Anthony LaRosa

11/21/20

Professor Williams

Final Project Summary

This was a great semester. Great is really an understatement and I gave that feedback to Professor Metzger in DSC550 aswell. I was very fortunate to have Professor Williams and Professor Metzger this term because my Python skills specifically with application to data science had the most drastic increase in the last (12) weeks than any term yet.

There was so much I learned in this project that I always wondered how to do.

First, the skills to clean, prepare, and normalize data starting off within a flat file.

Next, I always wondered how to do web scraping, and we got to learn it with BeautifulSoup. There’s one thing to learn it via reading, but it’s another level for me to be able to do it hands on with real data. I got to use Wikipedia’s auto data and tie it into my project.

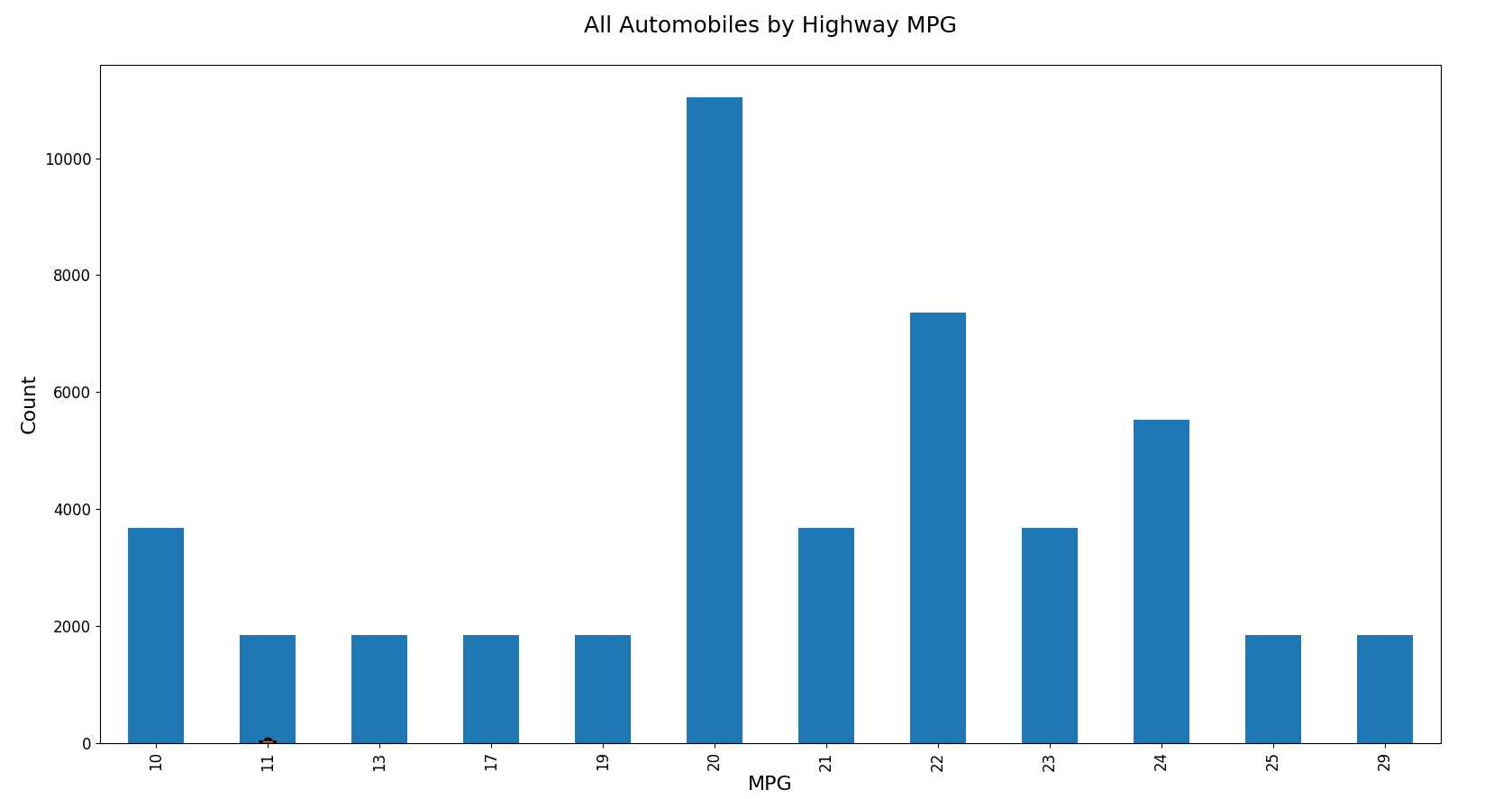
Finally, JSON API’s and Python are very smooth together. If I need data from another system using Python, at this point in my learning an API and especially one that returns JSON format would be my method of choice mostly because it’s clean, simple, and gets the job done.

