

# Electric Vehicle Data Analysis Assignment

**Name:** Kuppili Durga Maanasa

**Date:** 25-12-2025

**Course:** Data Analysis

## Introduction

This assignment analyzes the Electric Vehicle (EV) registration dataset provided by the Washington State Department of Licensing (DOL). The objective is to clean the data, explore trends, visualize insights, and build a linear regression model to predict electric range.

## Section 1: Data Cleaning

Missing values exist mainly in Base MSRP, Electric Range, Electric Utility, and CAFV Eligibility columns. Zero values in Base MSRP are treated as missing and replaced using median values grouped by model. Duplicate records are identified using VIN or DOL Vehicle ID and removed. VINs are anonymized using hashing to maintain privacy. Vehicle location data is cleaned by extracting latitude and longitude for geospatial analysis.

## Section 2: Data Exploration

Tesla is the most common EV make, with Model 3 and Model Y being the most registered models. King County has the highest number of EV registrations. EV adoption has increased significantly after 2018. The average electric range varies across makes and models, with premium brands offering higher ranges. Urban areas show higher EV adoption compared to rural regions.

## Section 3: Data Visualization

Bar charts are used to show the top EV makes and models. Line graphs display EV adoption trends by model year. Scatter plots illustrate the relationship between Base MSRP and Electric Range. Pie charts represent CAFV eligibility distribution. Geospatial maps visualize EV registrations based on vehicle location.

## Section 4: Linear Regression Model

Linear Regression is used to predict Electric Range using features such as Model Year, Base MSRP, Make, and Vehicle Type. Categorical variables are handled through one-hot encoding. The  $R^2$  score measures the model's prediction accuracy. Results indicate that higher Base MSRP generally corresponds to higher electric range. Model performance can be improved through feature engineering and advanced regression techniques.

## **Conclusion**

The analysis highlights rapid EV adoption, especially in urban counties. Tesla dominates the market, and electric range shows a positive relationship with vehicle pricing. Linear regression provides a foundational predictive model for estimating EV range.

## **Appendix**

The analysis uses Python libraries such as Pandas, Matplotlib, and Scikit-learn.