HW3

mpg
Target variable: Miles Per Gallon how fuel efficient the car is

cylinders (This could be a selected feature)
Number of cylinders in the engine More cylinders = usually more power less fule efficiency

displacement
Basically the volume of the engine... bigger = more gas-guzzling

horsepower (This could be a selected feature)
How powerful the engine is more horsepower = more acceleration less mpg

weight (This could be a selected feature)
Car's weight in pounds heavier cars usually burn more fuel

acceleration
Time it takes to go from 0 to 60 mph lower value means faster acceleration

model_year
The year the car was manufactured (70 = 1970) Can reflect changes in tech and regulations

origin (This could be a selected feature)

The region the car comes from, encoded as:

1 = USA

2 = Europe

3 = Japan

car_name
name of the car

First I chose to use cylinders, horsepower, weight and origin as the features to select I was hesitant on acceleration and model_year So if the accuracy is not what I like then I will add them in

Since I am using horse power as one of my features I decided to throw away the 6 missing data since it is only 1.5% of my data

I had to use delim_whitespace=True since the file is not a csv file first I converted all '?' in horse power to nan using numpy and then using dropna to remove the rows

When I plotted the features this is what I got
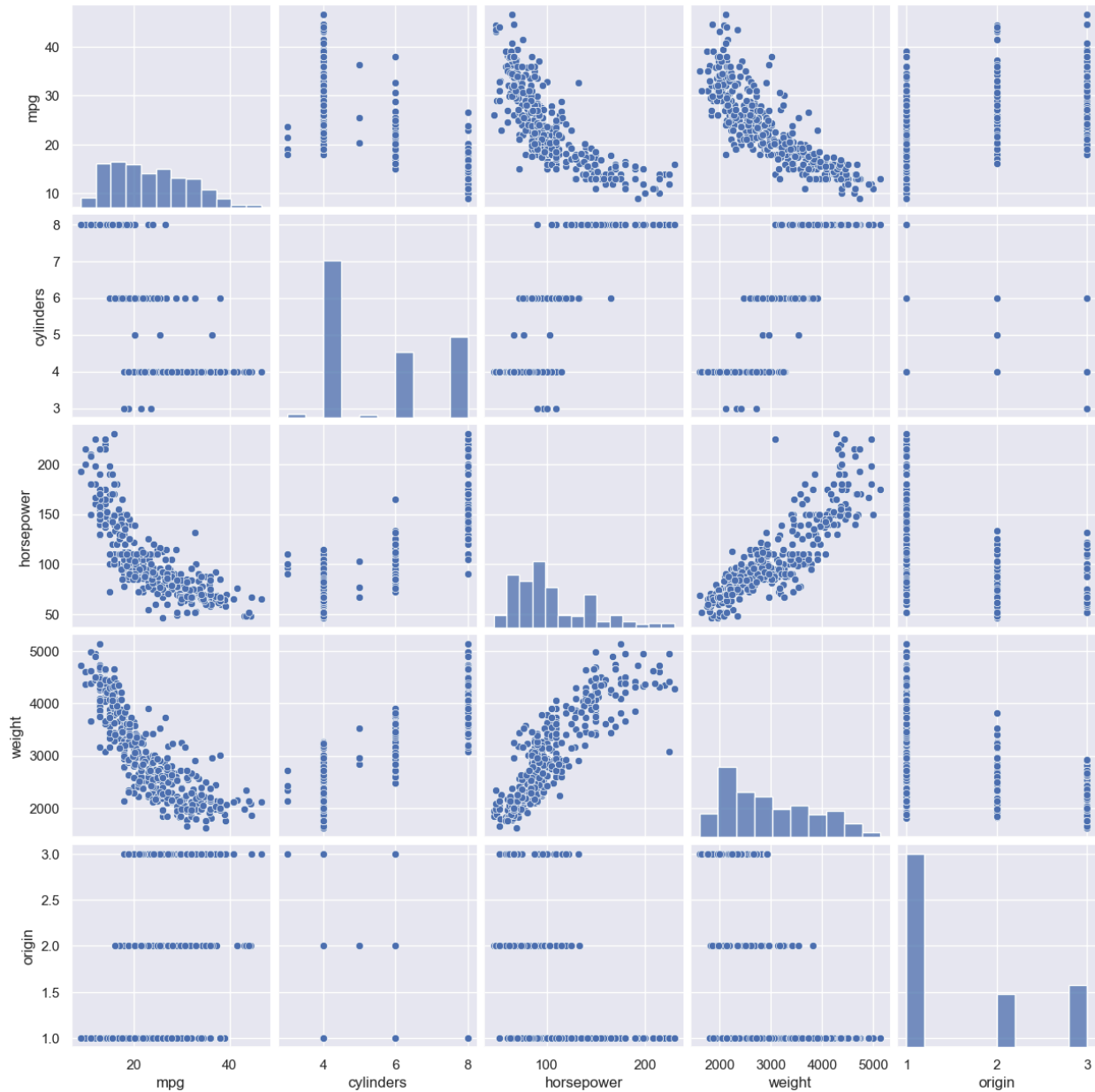
I decided to Take away cylinders since its correlated with weight and horse power

