

# ELECTRIC VEHICLES (EV's)

## IN WASHINGTON STATE

- **PART I:** Visual Data Exploration



- **PART II:** Linear Regression Model

Jesse Mazariegos  
Sheena Knott

- **PART III:** Random Forest Classification

# Understanding EV Adoption in State of Washington



The goal of this study is to create models to understand the practices and incentives of the population of the state of Washington regarding the use of EVs and to implement the main factors involved in predicting its future trend.

Understanding how various demographic, economic, and geographic factors influence EV ownership is key to predicting future adoption patterns and informing policy making.

## Data Choice

- CAFVs will be the driving choice of tomorrow
- Washington presents a ideal sample set for study
- Moderately Large dataset (~250k instances)
- Combination of Numerical & Categorical Features

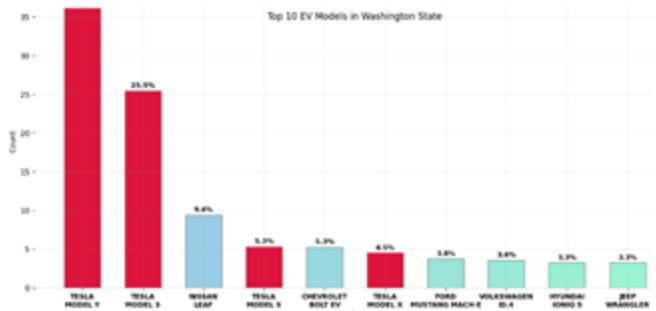
## Caveats

- Imbalanced Data
- Missing Data (~ Missing MSRP values)

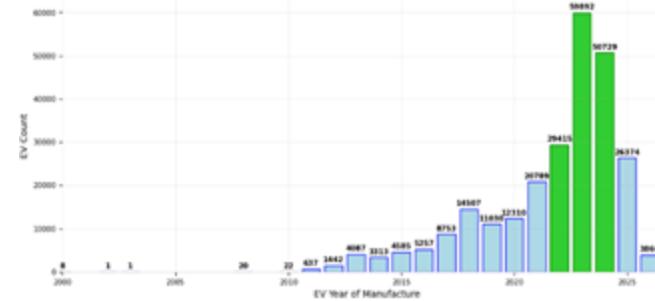
## Glossary

- BEV- Battery Electric Vehicle
- CFAV – Clean Alternative Fuel Vehicle
- EV- Electric Vehicle
- PHEV- Plug-in Hybrid Electric Vehicle
- MSRP- Manufacturer Suggested Retail Price

