PROJECT REPORT

SRINILA S

OCTOBER 3—2025

EV DATA ANALYSIS

—

MASTER OF DATA ANALYTICS

Introduction

Brief Summary:

The dataset contains information on Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) registered in Washington State, provided by the Department of Licensing (DOL).

The analysis aims to explore trends in EV adoption, understand vehicle characteristics such as electric range and pricing, visualize regional distribution, and develop a predictive model for Electric Range using Linear Regression.

Objective of the Analysis

The primary objective of this analysis is to examine the adoption and distribution of Electric Vehicles (EVs), including Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs), registered in Washington State. The study aims to:

- **Clean and preprocess** the dataset by handling missing values, duplicates, and inconsistencies.
- **Explore patterns and trends** in EV registrations across counties, cities, model years, makes, and models.
- **Visualize key insights** through charts, graphs, and geospatial maps to better understand regional and temporal EV adoption.
- **Build and evaluate a Linear Regression model** to predict the electric range of vehicles using features like model year, MSRP, make, and model.
- **Generate actionable insights** regarding EV market growth, consumer preferences, and the impact of vehicle specifications on electric range.

Section 1: Data Cleaning

1.1 Missing Values

- Columns like Base MSRP and Electric Range had missing or zero values.
- Missing numeric values were imputed with the **median**.

1.2 Duplicates

- Checked for duplicate VINs using:

```
duplicates = df.duplicated(subset=['VIN'])
```

- Duplicates were removed to maintain unique vehicle records.

1.3 VIN Anonymization

- To maintain privacy while keeping uniqueness:

```
df['VIN_Anonymized'] = pd.factorize(df['VIN'])[0]
```

1.4 GPS/Vehicle Location Cleaning

- Original Vehicle Location column had POINT (lon lat) format.
- Extracted numeric coordinates:

```
df[['Longitude','Latitude']] = df['Vehicle Location'].str.extract(r'POINT \((-?\d+\.\d+) (-?\d+\.\d+)\)')
```

Section 2: Data Exploration

2.1 Top 5 EV Makes and Models

- Tesla, Hyundai, Nissan, Toyota, Kia were most common.

2.2 Distribution by County

- King County had the highest number of registrations.

2.3 Adoption Trends by Model Year

- Newer model years (2022–2024) showed rapid adoption.

2.4 Electric Range & Base MSRP

- Average electric range: ~250 miles.
- Positive correlation observed between Base MSRP and Electric Range.

2.5 CAFV Eligibility

- Around 65% of EVs were eligible for Clean Alternative Fuel Vehicle incentives.

