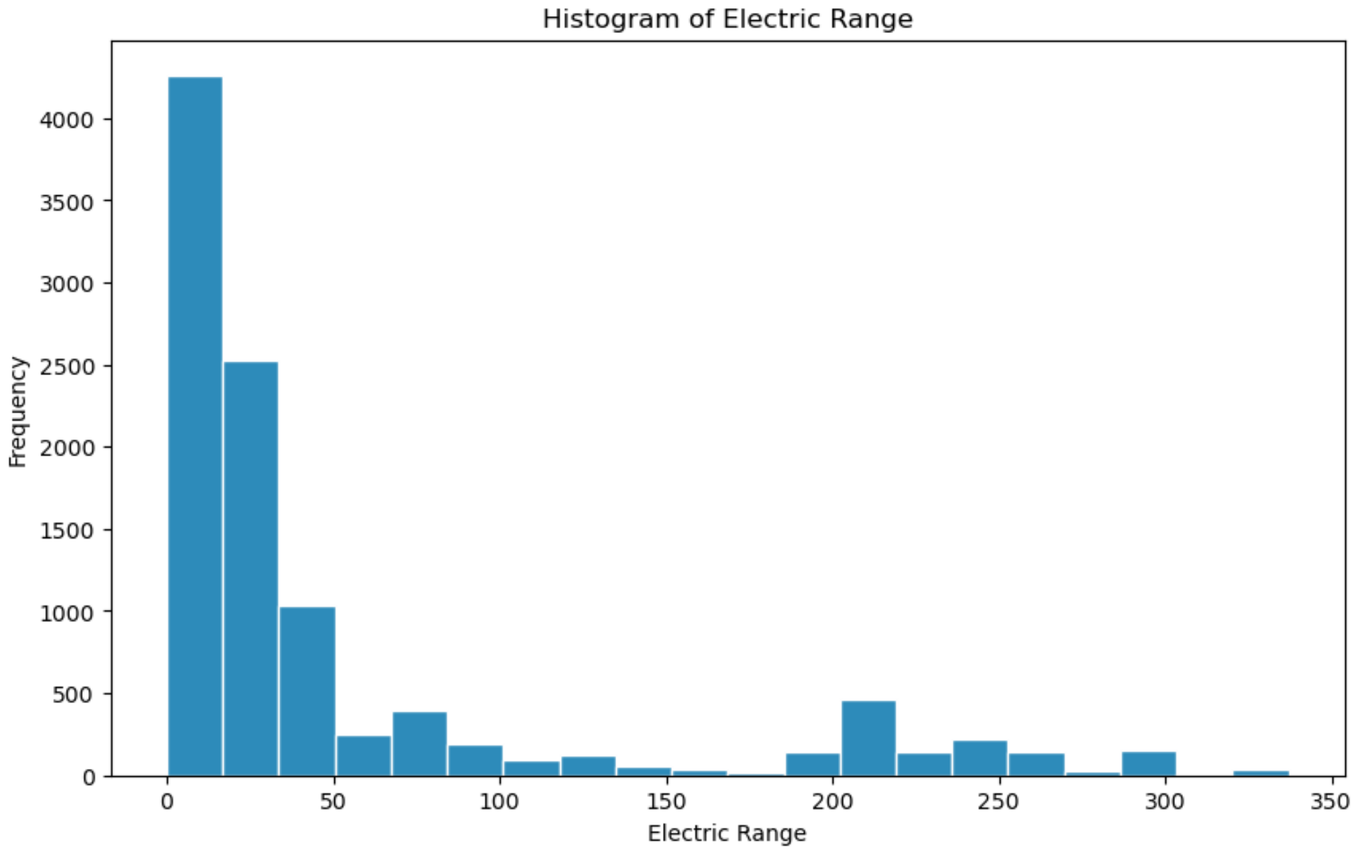
- **Data Loading:**
  - Loaded the dataset using Pandas.
  - Checked the first few rows to understand the data structure.
- **Data Exploration:**
  - Utilized **describe()** to generate statistical summaries for each column.
  - Used **info()** to inspect non-null records and data types.
- **Handling Missing Data:**
  - Identified and counted null records using **isnull().sum()**.
  - Filled null values with 'unknown' for initial assessment.
- **Duplicate Removal:**
  - Removed duplicate records based on the 'VIN (1-10)' column.
  - Ensured data cleanliness for accurate analysis.
- **Replacing 'unknown' with NaN:**
  - Replaced 'unknown' values with NaN for consistent handling.
  - Set the stage for proper numeric conversions.
- **Numeric Conversion:**
  - Converted columns ('Postal Code', 'Legislative District', '2020 Census Tract') to numeric.
  - Errors handled by coercing to NaN for data integrity.
- **NaN Handling:**
  - Dropped rows with NaN values after conversion.
  - Ensured that the dataset only includes records with valid numeric data.
- **Final Data Type Adjustment:**
  - Converted numeric columns to integers for uniformity and ease of analysis.
  - Resulting dataset is ready for meaningful exploration and visualization.