I rarely have done DBA work, but it was a nice touch at the end of the course to be able to interact with some foundational SQL.

For my project, once I had all the tables joined to one, I had to stop myself from keep tinkering and making scatterplots with the data. It was cool seeing all the insights that can be drawn.

Below are some examples with my thoughts:

One interesting piece of information I gleaned is that it is very expensive to own an Aston Martin even outside of the initial purchase price.

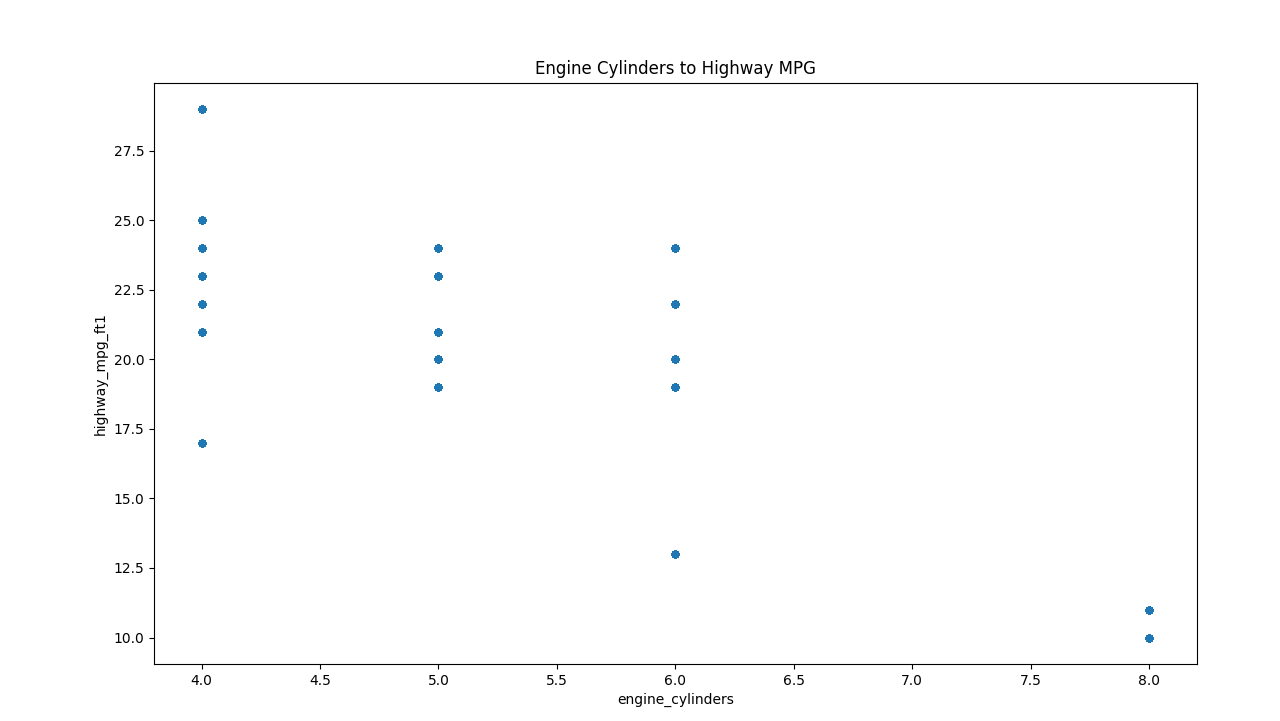


First I wanted to see the spread of the highway MPG counts.

