

Introduction to the Dataset

The dataset I'm using utilizes registered vehicle data in Washington state to identify what vehicles are currently registered. This includes useful data such as the model, the year, the vehicle type such as hybrid or full battery, and the range. There are 17 columns worth of features with at least 11,000 rows worth of unique data.

Questions

Some of the questions I'd like to answer include the correlation between vehicle models and their electric range. Some manufacturers are just entering the EV space and I feel that some aren't taking advantage of the form-factor that they provide when it comes to battery space, and will likely have lower range as a result. I'd like to test whether or not this is true. I also want to take a look at popularity trends and see which make/models are the most popular. Are people sticking to brand loyalty from their gas-powered alternatives? Or are they looking at different manufacturers with more experience in the EV space? Lastly I want to see where in the state these vehicles are the most popular. While this only covers Washington state, it could still prove to be helpful data to see if EVs are more popular in populated cities or if they are more spaced out. This could help with understanding the trends for other states and if people take these sorts of things into account when they are purchasing an EV.

Pre-Processing Data

During the Pre-Processing step, I ran into a few parts of the dataset that I wanted to avoid using for the final visualizations. The first step I took was to remove any null values that were included in the data, or rather any rows of data that contained null values. I opted to outright remove these rows as this is a dataset containing over 100,000 rows, and the amount of null values were only in the low hundreds. At this point, I also went ahead and dropped some of the columns that were not providing useful data that I wanted to analyze. These included things like very specific vehicle locations, legislation data, census tract, etc. After this, I realized that some of the data may not be very accurate, and searched to see if there were duplicates in the data. Since this was vehicle data, the VIN number for each registered vehicle had to be different so I was able to remove duplicates by identifying which VINs were repeating. This narrowed down the results to be around 11,000 or so cars registered, which was much more manageable.

Visualization #1: Line Plot measuring EV Range over the years

The first visualization lets us see that while the range that an EV can provide has been steadily increasing over the years, it has also already seen a peak in the year 2020. Tesla has been at the forefront of the EV market for years, so it is not a huge surprise to see them reaching the top of the chart in just about every year. What is surprising is to see how other EVs are progressing steadily towards the same range, and that these vehicles are popular enough to be on these charts. Another to keep in mind while looking at these charts is that Hybrid vehicles are also included, and the mileage information they provide is referring to their electric range. So another good

trend to see here is that even though some of these vehicles appear to be on the low end of range, they have been steadily improving and having more range, likely in that Hybrid form factor. This could mean that we're seeing more electrically running cars on daily commuters that aren't even touching the gas tank, as they don't require that much range to get to work and back to recharge overnight and do it all again in the morning. These are all very nice trends to be seeing here.

Line Chart