I kept horse power and weight since they have a parabola

Origin is a weak predictor but japanese cars TEND to have more fuel efficient cars which doesn't really mean much

I then looked at displacement and acceleration decided to keep displacement and drop acceleration
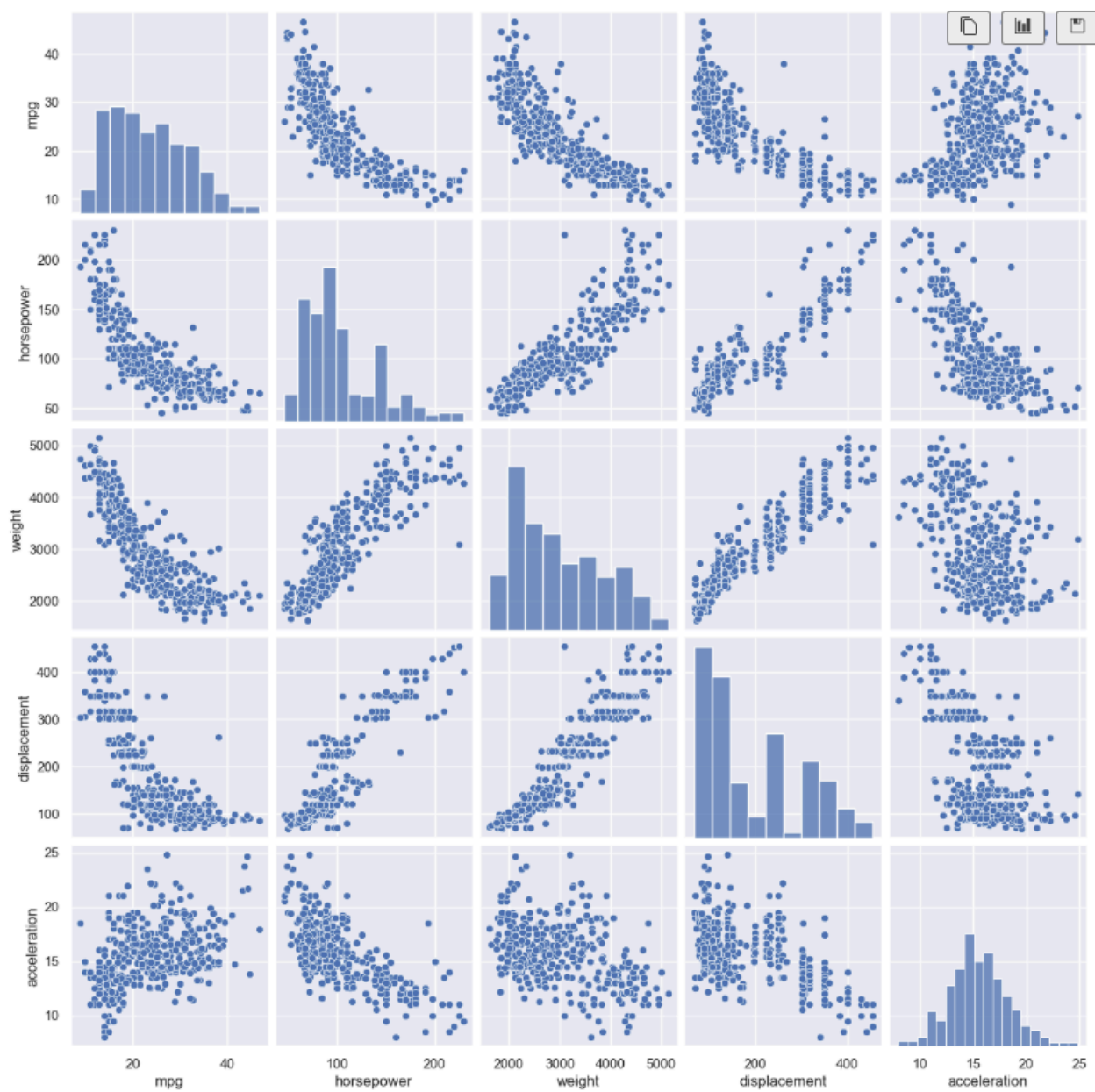
I decided to only use horsepower weight and displacement

The RMSE scores were between 3.1 and 5.0 so the model was usually off by about 3 to 5 miles per gallon... Out of all the features, weight had the biggest effect

heavier cars clearly got worse mpg horsepower and displacement also brought mpg down but not as much

	Fold	Horsepower Coef	Weight Coef	Displacement Coef	RMSE
0	1	−1.528463	−4.469687	−0.791723	4.170198
1	2	−1.735667	−4.477922	−0.709901	4.162047
2	3	−1.794711	−4.369634	−0.650546	4.353491
3	4	−1.857064	−4.618458	−0.270550	4.709866
4	5	−1.265840	−4.566043	−0.905701	4.398953
5	6	−1.610826	−4.571431	−0.480464	5.025100
6	7	−1.114528	−4.754356	−0.784364	4.821236
7	8	−1.838471	−4.556229	−0.350488	3.107398
8	9	−1.459254	−4.577988	−0.637044	3.322329
9	10	−1.798226	−4.440940	−0.465896	4.223758