In [74]:
```
#1. Import following libraries:
#a. Numpy and Pandas
#b. Matplotlib.pyplot
#c. Seaborn



import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [124]:
```
#2. Import the dataset using the following command in your Jupyter Noteboo


dataset = pd.read_csv('CarPrice_Assignment.csv')
dataset
```

Out[124]:

	car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	e
0	1	3	alfa-romero giulia	gas	std	two	convertible	rwd	
1	2	3	alfa-romero stelvio	gas	std	two	convertible	rwd	
2	3	1	alfa-romero Quadrifoglio	gas	std	two	hatchback	rwd	
3	4	2	audi 100 ls	gas	std	four	sedan	fwd	
4	5	2	audi 100ls	gas	std	four	sedan	4wd	
...	...	...	...	...	...	...	...	...	...
200	201	-1	volvo 145e (sw)	gas	std	four	sedan	rwd	
201	202	-1	volvo 144ea	gas	turbo	four	sedan	rwd	
202	203	-1	volvo 244dl	gas	std	four	sedan	rwd	
203	204	-1	volvo 246	diesel	turbo	four	sedan	rwd	
204	205	-1	volvo 264gl	gas	turbo	four	sedan	rwd	

205 rows × 26 columns

In [77]:
```
1 dataset.isnull()
```

Out[77]:

	car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	engin
0	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	
200	False	False	False	False	False	False	False	False	
201	False	False	False	False	False	False	False	False	
202	False	False	False	False	False	False	False	False	
203	False	False	False	False	False	False	False	False	
204	False	False	False	False	False	False	False	False	

205 rows × 26 columns

In [78]:
```
1 dataset.nunique()
```

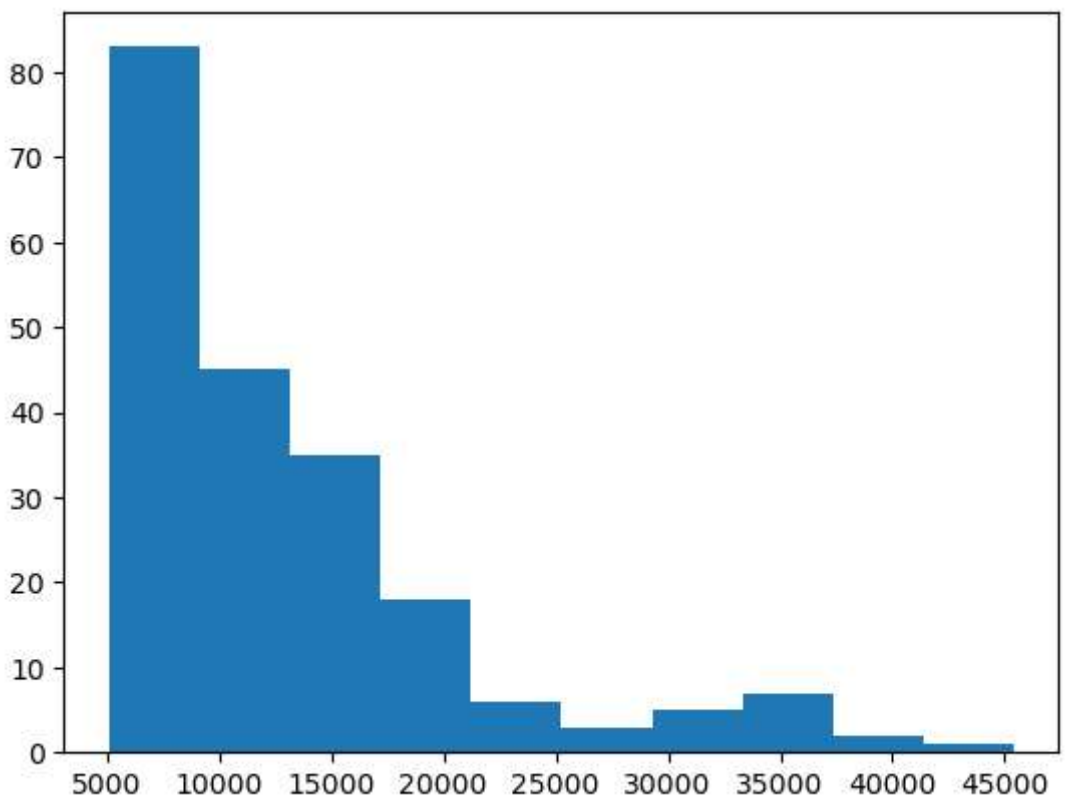
Out[78]:
```
car_ID              205
symboling             6
CarName             147
fueltype              2
aspiration            2
doornumber            2
carbody               5
drivewheel            3
enginelocation        2
wheelbase            53
carlength            75
carwidth             44
carheight            49
curbweight          171
enginetype            7
cylindernumber        7
enginesize           44
fuelsystem            8
boreratio            38
stroke               37
compressionratio     32
horsepower           59
peakrpm              23
citympg              29
highwaympg           30
price               189
dtype: int64
```

In [79]:
```
#4. Check for duplicate values in the dataset

duplicates = dataset.duplicated()
duplicates
```

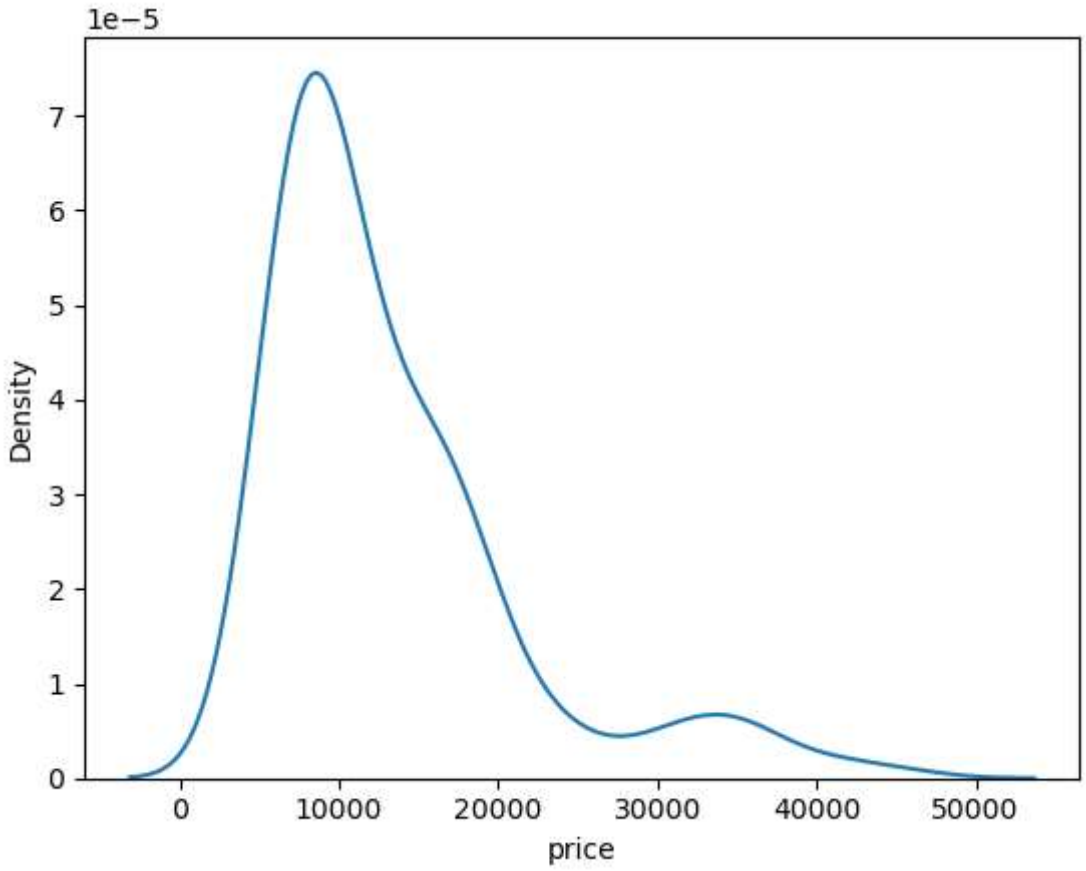
Out[79]:
```
0      False
1      False
2      False
3      False
4      False
       ...
200    False
201    False
202    False
203    False
204    False
Length: 205, dtype: bool
```

In [80]:
```
#5. Create a Histogram, KDE plot and box plot for variable 'price'

plt.hist(dataset['price'])
```

Out[80]:
```
(array([83., 45., 35., 18.,  6.,  3.,  5.,  7.,  2.,  1.]),
 array([ 5118. ,  9146.2, 13174.4, 17202.6, 21230.8, 25259. , 29287.2,
        33315.4, 37343.6, 41371.8, 45400. ]),
 <BarContainer object of 10 artists>)
```

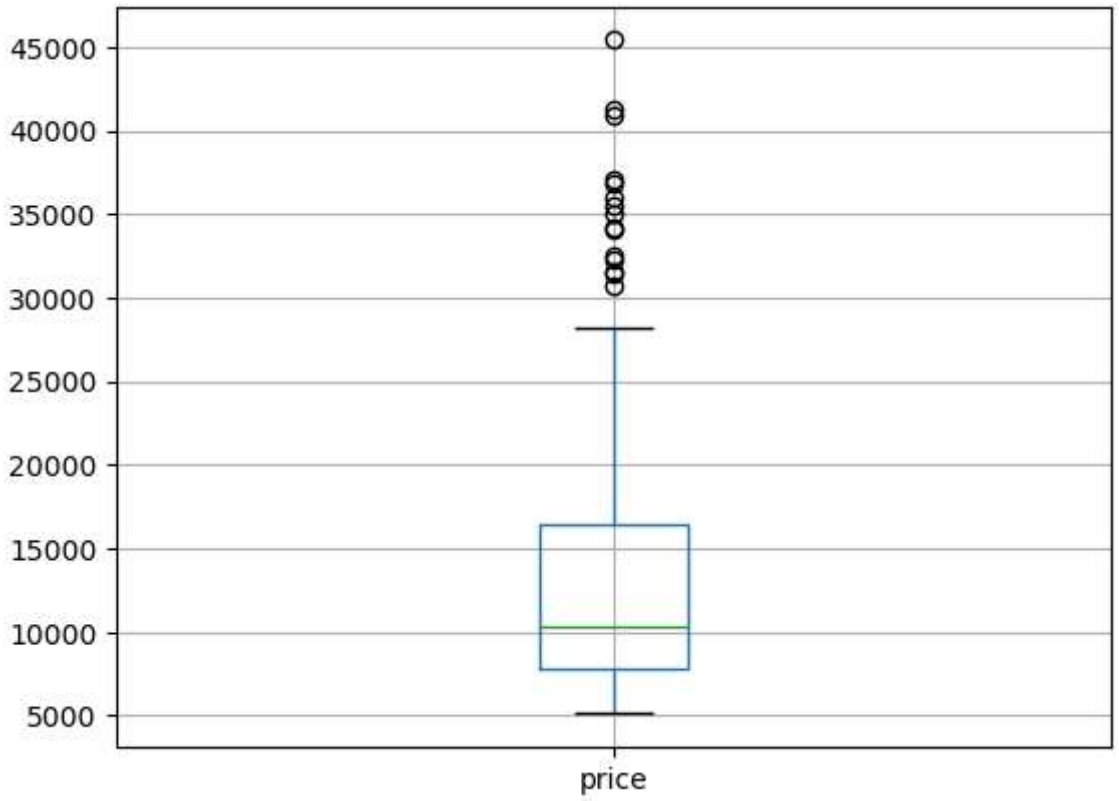


In [81]:
```
#5. Create a Histogram, KDE plot and box plot for variable 'price'