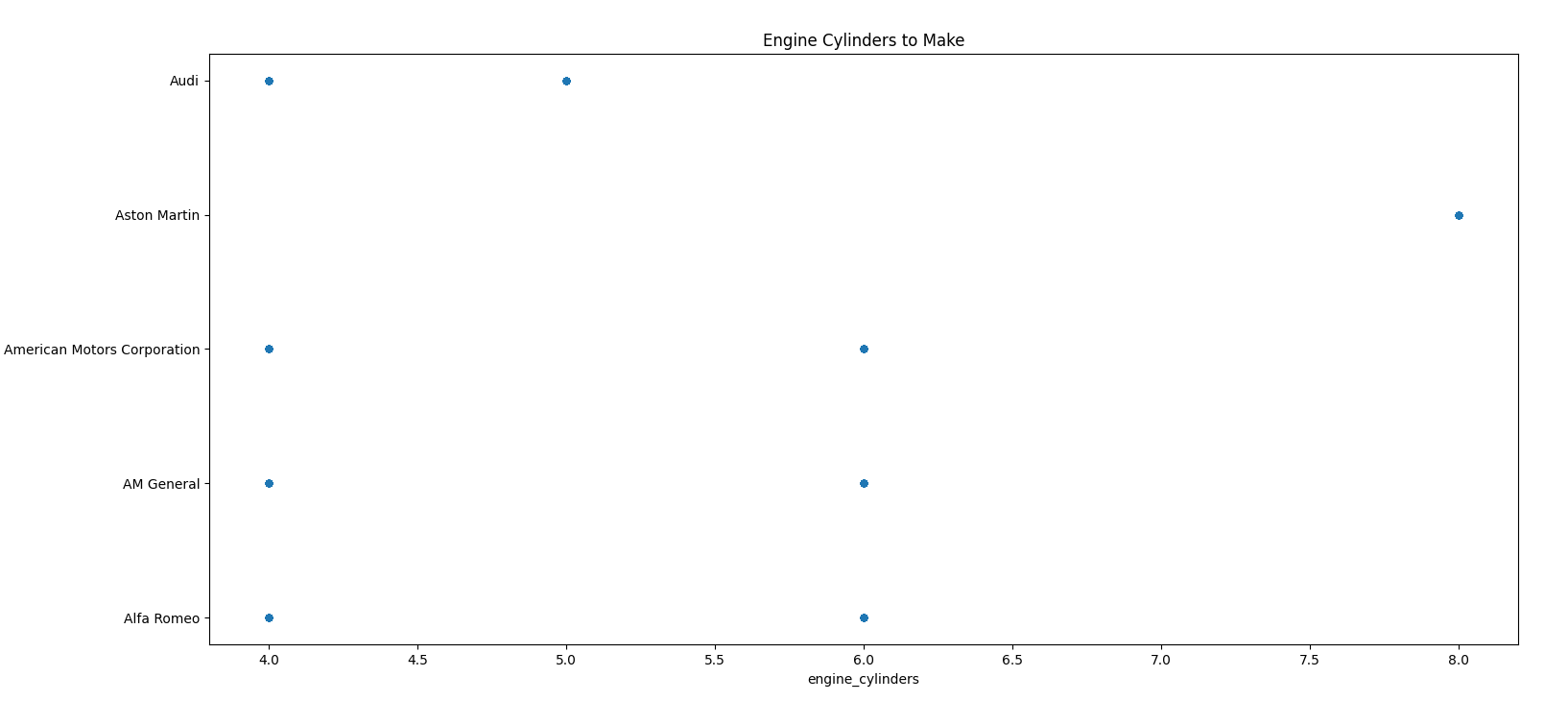
It can be seen that the most common count is 20 MPG.

Next, I would expect that the more cylinders that an engine has, the worse its highway MPG would get. It’s interesting that between 4, 5, and 6 cylinders it’s rather consistent.