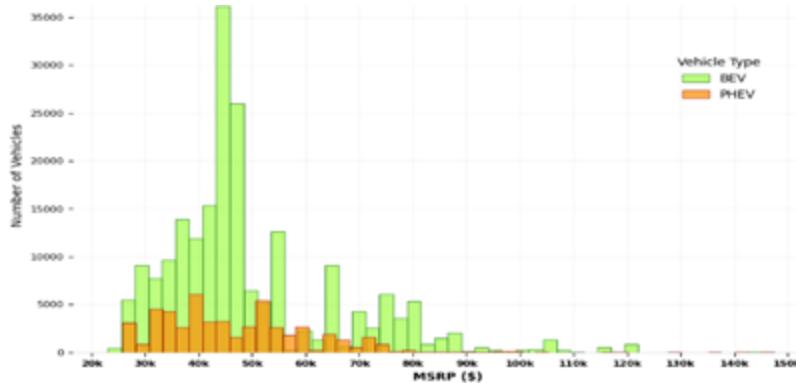
## Top 10 EV Models in WA



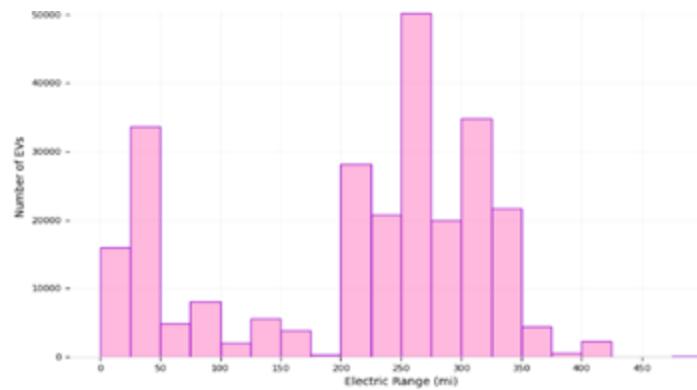
## EV Count by Year in WA



## Base MSRP Distribution



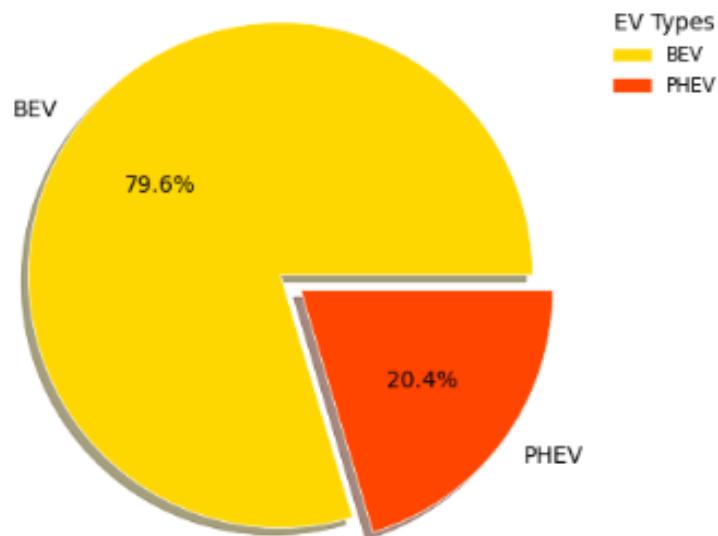
## EV Electric Range Distribution



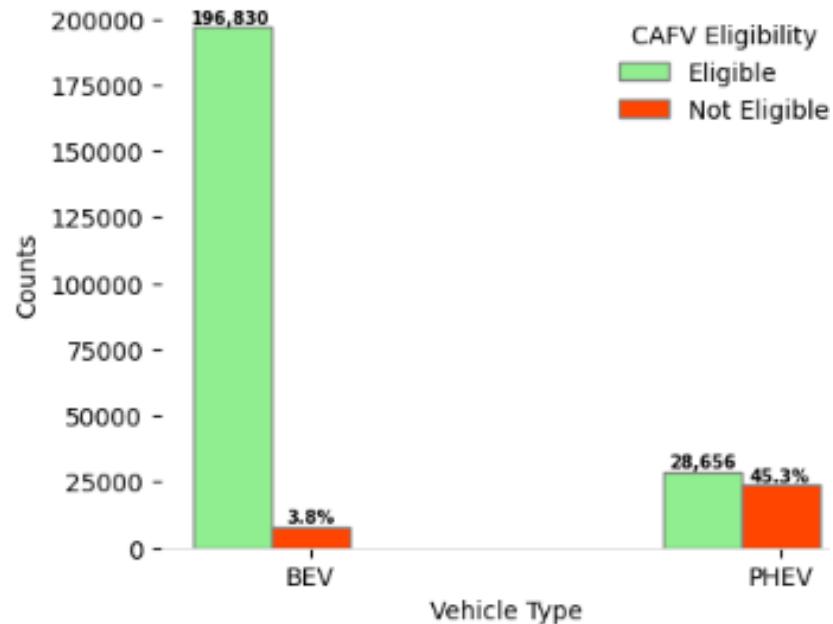
## PART I -Understanding Data

# Percentages of EV Types, CAFV Eligibility

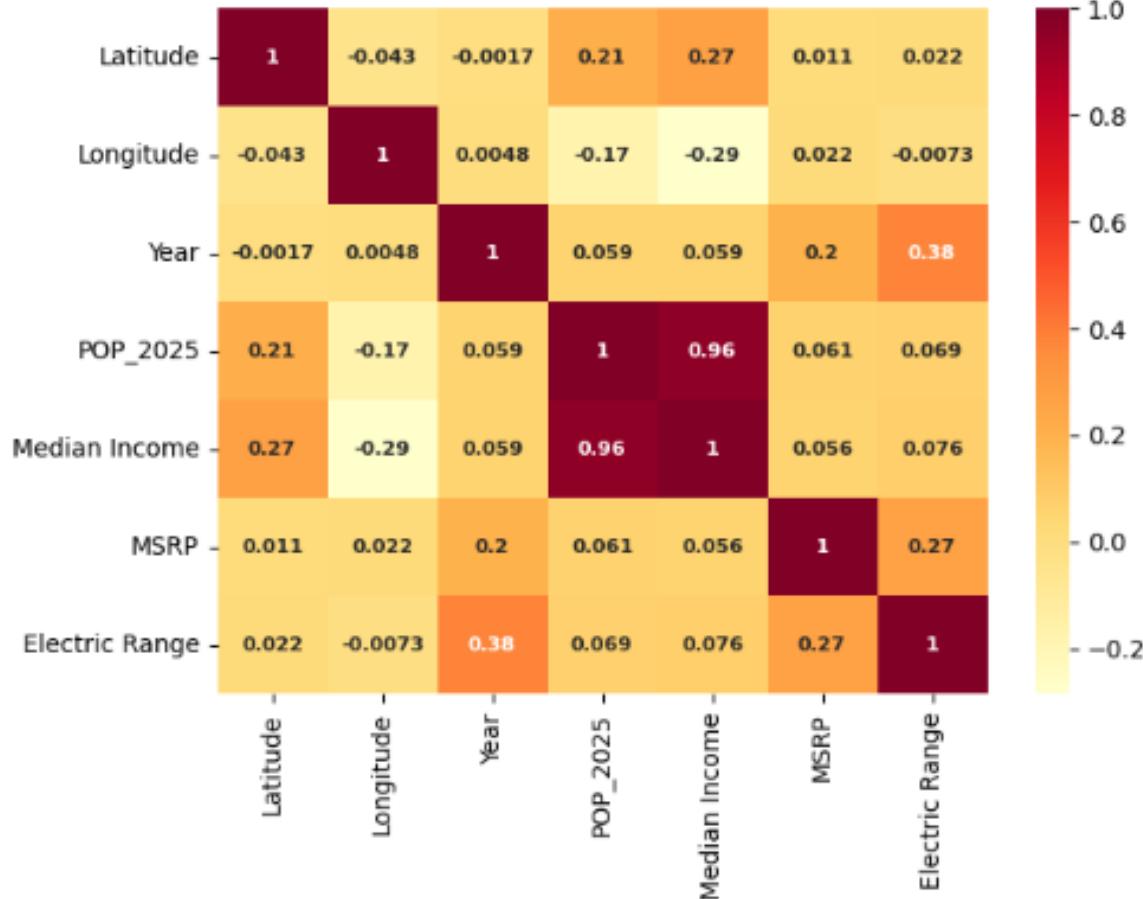
Battery EVs and Plug-In Hybrid EVs



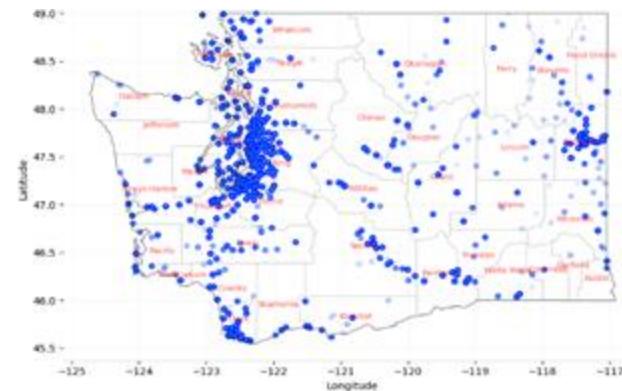
BEV, PHEV counts & CAFV Eligibility



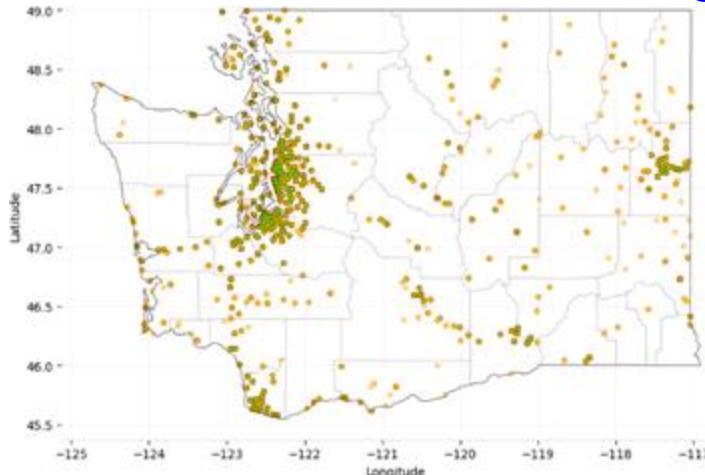
## Correlation Heatmap for Features in EV Dataset



