Hyperiondev

TASK

Exploratory Data Analysis on the Automobile Data Set

Visit our website

Introduction

Automobile's data set summarized here:

Table 1: Table data summary

	wheel base	Curb weight	Engine size	Compression ratio	horsepower	Peak rpm	City mpg	Highway mpg	price
count	190.00	190.00	190.00	190.00	190.00	190.00	190.00	190.00	190.00
mean	98.95	2559.45	127.04	10.10	103.47	5116.32	25.21	30.65	13196.41
std	6.09	526.71	42.07	3.93	37.43	475.12	6.48	6.90	7988.68
min	86.60	1488.00	61.00	7.00	48.00	4150.00	13.00	16.00	5118.00
25%	94.50	2145.00	97.25	8.50	70.00	4800.00	19.00	25.00	7747.25
50%	97.00	2414.00	110.50	9.00	95.00	5200.00	24.00	30.00	10246.50
75%	102.40	2947.75	144.00	9.40	116.00	5500.00	30.00	34.00	16502.25
max	120.90	4066.00	326.00	23.00	262.00	6600.00	49.00	54.00	45400.00

Was used for the analysis, containing 205 rows and 26 columns

DATA CLEANING

The following columns were dropped by passing a list of columns in a loop to a method which called the dro() method:

Normalized-losses, Symboling, engine-location, fuel-system, bore, stroke, length, width and height.

Duplicated() method to find and drop_duplicates() to drop the rows:

0 Duplicate columns were found and dropped
7 Duplicate rows were found and dropped

The following column datatypes were of object type and casted to numerical datatype:

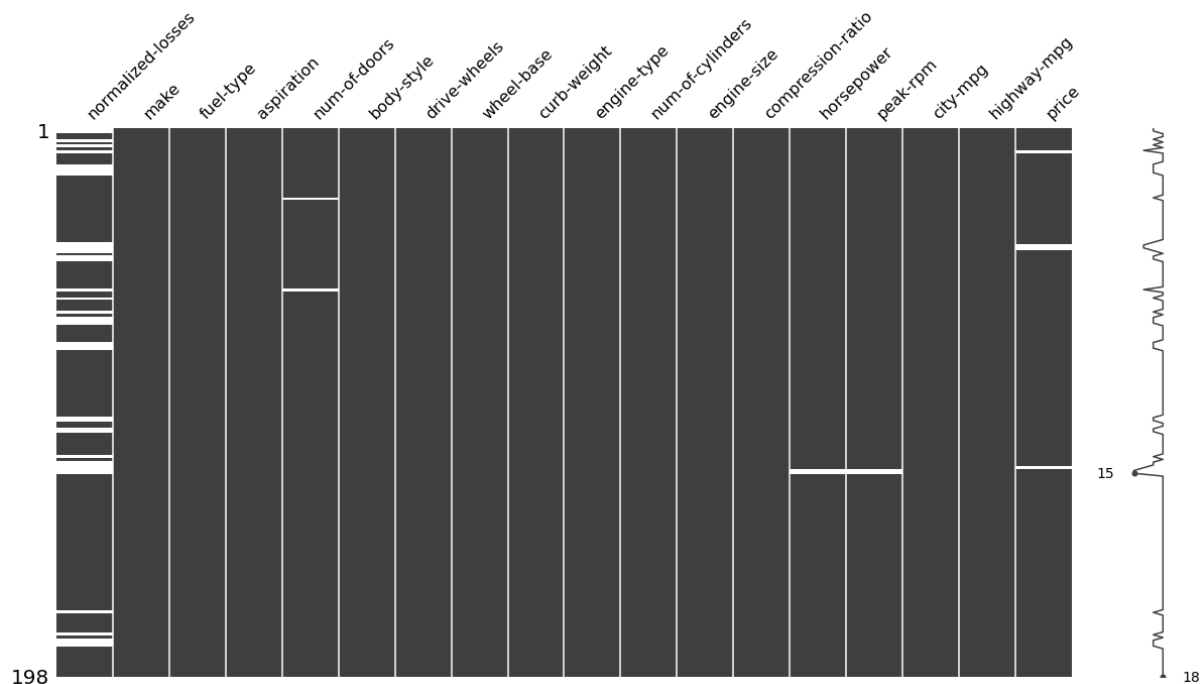Horsepower, peak-rpm and price

MISSING DATA

Figure 1: Missing data matrix graph

The data was first checked for irregularities:
'?' were found and converted to 'NAN' using Replace method

The amount of missing data = 1.38 %

The missing data matrix (fig.1 created using missingno) indicated a lot of data missing and the column 'Normalized-losses' = 24.75 %
For this reason, the column was dropped (drop() method) and excluded from the analysis.