8 cylinders are when it really drops off.



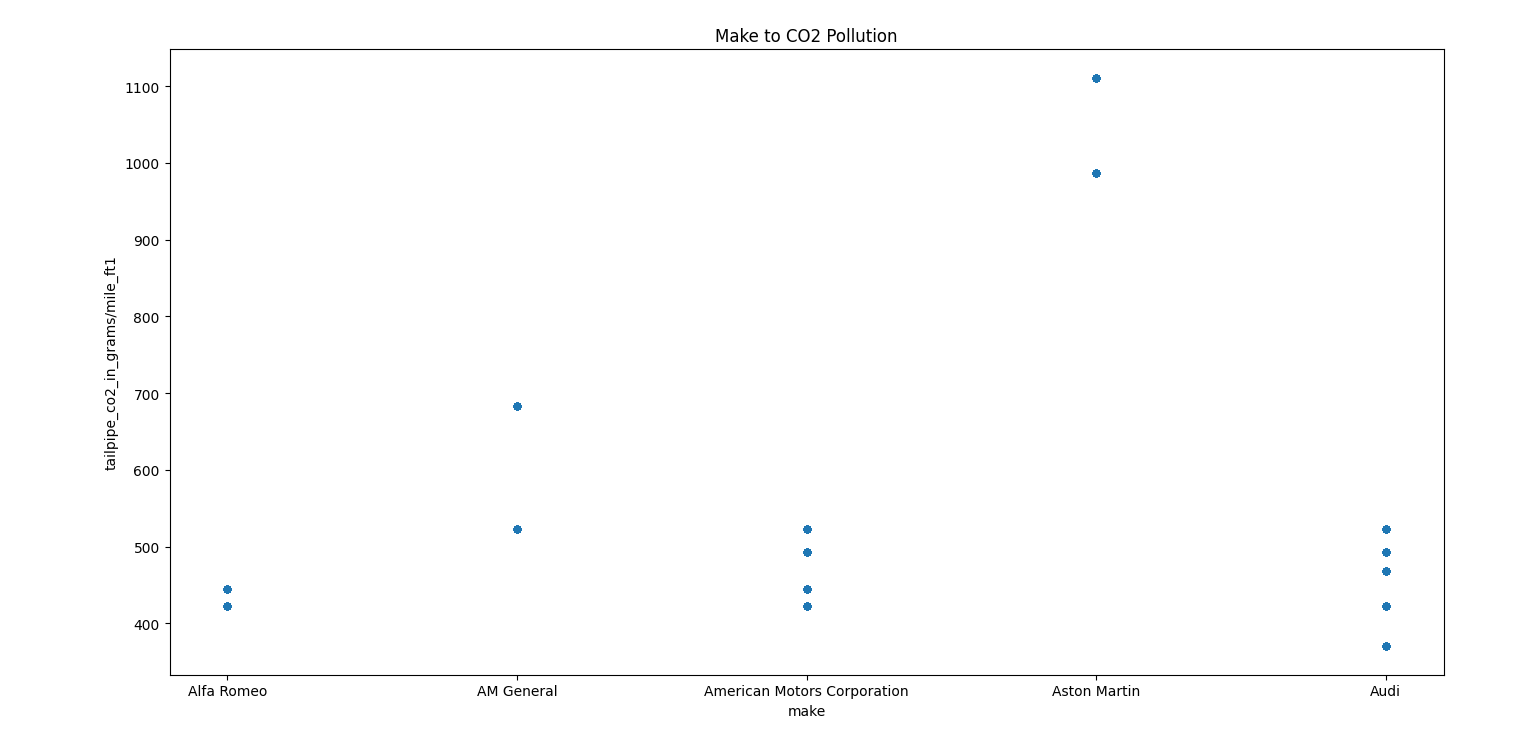
I was then curious to see out of the sample of “A” automakers what engine cylinders were being produced. Aston Martin is holding down the 8 cylinders in this sample.



Next, let’s cross reference what we expect out of the cylinders and highway mpg.