price = dataset["price"]
sns.kdeplot(price)
plt.show()
```



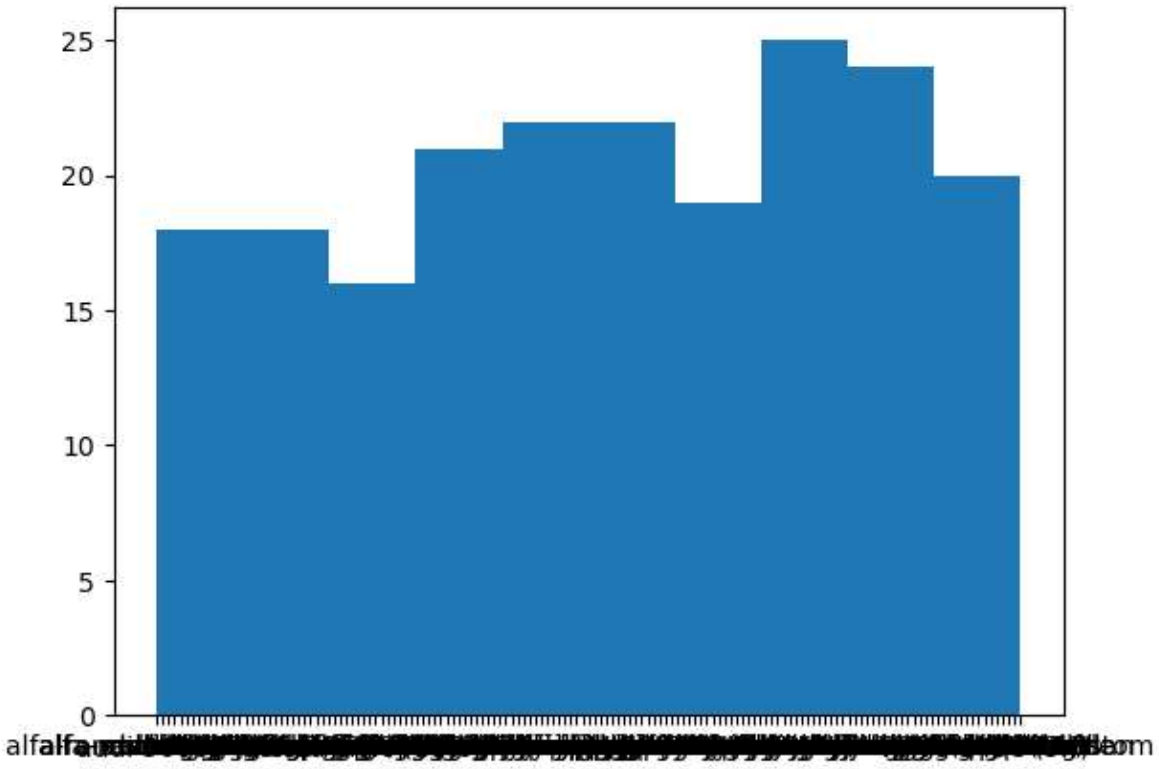
In [82]:
```
#5. Create a Histogram, KDE plot and box plot for variable 'price'

dataset.boxplot("price")
plt.show
```

Out[82]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



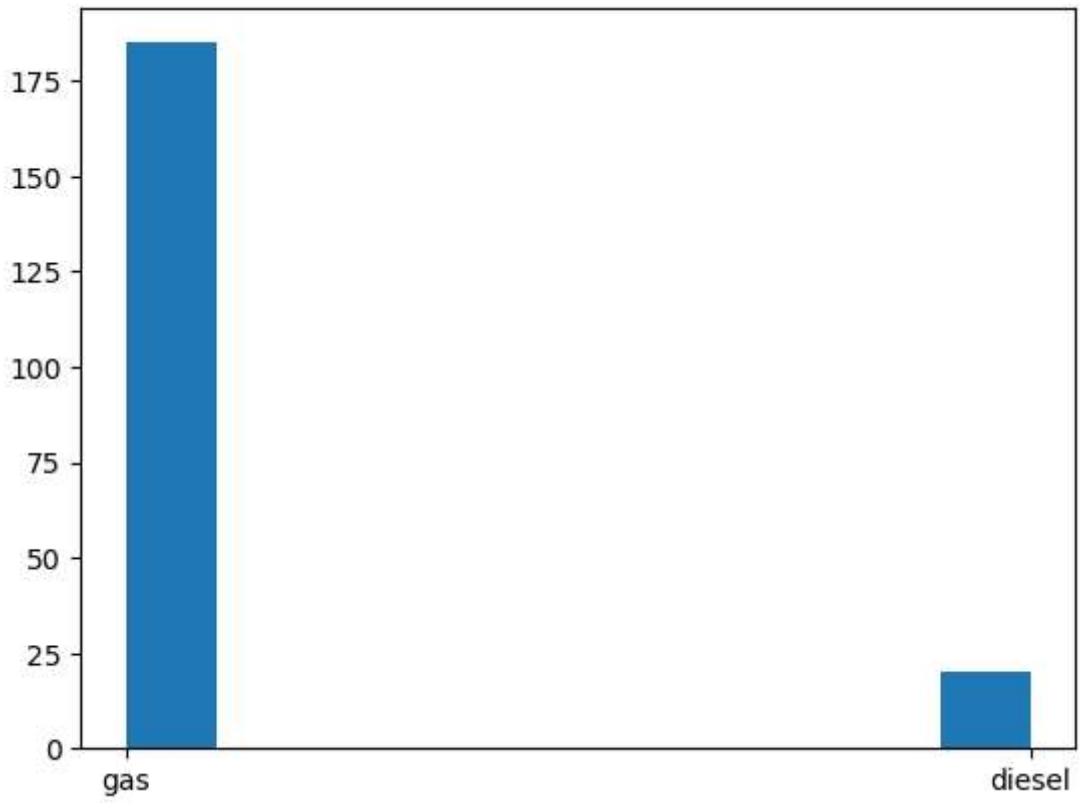
In [83]:
```
#6. Create Histograms for Car Company, Fuel Type, Car Type, Symboling, Eng

plt.hist(dataset['CarName'])
```

Out[83]:
```
(array([18., 18., 16., 21., 22., 22., 19., 25., 24., 20.]),
 array([  0. ,  14.6,  29.2,  43.8,  58.4,  73. ,  87.6, 102.2, 116.8,
        131.4, 146. ]),
 <BarContainer object of 10 artists>)
```



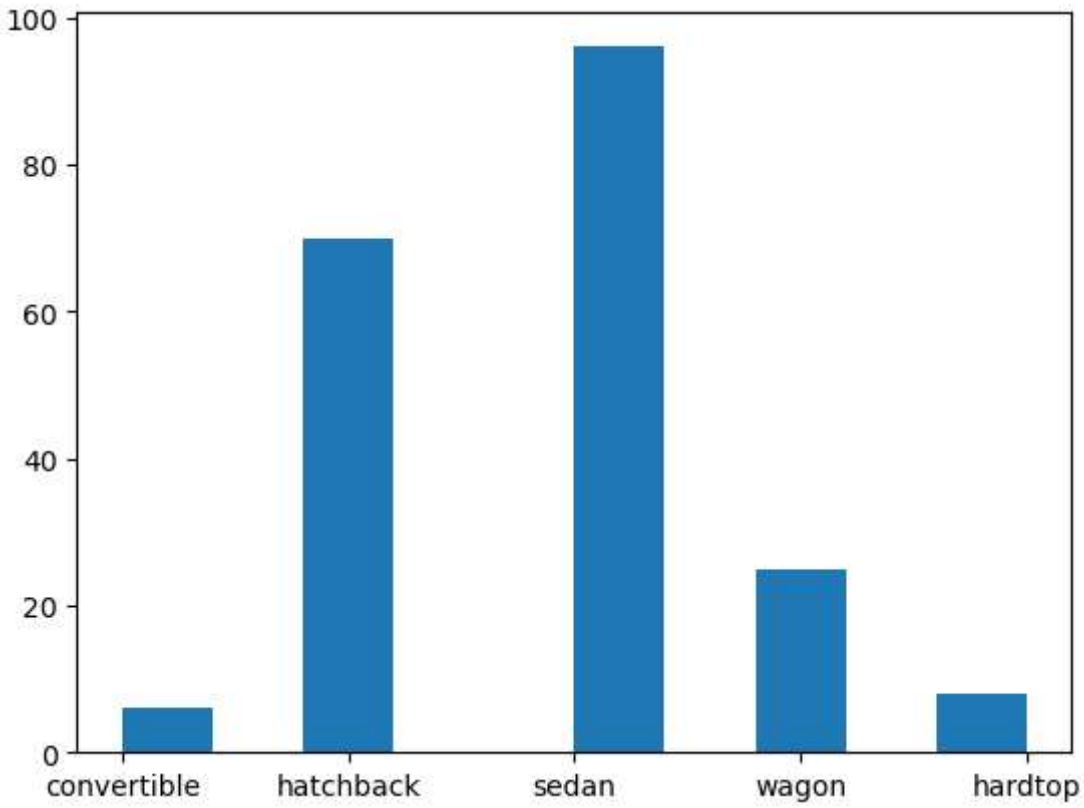
In [84]:
```
#6. Create Histograms for Car Company, Fuel Type, Car Type, Symboling, Eng

plt.hist(dataset['fueltype'])
```

Out[84]:
```
(array([185.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,  20.]),
 array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <BarContainer object of 10 artists>)
```



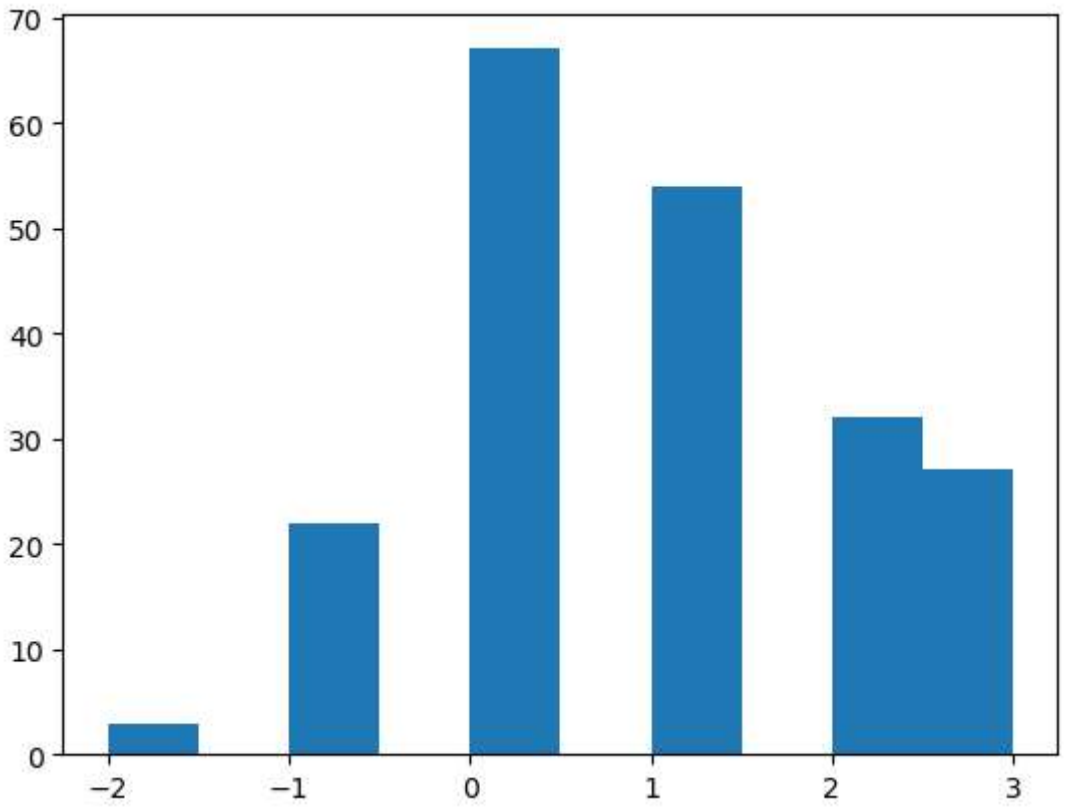
In [85]:
```
#6. Create Histograms for Car Company, Fuel Type, Car Type, Symboling, Eng

