Car Sales Prediction

Lachandra Ash

November 14, 2022

Topic

Consumer Reports conducted a representative nationwide poll of automobile buyers in the spring of 2022, and their findings show that rising costs for both new and used vehicles have drastically altered consumers' preferences. The demand for pre-owned vehicles has decreased while the demand for new vehicles has increased, and more individuals are contemplating purchases in higher price brackets than in 2019.

Business Problem

The used car dealerships want to increase the sales of their used cars. The dealers want to predict their sales to increase their revenues.

Background | History

Although the Germans and the French had the honor of inventing and perfecting the vehicle, it was the Americans who would go on to dominate the industry for the better part of the twentieth century (History, 2018).. By the 1920s, Ford, General Motors, and Chrysler had established themselves as the three major automakers, thanks in large part to the mass manufacturing methods pioneered by Henry Ford (History, 2018).. When the war ended, manufacturing of automobiles in Europe and Japan went through the roof to satisfy surging demand (History, 2018).

Data Explanation

I retrieved this dataset from the Kaggle website. The Analytixlabs dataset is being used for statistical forecasting. There are two things to consider here. They had to determine which option significantly affects automobile sales. The next thing to consider is they need to examine how well the classifier has learned to forecast automobile sales. The Price_in_thousands is the target variable. The dataset features are:

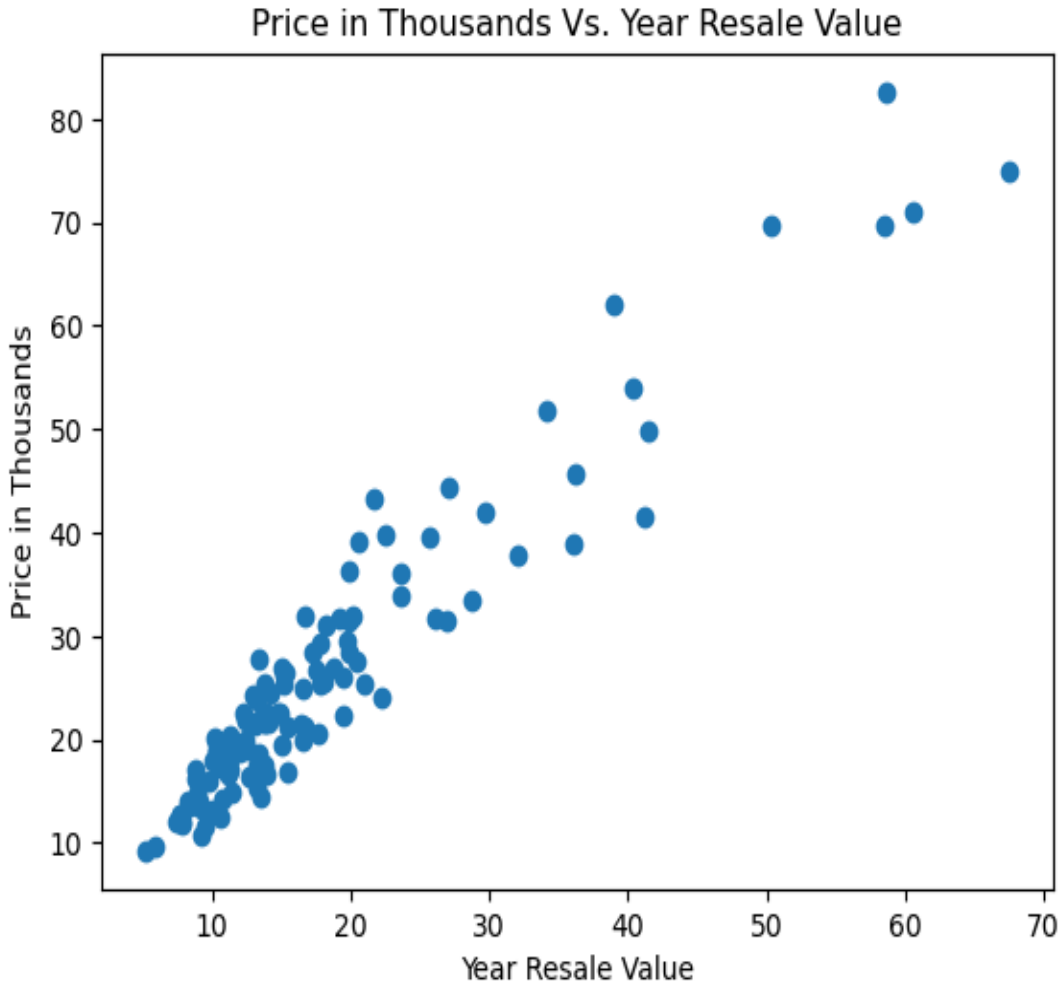
- Manufacturer
- Model
- Sales in thousands
- Year resale value
- Vehicle type
- Price in thousands
- Engine size
- Horsepower
- Wheelbase
- Width
- Length
- Curb weight
- Fuel capacity
- Fuel efficiency
- Latest launch
- Power perf Factor (GaganBhatia, n.d.).

Methods

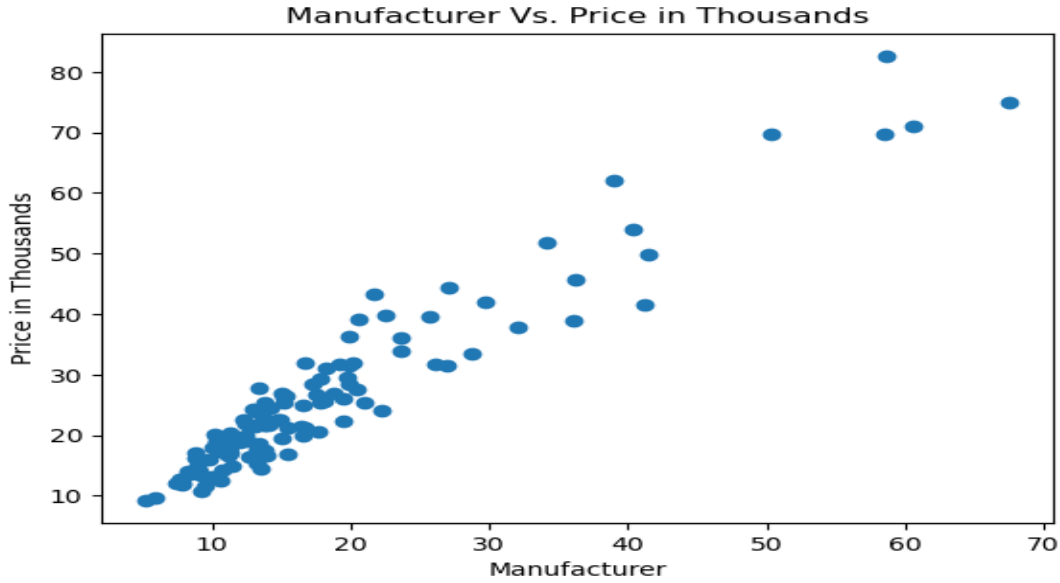
I used the Jupyter Python Notebook 3 to examine the dataset. I imported the libraries and modules needed for the analysis. I read the Car sales csv file into the dataframe and viewed the dataframe records. I cleaned the dataset using the isna, replace, and duplicates methods. I studied the dataframe's index, dtypes, information, sum, shape, describe, correlation, covariance, and value counts.

Analysis

1) Price in Thousands vs. Year Resale Value: The year resale value increased in price by the thousands as the year resale value increased.