## EV Locations in Washington State



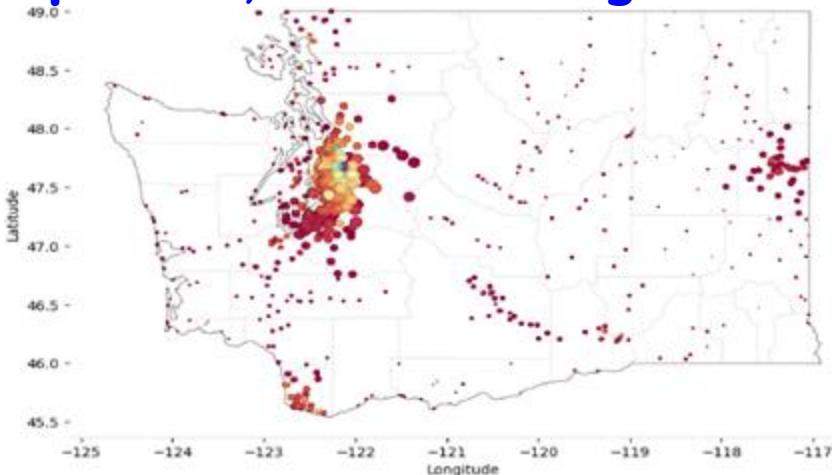
## BEV, PHEV Locations in Washington State



## Counties in Washington State



## Population, EVs in Washington State



# PART II - Linear Regression



I selected Linear Regression because it provides a clean, interpretable baseline to compare against more complex models.

- **R<sup>2</sup>**: Proportion of variance explained. Higher is better.
  - Formula:  $R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$
  - **RMSE**: Mean Squared Error penalizing larger errors.
  - Formula:  $RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2}$
  - **MAE**: Mean Absolute Error; easier to interpret in miles.
  - Formula:  $MAE = \frac{1}{n} \sum_i |y_i - \hat{y}_i|$

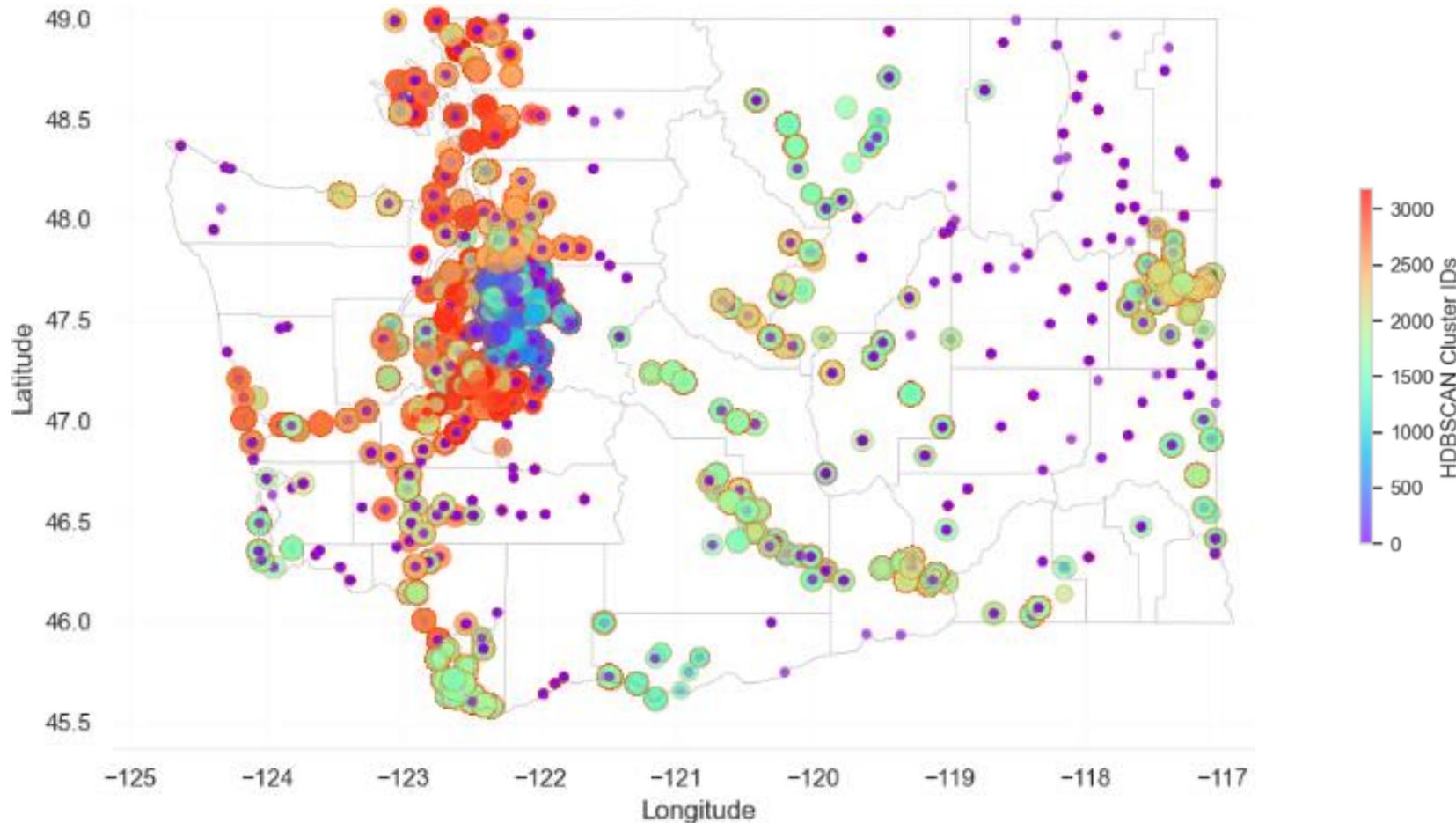
Fold	R <sup>2</sup> Score	RMSE	MAE	MSE	Support
1	0.8955	35.5948	26.0613	1266.99	41,126
2	0.8953	35.6228	26.1126	1268.98	41,126
3	0.8984	35.2379	26.0702	1241.71	41,126
4	0.8968	35.4948	26.1828	1259.88	41,126
5	0.8965	35.4527	26.1587	1256.89	41,126