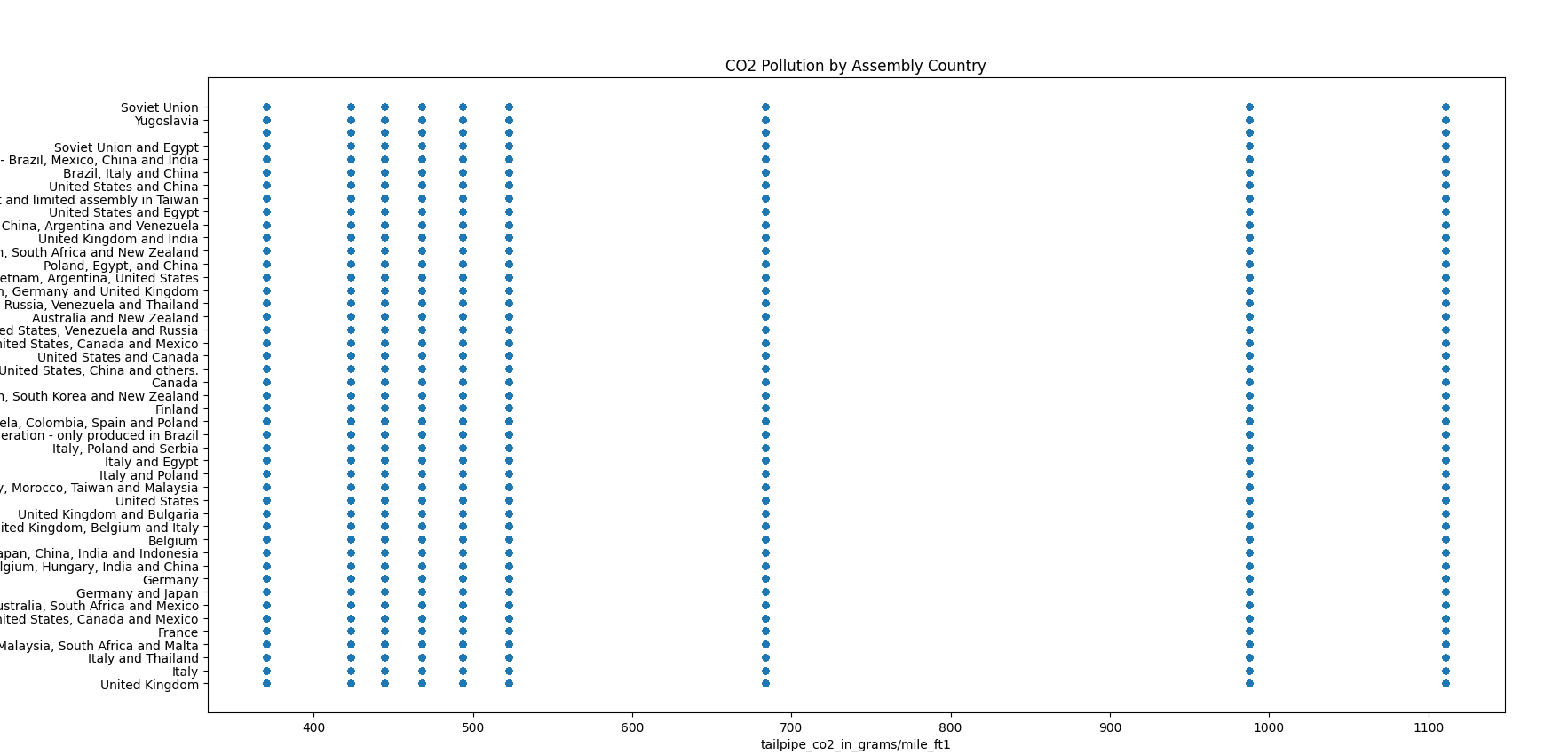
We can do this by creating a plot of the automaker to the pollution level.

Aston Martin is our top polluter in this sample.



This made me curious, when a car is assembled, and it can be assembled in multiple countries, is there a certain location that creates cars that pollute to a greater extent?

The answer seems to be no, they are all in the arena. What was interested to me is the consistency that they jump between pollution levels. They seem to be under 700 grams or almost 1,000 and up.