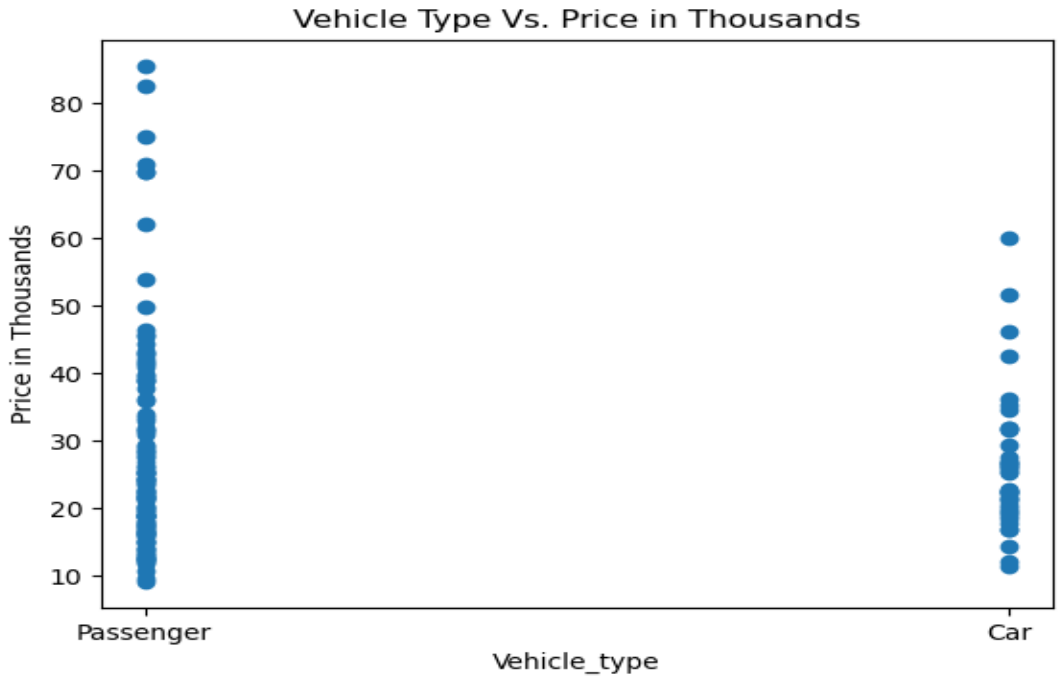


2)Manufacturer vs. Price in Thousands: Price increased among the various types of manufacturers.

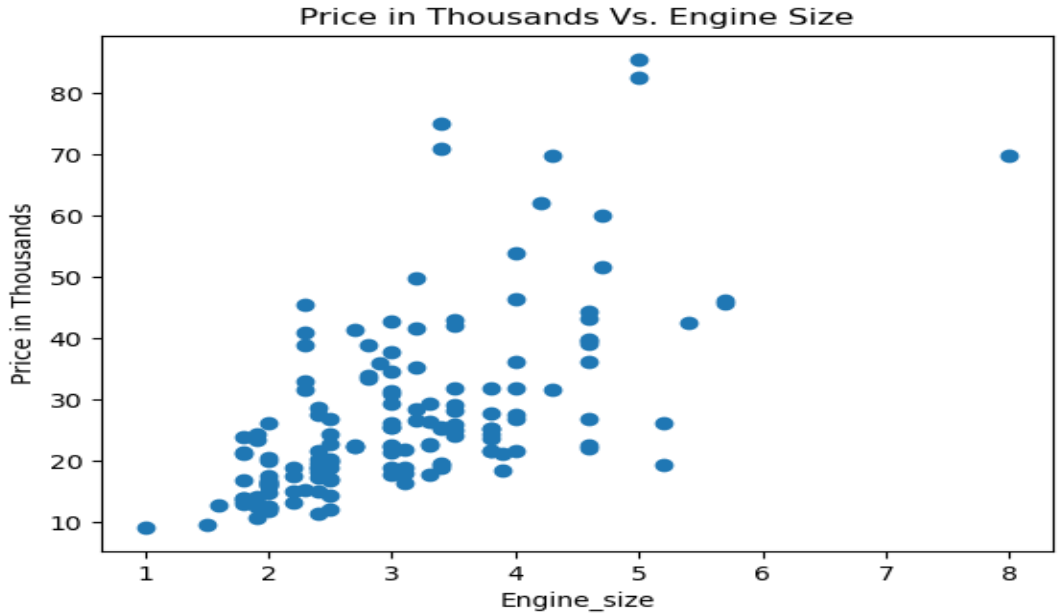


---

3)Vehicle Type vs. Price in Thousands: There are more passenger vehicles versus cars.

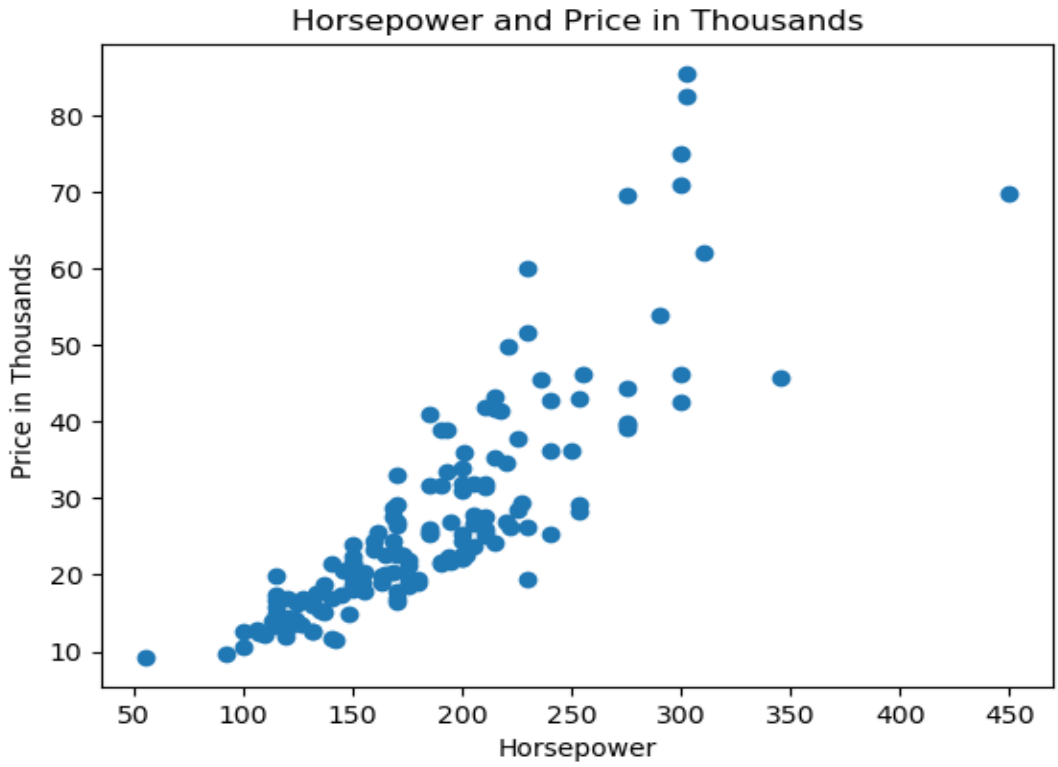


4)Engine size vs. Price in thousands: The bigger the engine size the more expensive the car. Vehicles with engine sizes greater than 5 experienced a price drop. There were no vehicles that had engine sizes between 6-8.

