

DSML Capstone Project - README

****Project Title****: Electric Vehicle Population Analysis & Classification

****Dataset****: Washington State Electric Vehicle Population Dataset

****Objective****: Analyze and predict the type of electric vehicle (BEV or PHEV) using various features such as make, model year, MSRP, and electric range.

****Phase 1: Problem Definition****

- Identified business objective: Predict electric vehicle type (BEV or PHEV)
- Chose classification as the modeling task

****Phase 2: Data Cleaning & Preprocessing (Excel)****

- Cleaned and standardized columns
- Removed missing and duplicate rows
- Transformed inconsistent values
- Generated 'cleaned' Excel file for further analysis

****Phase 3: Data Analysis & Visualization (Power BI)****

- Created KPIs: Total EVs, EVs by County, Top Manufacturers
- Used bar charts, pie charts, line graphs for trend analysis
- Built interactive dashboard with dropdown slicers

****Phase 4: Model Building & Performance Evaluation (Python)****

- Used Jupyter Notebook and scikit-learn
- Preprocessed data: feature selection, encoding, train-test split
- Trained Logistic Regression and Decision Tree models
- Evaluated models using accuracy, confusion matrix, classification report
- Visualized distribution, correlations, confusion matrix, and accuracy comparison

****Tools Used**:**

- Excel for data cleaning
- Power BI for dashboarding
- Python (pandas, scikit-learn, matplotlib, seaborn) for modeling

****Final Deliverables**:**

- Cleaned Excel dataset
- Power BI dashboard
- Python model notebook (.ipynb)
- This README PDF