plt.hist(dataset['carbody'])
```

Out[85]:
```
(array([ 6.,  0., 70.,  0.,  0., 96.,  0., 25.,  0.,  8.]),
 array([0. , 0.4, 0.8, 1.2, 1.6, 2. , 2.4, 2.8, 3.2, 3.6, 4. ]),
 <BarContainer object of 10 artists>)
```



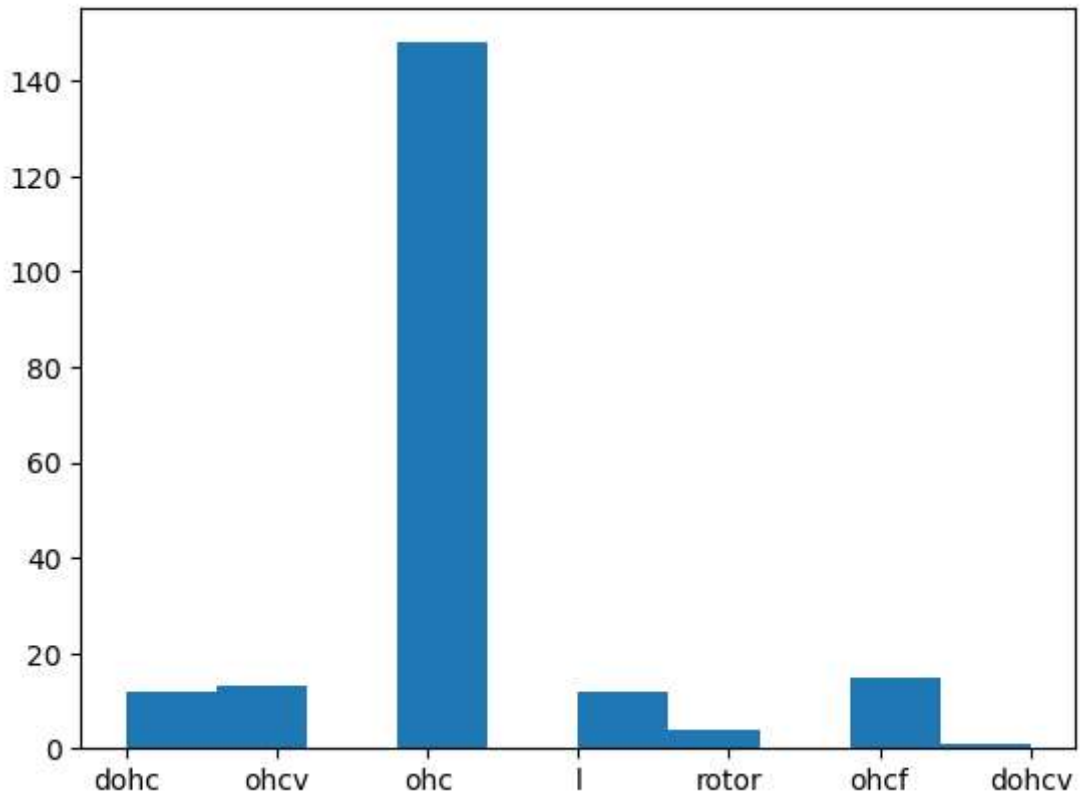
In [86]:
```
#6. Create Histograms for Car Company, Fuel Type, Car Type, Symboling, Eng

plt.hist(dataset['symboling'])
```

Out[86]:
```
(array([ 3.,  0., 22.,  0., 67.,  0., 54.,  0., 32., 27.]),
 array([-2. , -1.5, -1. , -0.5,  0. ,  0.5,  1. ,  1.5,  2. ,  2.5,  3. ]),
 <BarContainer object of 10 artists>)
```



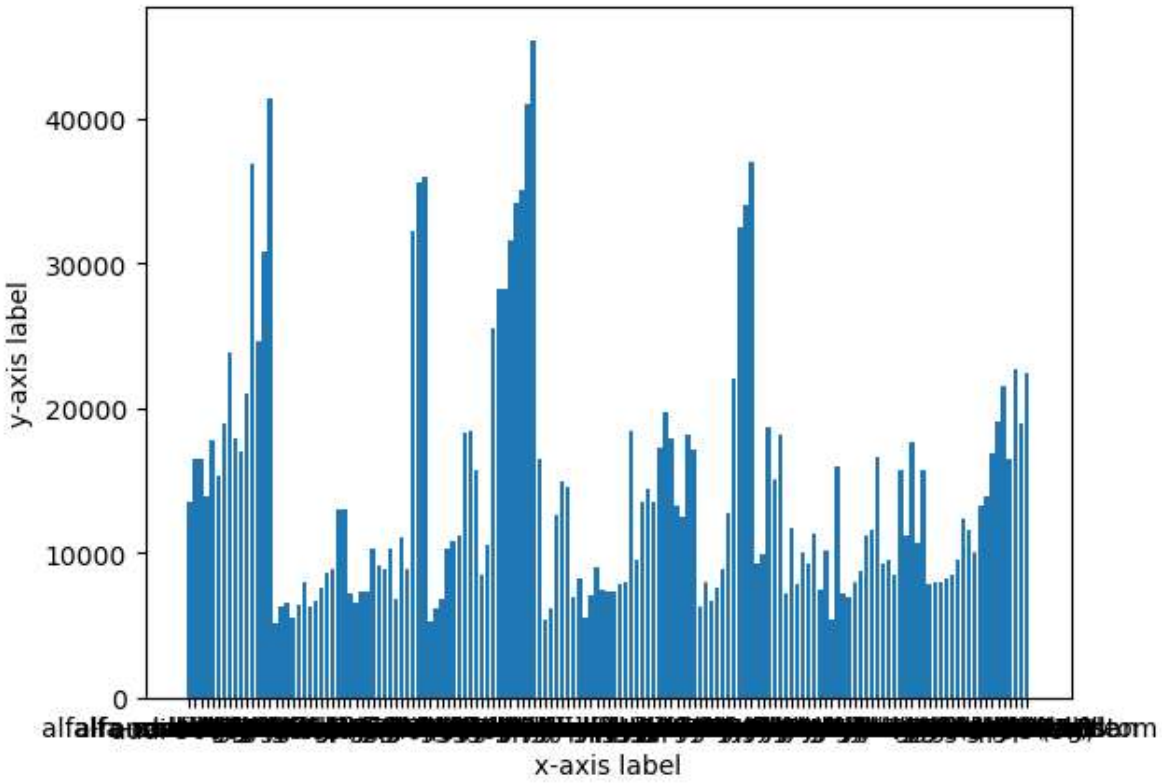
In [87]:
```
#6. Create Histograms for Car Company, Fuel Type, Car Type, Symboling, Eng

plt.hist(dataset['enginetype'])
```

Out[87]:
```
(array([ 12.,  13.,   0., 148.,   0.,  12.,   4.,   0.,  15.,   1.]),
 array([0. , 0.6, 1.2, 1.8, 2.4, 3. , 3.6, 4.2, 4.8, 5.4, 6. ]),
 <BarContainer object of 10 artists>)
```



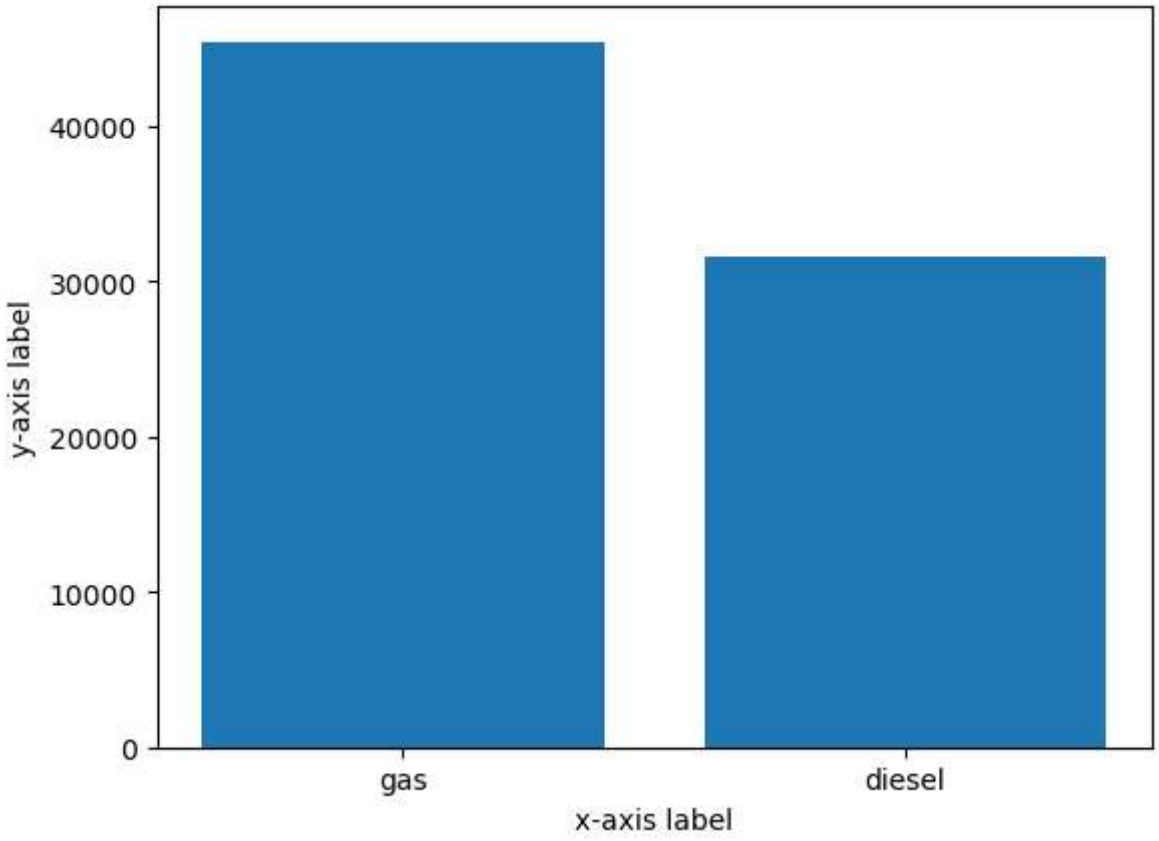
In [88]:

```
#7. Create bar plots for the following:
#a. Company name vs Average Price


x = dataset['CarName']
y = dataset['price']

plt.bar(x, y)

plt.xlabel('x-axis label')
plt.ylabel('y-axis label')

plt.show()
```



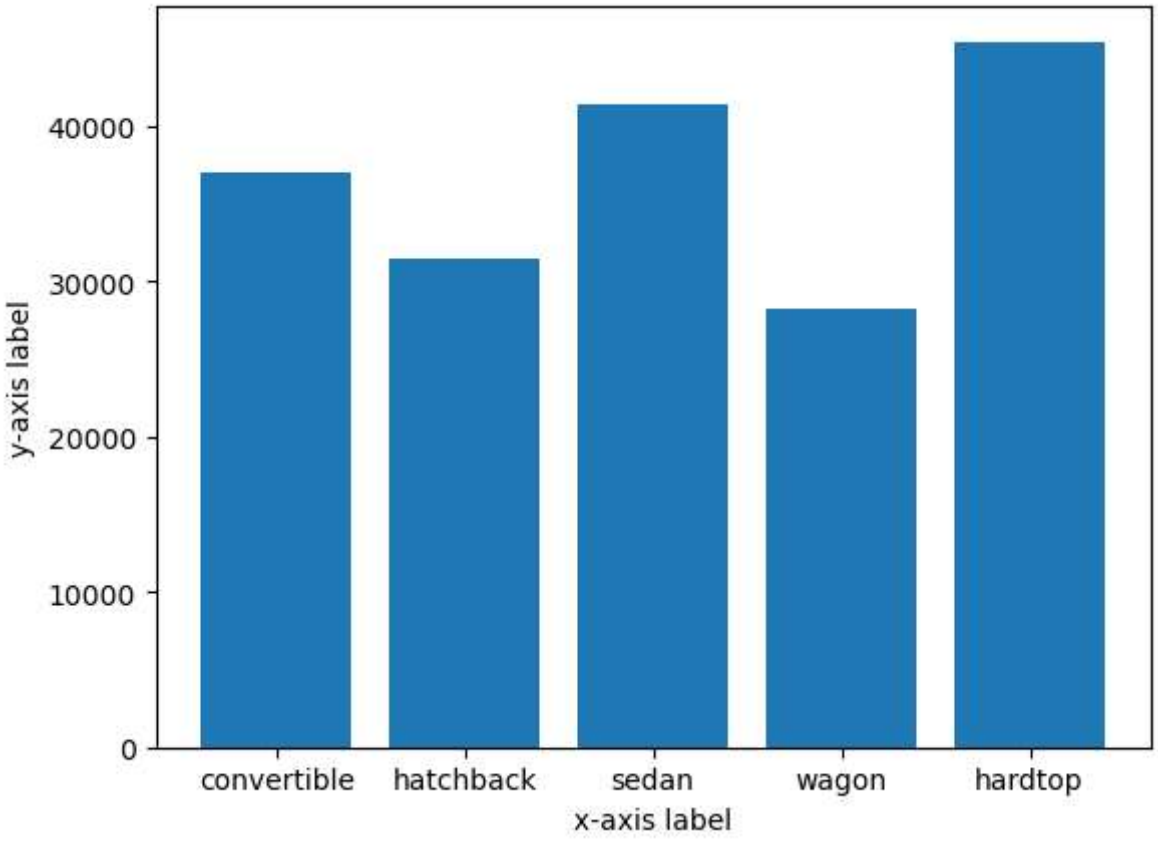
In [89]:

```
#b. Fuel type vs Average Price


x = dataset['fueltype']
y = dataset['price']

plt.bar(x, y)

plt.xlabel('x-axis label')
plt.ylabel('y-axis label')

plt.show()
```



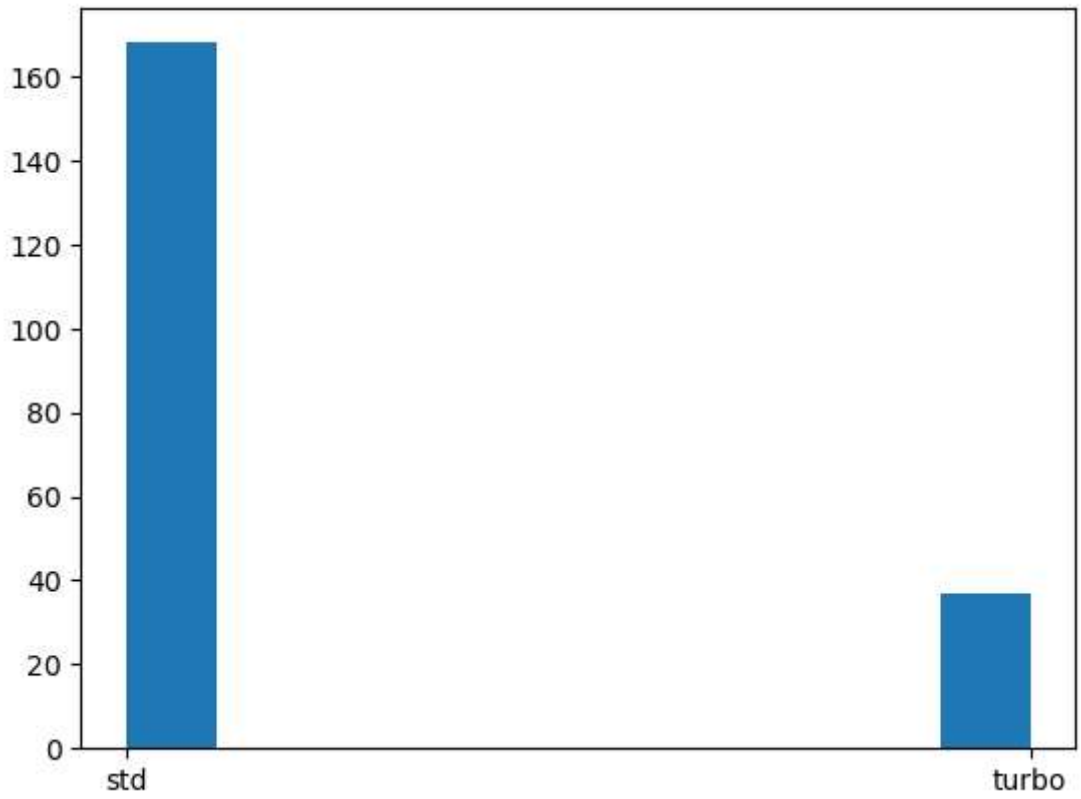
In [90]:

```
#c. Car Type vs Average Price


x = dataset['carbody']
y = dataset['price']

plt.bar(x, y)

plt.xlabel('x-axis label')
plt.ylabel('y-axis label')

plt.show()
```



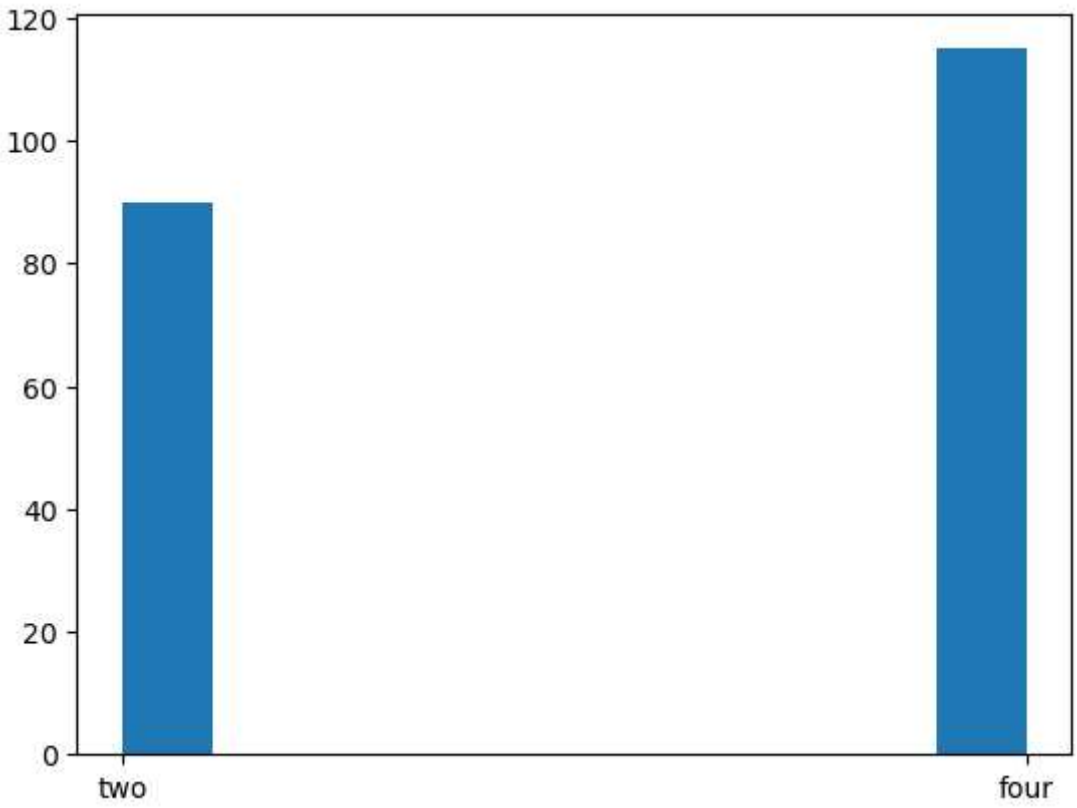
In [91]:
```
#8. Create Histograms for rest of the variables (excluding the ones that 


plt.hist(dataset['aspiration'])
```

Out[91]:
```
(array([168.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,  37.]),
 array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <BarContainer object of 10 artists>)
```



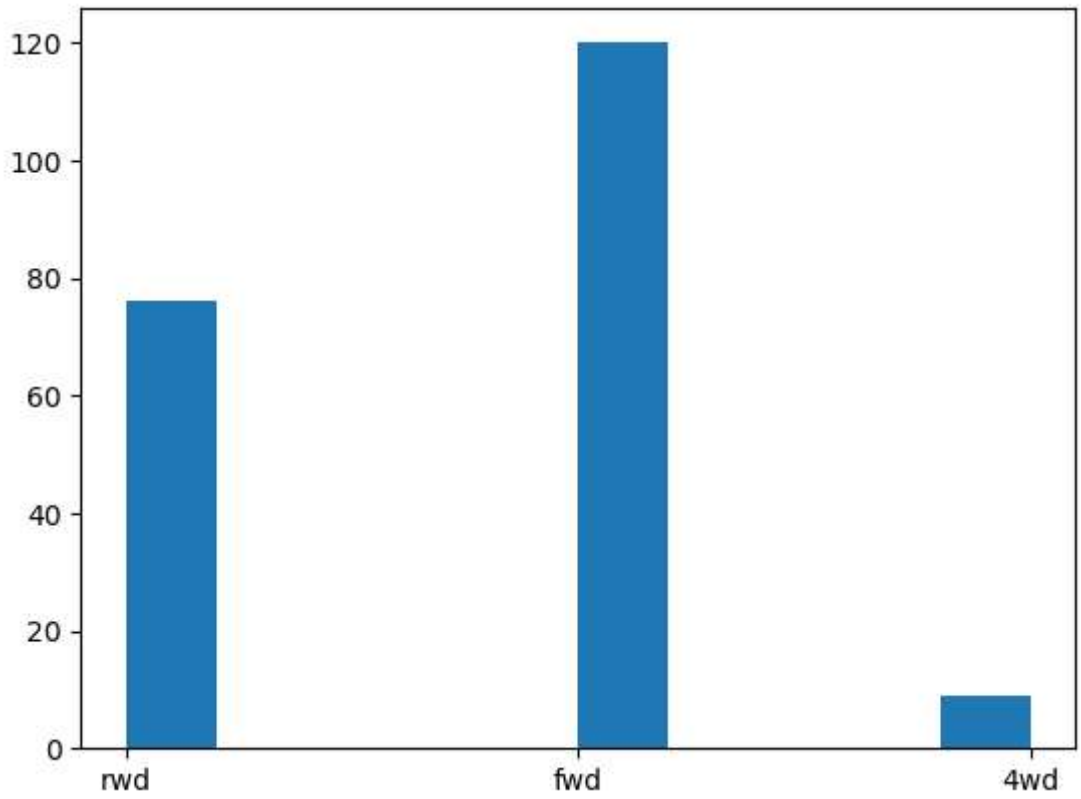
In [92]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['doornumber'])
```