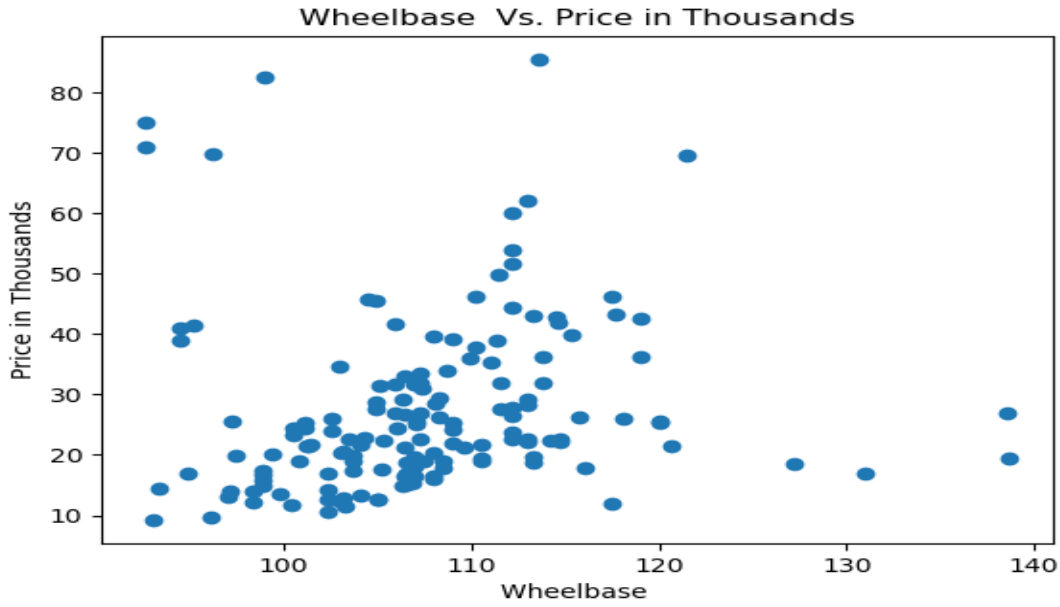


---

5)Horsepower and Price in thousands: The more horsepower that a vehicle has, the more expensive it is.

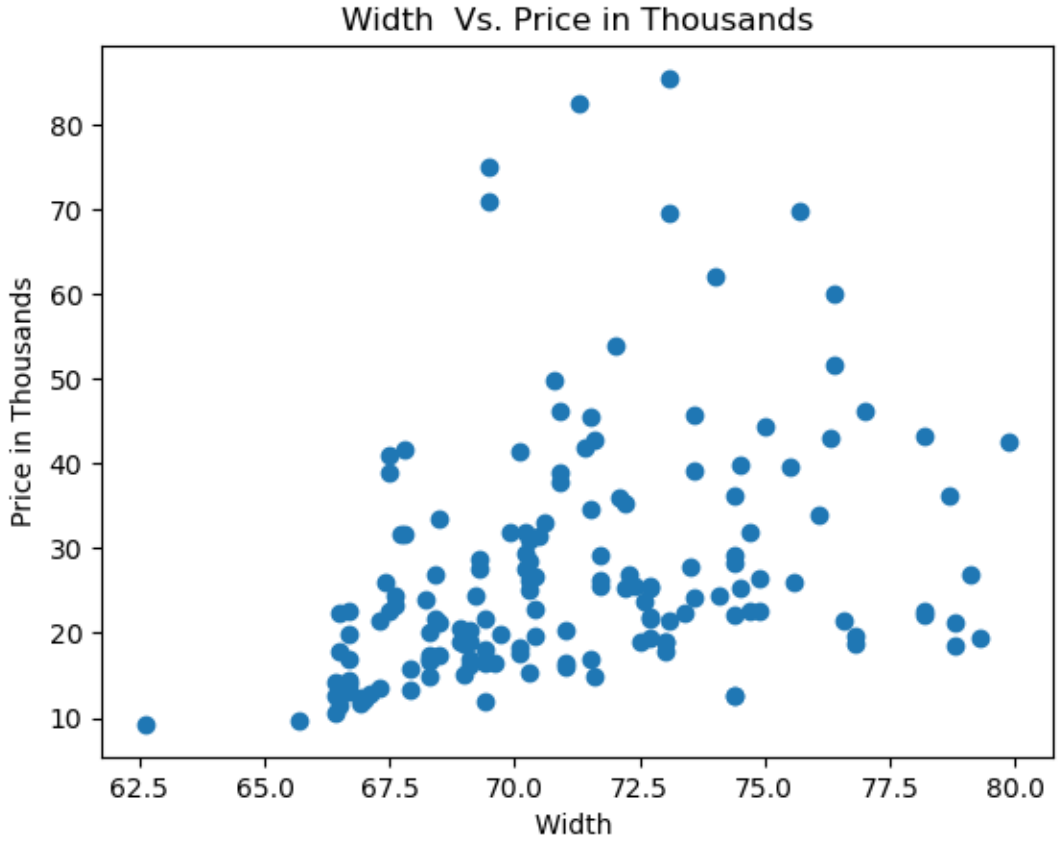


6)Wheelbase vs. Price in thousands. The wheelbase sizes approximately between 100-115 were more expensive versus the wheelbase sizes from 115-140.

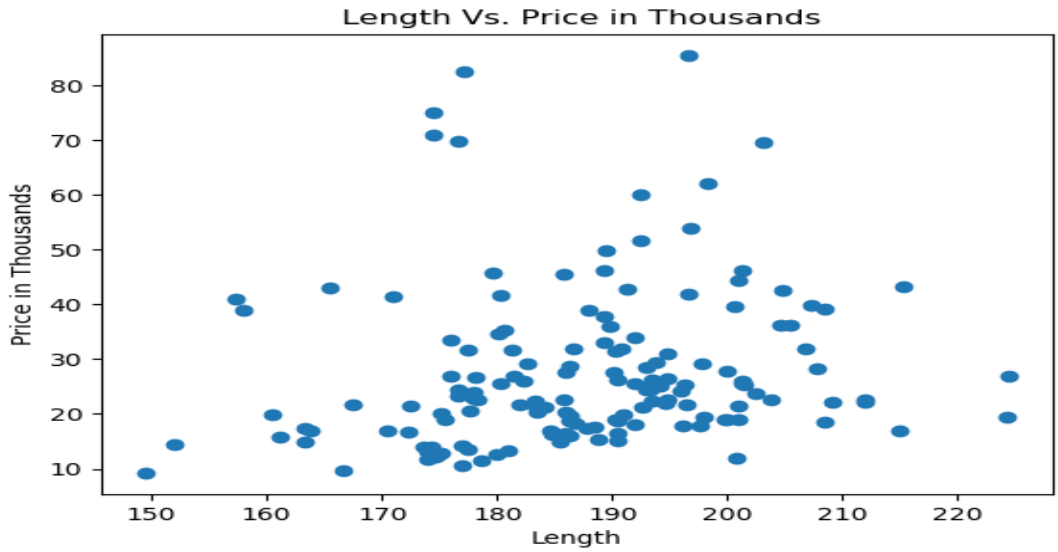


---

7)Width vs Price in thousands: The greater the width of the wheelbase, the more expensive it is..

