

Sean McLean

ALY 6140

Module 2

Capstone Project Proposal

Dimension of the Dataset

The dataset that was located on data.gov is called 'Electric Vehicle Population Data' that was uploaded on September 20th, 2024. The dimensions of the dataset are 17 attributes and 205,440 rows.

Rationale for Dataset

I chose this dataset because of the increase in recent years of electric vehicles and how climate change has had a major impact on more people purchasing these types of vehicles. By using this dataset from one state in the country, I will have the opportunity to identify trends and patterns that could be a microcosm of the country's vehicle purchasing behaviors. It also provides me the opportunity to use different types of modeling that will help me solve my questions that I will be investigating for the project.

Questions to Investigate

- What cities and counties have the most electric vehicles in the state of Washington as of 2024?
- How do legislative districts play a role in the amount of electric vehicle owners per district?
- What are the most popular makes and models of electric vehicles?
- What are the most common electric vehicle types in Washington state?
- How does Clean Alternative Fuel Vehicle (CAFV) eligibility affect electric vehicle sales?
- What are the most common electric vehicles by their electric range?
- How can the results from the previous questions help decide where to add more charging stations in the state?

Models

Logistic Regression: Logistic regression can analyze the relationship between categorical variables, like policy presence (e.g., CAFV eligibility), and their impact on EV adoption. It will help assess how legislative districts and policy incentives influence electric vehicle sales.

Descriptive Statistics & Data Visualization: Descriptive statistics and visualizations like bar charts and heatmaps can provide a summary of EV distribution across regions and show the popularity of various makes and models. This approach is crucial for identifying key trends and offering an overall view of the dataset.

Geospatial Analysis: Geospatial analysis can determine optimal locations for charging stations by mapping EV ownership alongside existing infrastructure. It is essential for identifying areas

with growing demand and can guide decisions on resource allocation and infrastructure development.

References

U.S. Department of Energy. (2024, Sept. 20). *Electric vehicle population data*. Data.gov.
<https://catalog.data.gov/dataset/electric-vehicle-population-data>