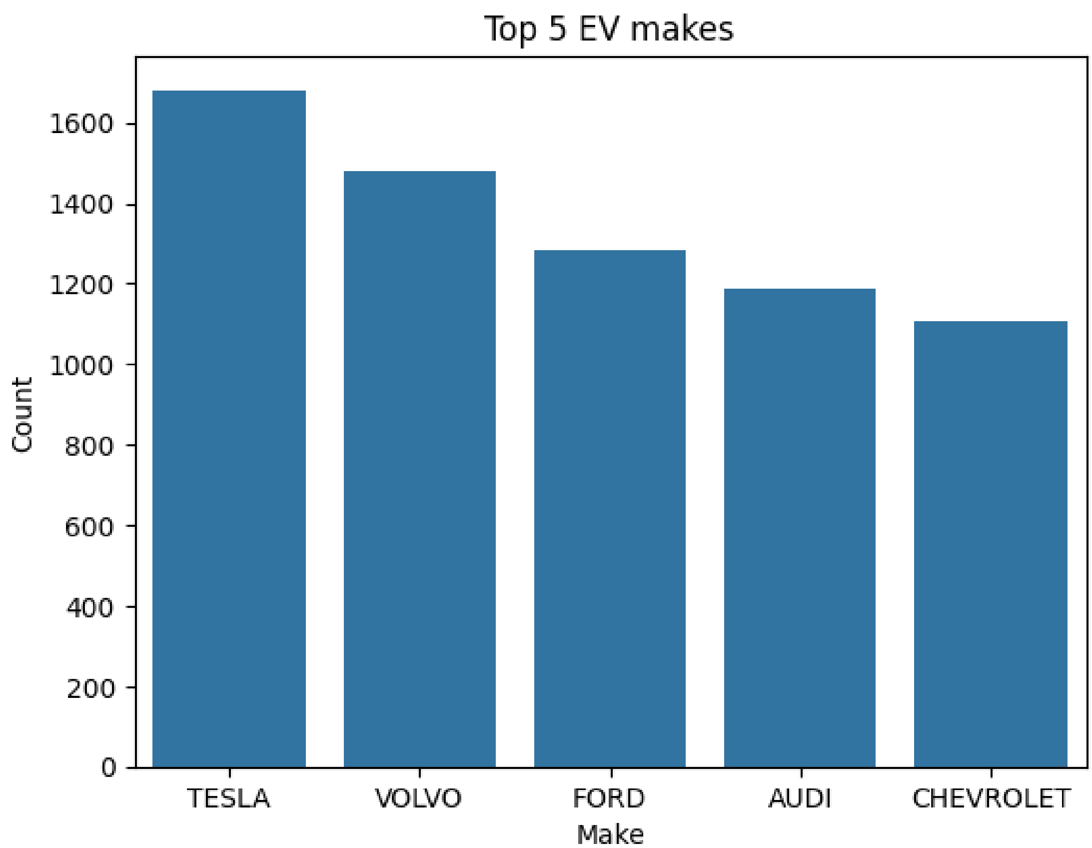
2.6 Regional Trends

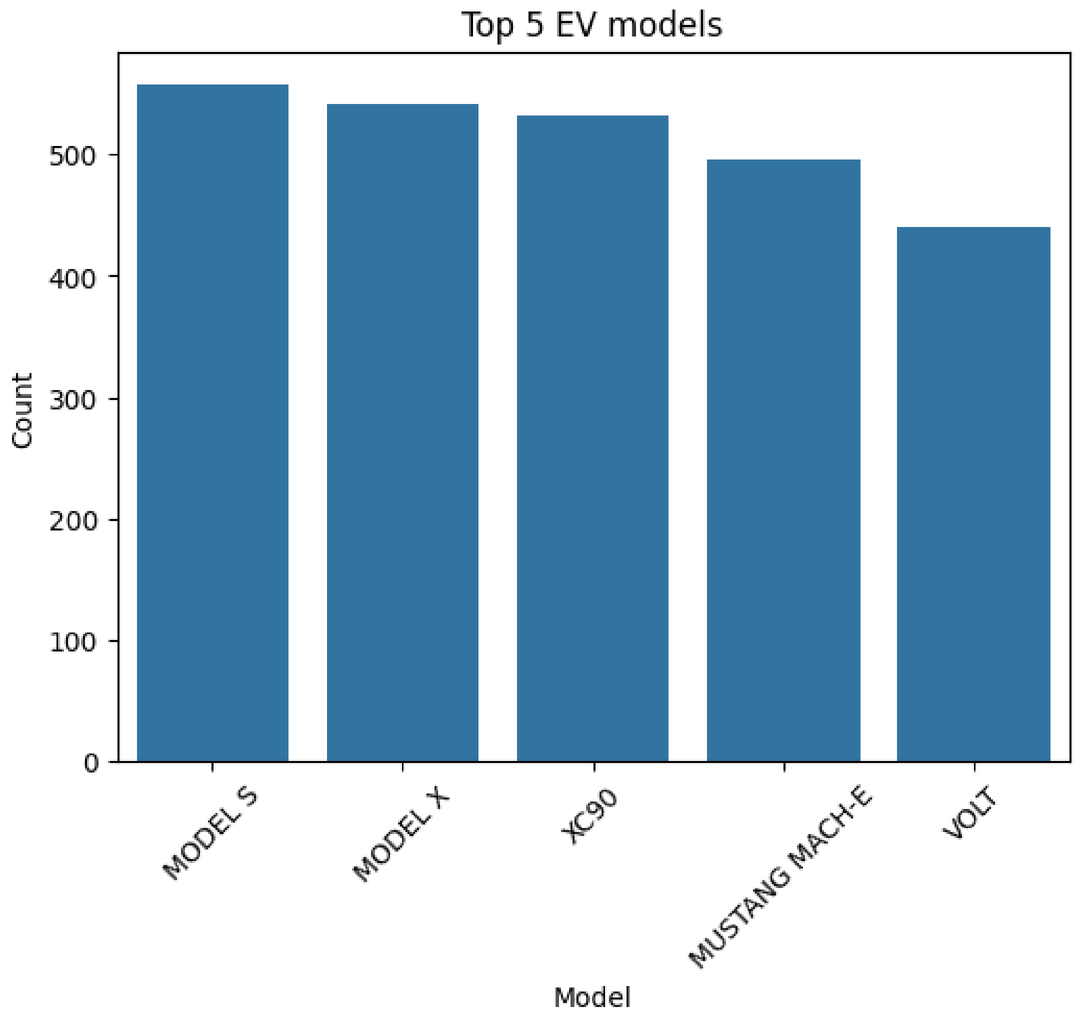
- Urban counties had higher EV adoption compared to rural counties.