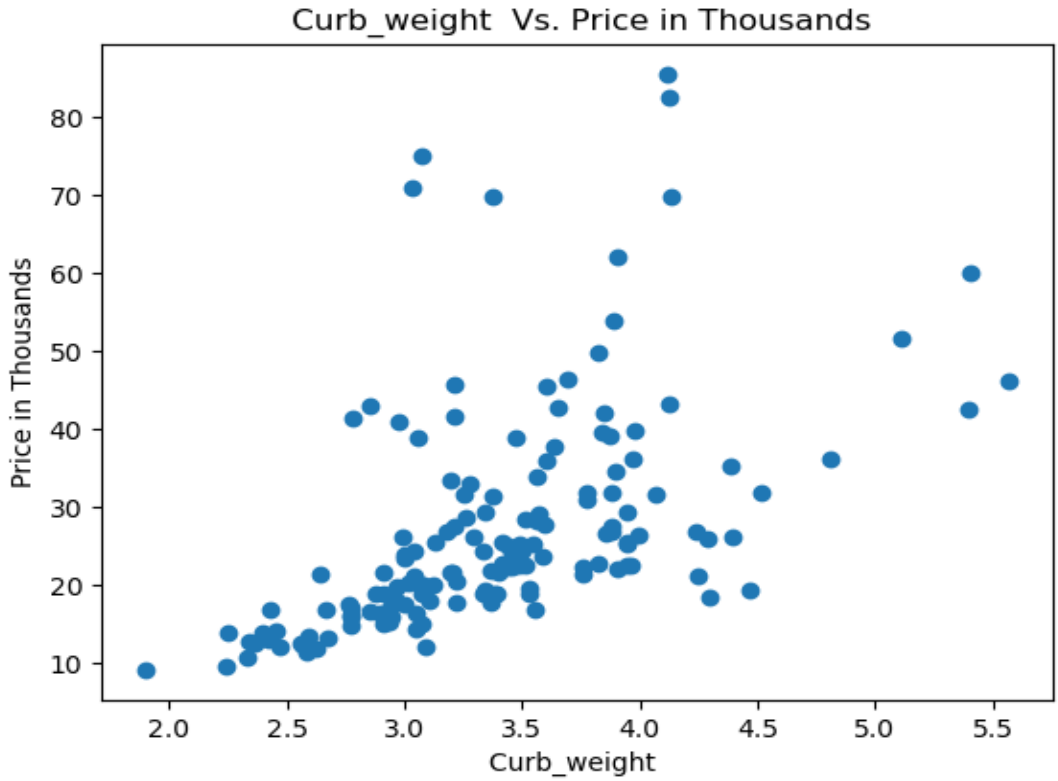


8) Length and price in the thousands. The greater the length of the wheelbase, the more expensive it is. After the length of 20, the prices dropped.

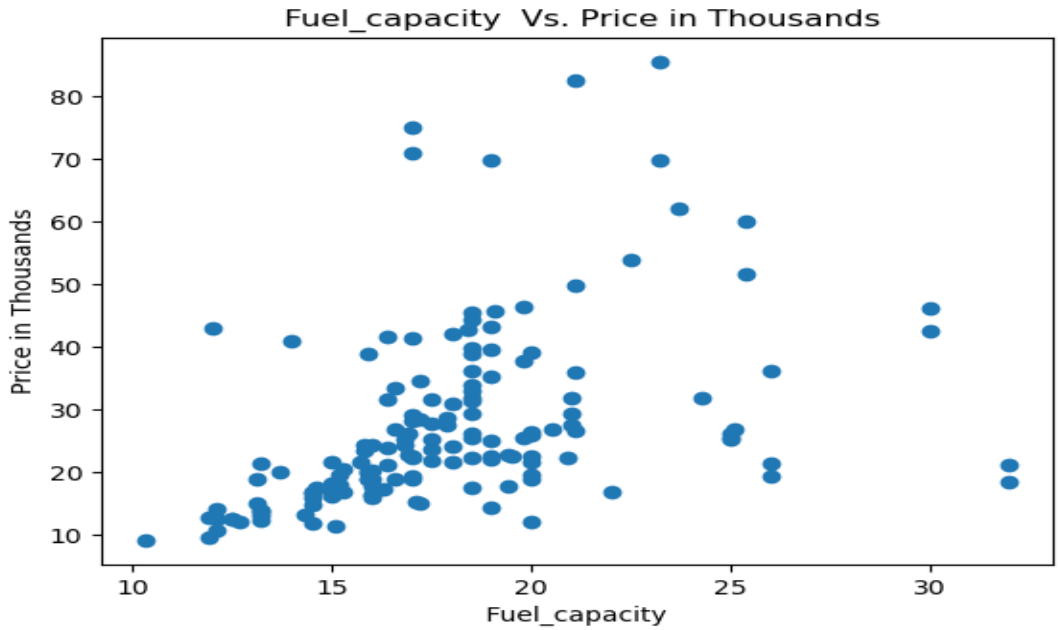


---

9)Curb weight vs. price in the thousands: As the curb weight increased, the price increased.

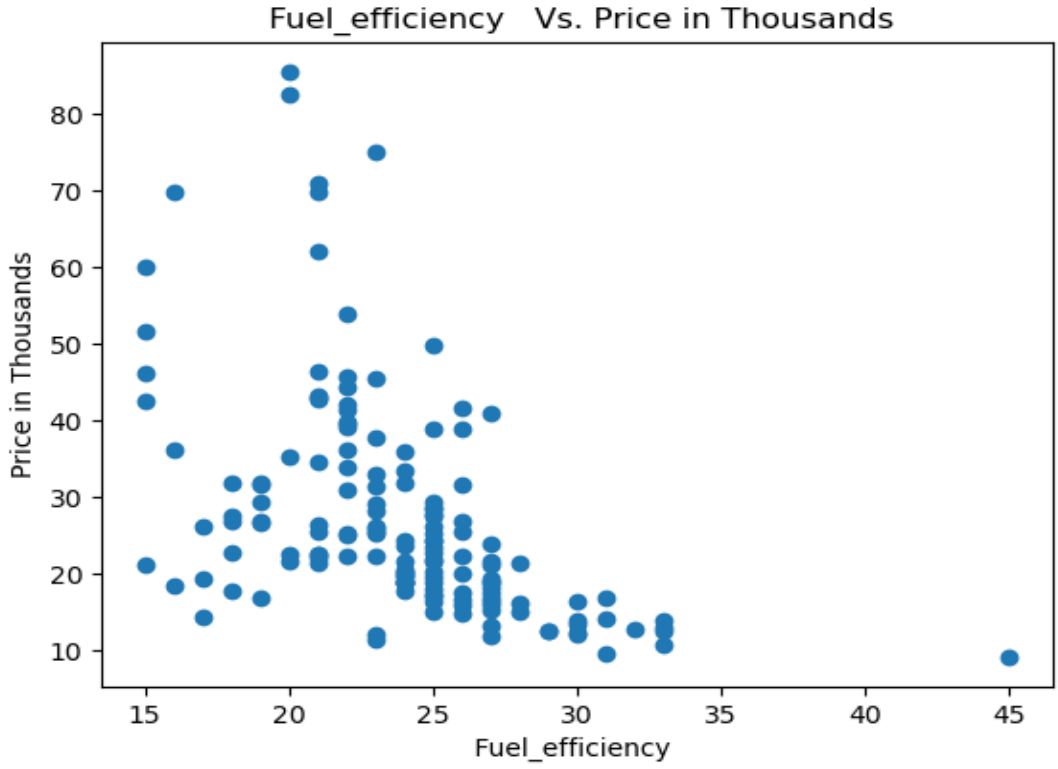


10)Fuel capacity and price in thousands: The vehicles with greater fuel capacity were more expensive.