The remaining rows with missing data were dropped (using drop(axis=0) method) for their small number and impact on the analysis.

DATA STORIES AND VISUALIZATIONS

The most important column the dataset was the price. So, everything must be around the price concept. So, I believe that we are basically asking here, **"What affects the price of cars?"** and how each of those characteristics affect the cost and average cost of those vehicles.
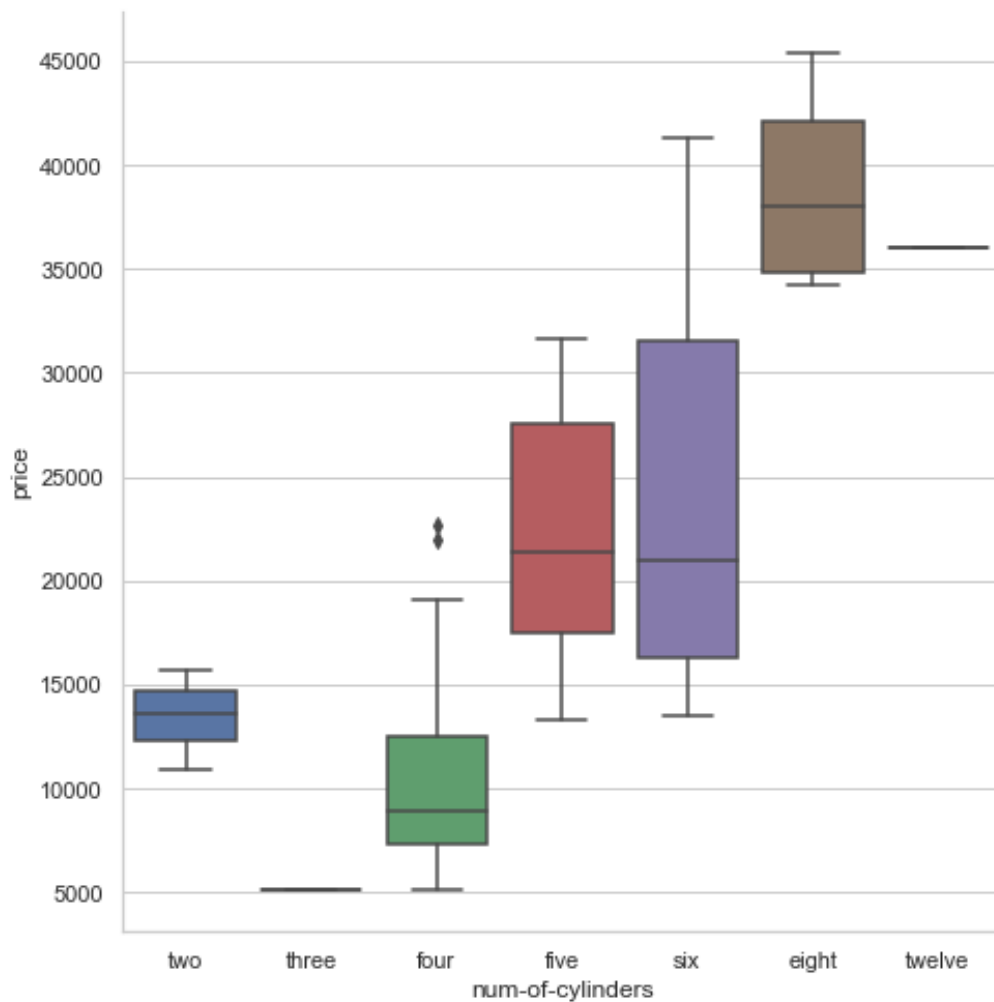