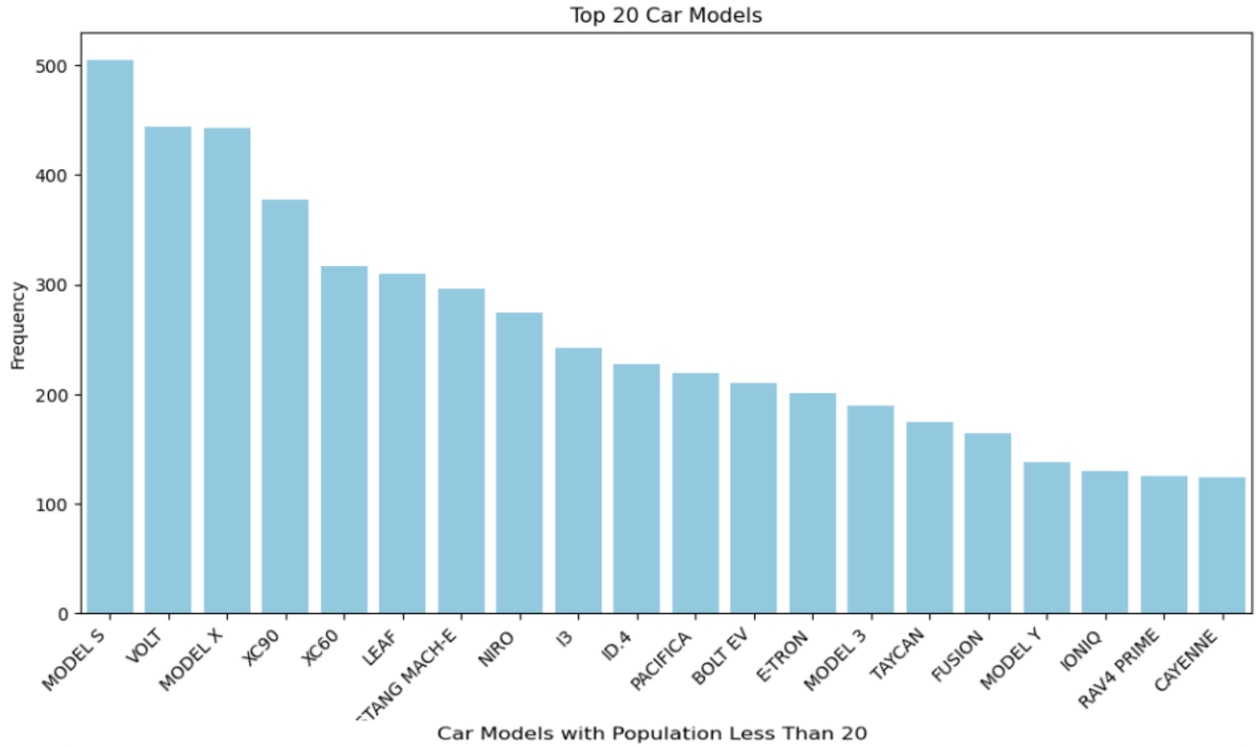
Exploratory Data Analysis (EDA) with a Focus on Static Visualization

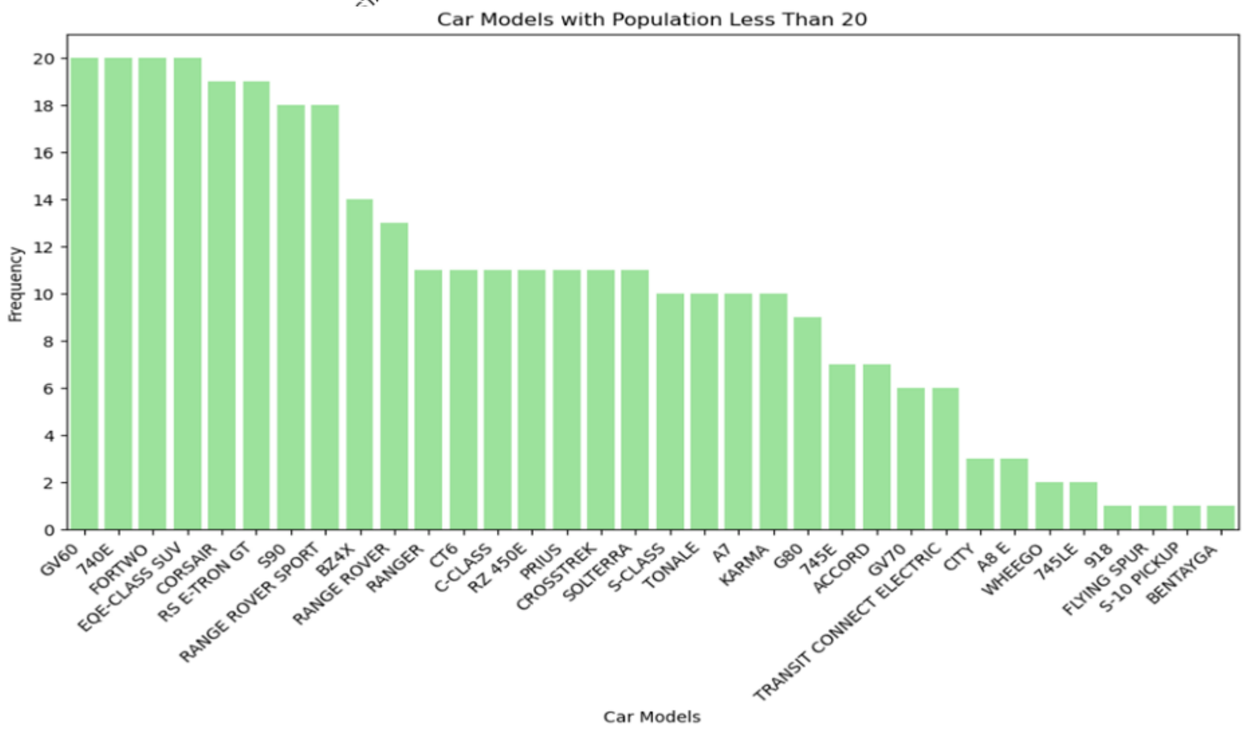
Histogram of Electric Range

The histogram of electric range provides a visual representation of the distribution, offering insights into the prevalent and frequent electric range values among registered vehicles, crucial for understanding consumer preferences in the electric vehicle landscape.