Out[92]:
```
(array([ 90.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0., 115.]),
 array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <BarContainer object of 10 artists>)
```



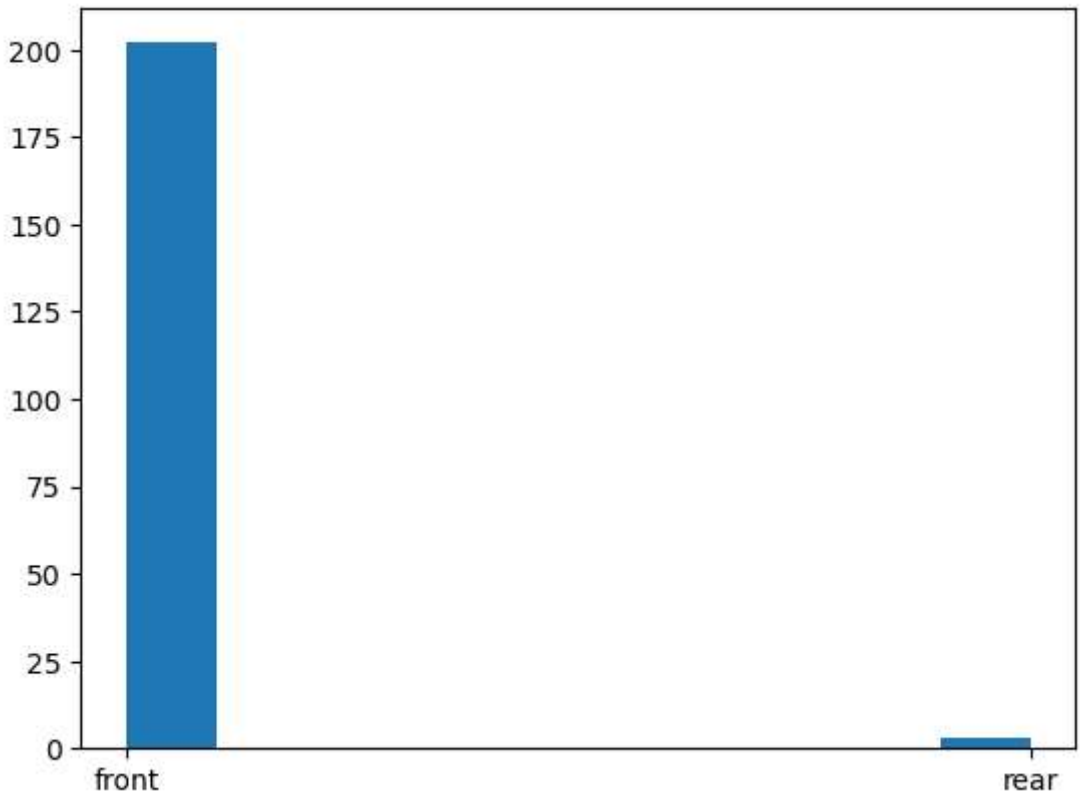
In [93]:
```
#8. Create Histograms for rest of the variables (excluding the ones that 


plt.hist(dataset['drivewheel'])
```

Out[93]:
```
(array([ 76.,   0.,   0.,   0.,   0., 120.,   0.,   0.,   0.,   9.]),
 array([0. , 0.2, 0.4, 0.6, 0.8, 1. , 1.2, 1.4, 1.6, 1.8, 2. ]),
 <BarContainer object of 10 artists>)
```



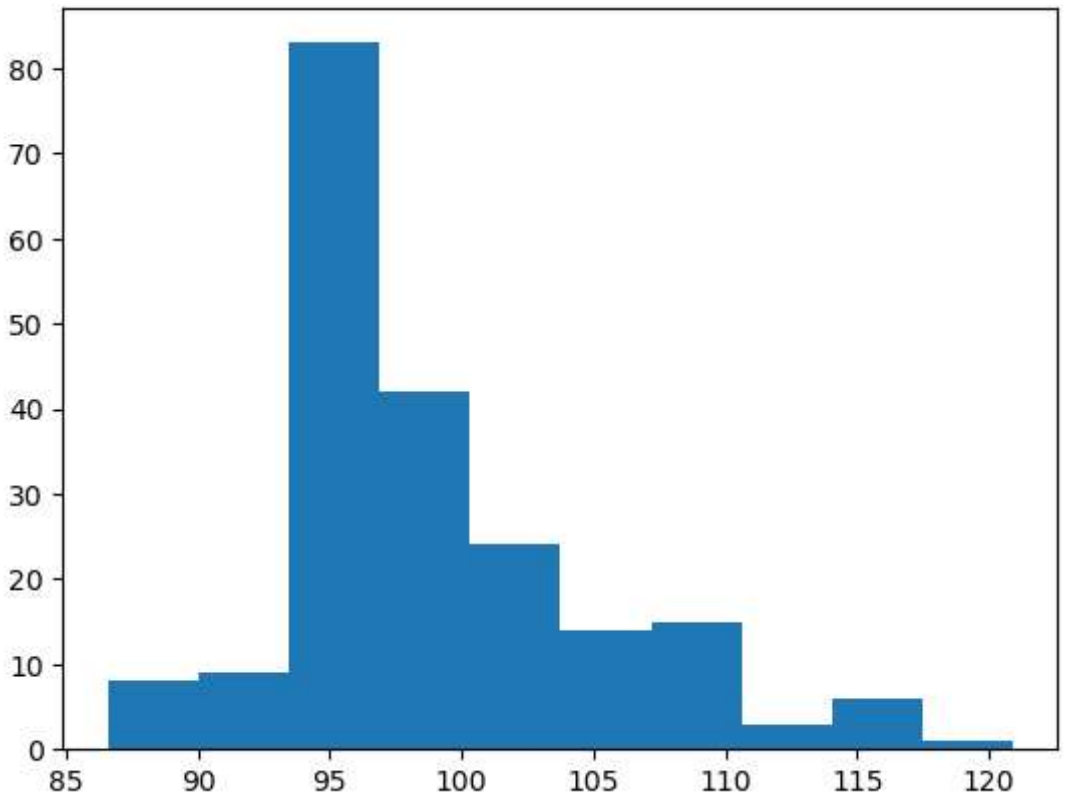
In [94]:
```
#8. Create Histograms for rest of the variables (excluding the ones that 


plt.hist(dataset['enginelocation'])
```

Out[94]:
```
(array([202.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   0.,   3.]),
 array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <BarContainer object of 10 artists>)
```



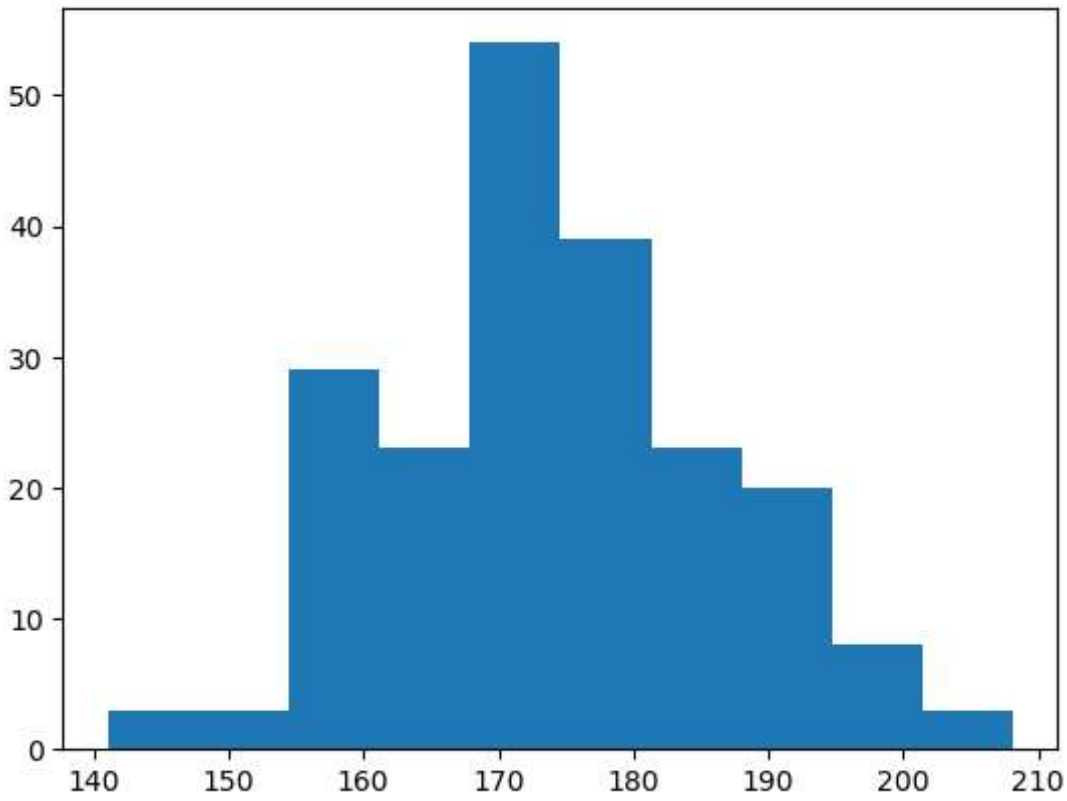
In [95]:
```
#8. Create Histograms for rest of the variables (excluding the ones that 


plt.hist(dataset['wheelbase'])
```

Out[95]:
```
(array([ 8.,  9., 83., 42., 24., 14., 15.,  3.,  6.,  1.]),
 array([ 86.6 ,  90.03,  93.46,  96.89, 100.32, 103.75, 107.18, 110.61,
        114.04, 117.47, 120.9 ]),
 <BarContainer object of 10 artists>)
```