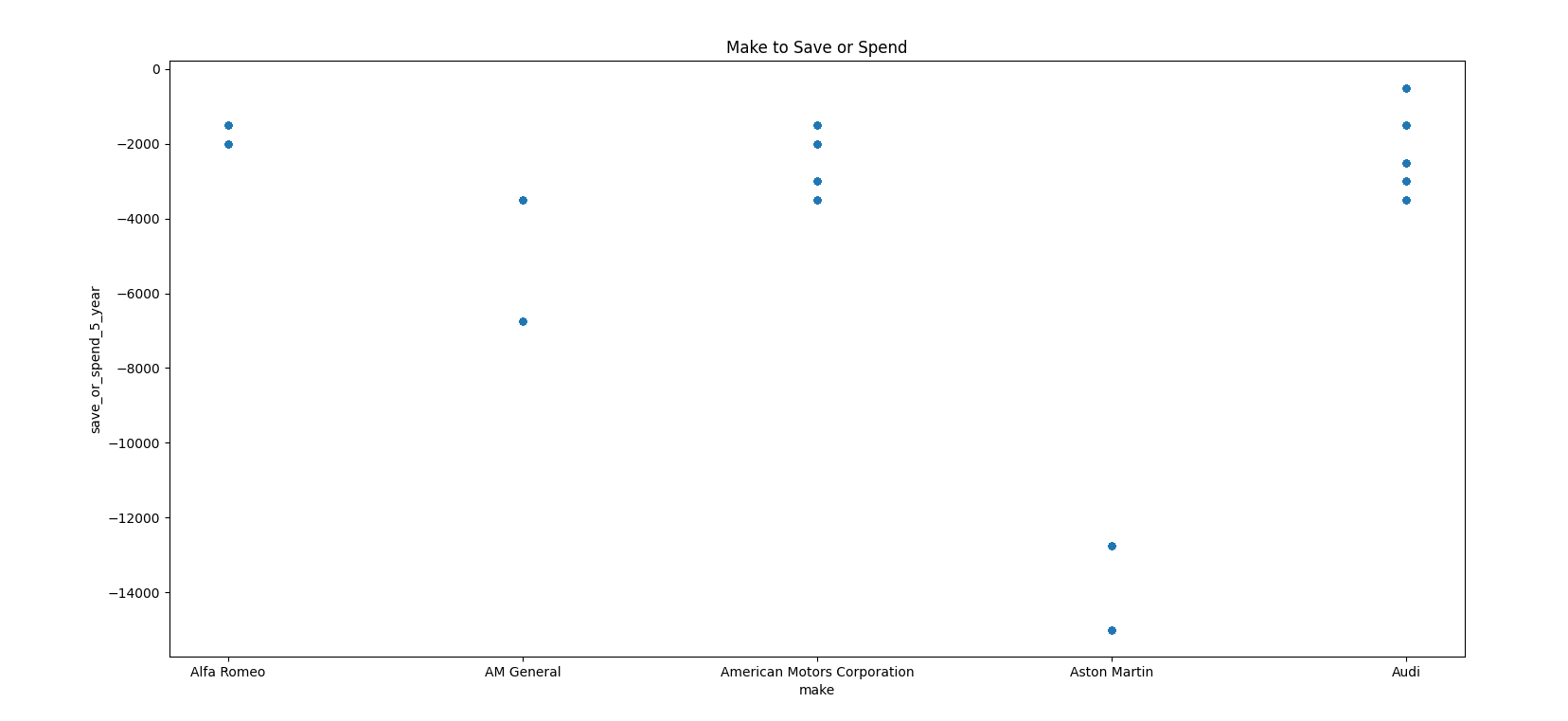


Finally, and this is my favorite chart of the five, is what affect in our “A” company sample is it having on your wallet in 5 years in additional costs.

Two things I found surprising.

Audi, which can be known as expensive to own, wasn’t that bad.

Aston Martin, was really costly to the point of being viewed as an outlier.



Overall I enjoyed this project and thank you for the opportunity and guidance.