- Majority of electric vehicles have a range of 0 to 50 miles.
- Most common electric range is approximately 16.85 miles.
- Notable peaks observed around 0, 16.85, 33.70, and 202.20 miles.
- Significant portion of vehicles falls within 0 to 100 miles range.
- Electric ranges beyond 200 miles are less common.



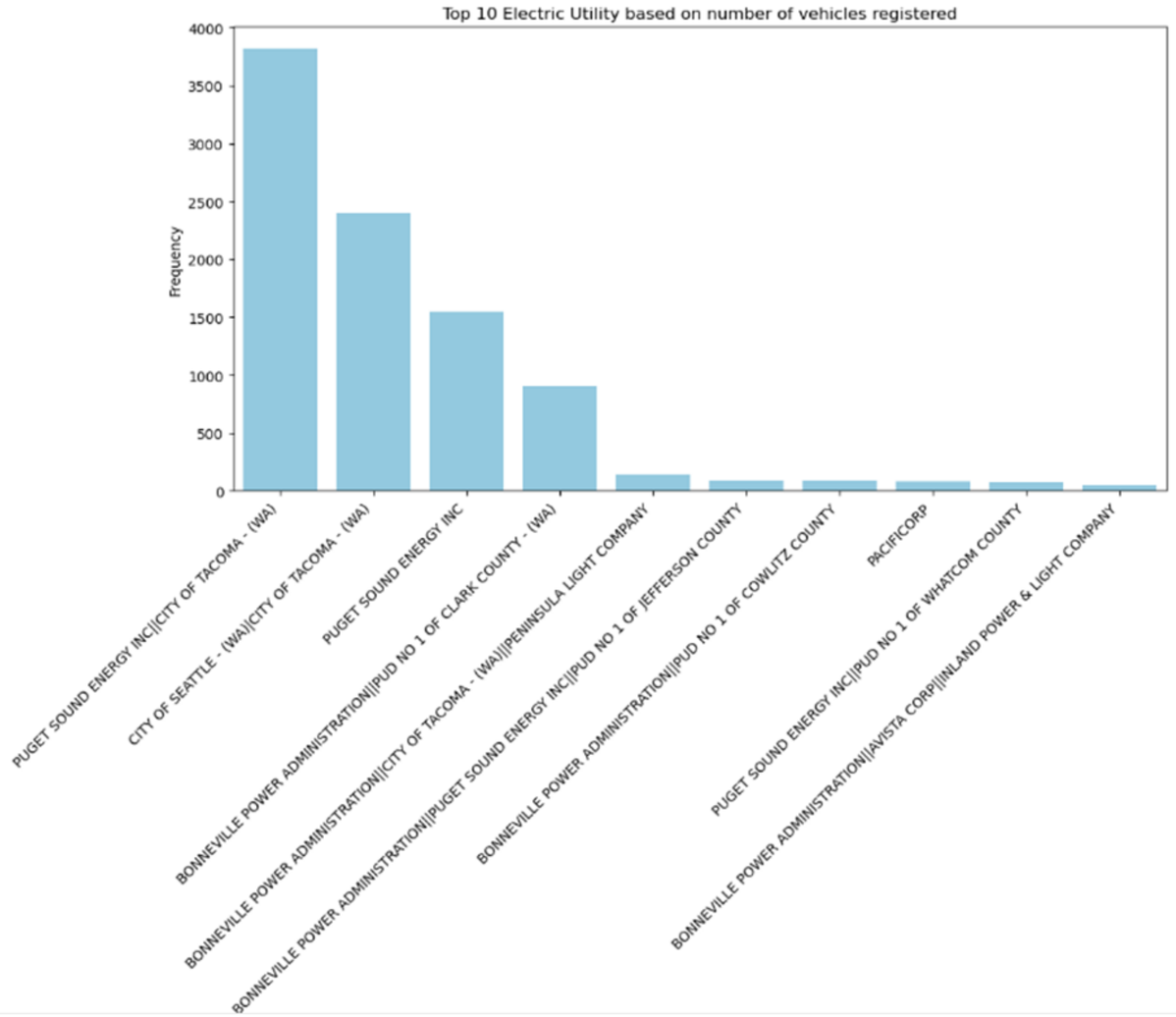




·There are over 120 unique car models registered in Washington State Department of Licensing.

·MODEL S has the highest population of 505, followed by VOLT and MODEL X with 444 and XC90 with 378.