Figure 2: Price shift by Number of cylinders

The plot represented by figure 2, indicates that the cost of vehicles increase as the cylinder count does, and this is a realistic factor as increased cylinder increases power in the vehicle.
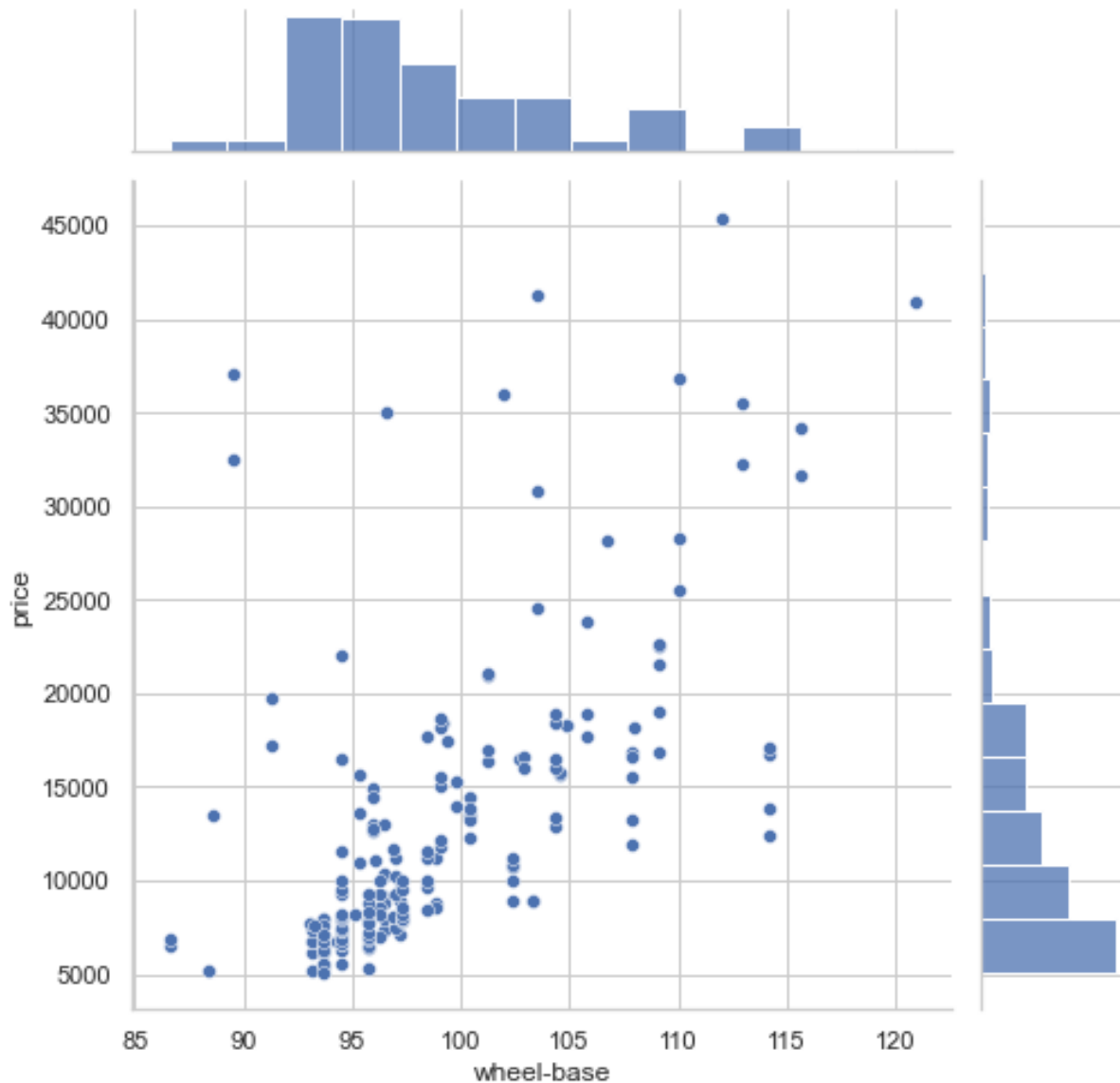
Figure 3: Price shift characterised by wheel base length