# PART III

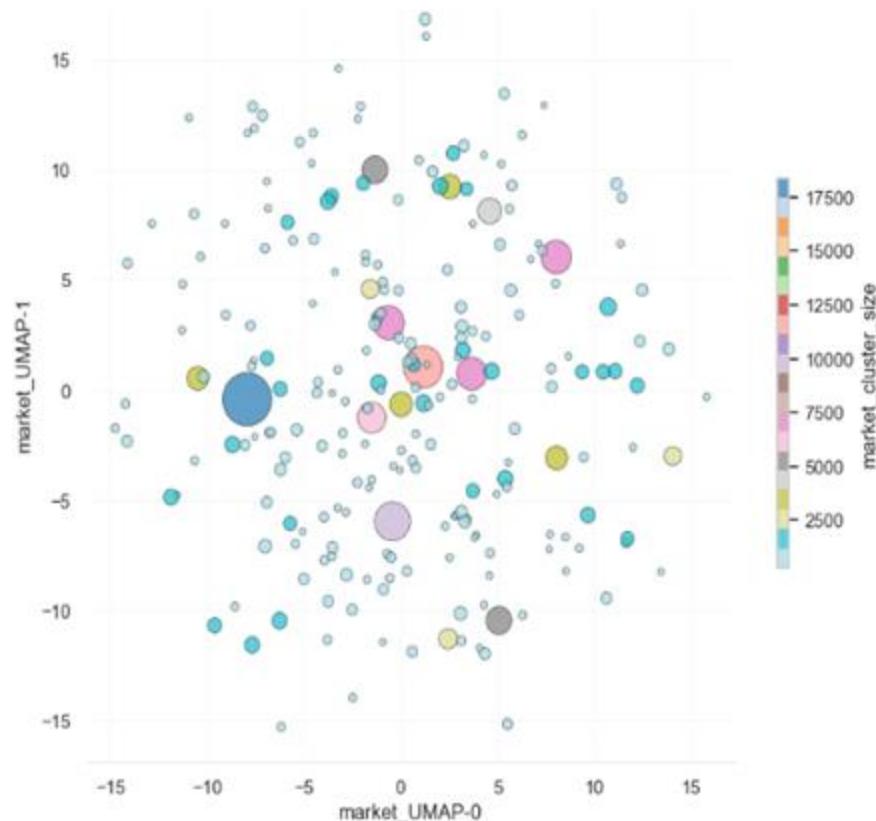
**Clustering  
Using  
Random Forest Classification**

# HDBSCAN Clusters : 3184



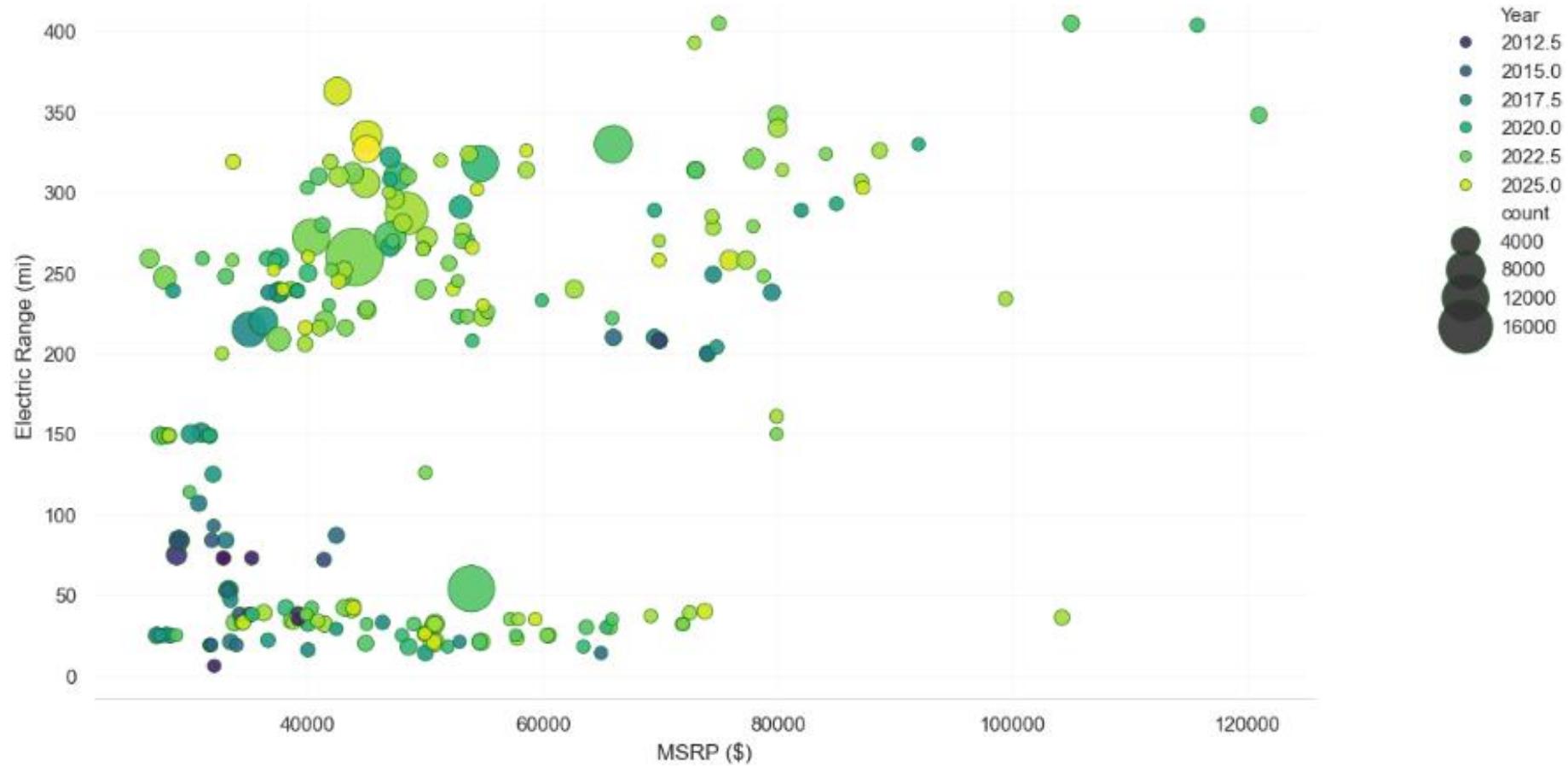
## Market Segmentation - based on Vehicle Characteristics