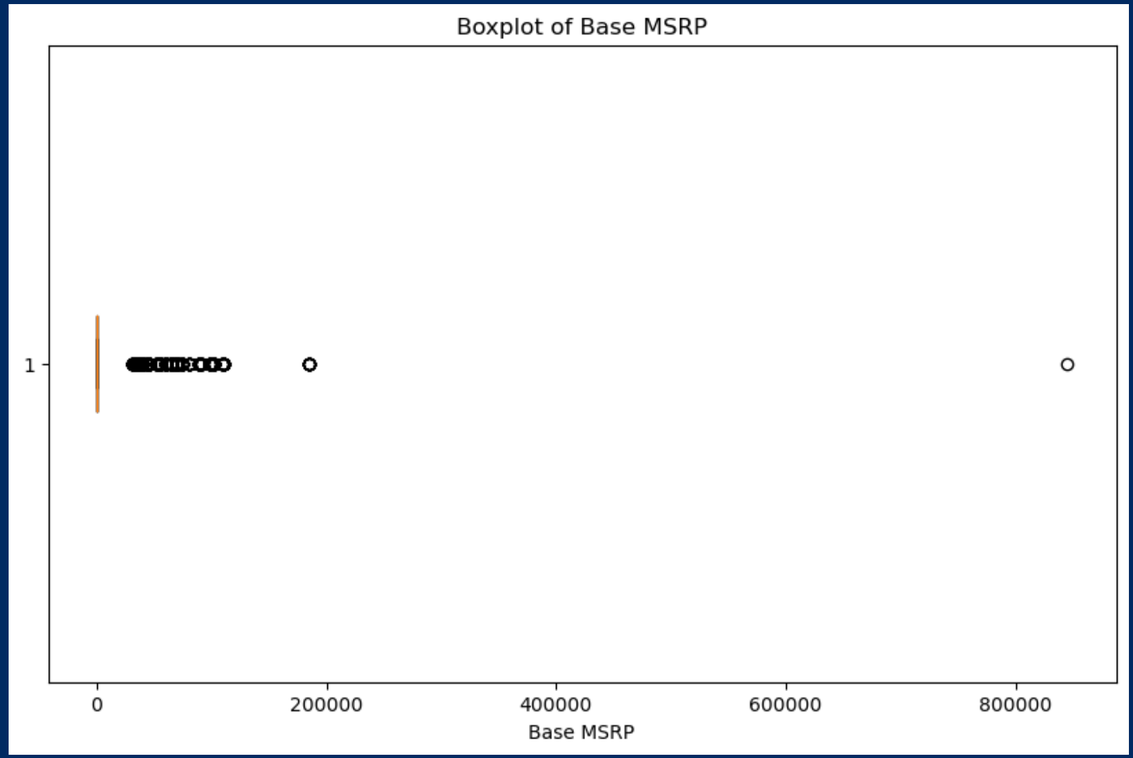
·There are over 25 car models that have a population lower than 20 with 918, FLYING SPUR, S-10 PICKUP and BENTAYGA with the population of 1.



- 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.

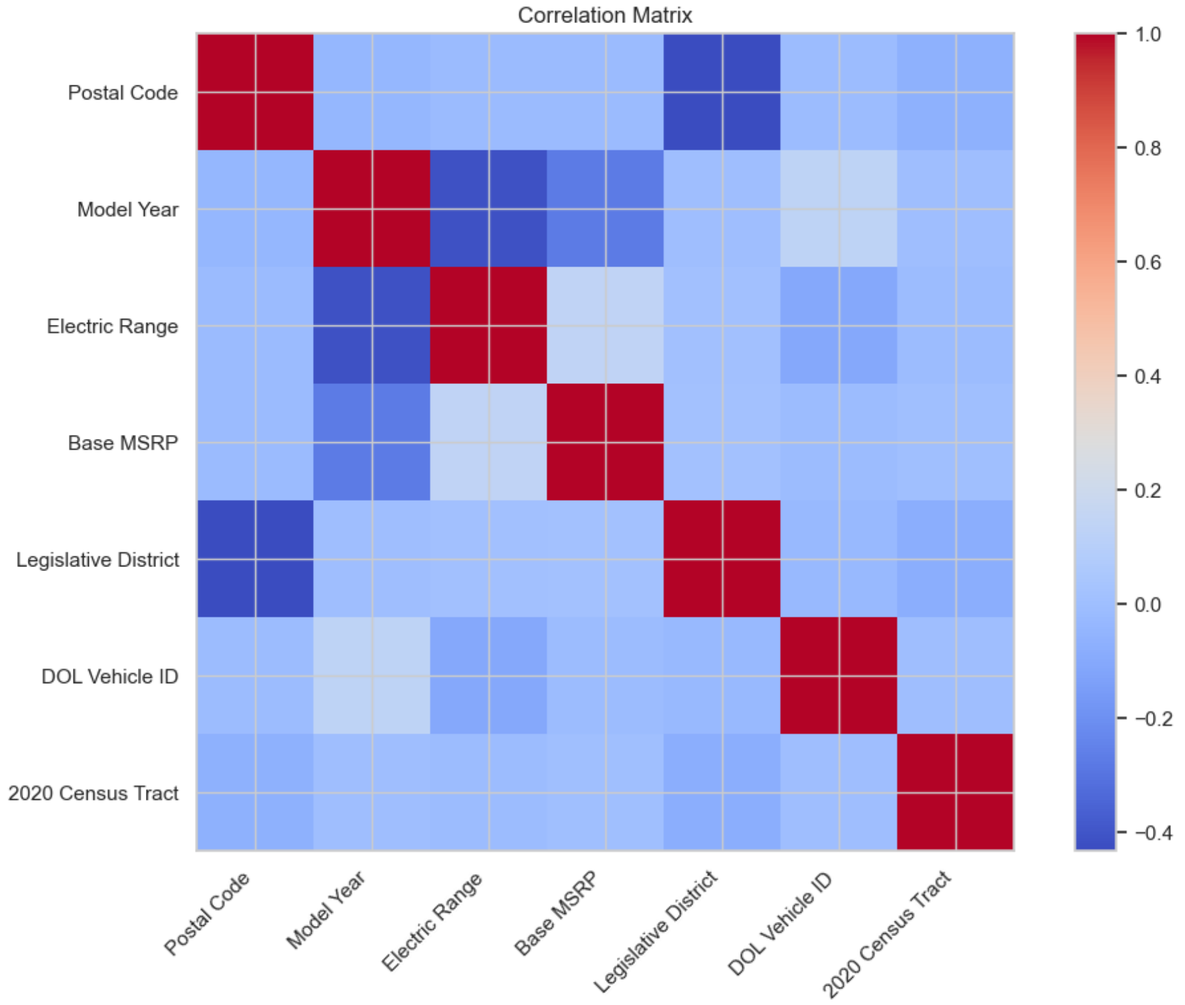
The boxplot of Base Manufacturer's Suggested Retail Price (MSRP) is useful for visually assessing the distribution and variability of pricing within the dataset, aiding in the identification of central tendencies, spread, and potential outliers.

