

# **Open Source Project: Electric Vehicles in Washington**

## **Data Source**

The data set is sourced from data.gov, however it is a non-federal dataset with different terms of use. Last updated on March 14, 2025. It is state data from a trustworthy source.

<https://catalog.data.gov/dataset/electric-vehicle-population-data>

Landing page has some explanations about the data.

[https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2/about\\_data](https://data.wa.gov/Transportation/Electric-Vehicle-Population-Data/f6w7-q2d2/about_data)

## **Data Collection**

The administrative data is collected by Washington State Department of Licensing (DOL) from its own records as well as sourcing data externally. It shows the number of electric vehicles that were registered by the DOL each month. DOL integrates National Highway Traffic Safety Administration (NHTSA) data and the Environmental Protection Agency (EPA) fuel efficiency ratings with DOL titling and registration data to create this information.

## **Interest in the Data**

There is much discussion about electric vehicles being the future or already being a thing of the past. With predictions that half our cars will be EVs by 2035, does the current uptake suggest this will be true or is the hype suggestive? The Washington State data can serve as a sample to test these assumptions.

## **Data Profile**

### **Data Cleaning**

#### *Mixed data types*

Changed mixed data types on four columns ['County'], ['City'], ['Vehicle Location'], ['Electric Utility'] to 'str'

#### *Missing values*

Column Name	Count of missing values	Changes made
Postal Code	3	Removed these 3 rows
Electric Range	36	Imputed ER 50
Base MSRP	36	Imputed 88,392
Legislative District	494	Removed column as not relevant to analysis
2020 Census Tract	3	*same rows as Postal Code

#### *Notes on Imputed values*

The missing values in the Electric Range column corresponded with missing Base MSRP values. See Table for breakdown of the 36 missing values. Open research was conducted to find values to impute because the mean, median from the df are not fit for purpose. Imputing 36 values should have minimum impact on the analysis results.

<i>Count</i>	<i>Model Year</i>	<i>Make</i>	<i>Model</i>	<i>Electric Range *</i>	<i>Base MSRP*</i>
9	2025	PORSCHE	Panamera	30	\$104,795
21	2025	PORSCHE	Cayenne	60	\$86,695.00
1	2025	MERCEDES	SL-Class	7	\$113,100
5	2025	MERCEDES	GLC-Class	54	\$61,050
36	-	-	MEAN	50	\$88,392

\*Values from caranddriver.com

*Duplicate values*

No duplicate values were found.

## Data profile

### Numerical

Column name	Electric Range	Base MSRP
Value count	235689	235689
Mean	46.262812	800.432846
Std	84.039283	7257.550227
Min	0.00	0.00
25%	0.00	0.00
50%	0.00	0.00
75%	38	0.00
Max	337	845000*
Data Type	float64	float64

\*Outlier suspect should be \$84,500.

## Categorical

Column name	Description	Data type
VIN (1-10)	Vehicle Identity Number, unique values.	object
County	County name where car was previously registered.	object
City	City name where the car was previously registered.	object
State	Two letter code for State where car was previously registered.	object
Postal Code	Zip Code where the car was previously registered.	float64
Model Year	Year car manufactured from 2000 to 2025	int64
Make	Name of car brand (46 different makes)	object
Model	Model name	object
Electric Vehicle Type	Battery Electric Vehicle (BEV) or Plug-in Hybrid Electric Vehicle (PHEV)	object
Clean Alternative Fuel Vehicle (CAFV) Eligibility	Eligibility unknown as battery range has not been researched OR Clean Alternative Fuel Vehicle Eligible OR Not eligible due to low battery range	object
Legislative District	Location by code number	float64
DOL Vehicle ID	Department of Licensing Vehicle ID number, unique values.	int64
Vehicle Location	Geopin/ geographic coordinates	object
Electric Utility	Name of Electric provider (76 utilities)	object
2020 Census Tract	11 digit Census Tract code	float64

## Data Limitations and Ethics

1. Incomplete Data Entry: Some values unavailable to the data provider have not been recorded or have been set to 0 (Base MSRP).

2. Missing or Inconsistent Data: Although the dataset contains car Model Years it omits the date of registration which means the market segmentation of new and used is not available. We can't determine how many EVs are registered each year without registration dates. We also don't know how many electric vehicles are deregistered.

3. Collection Bias: Data contains only Electric Vehicles and therefore lacks any control variables (the project can't compare trends with the wider market (eg. if there is a fall in fossil-fuel cars or an increase in public transport use.) The dataset contains vehicles for personal use only omitting insights into logistics, trucking, and bus transportation.

4. Ethical considerations: This project is being conducted for educational purposes and uses open source data. The analysis will need to be mindful of the bias outlined above in order to report findings within the data limitations and remain transparent.

### **Questions to guide the analysis**

Have the number of EVs in Washington increased over time?

Have the makes of EVs changed over time?

Where are the EVs located? Are these places urban, suburban or rural?

What has changed over time?

Where do the cars that are not first registered in Washington come from?

What is the age range of the EVs?

Where are the oldest EVs located? Where are the newest?

Are some electric providers in areas with more EVs than others?

Is there a correlation between affluent areas and the number of EVs?

Why does King county have the most EVs?

Can we use this data or combine it with another data reference to infer the socioeconomic status of EV owners?

Has the average MSRP changed over time?

Have there been improvements to Electric Range over time?

Which Makes have the longest Electric Range? Which Makes have the lowest MSRP on average? Which have the highest?

Has the CAFV eligibility changed over time? Are newer models more likely to be eligible?