market_cluster	MSRP	Electric Range	Year	count
29	43990.0	260.0	2023.0	18371
-1	53900.0	54.0	2022.0	11603
4	48365.0	287.0	2024.0	9800
57	65990.0	330.0	2022.0	7496
17	40240.0	272.0	2023.0	7150
36	54657.0	318.0	2021.0	6951
152	35000.0	215.0	2018.0	6212
47	46990.0	272.0	2022.0	4982
10	44990.0	335.0	2025.0	4971
6	44865.0	306.0	2024.0	4272
39	36200.0	220.0	2019.0	3898
31	47657.0	310.0	2021.0	3883
24	42490.0	363.0	2025.0	3662
2	44990.0	327.0	2026.0	3539
51	37495.0	209.0	2023.0	2692

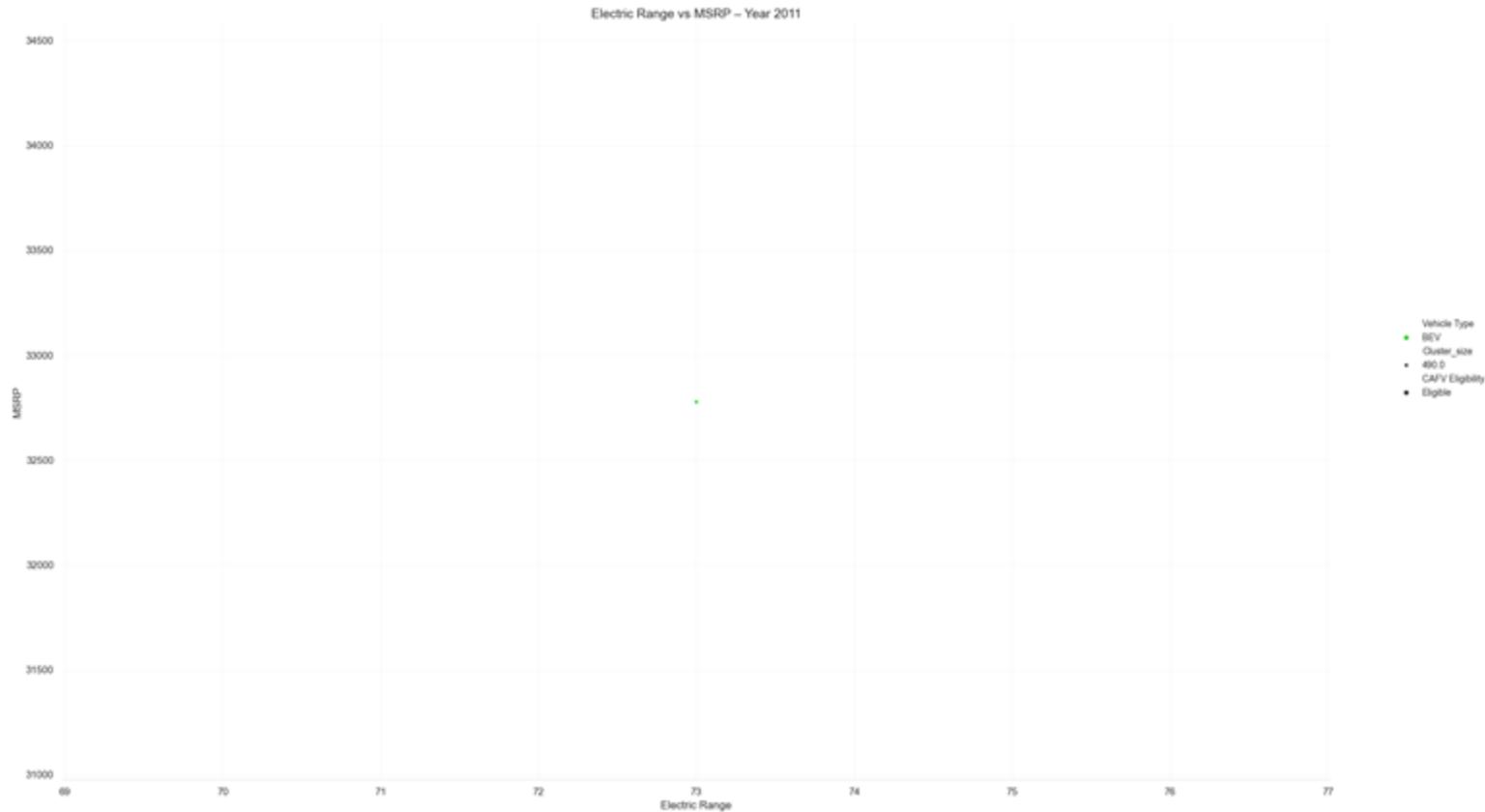


Market Segmentation Cluster around Centroids in UMAP Space

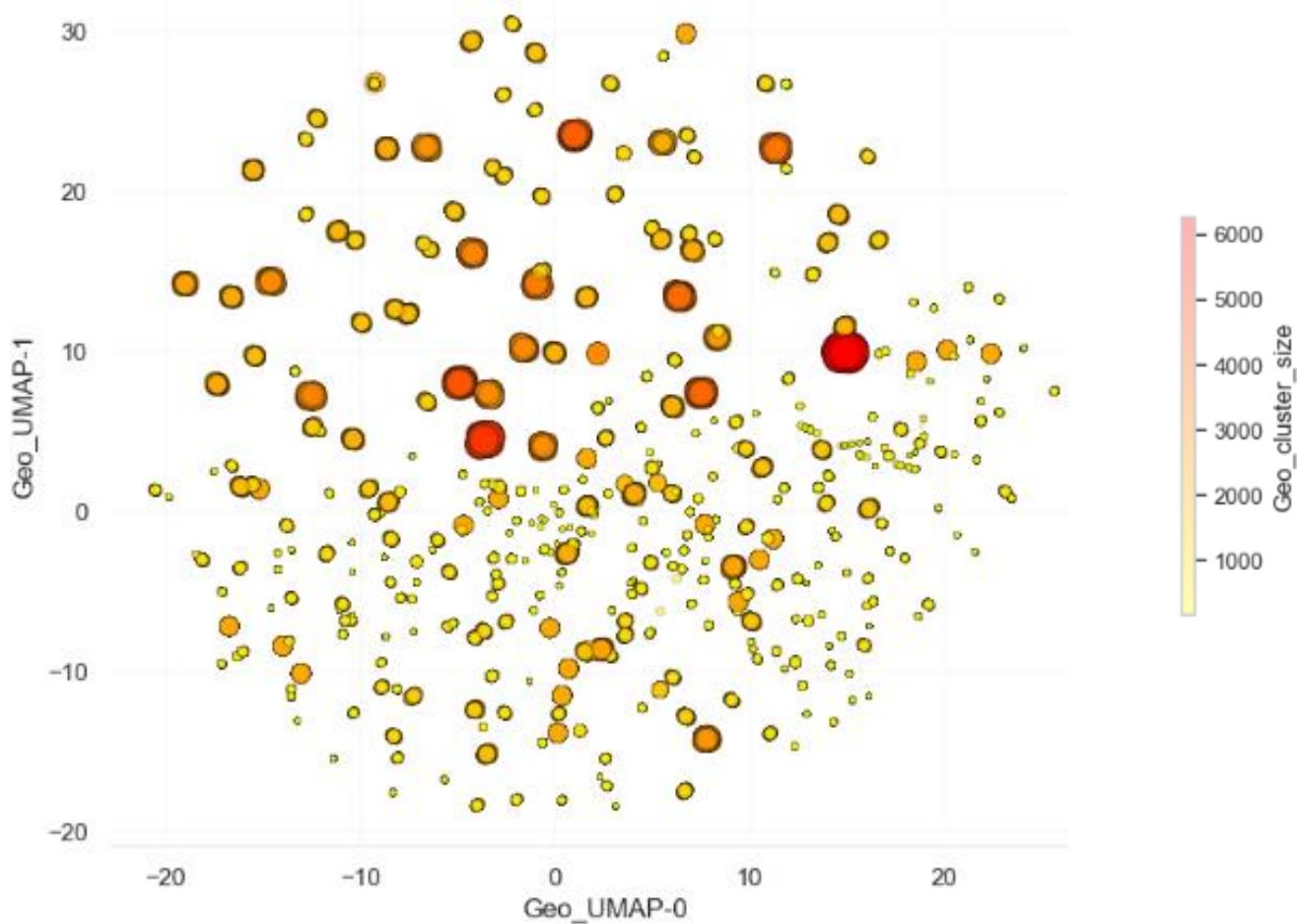