- Base MSRP ranges from $0 to $845,000, showcasing pricing diversity.
- Understanding the reasons behind these extreme MSRP values can provide valuable insights into unique pricing strategies or exceptional features.



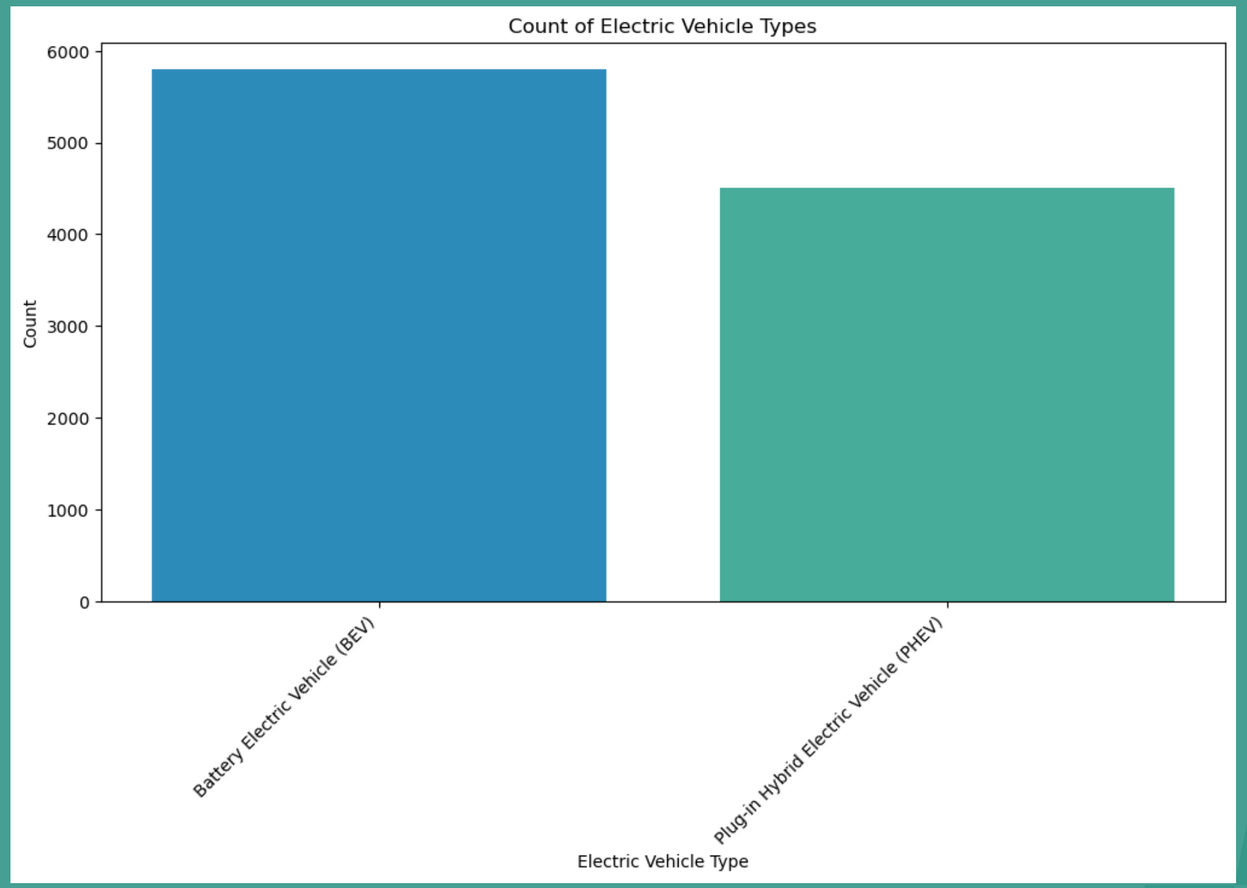
Correlation HeatMap

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

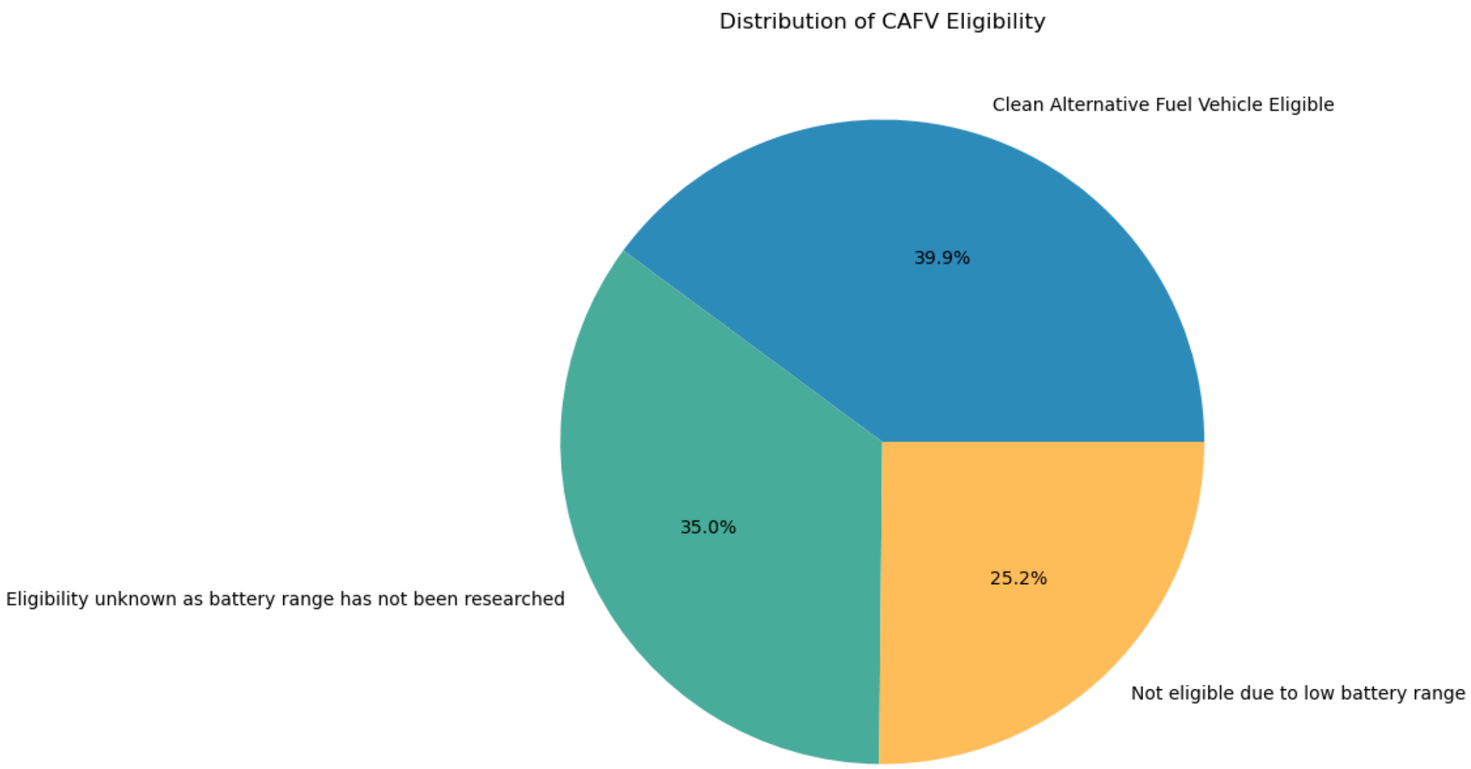


Visualization of Count of Electric Vehicle Types

- 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.

