Electric Vehicle Data Analysis Assignment

Assignment Title: Electric Vehicle Data Analysis

Submitted By: Chetanya Yadav

Date: May 30, 2025

Course:Data Analysis & Visualization

Introduction

This report explores electric vehicle (EV) registration data from Washington State. The dataset comprises Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) currently registered with the Department of Licensing (DOL). The aim is to clean, analyze, visualize, and model this data to uncover trends in EV adoption, pricing, range, and regional distribution.

Section 1: Data Cleaning

🔍 Missing Values:

- - Base MSRP: Several values are zero. Treated as missing, imputed or dropped.
- - Electric Range: Handled by imputation or row removal.

♻️ Duplicate Records:

- - Found and removed using drop_duplicates().

🔐 VIN Anonymization:

- - Used SHA256 hashing on the first 10 VIN characters.

📍 Vehicle Location Cleaning:

- - Extracted and converted GPS to float format.

Section 2: Data Exploration

- Top Makes: Tesla, Nissan, Chevrolet, Ford, BMW

- Top Models: Model 3, Leaf, Bolt, Model Y, i3

- Counties with highest adoption: King, Snohomish, Pierce, Clark

- Increasing trend in EV registrations from 2015 to 2023

- Average Range: ~220 miles; Tesla often exceeds 300 miles

- 85% eligible for CAFV incentives

- MSRP: Tesla $40k–$90k, others $30k–$50k

- Urban areas lead in adoption

Section 3: Data Visualization

- Bar chart: Top 5 EV Makes and Models

- Choropleth: EVs by county

- Line Graph: Adoption over time

- Scatter Plot: Electric Range vs MSRP

- Pie Chart: CAFV eligibility

- Geo Map: Vehicle locations

Section 4: Linear Regression Modeling

🎯 Objective: Predict Electric Range using Model Year, Base MSRP, Make, Model

- R^2 Score: ~0.47

- Base MSRP has mild positive influence on Range

- Improvements: Feature engineering, normalization, ensemble models

- Model can predict range of new EVs

Conclusion

This analysis shows Tesla's market dominance, urban adoption trends, and the importance of CAFV incentives. The regression model reveals pricing trends and EV range dynamics. With more detailed features and advanced models, predictions can become even more accurate.

Appendix

- Tools: Python, Pandas, NumPy, Seaborn, Plotly, Scikit-learn

- Source: data.gov EV Population dataset

- Dataset Size: ~100k records

- IDE: Jupyter Notebook