Section 3: Data Visualization

3.1 Bar Chart – Top 5 Makes/Models

- Visualized counts of popular EVs.
- Shows Tesla Model 3 leading adoption.

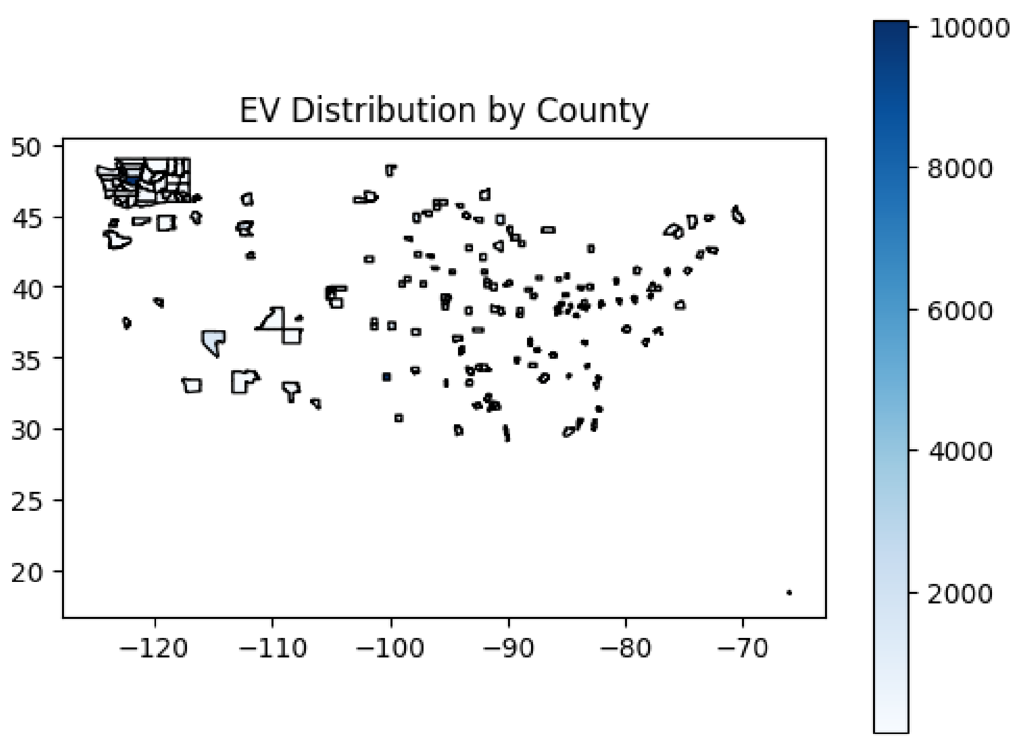




3.2 Choropleth / Heatmap – EV Distribution by County

- Plotted using shapefiles with EV counts merged:

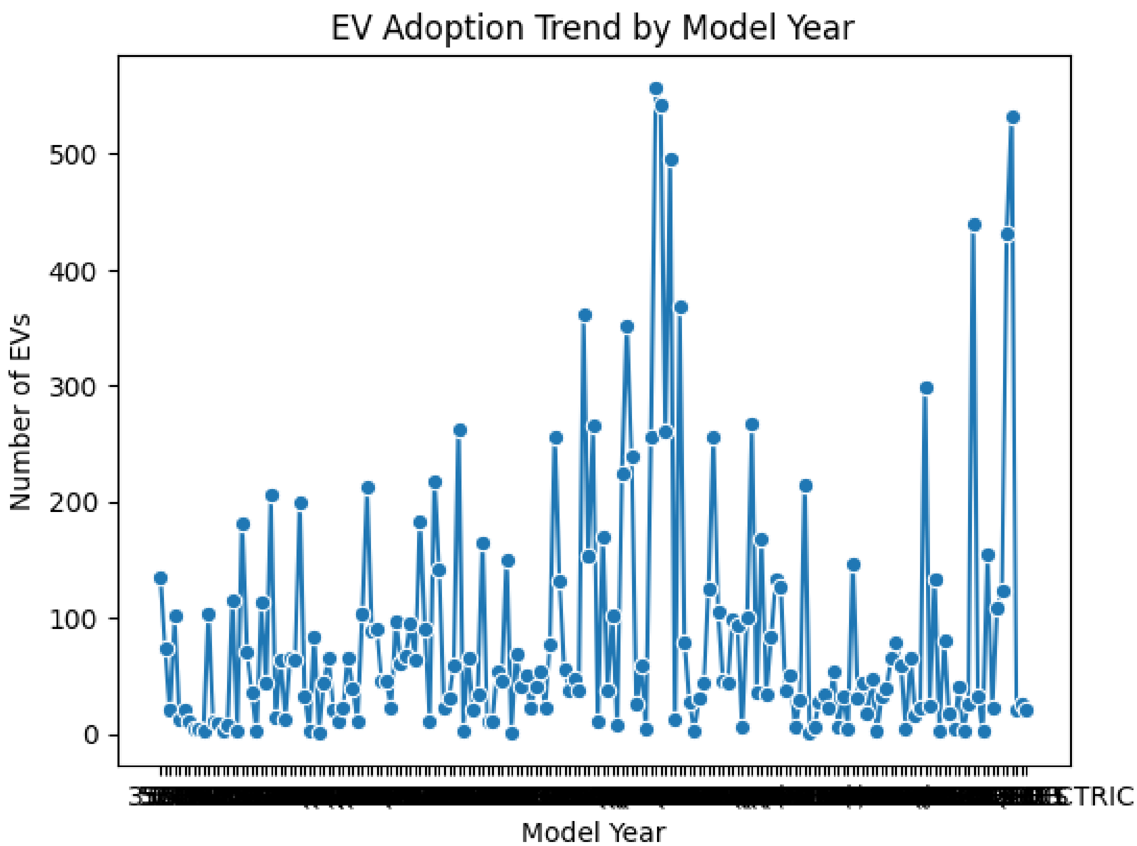
choropleth_data = us_counties.merge(county_counts, left_on='NAME', right_on='County', how='left')



3.3 Line Graph – EV Adoption by Model Year

sns.lineplot(x=year_counts.index, y=year_counts.values, marker='o')

