**Electric Vehicle Data Analysis**

**Data**

Market Size Analysis is the process of estimating the potential sales for a product or service within a particular market segment. In the context of electric vehicles (EVs), it involves assessing the total volume of EV registrations to understand the growth of the market, forecast future trends, and help stakeholders make informed decisions regarding production, infrastructure development, and policy-making.

The provided dataset contains the following columns, each representing different aspects of the electric vehicle (EV) population in the United States:

* **VIN (1-10):** Partial Vehicle Identification Number.
* **County:** The county in which the vehicle is registered.
* **City:** The city in which the vehicle is registered.
* **State:** The state in which the vehicle is registered. It appears that this dataset may be focused on Washington (WA) state.
* **Postal Code:** The postal code where the vehicle is registered.
* **Model Year:** The year of the vehicle model
* **Make:** The manufacturer of the vehicle.
* **Model:** The model of the vehicle.
* **Electric Vehicle Type:** The type of electric vehicle, e.g., Battery Electric Vehicle (BEV).
* **Clean Alternative Fuel Vehicle (CAFV) Eligibility:** Eligibility status for clean alternative fuel vehicle programs.
* **Electric Range:** The maximum range of the vehicle on a single charge (in miles).
* **Base MSRP:** The Manufacturer’s Suggested Retail Price.
* **Legislative District:** The legislative district where the vehicle is registered.
* **DOL Vehicle ID:** Department of Licensing Vehicle Identification.
* **Vehicle Location:** Geographic coordinates of the vehicle location.
* **Electric Utility:** The electric utility service provider for the vehicle’s location.
* **2020 Census Tract:** The census tract for the vehicle’s location.

**Basic Steps:**

* + 1. Display the top 5 rows.
    2. Display the last 5 rows
    3. Check the shape of dataset.
    4. Check the datatypes of each feature.
    5. Check the Statistical summary
    6. Check the null values
    7. Check the duplicate values
    8. Check the anomalies or wrong entries.
    9. Check the outliers and their authenticity.
    10. Do the necessary data cleaning steps like dropping duplicates, unnecessary columns, null value imputation, outliers treatment etc.

The primary objective of this analysis is to leverage historical EV registration data to understand the current market penetration of EVs, predict future market growth, and identify key trends and factors driving market expansion. The specific goals include:

1. **Descriptive Statistics:**
   * What are the mean, median, and standard deviation of the base MSRP for the vehicles in the dataset?
2. **Data Distribution:**
   * What is the distribution of vehicle makes in the dataset? Represent it using a bar chart.
3. **Model Year Analysis:**
   * What are the most common model years in the dataset? Provide a frequency table and histogram.
4. **Electric Vehicle Type:**
   * What is the proportion of Battery Electric Vehicles (BEV) versus other types of electric vehicles?
5. **Electric Range Analysis:**
   * What is the average electric range for vehicles of different makes? Provide a summary table.
6. **County Distribution:**
   * How are vehicles distributed across different counties in Washington state? Represent the distribution using a pie chart.
7. **Price Analysis:**
   * Compare the average base MSRP of vehicles eligible for the Clean Alternative Fuel Vehicle (CAFV) program versus those that are not.
8. **Geographical Analysis:**
   * How does the base MSRP vary across different cities in Washington state?
9. **Legislative Districts:**
   * Which legislative districts have the highest number of registered electric vehicles? Provide a ranked list.
10. **Electric Utility Providers:**
    * What is the distribution of electric utility service providers for the vehicles in the dataset?
11. **Census Tract Analysis:**
    * How are vehicles distributed across different 2020 Census Tracts? Provide insights based on vehicle counts per tract.
12. **Electric Range Correlation:**
    * Is there a correlation between the electric range and the base MSRP of the vehicles? Provide the correlation coefficient and interpret the result.
13. **VIN Analysis:**
    * Identify any patterns or commonalities in the VIN (1-10) for the vehicles. Are there any frequent prefixes or suffixes?
14. **Eligibility Status:**
    * What percentage of vehicles are eligible for the Clean Alternative Fuel Vehicle (CAFV) program?
15. **Model Popularity:**
    * Which vehicle models are the most popular in the dataset? Provide a frequency table of the top 10 models.
16. **Postal Code Distribution:**
    * How are vehicles distributed across different postal codes? Provide a heatmap or density plot.
17. **Vehicle Location Analysis:**
    * Analyze the geographic coordinates to determine any clusters of electric vehicles in certain areas of Washington state.
18. **Model Year Trend:**
    * Analyze the trend in the number of registered electric vehicles by model year. Provide a line chart to show any increase or decrease over the years.
19. **Range vs. Year:**
    * Is there a trend between the model year and the electric range of the vehicles? Provide a scatter plot and analyze the trend.
20. **Legislative District and MSRP:**
    * How does the average base MSRP vary across different legislative districts?