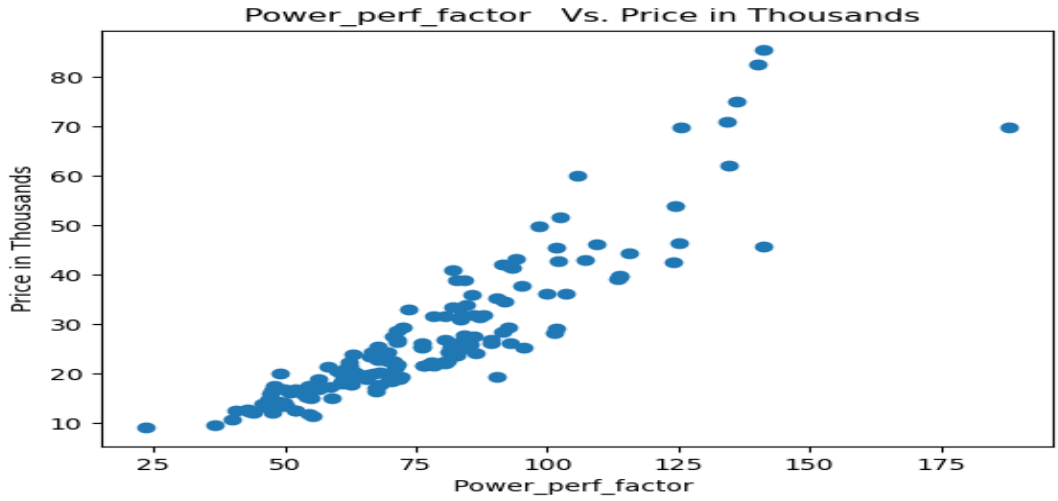


---

11)Fuel efficiency vs. price in thousands: The vehicles with greater fuel efficiency were more expensive till approximately 23, then the price began to decline from 23 to 45.

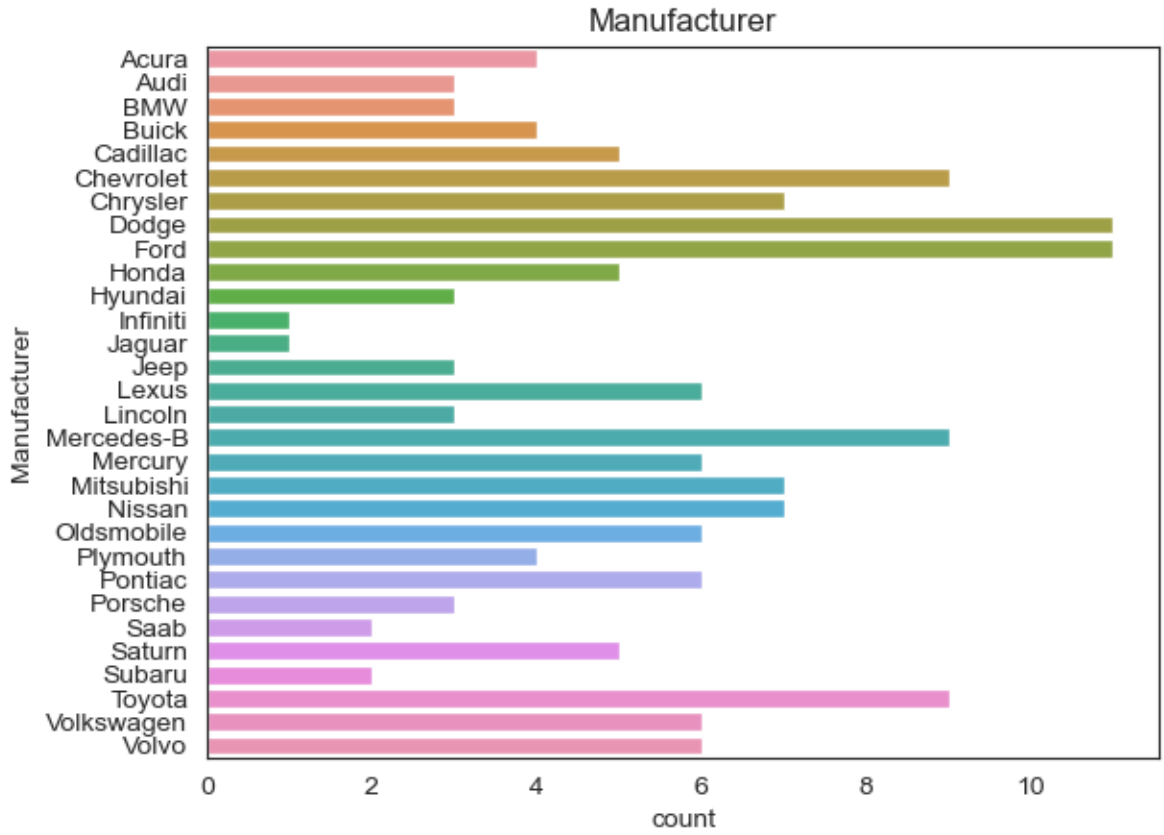


12) The power perfection factor vs. price in thousands: The higher the vehicle's power perfection factor is, the higher the price of the vehicle.

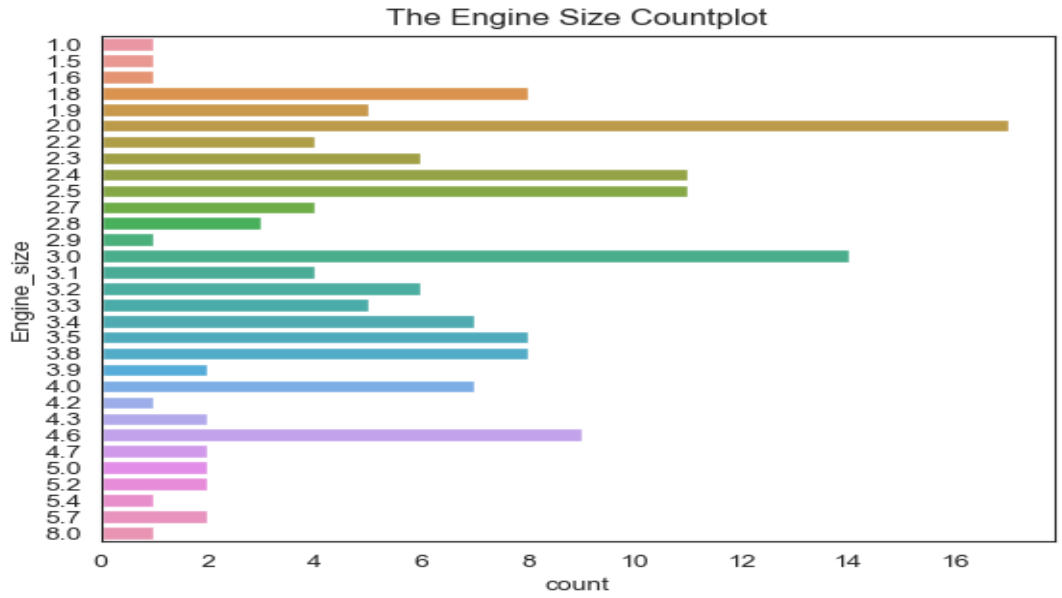


---

13)Manufacturer: Dodge and Ford had the highest number of vehicles. Chevrolet, Mercedes-Benz, and Toyota had the second highest number of vehicles. Infiniti and Jaguar have the least number of vehicles.