Wheelbase has an effect on the weight distribution of a vehicle and the dimensions are crucial in vehicle balance and steering characteristics. It is found that in high performance vehicles the wheelbase is much longer which allows for greater stability and aerodynamics. This plot indicates that vehicles with a longer wheelbase have an increasing price.
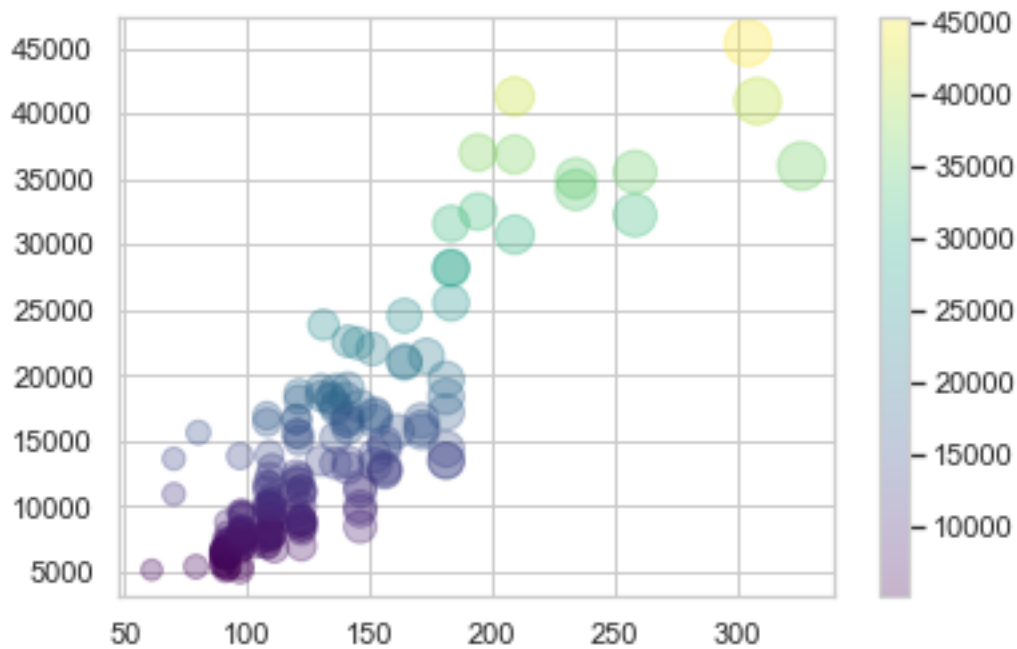
Figure 4: Price shift characterized by engine size

Engine size plays an important role in power output, and figure 4 indicates a direct proportionality to vehicle cost and engine size.
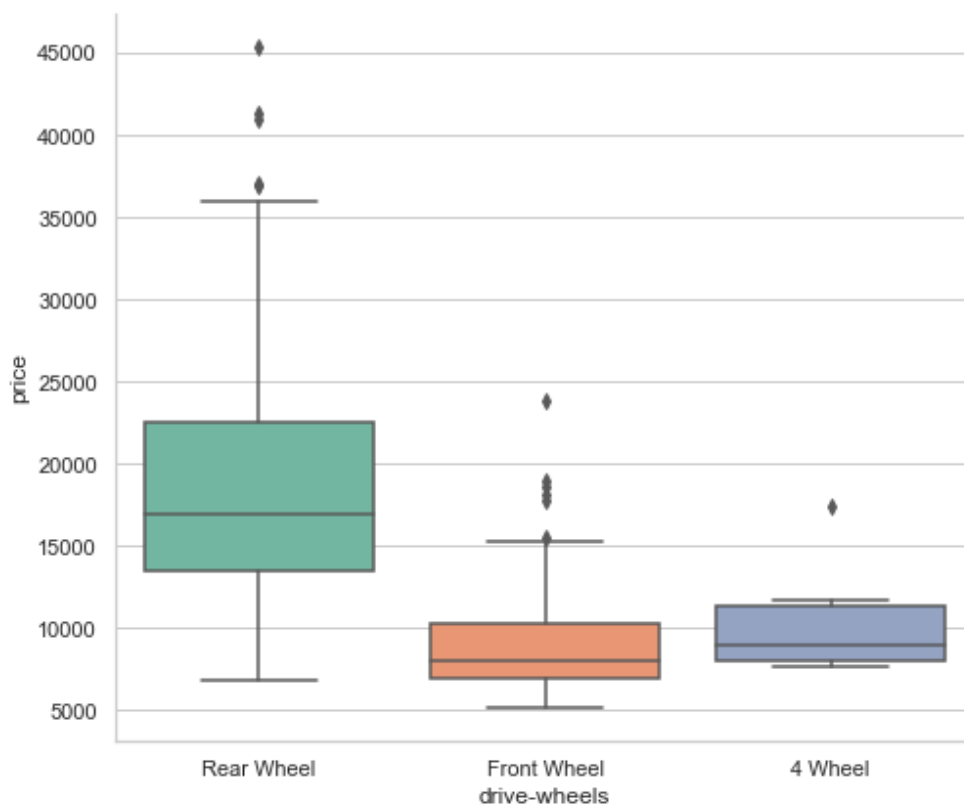

