Project Title

Analyzing Electric Vehicle Adoption Trends and Performance Metrics

---

Dataset Overview

Source: [Washington State Open Data Portal](#)

- **Rows**: ~15,000 (A subset of the full Washington State EV registration dataset will be used).

- **Columns**: 17 features including:

    o   VIN, County, City, State, Postal Code

    o   Model Year, Make, Model, Electric Vehicle Type

    o   Electric Range, Base MSRP, Legislative District

    o   Vehicle Location (geospatial coordinates)

Project Objectives

1. **Data Exploration**: Analyze geographic, brand-level, and vehicle-type trends in EV adoption.

2. **Performance Analysis**: Examine how electric range varies with vehicle age and MSRP.

3. **Dashboard Creation**: Build an interactive Streamlit app to visualize key insights.

4. **Predictive Modeling**: Predict Electric Range using regression techniques.

Project Scope

1. Data Loading & Initial Exploration

- Import libraries (Pandas, NumPy, Matplotlib/Seaborn/Plotly, Streamlit).

- Load the dataset and analyze structure (shape, info, describe).

- Check for missing values, duplicates, and outliers (e.g., unrealistic Electric Range values).

2. Exploratory Data Analysis (EDA)

- **Univariate Analysis**:

  o Distribution of Electric Range, Model Year, and Make.

  o Most common EV types (e.g., Battery vs. Plug-in Hybrid).

- **Bivariate Analysis**:

  o Electric Range vs. Model Year (Are newer EVs more efficient?).

  o Make vs. average Electric Range (Which brands perform best?).

  o Geographic trends: EV adoption by County or City.

- **Multivariate Analysis**:

  o Interaction between Model Year, Make, and Electric Range.

  o Correlation matrix (e.g., Base MSRP vs. Electric Range).

3. Data Cleaning

- Handle missing values (e.g., impute Electric Range using median/mean).

- Remove irrelevant columns (e.g., VIN, Legislative District).

- Encode categorical variables (e.g., Make, Electric Vehicle Type).

4. Visualization

- Create histograms, box plots, and heatmaps.

- Use geospatial plots to map EV density by region.

- Compare distributions across categories.

5. Streamlit Dashboard

- **Interactive Filters**: Allow users to select Make, County, or Model Year.

- **Visualizations**:

  o Bar charts: Top 10 EV brands by adoption.

  o Scatter plots: Electric Range vs. Model Year.

  o Geo-map: EV registrations by city/county.

- **Summary Statistics**: Display average range, price, and adoption trends.

6. Data Preprocessing

- **Goal**: Predict Electric Range based on features like Make, Model Year, and Vehicle Type.

- **Steps**:

  o Encode categorical variables.

  o Split data into input (X) and target (y).

  o Train-test split (80-20) %.

Expected Deliverables

1. **Cleaned Dataset**: Processed CSV file after handling missing values and outliers.

2. **Jupyter Notebook**: Code for EDA, visualization, and data preprocessing (with markdown explanations).

3. **Streamlit App**: Interactive dashboard deployed locally or on Streamlit Cloud.

4. **Report**: PDF/PowerPoint summarizing insights, and conclusion.

5. **Video:** Short video illustrates code, insights, and challenges.

Tools & Libraries

- **Python**: Pandas, NumPy, Matplotlib, Seaborn, Plotly.

- **Machine Learning**: Scikit-learn (for regression).

- **Dashboard**: Streamlit.