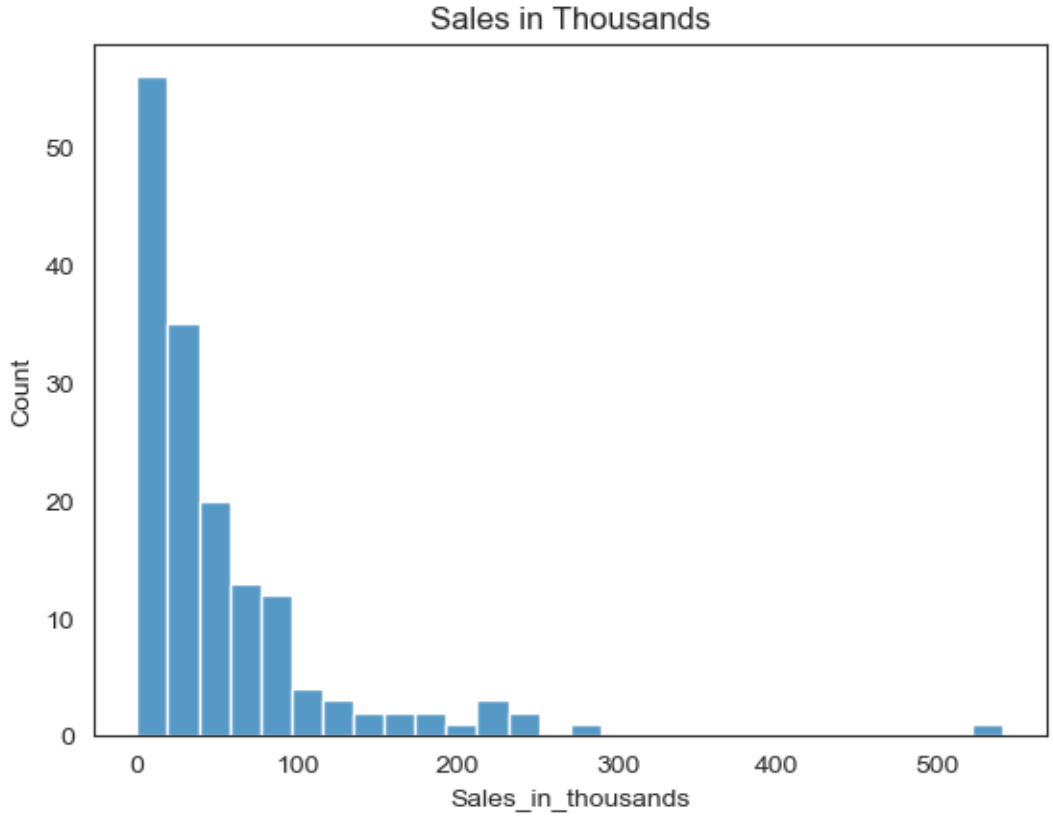


14)Engine Size: The 2.0 engine size has the greater number of engines of that size. The 3.0 engine size has the second greatest number of engines of that size.

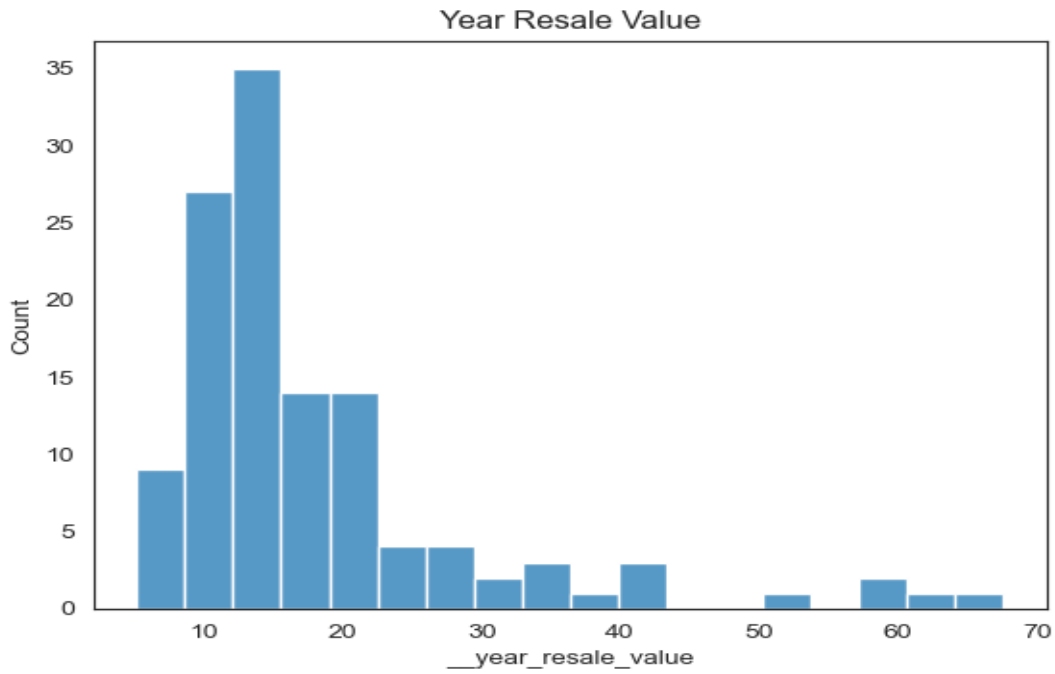


---

15)Sales in Thousands: The number of vehicle sales are greater in the beginning of the sales, but as the sales decreased, the number of sales decreased.

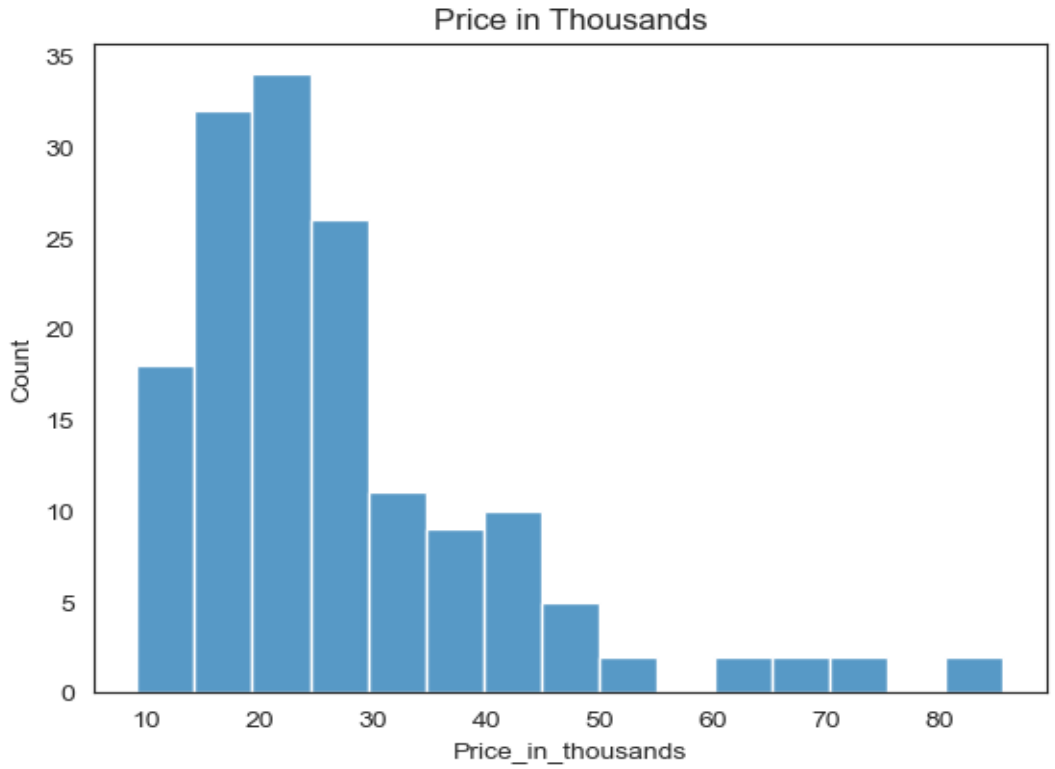


16)Year resale value: The vehicle year resale value was greater between 10-15 years of age. After 15 years of vehicle age, the resale value begins to decrease.