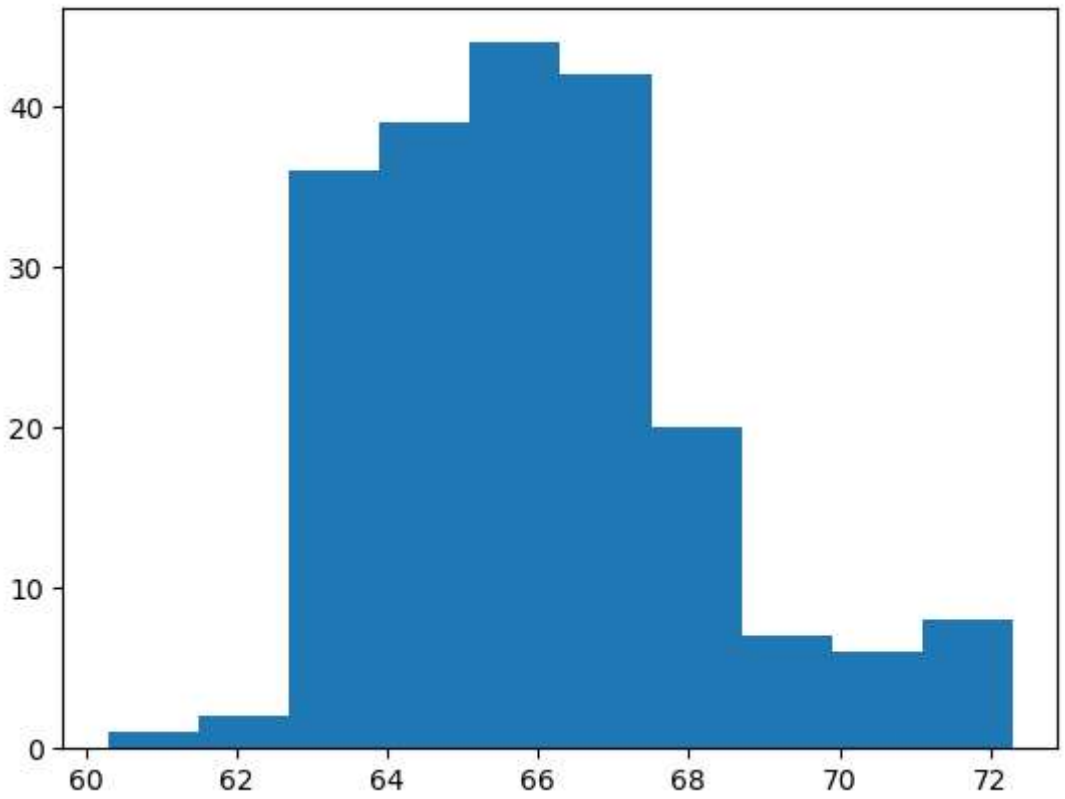
In [96]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['carlength'])
```

Out[96]:
```
(array([ 3.,  3., 29., 23., 54., 39., 23., 20.,  8.,  3.]),
 array([141.1, 147.8, 154.5, 161.2, 167.9, 174.6, 181.3, 188. , 194.7,
        201.4, 208.1]),
 <BarContainer object of 10 artists>)
```



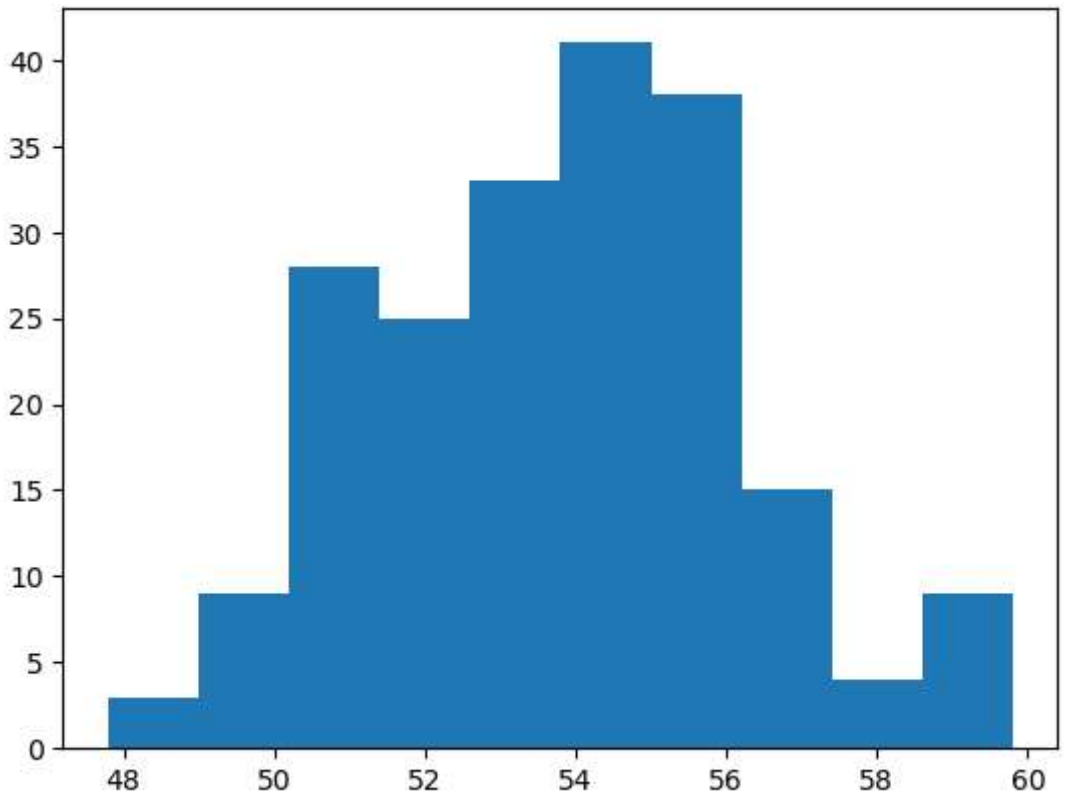
In [97]:
```
#8. Create Histograms for rest of the variables (excluding the ones that 


plt.hist(dataset['carwidth'])
```

Out[97]:
```
(array([ 1.,  2., 36., 39., 44., 42., 20.,  7.,  6.,  8.]),
 array([60.3, 61.5, 62.7, 63.9, 65.1, 66.3, 67.5, 68.7, 69.9, 71.1, 72.3]),
 <BarContainer object of 10 artists>)
```



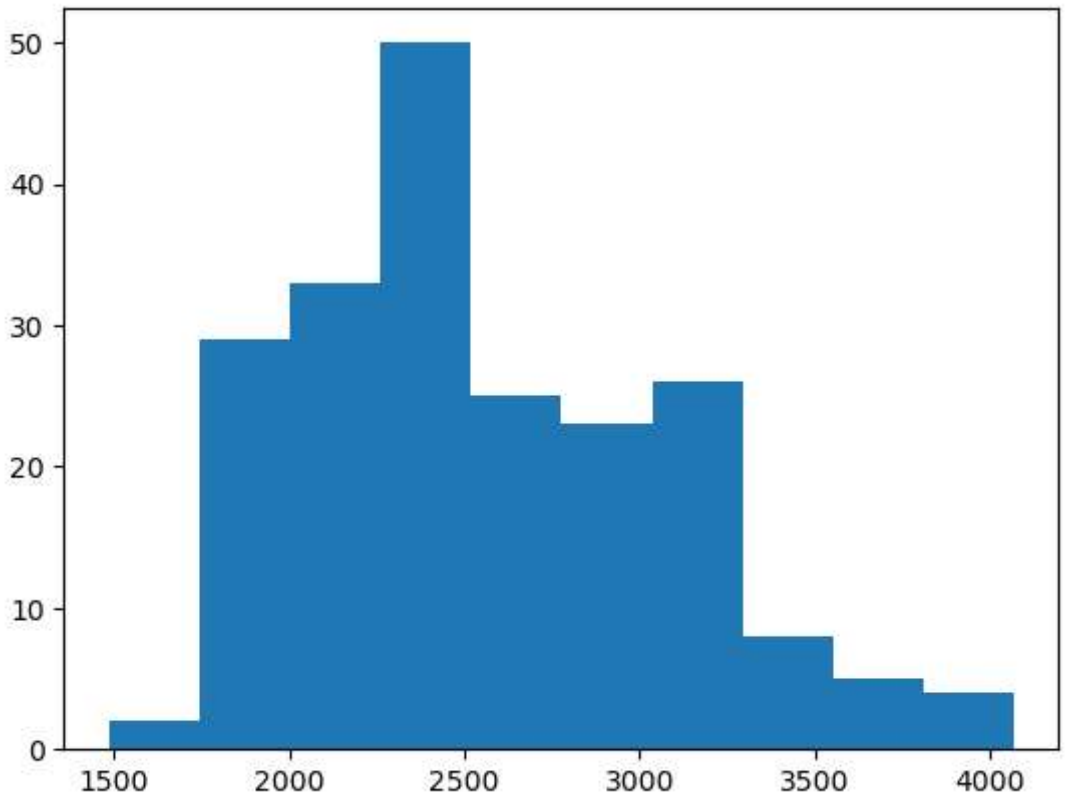
In [98]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['carheight'])
```

Out[98]:
```
(array([ 3.,  9., 28., 25., 33., 41., 38., 15.,  4.,  9.]),
 array([47.8, 49. , 50.2, 51.4, 52.6, 53.8, 55. , 56.2, 57.4, 58.6, 59.8]),
 <BarContainer object of 10 artists>)
```



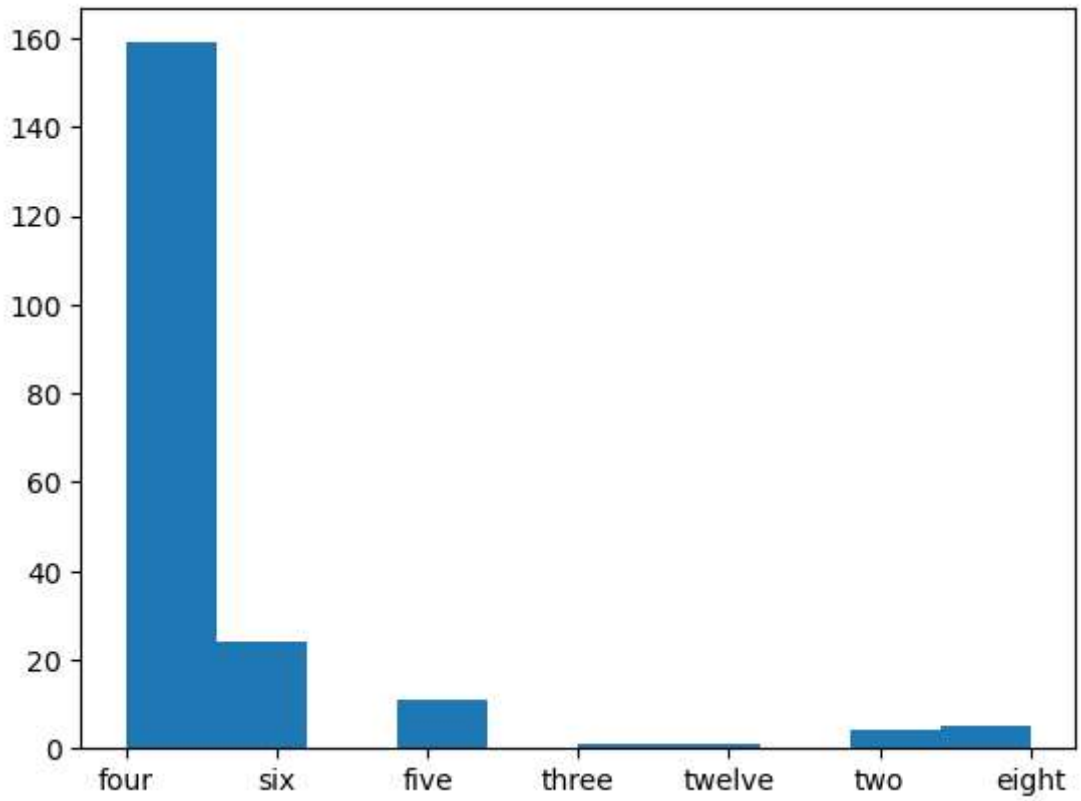
In [99]:
```
#8. Create Histograms for rest of the variables (excluding the ones that 


plt.hist(dataset['curbweight'])
```