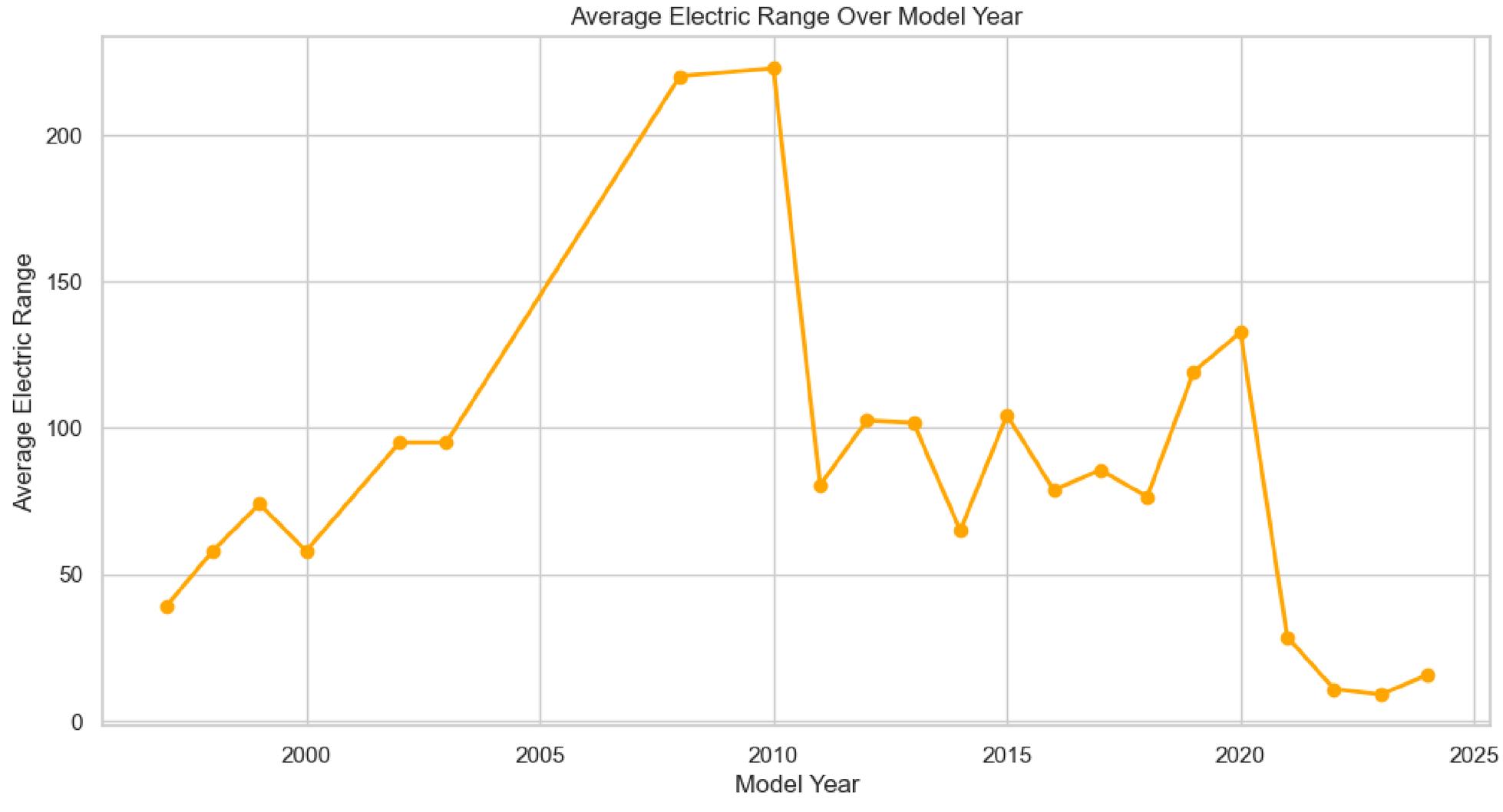


Distribution of CAFV Eligibility



- 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.