# Clusters: MSRP vs. Electric Range



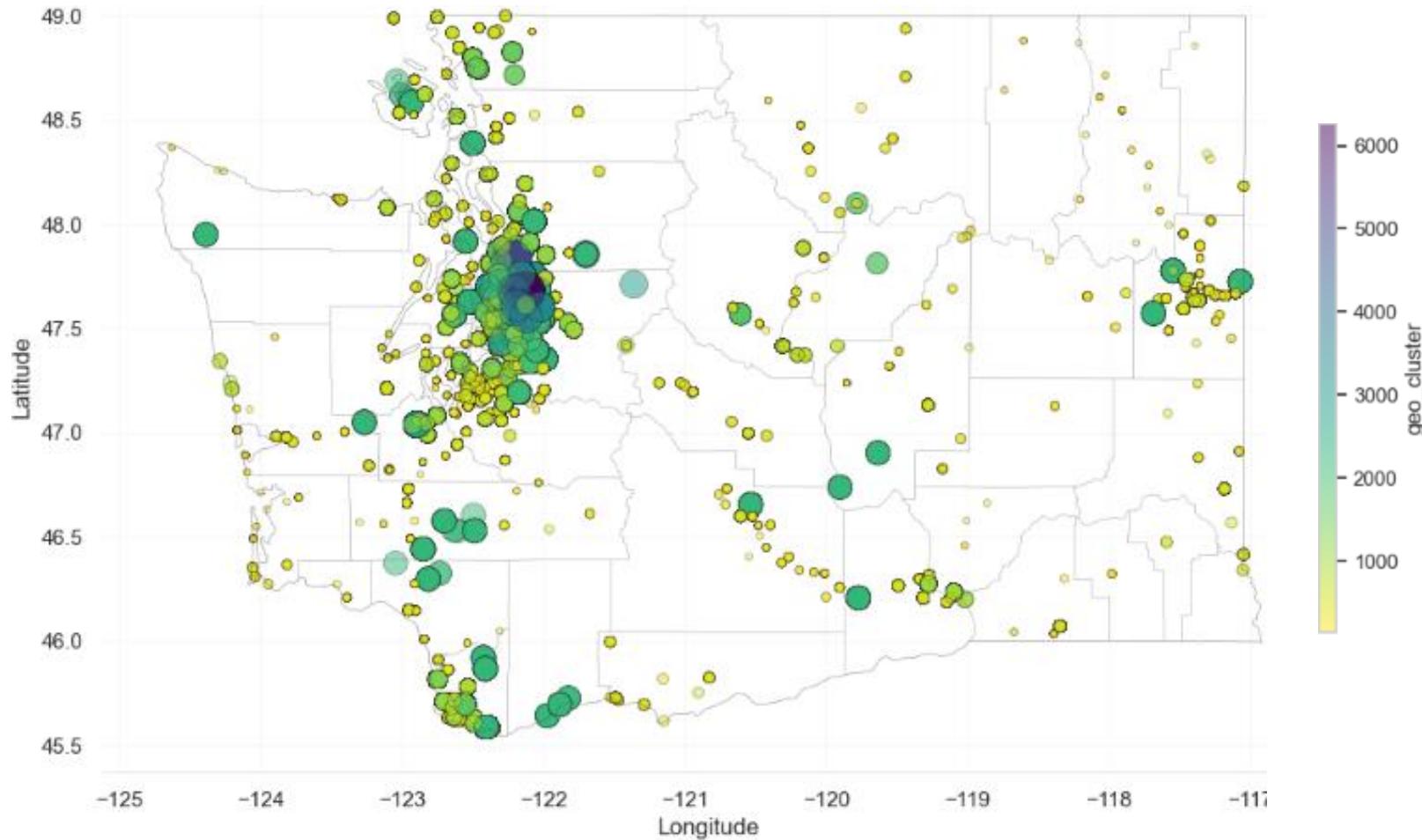
# Cluster MSRP vs. Electric Range



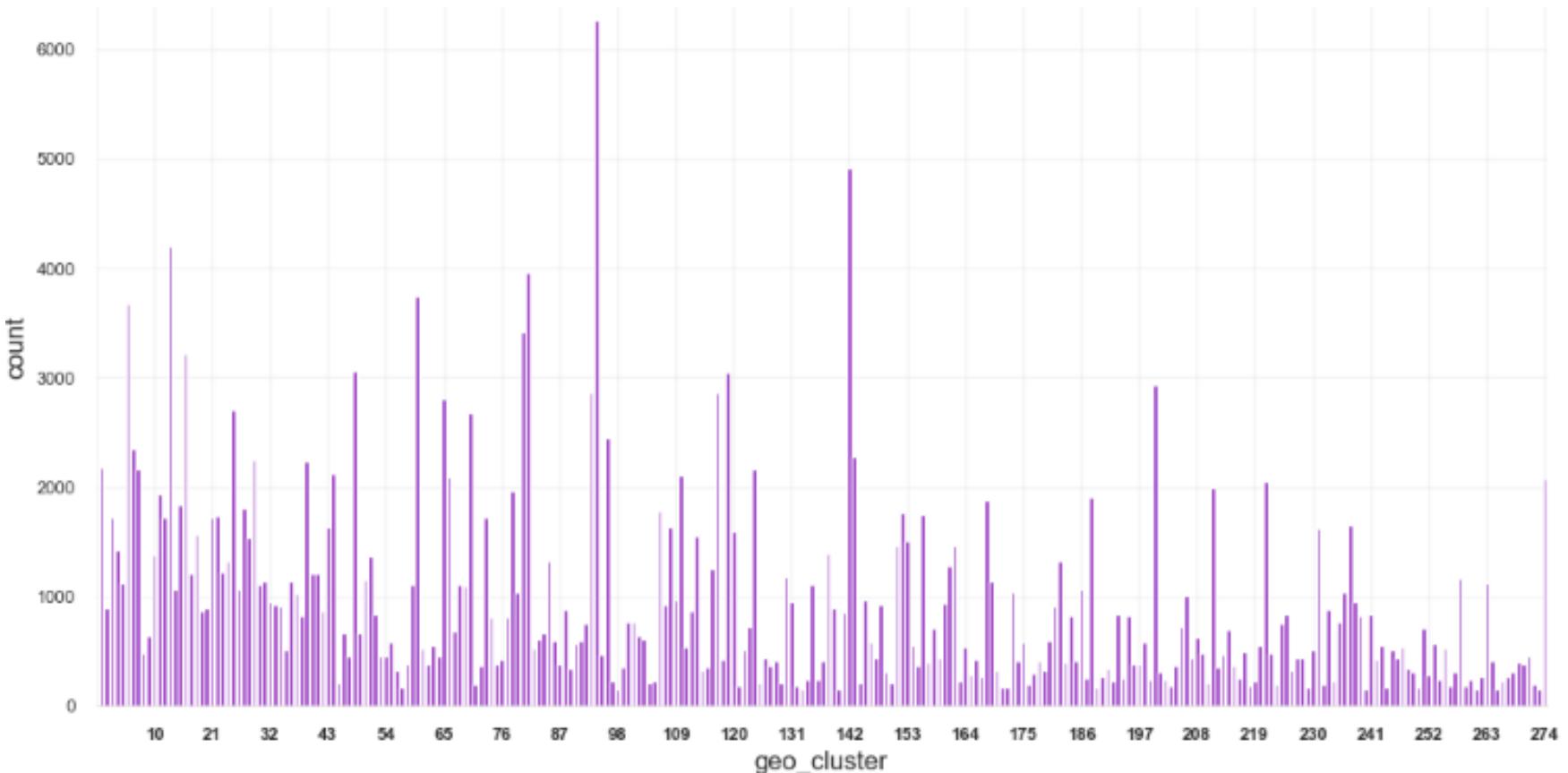
# Geographic Clusters in UMAP Space



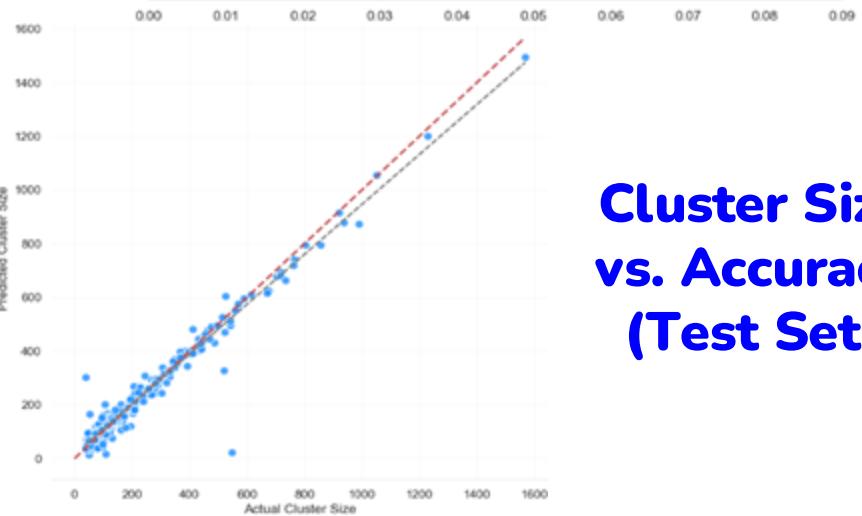
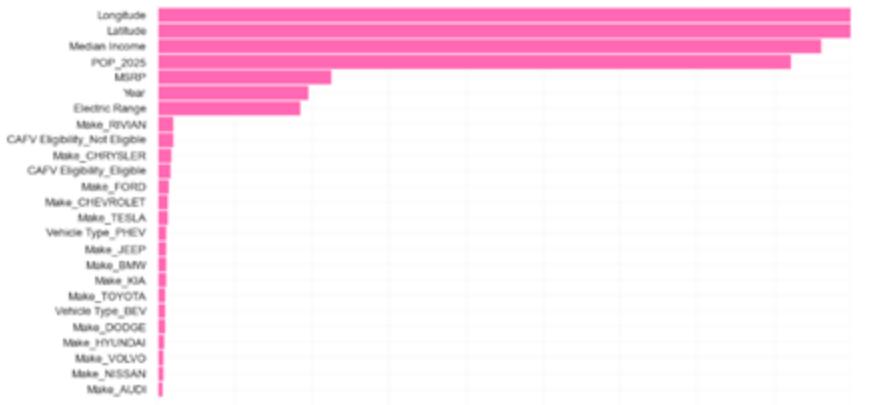
# Geo Clusters in Physical Space