Out[99]:
```
(array([ 2., 29., 33., 50., 25., 23., 26.,  8.,  5.,  4.]),
 array([1488. , 1745.8, 2003.6, 2261.4, 2519.2, 2777. , 3034.8, 3292.6,
        3550.4, 3808.2, 4066. ]),
 <BarContainer object of 10 artists>)
```



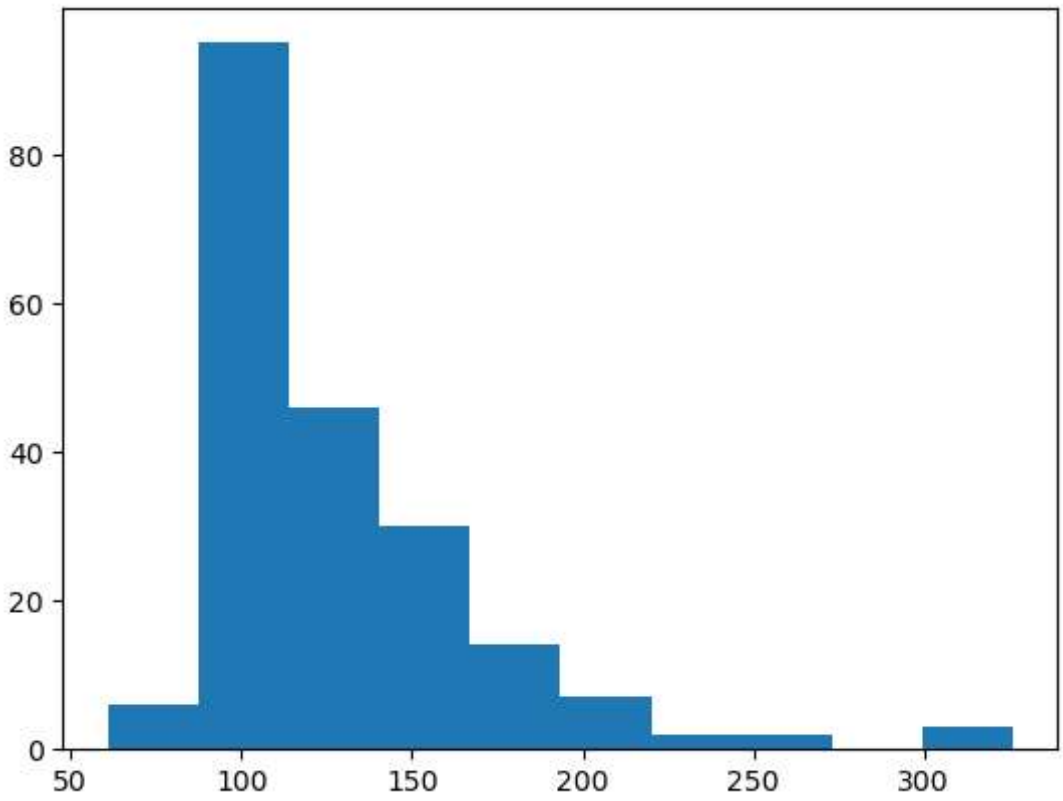
In [100]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['cylindernumber'])
```

Out[100]:
```
(array([159.,  24.,   0.,  11.,   0.,   1.,   1.,   0.,   4.,   5.]),
 array([0. , 0.6, 1.2, 1.8, 2.4, 3. , 3.6, 4.2, 4.8, 5.4, 6. ]),
 <BarContainer object of 10 artists>)
```



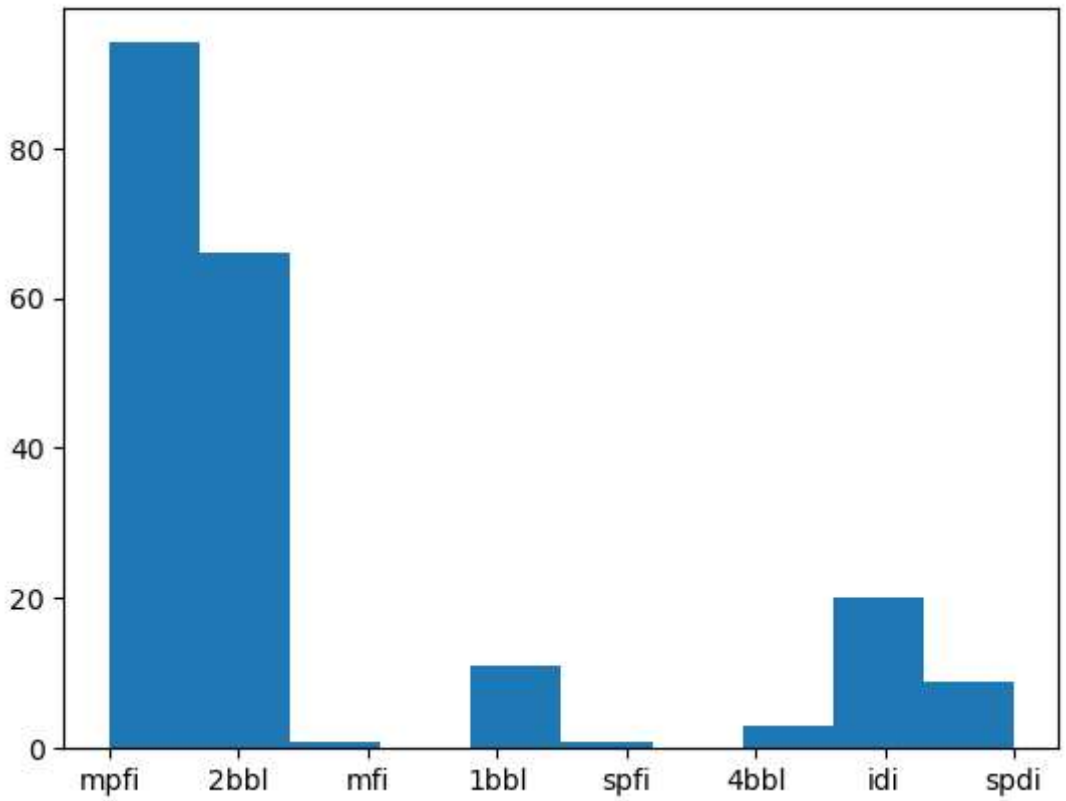
In [101]:
```
#8. Create Histograms for rest of the variables (excluding the ones that 


plt.hist(dataset['enginesize'])
```

Out[101]:
```
(array([ 6., 95., 46., 30., 14.,  7.,  2.,  2.,  0.,  3.]),
 array([ 61. ,  87.5, 114. , 140.5, 167. , 193.5, 220. , 246.5, 273. ,
        299.5, 326. ]),
 <BarContainer object of 10 artists>)
```



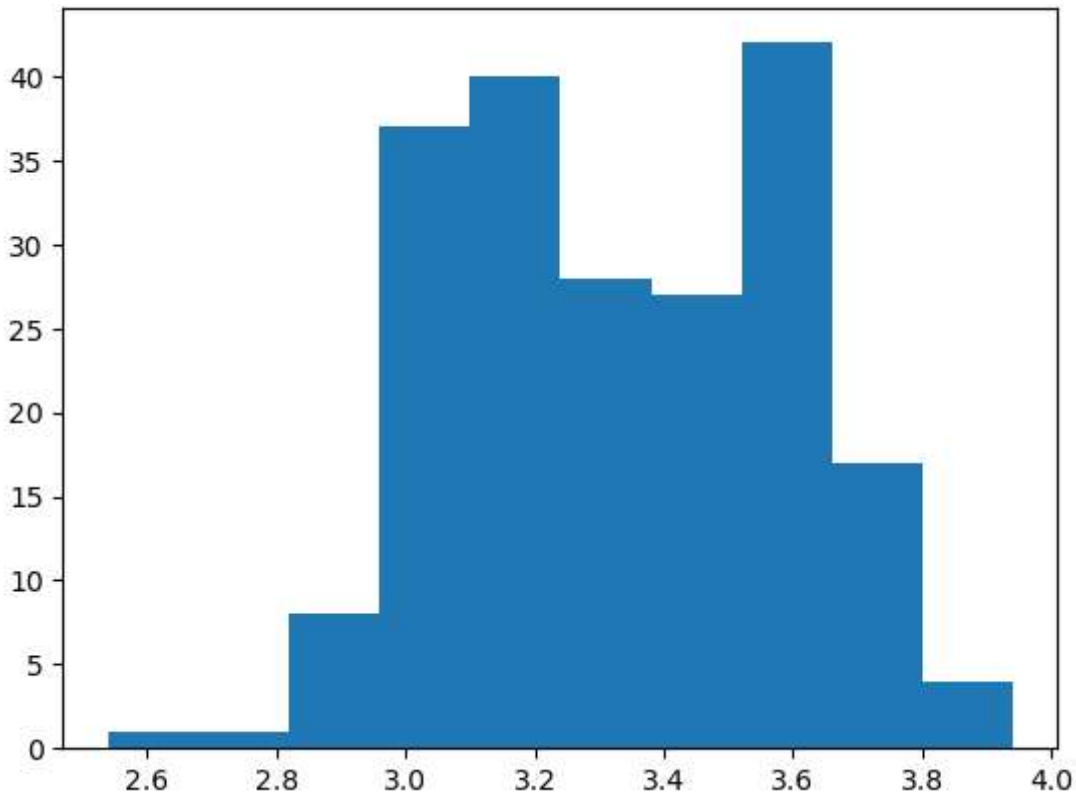
In [102]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['fuelsystem'])
```

Out[102]:
```
(array([94., 66.,  1.,  0., 11.,  1.,  0.,  3., 20.,  9.]),
 array([0. , 0.7, 1.4, 2.1, 2.8, 3.5, 4.2, 4.9, 5.6, 6.3, 7. ]),
 <BarContainer object of 10 artists>)
```