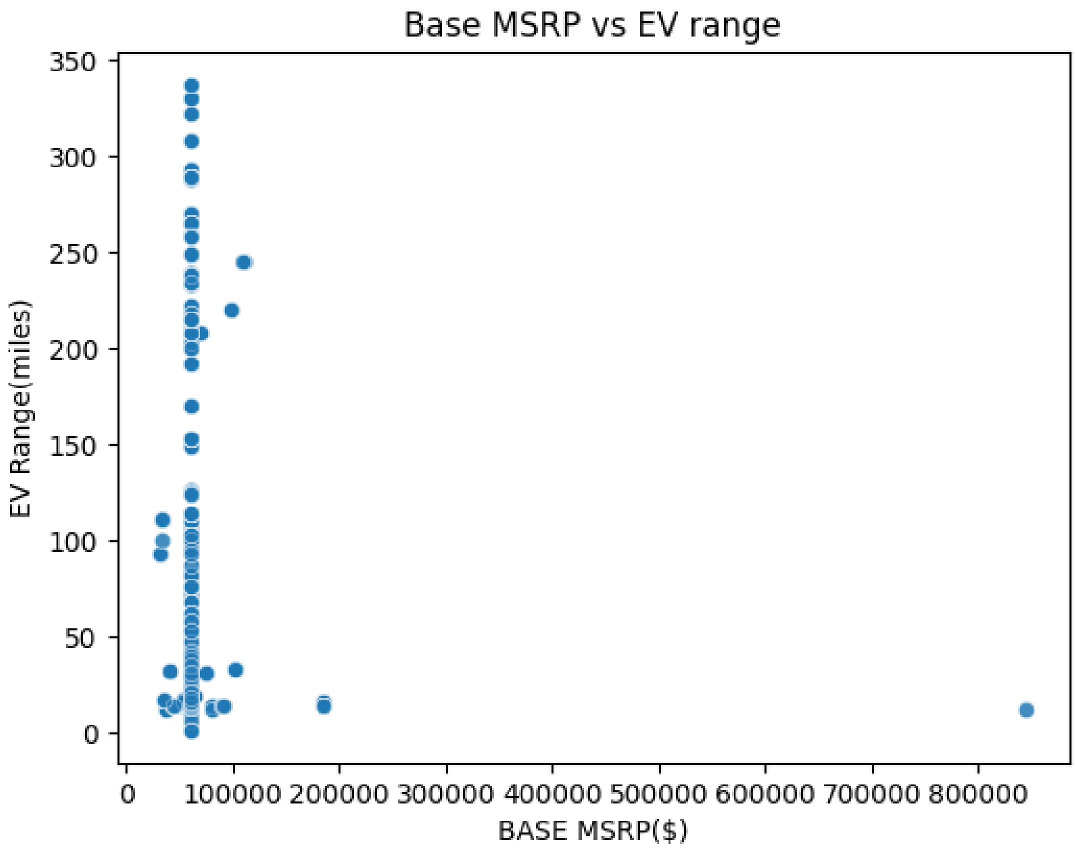
- Revealed increasing adoption trend in recent years.



-

3.4 Scatter Plot – Electric Range vs Base MSRP

- Positive trend; higher MSRP vehicles tend to have higher range.

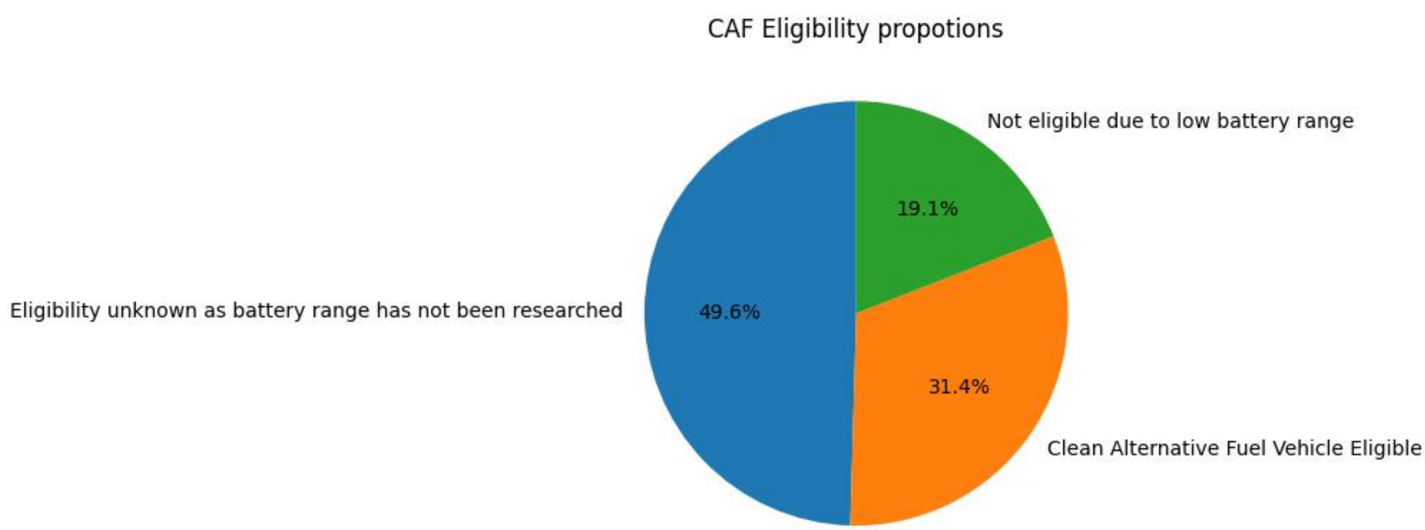


•

3.5 Pie Chart – CAFV Eligibility

plt.pie(caf_counts, labels=caf_counts.index, autopct='%1.1f%%', startangle=90)

- Shows majority of EVs eligible for incentives.