# HDBSCAN Cluster Labels, Size Distribution for Geographic Segmentation

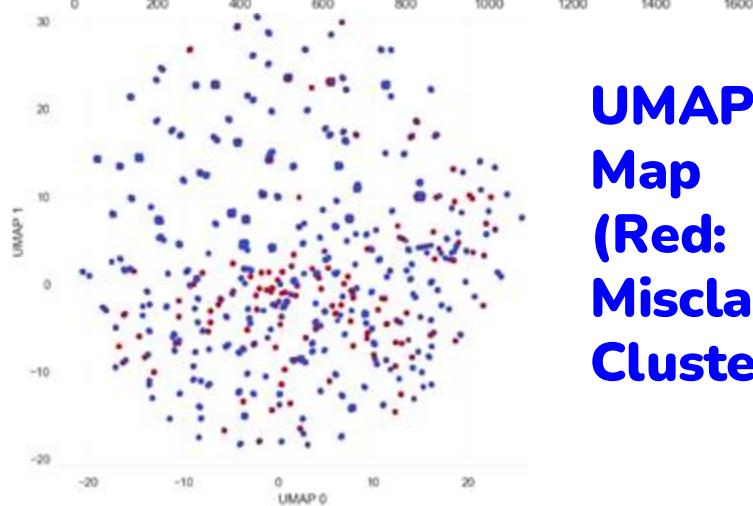
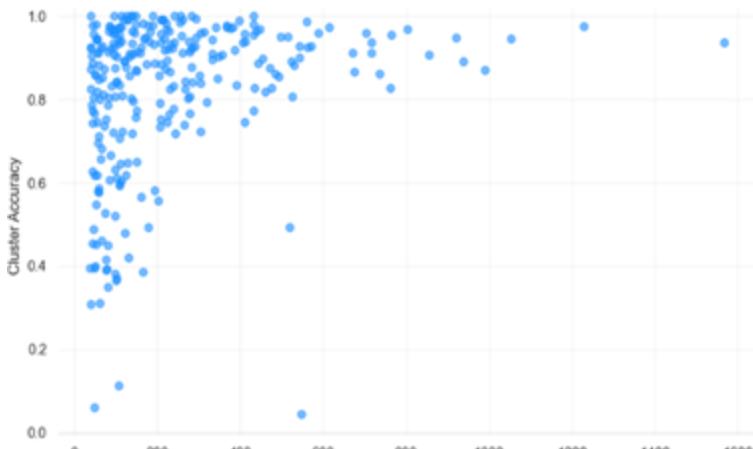


## Top 25 Features by Importance



**Cluster Size  
vs. Accuracy  
(Test Set)**

## Cluster Size vs. Accuracy (Test Set)



**UMAP Error  
Map  
(Red:  
Misclassified  
Clusters)**

# Random Forest Classification Report

## Cross Validation + Testing

# Linear Regression Report

		Precision	Recall	F1-Score	Support
Cross Validation Fold 1	Accuracy			0.88	38556
	Macro Avg	0.81	0.83	0.81	38556
	Weighted Avg	0.89	0.88	0.87	38556
Cross Validation Fold 1	Accuracy			0.87	38556
	Macro Avg	0.81	0.83	0.81	38556
	Weighted Avg	0.89	0.87	0.87	38556
Cross Validation Fold 1	Accuracy			0.87	38556
	Macro Avg	0.80	0.82	0.80	38556
	Weighted Avg	0.89	0.87	0.87	38556
Cross Validation Fold 1	Accuracy			0.87	38556
	Macro Avg	0.80	0.83	0.80	38556
	Weighted Avg	0.88	0.87	0.87	38556
Cross Validation Fold 1	Accuracy			0.88	38556
	Macro Avg	0.81	0.83	0.81	38556
	Weighted Avg	0.89	0.88	0.88	38556
Test Set	Accuracy			0.88	64260
	Macro Avg	0.82	0.84	0.82	64260
	Weighted Avg	0.90	0.88	0.88	64260

Fold	R² Score	RMSE	MAE	MSE	Support
1	0.8955	35.5948	26.0613	1266.99	41,126
2	0.8953	35.6228	26.1126	1268.98	41,126
3	0.8984	35.2379	26.0702	1241.71	41,126
4	0.8968	35.4948	26.1828	1259.88	41,126
5	0.8965	35.4527	26.1587	1256.89	41,126



# CONCLUSION

- RF Classification has high accuracy: 0.88
  - Clusters discovered by HDBSCAN are learnable.
  - Small sized error clusters are spread across embedding.
  - Geo-Socio-Economic factors dominate EV adoptions.
  - BEVs dominate the market.
- 
- Linear Regression has high accuracy: 0.8956
  - Year and MSRP strongly influence predictions
  - Predicted ranges vary meaningfully across different locations

# RESOURCES



- [catalog.data.gov – Electric Vehicle Population Data](https://catalog.data.gov/dataset/electric-vehicle-population-data)
- [catalog.data.gov – April 1 Population by State, County and City \(1990–Present\)](https://catalog.data.gov/dataset/april-1-population-by-state-county-and-city-1990-present)
- [hdpulse.nimhd.nih.gov](https://hdpulse.nimhd.nih.gov)
- [insideevs.com – Electric Car Prices in the U.S.](https://insideevs.com/electric-car-prices-in-the-u-s/)
- <https://afdc.energy.gov/laws/6216#:~:text=Washington%20adopted%20the%20California%20vehicle%20emissions%20standards,see%20the%20Washington%20Clean%20Car%20Standards%20website>