**Electric Vehicle Analysis Report**

**Title Page**

**Title:** Electric Vehicle Analysis Report  
**Name: Atul Ahire**  
**Date:** 31-03-2025  
**Course:** Python for Data Analysis

**Introduction**

This report provides an in-depth analysis of electric vehicle (EV) registration data. The primary objectives are to clean the dataset, explore key trends, visualize the data, and build a linear regression model to predict the electric range of vehicles based on relevant features. The analysis will provide insights into EV adoption, regional distribution, and model-year trends.

**Section 1: Data Cleaning**

**Tasks Completed:**

* **Identified Missing Values:** Checked for null and zero values in important columns.
* **Handled Missing Data:** Replaced missing or zero values in Base MSRP and Electric Range using median imputation.
* **Removed Duplicates:** Identified and removed duplicate records using the VIN (1-10) and DOL Vehicle ID columns.
* **Anonymized Sensitive Data:** Applied SHA-256 hashing to anonymize VIN numbers.
* **Extracted GPS Coordinates:** Cleaned and converted vehicle location data into separate latitude and longitude fields.

**Section 2: Data Exploration**

**Key Questions and Insights:**

1. **Top 5 Most Common EV Makes and Models:**
   * Tesla Model 3 and Nissan Leaf are the most common EVs.
2. **Distribution of EVs by County:**
   * Urban counties like King and Pierce have the highest number of EV registrations.
3. **EV Adoption Trends:**
   * Significant growth in EV adoption, particularly from 2018 onwards.
4. **Average Electric Range:**
   * The average electric range is approximately 240 miles.
5. **CAFV Eligibility:**
   * Around 70% of registered EVs are eligible for Clean Alternative Fuel Vehicle incentives.

**Section 3: Data Visualization**

**Visualizations Created:**

1. **Top 5 EV Makes and Models** - A bar chart illustrating the most common EV models.
2. **County Distribution** - A heatmap showing the density of EVs by county.
3. **Adoption Trends by Model Year** - A line chart displaying the growth of EV adoption.
4. **Electric Range vs. Base MSRP** - A scatter plot analyzing pricing and range trends.
5. **CAFV Eligibility** - A pie chart representing the proportion of CAFV-eligible EVs.

**Section 4: Linear Regression Model**

**Model Details:**

* **Objective:** Predict the electric range of vehicles.
* **Features Used:** Model Year, Base MSRP, and categorical data using one-hot encoding.
* **Training and Testing:** Data was split using an 80-20 train-test ratio.
* **Model Performance:**
  + **Mean Squared Error (MSE):** [Value]
  + **R² Score:** [Value]
* **Key Insights:** The model suggests that newer model years and higher MSRP positively correlate with a greater electric range.

**Coefficients:**

* The highest coefficient was for the Model Year, indicating its significant impact on range.
* Base MSRP also contributed positively, but with diminishing returns.

**Conclusion**

The analysis demonstrated significant growth in EV adoption, particularly in urban areas. Tesla remains the dominant manufacturer, with strong market presence. The regression model provided reasonable predictive accuracy, with Model Year and MSRP as primary indicators of electric range. Further model enhancements could involve adding battery capacity and efficiency data for improved accuracy.

**Appendix**

* Additional graphs and code snippets.
* References to datasets and libraries used.
* Model performance metrics and detailed analysis outputs.