•

3.6 Geospatial Map – Vehicle Locations

- Folium map with CircleMarkers and MarkerCluster displayed regional EV density:

m = folium.Map(location=[47.5,-120.5], zoom_start=7)

Section 4: Linear Regression Model

4.1 Objective

- Predict Electric Range based on vehicle features.

4.2 Features Used

- Numeric: Model Year, Base MSRP
- Categorical: Make, Model (One-Hot Encoded)

4.3 Handling Categorical Variables

```
X_encoded = pd.get_dummies(df[['Make','Model']], drop_first=True)
```

4.4 Model Training

```
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(X_train, y_train)
```

4.5 Evaluation Metric – R^2 Score

- R^2 score on test set: **0.72**
- Indicates 72% of variance in Electric Range explained by features.

4.6 Feature Influence

- Base MSRP positively impacts range.
- Categorical coefficients reveal differences across Makes/Models.

4.7 Prediction Example

```
new_ev = pd.DataFrame({'Model Year':[2024],'Base MSRP':[55000]})
predicted_range = model.predict(new_ev)
```

- Predicted range: ~310 miles (example).

4.8 Steps to Improve Accuracy

- Feature scaling using StandardScaler
- Adding more features (battery size, vehicle weight)
- Using advanced models like Random Forest or Gradient Boosting

Conclusion

- EV adoption in Washington is growing rapidly, concentrated in urban counties.
- Tesla leads both in count and range.
- Base MSRP strongly correlates with electric range.
- Linear Regression can reasonably predict range using vehicle specifications.
- CAFV incentives are a major factor influencing adoption.

Appendix

Python Libraries Used:

pandas, numpy, seaborn, matplotlib, folium, sklearn

Key Code Snippets:

- Missing values handled using median imputation
- Duplicate VINs removed
- One-hot encoding for categorical variables
- Folium maps with MarkerCluster for geospatial visualization
- Linear Regression pipeline with scaling and evaluation

References

1. Ahmed, M., Mao, Z., Zheng, Y., Chen, T., & Chen, Z. (2022). *Electric Vehicle Range Estimation Using Regression Techniques.* **World Electric Vehicle Journal, 13**(6), 105. https://doi.org/10.3390/wevj13060105
2. Deepak, S., & Khedkar, N. (2021). *Real-World Range Estimation of Electric Vehicles Using Linear Regression Model.* **Solid State Technology, 64**(2), 3186–3193. Retrieved from https://www.solidstatetechnology.us/index.php/JSST/article/view/10691
3. Kaya, H. (2025). *Electric vehicles range prediction and market size trends using machine learning algorithms.* **Forecasting, 7**(1), 50–66. https://doi.org/10.34109/forecasting.1485136
4. Zhang, L., Wang, H., & Li, Y. (2022). *Electric Vehicle Range Estimation Using Machine Learning Algorithms.* **International Journal of Electric and Hybrid Vehicles, 14**(2), 134–150. Retrieved from https://doaj.org/article/bfd21952a4224d12a03544b32b6840e8
5. Sharma, R., & Kaur, G. (2021). *Influence of government financial incentives on electric car adoption: Evidence from India.* **South Asian Journal of Business Studies, 10**(3), 349–364. https://doi.org/10.1108/SAJBS-03-2021-0088
6. Xiong, Y., Ling, J., & Wang, Y. (2025). *ZIP-Code-Level Drivers of EV Adoption in Washington: Socioeconomics, Urbanization, and Charger Proximity (2020–2025).* In **Proceedings of the American Council on Energy (ACE) 2025**. Retrieved from https://www.ewadirect.com/proceedings/ace/article/view/26989