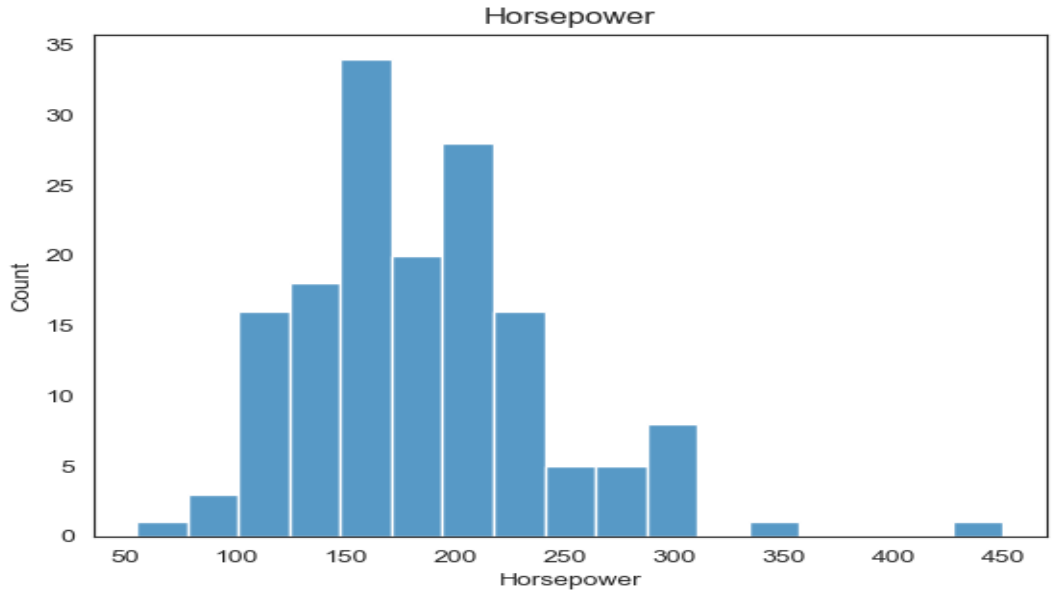


---

17)Price in thousands. The prices between 15,000 to 30,000 had the highest number of vehicles.

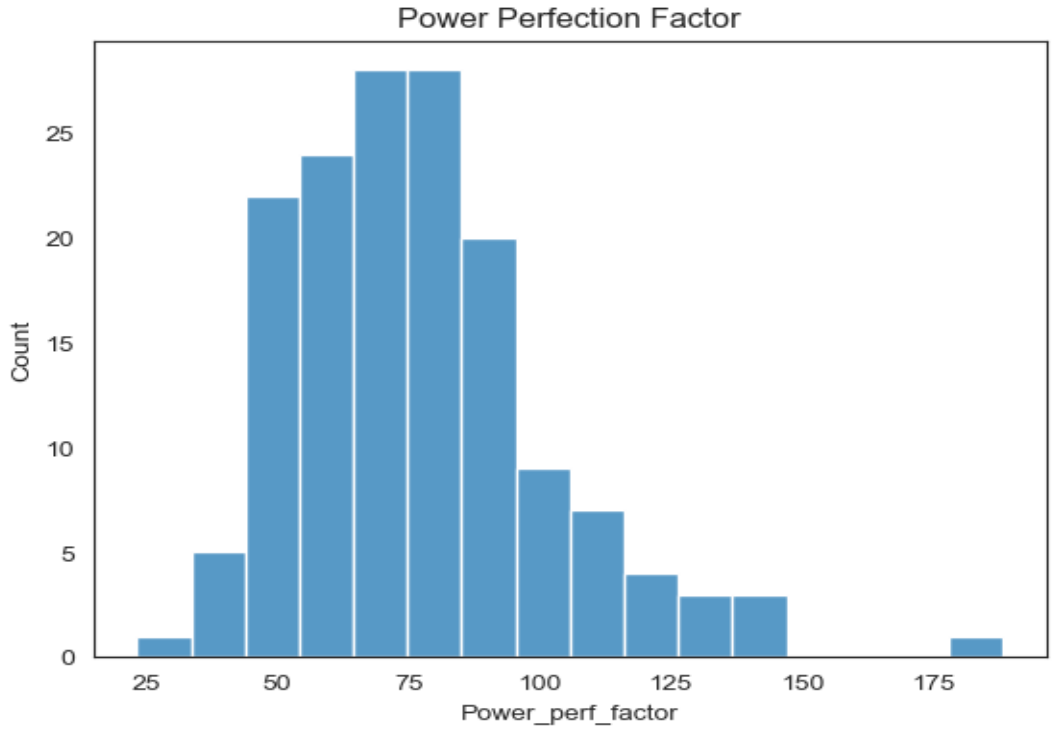


18)Horsepower: The engines with 150 and 200 horsepower have the greatest number of vehicles with the same number of vehicles. The engines with 50, 100, 350, and close to 450 horsepower has the least number of vehicles with that horsepower.