Figure 5: Price shift characterized by drive wheel

It appears from figure 5, that rear wheel vehicles are much more expensive than front wheel or four-wheel drive vehicles.
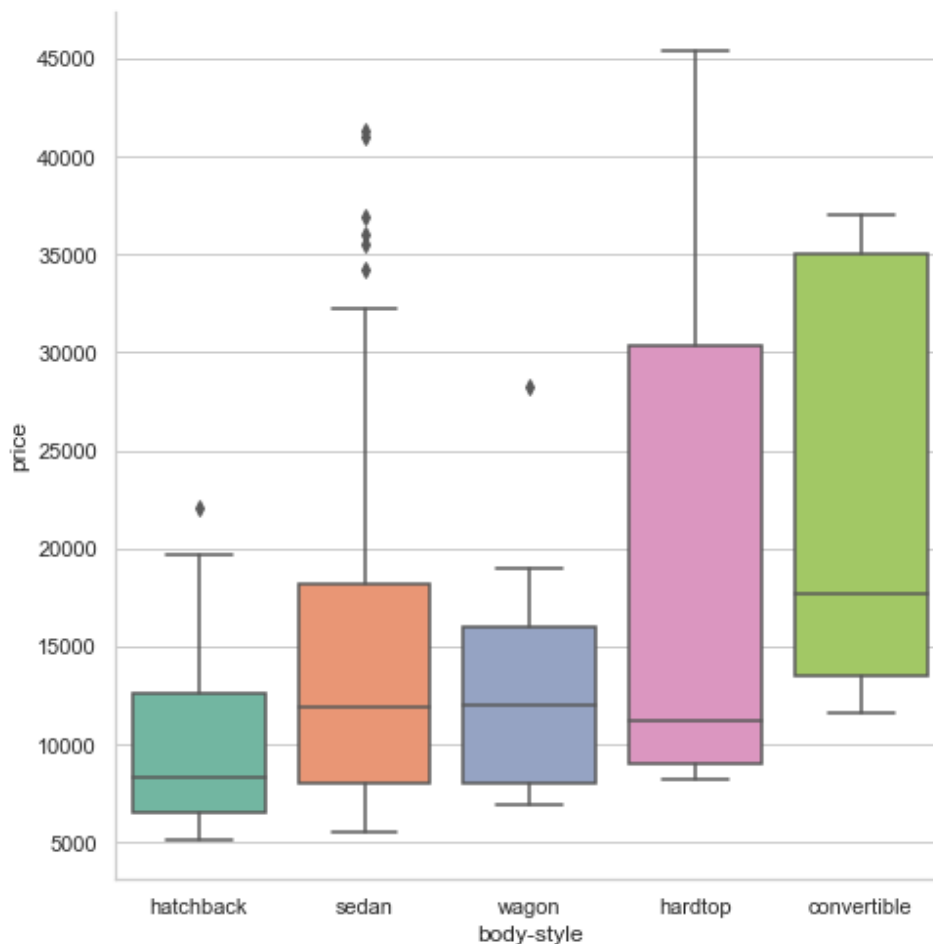
Figure 6: Price shift characterized by body style

When comparing body styles of vehicles and their price, convertibles are found at the high end of the price scale, with an average cost lying somewhere close the max cost of a vehicle that has a sedan body type. However generally the trend here in increasing cost is Hatchback, sedan, wagon, hardtop and convertible. The hardtop and convertibles even though at the high end of the price scale, have a large upper quartile indicating that majority of the vehicles cost are much higher than the average cost for those vehicles.

There are some outliers represented by dots above which indicate extreme costs from the normal pricing of those types of vehicles. These may be due to other factors such as engine size, and horsepower as well as wheelbase length, and other contributing factors not discussed in this dataset.

Table 2: Average characteristics for each Wheel drive

	engine-size	wheel-base	horsepower	price
4 Wheel Drive	107.50	96.49	87.25	10241.00
Front Wheel Drive	108.61	96.45	86.51	9224.25
Rear Wheel Drive	158.06	103.13	131.82	19739.41

Table 2 , gives the mean values for engine size, wheelbase, horsepower and price for these categorized by wheel drive. It is apparent that vehicles with rear-wheel-drive average costs are greater than the other two, and for good reason, as you can see greater engine size, longer wheelbase and much greater horsepower in such vehicles. This would stipulate significantly for the difference in price.