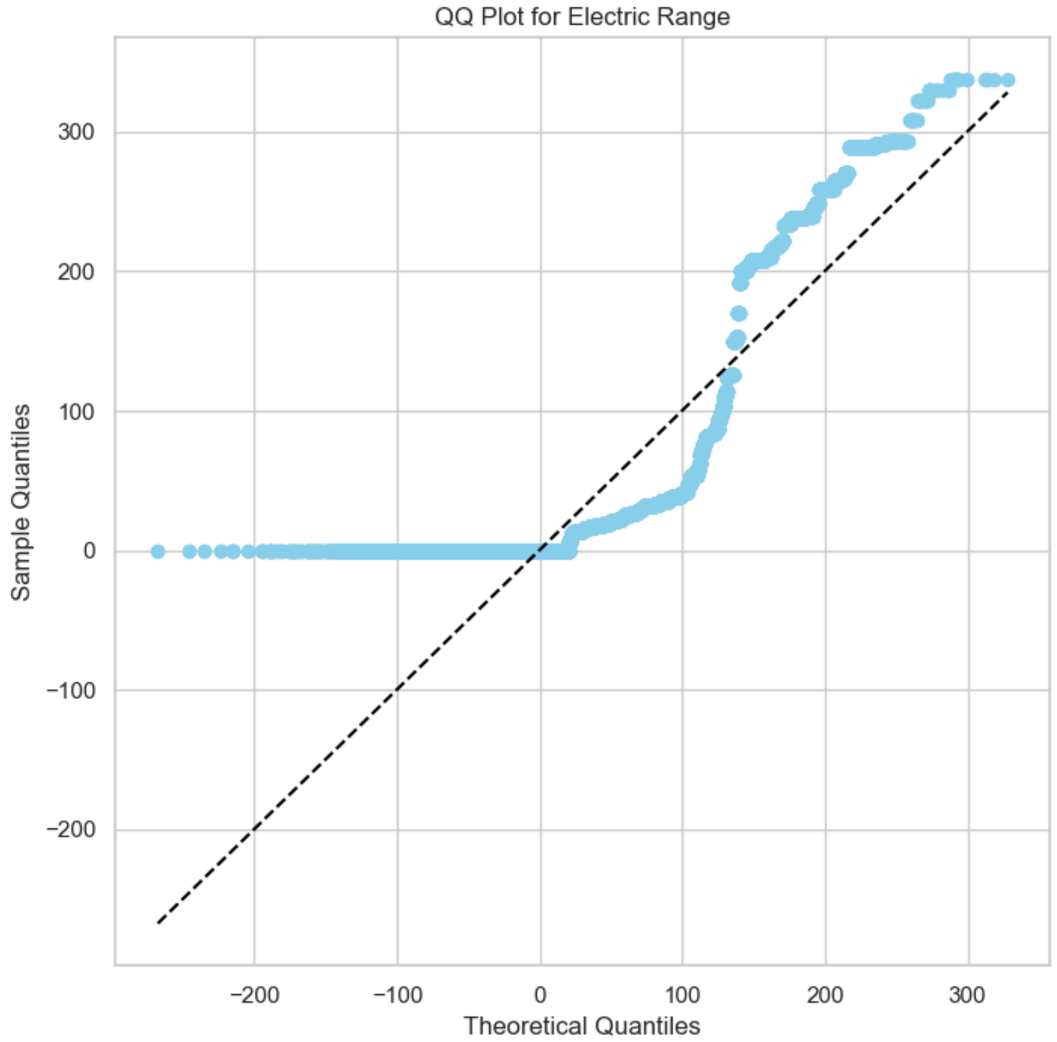
Average Electric Range over Model Year



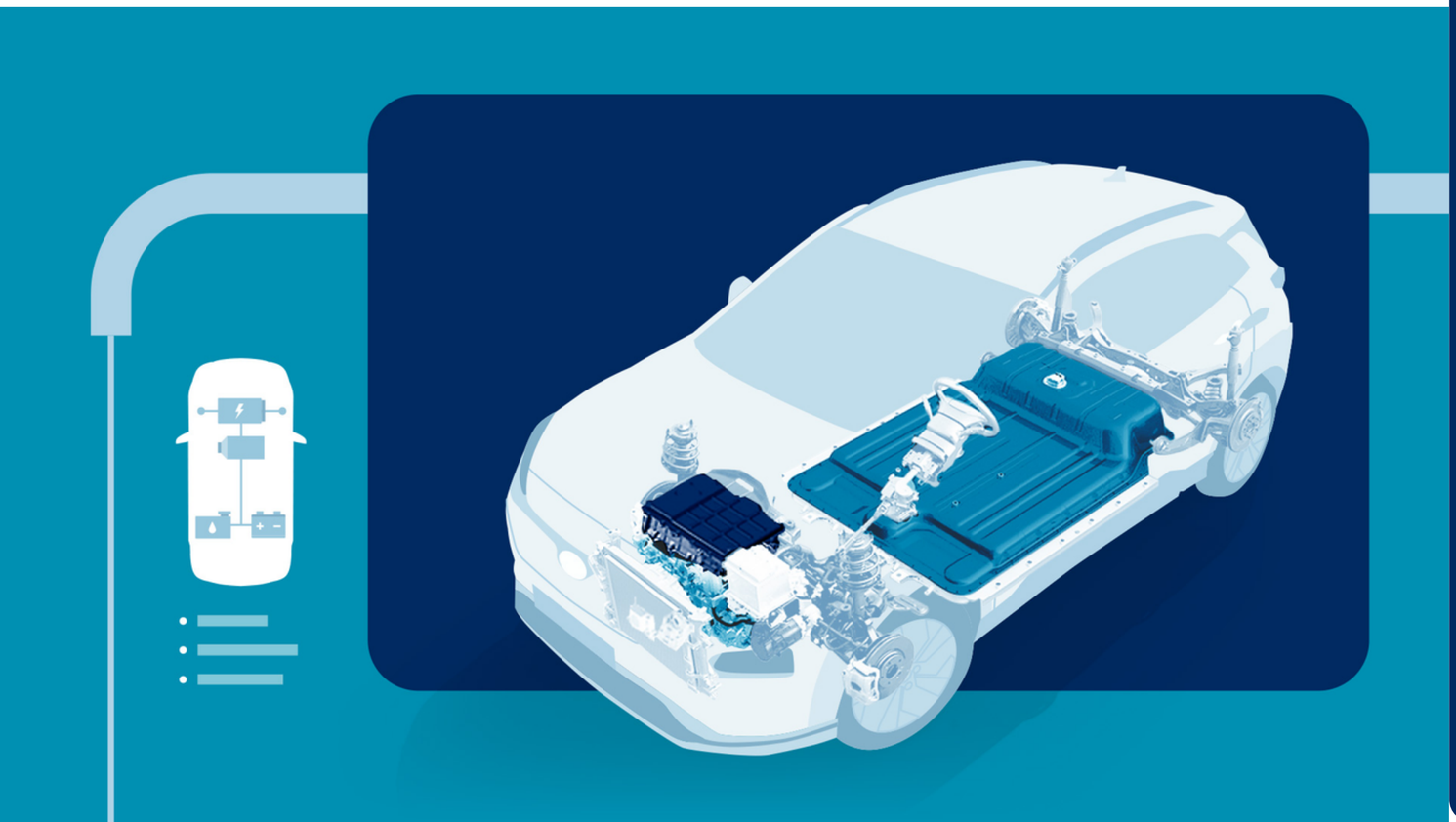
- The data for this plot is derived from grouping the dataset by 'Model Year' and calculating the mean electric range for each year.
- Each point on the line represents the average electric range for a specific model year.
- The highest average range is between the year 2007-2010 approx.
- The average would decline to a very low range post 2024 followed by a slight increase. Overall, there is an irregular distribution of electric range.

QQ Plot for Electric Range

- 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.



Statistical Analysis - T-test Results



Electric Range Comparison:

- 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.


FINISH