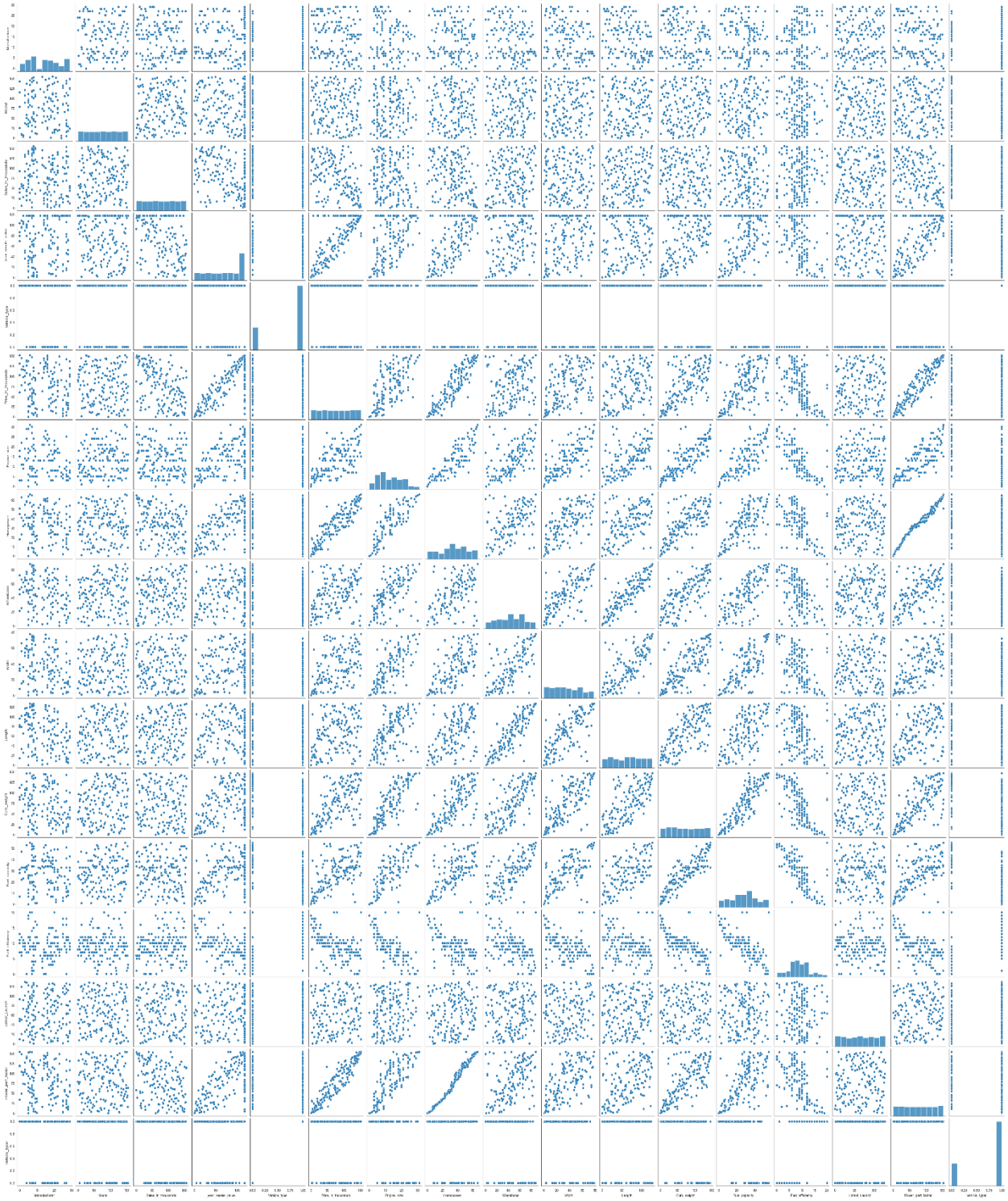


---

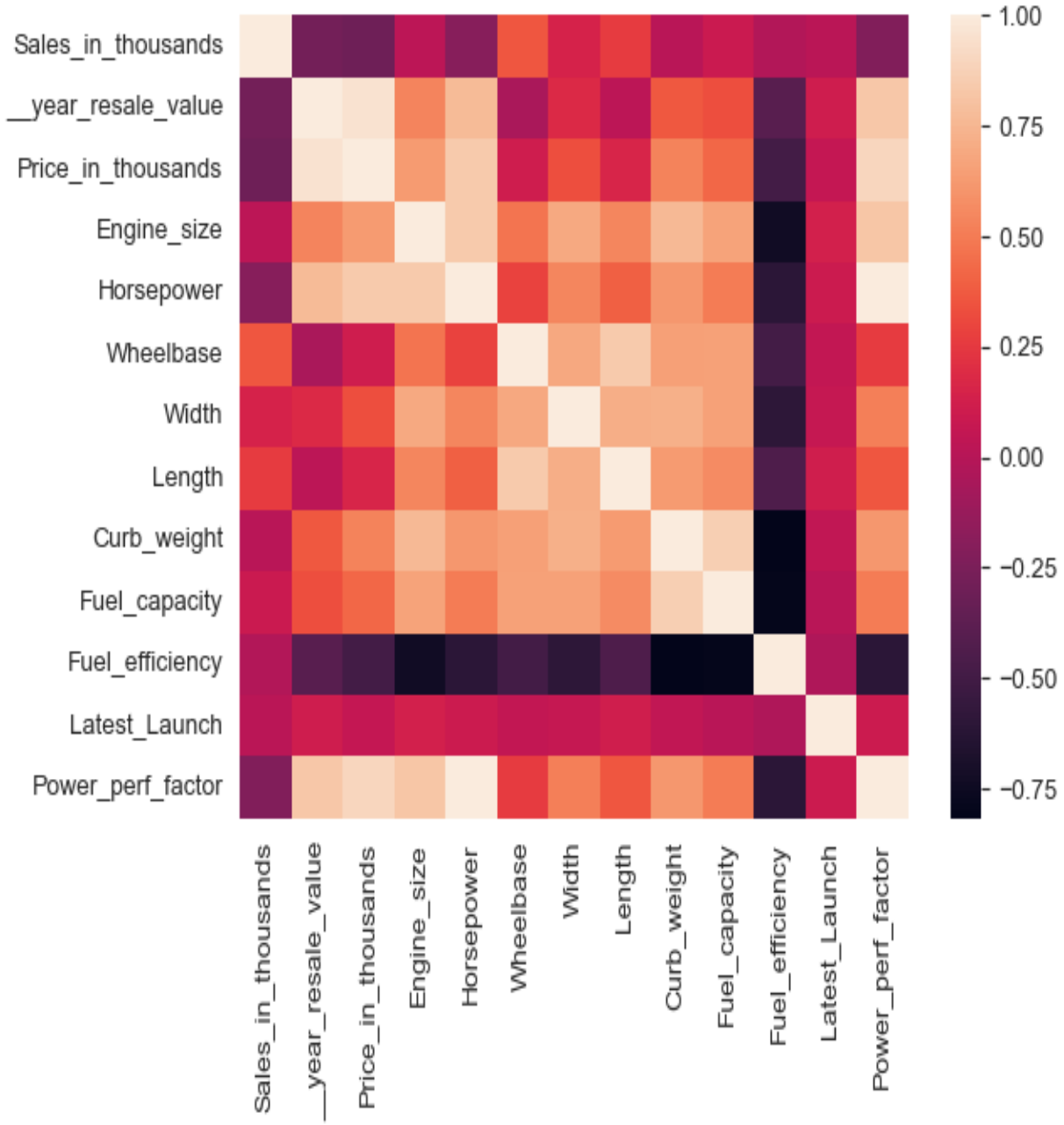
19)Power Perfection Faction: The power perfection factor with the greatest number of vehicles was approximately 75. The power perfection factor with the least number of vehicles was approximately 25 and approximately more than 175.



20)Dataframe Pairplot: The dataframe's pairplot includes are the features of the dataframe.



21) Heatmap: The heatmap correlation displays the correlation between each feature within the dataframe.



Naïve Bayes Model

I dropped the Price in thousands column before starting the model. I defined the x and y, and then I trained, tested, and split the dataframe. I fit the model and The model results revealed zero percent in accuracy.

View Appendix: Table 1, Naïve Bayes Model

Conclusion

Used car dealerships can increase their sales by dealing with high quality vehicles with good year resale value. The engine size causes the vehicle to be more expensive. The vehicles with horsepower between 100-200 were priced between $15,000 – $40.000. I did a lot of research on this data set, and it did not have enough explanation for the dataset. I was still interested in exploring the dataset.

References

Gaaganbhatia (n.d.). Car Sales Dataset. Retrieved from Car sales | Kaggle, November 15, 2022.

History (2010). Automobile History. Retrieved from Automobile History - HISTORY, on November 14, 2022.

Appendix:

Table 1, Naïve Bayes Model

```
[[0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 [0 0 0 ... 0 0 0]]
9.0
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1.0
7	0.00	0.00	0.00	1.0
11	0.00	0.00	0.00	1.0
17	0.00	0.00	0.00	1.0
24	0.00	0.00	0.00	1.0
26	0.00	0.00	0.00	1.0
32	0.00	0.00	0.00	1.0
35	0.00	0.00	0.00	1.0
37	0.00	0.00	0.00	1.0
40	0.00	0.00	0.00	1.0
41	0.00	0.00	0.00	1.0
43	0.00	0.00	0.00	0.0
46	0.00	0.00	0.00	1.0
49	0.00	0.00	0.00	1.0
53	0.00	0.00	0.00	1.0
54	0.00	0.00	0.00	1.0
57	0.00	0.00	0.00	1.0
59	0.00	0.00	0.00	1.0
66	0.00	0.00	0.00	1.0
74	0.00	0.00	0.00	1.0
89	0.00	0.00	0.00	1.0
93	0.00	0.00	0.00	1.0
94	0.00	0.00	0.00	1.0
97	0.00	0.00	0.00	1.0
98	0.00	0.00	0.00	1.0
108	0.00	0.00	0.00	1.0
109	0.00	0.00	0.00	1.0
111	0.00	0.00	0.00	1.0
117	0.00	0.00	0.00	1.0
124	0.00	0.00	0.00	1.0
146	0.00	0.00	0.00	1.0
150	0.00	0.00	0.00	1.0
152	0.00	0.00	0.00	1.0
accuracy			0.00	32.0
macro avg	0.00	0.00	0.00	32.0
weighted avg	0.00	0.00	0.00	32.0