Table 3: Average cost per wheel type based on the body-type characteristics

	convertible	hardtop	hatchback	sedan	wagon
4 Wheel Drive	0	0	7603.00	12647.33	9095.75
Front Wheel Drive	11595.0	8249.00	8317.72	9851.80	10061.18
Rear Wheel Drive	25812.0	22565.17	14484.41	21819.31	16994.22

When it comes to comparison of body types and cost, yet again rear wheel vehicles are more costly. There was no data on 4-wheel drive vehicles with convertibles and hardtops, and this may be because those vehicles were not made, or sold by the company. A characteristic increase in cost can be seen in ascending order from; convertibles, hardtop, hatchback, sedan and wagon.
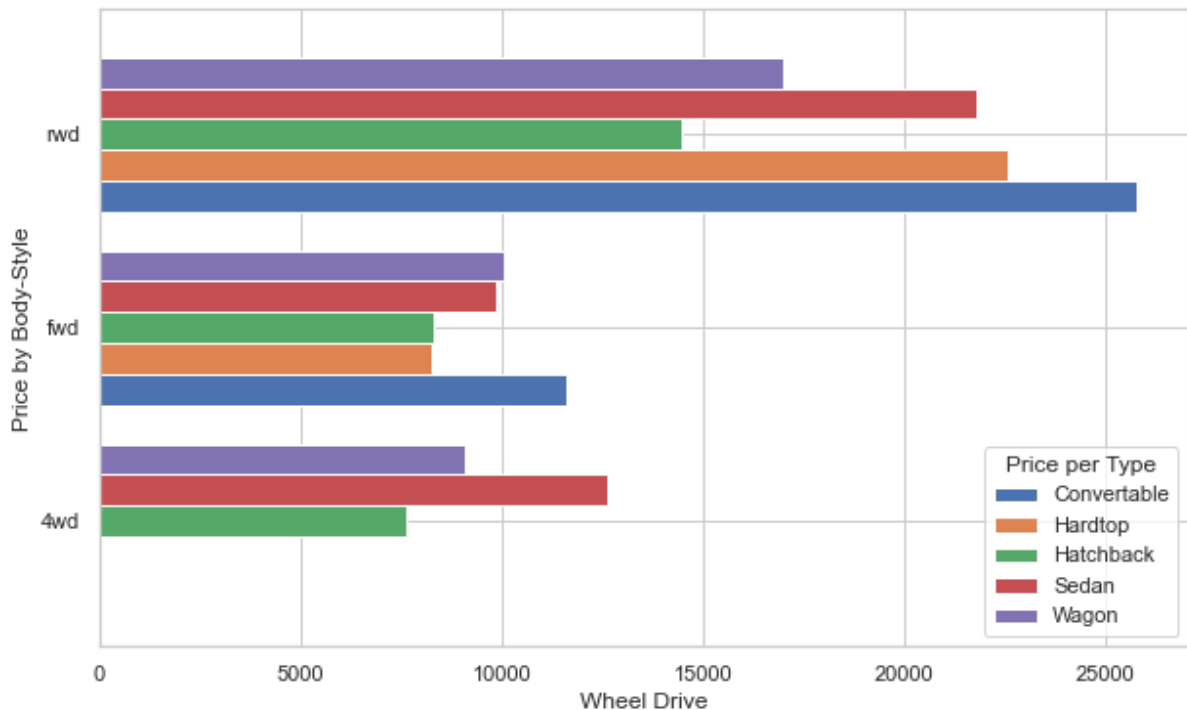
Figure 7: Price shift characterized by wheel drive and grouped by body style

To confirm this, figure 7 indicates that rear wheel vehicles are more costly irrespective of the body type, and of the body types it is the convertible which are most expensive in both rear and front wheel drive vehicles. Further more we saw that in fig 6, the convertible type displayed a rather large 3rd Quartile range.

So, in conclusion vehicles price range in ascending price order based on the body type with the costliest ones being hardtop and convertible. Front wheel and four-wheel drive vehicles don't have a large significant difference in value, rather the characteristics of engine type, wheel base, horsepower play an important role in increasing costs per vehicle. Thus it can be concluded that vehicle characteristics such as the ones discussed here are major factors that influence the cost of vehicles.

THIS REPORT WAS WRITTEN BY : RAMAN SEWJUGATH