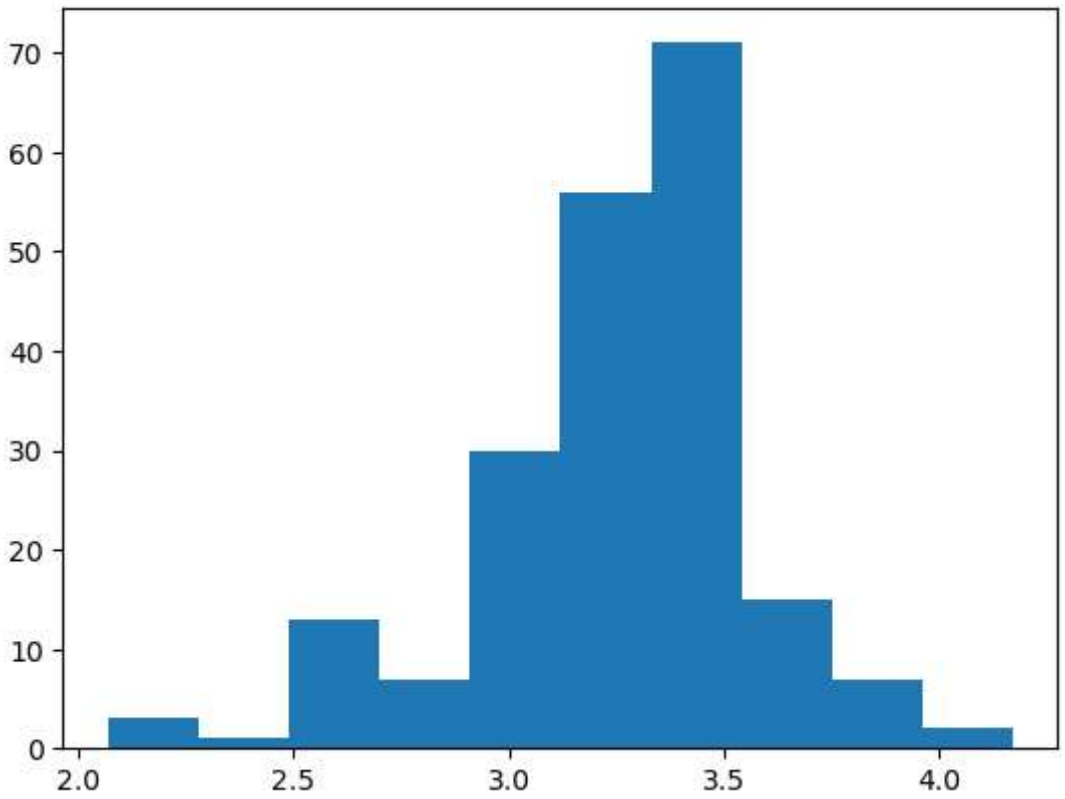
In [103]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['boreratio'])
```

Out[103]:
```
(array([ 1.,  1.,  8., 37., 40., 28., 27., 42., 17.,  4.]),
 array([2.54, 2.68, 2.82, 2.96, 3.1 , 3.24, 3.38, 3.52, 3.66, 3.8 , 3.94]),
 <BarContainer object of 10 artists>)
```



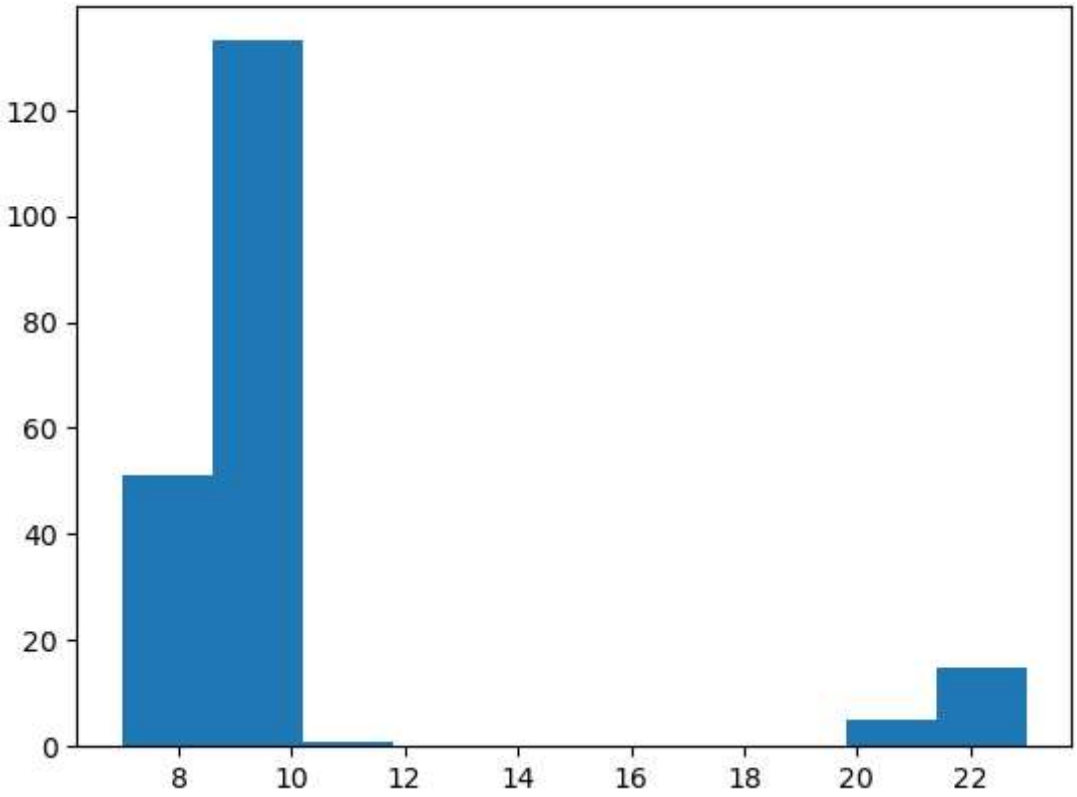
In [104]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['stroke'])
```

Out[104]:
```
(array([ 3.,  1., 13.,  7., 30., 56., 71., 15.,  7.,  2.]),
 array([2.07, 2.28, 2.49, 2.7 , 2.91, 3.12, 3.33, 3.54, 3.75, 3.96, 4.17]),
 <BarContainer object of 10 artists>)
```



In [105]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['compressionratio'])
```

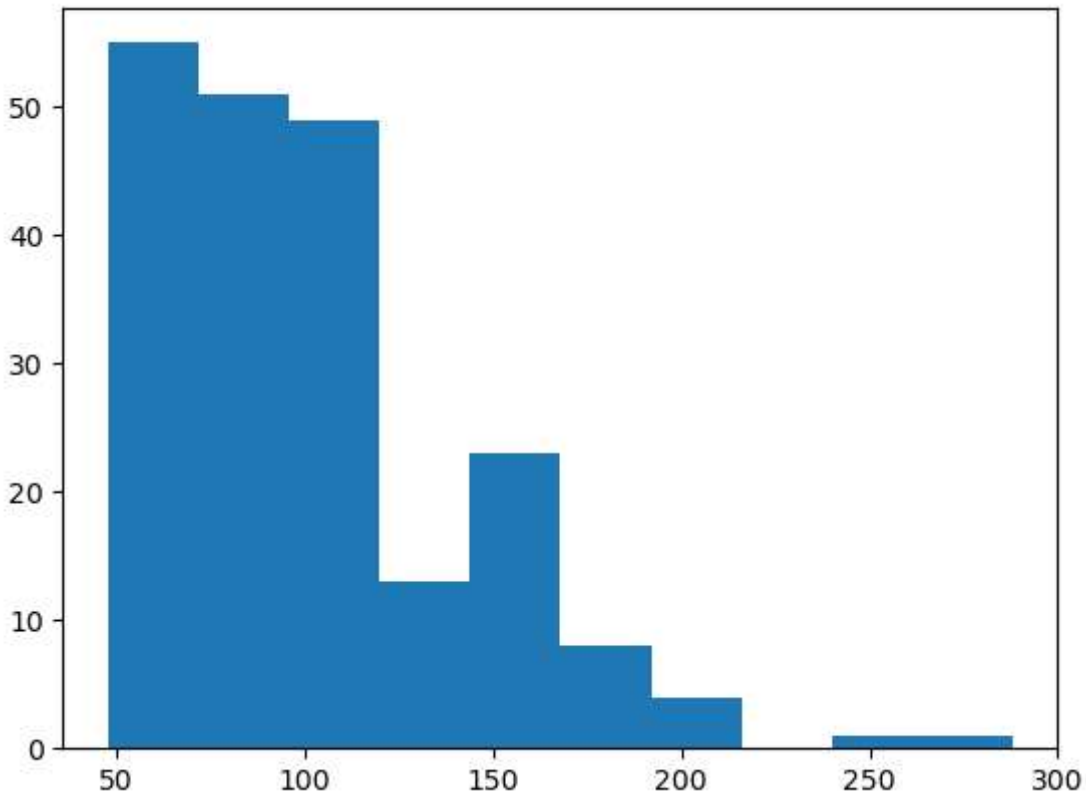
Out[105]:
```
(array([ 51., 133.,   1.,   0.,   0.,   0.,   0.,   0.,   5.,  15.]),
 array([ 7. ,  8.6, 10.2, 11.8, 13.4, 15. , 16.6, 18.2, 19.8, 21.4, 23. ]),
 <BarContainer object of 10 artists>)
```



In [106]:
```
#8. Create Histograms for rest of the variables (excluding the ones that
```

```

plt.hist(dataset['horsepower'])
```

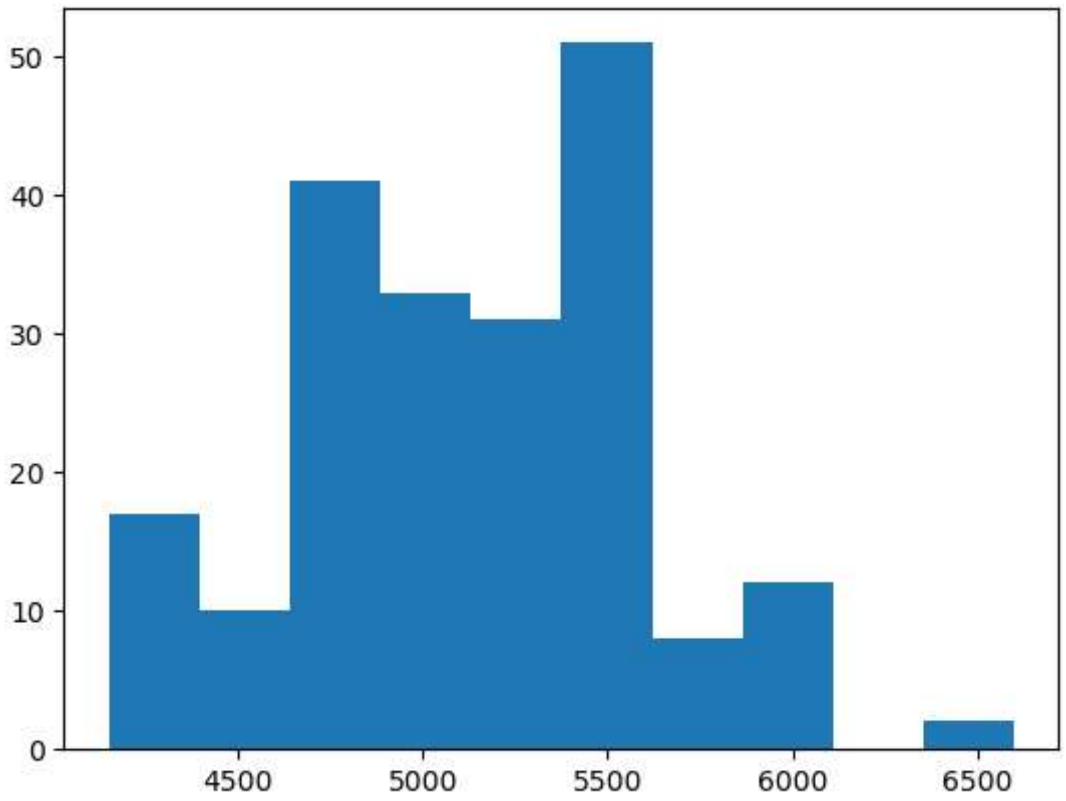
Out[106]:
```
(array([55., 51., 49., 13., 23.,  8.,  4.,  0.,  1.,  1.]),
 array([ 48.,  72.,  96., 120., 144., 168., 192., 216., 240., 264., 288.]),
 <BarContainer object of 10 artists>)
```



In [107]:
```
#8. Create Histograms for rest of the variables (excluding the ones that
```

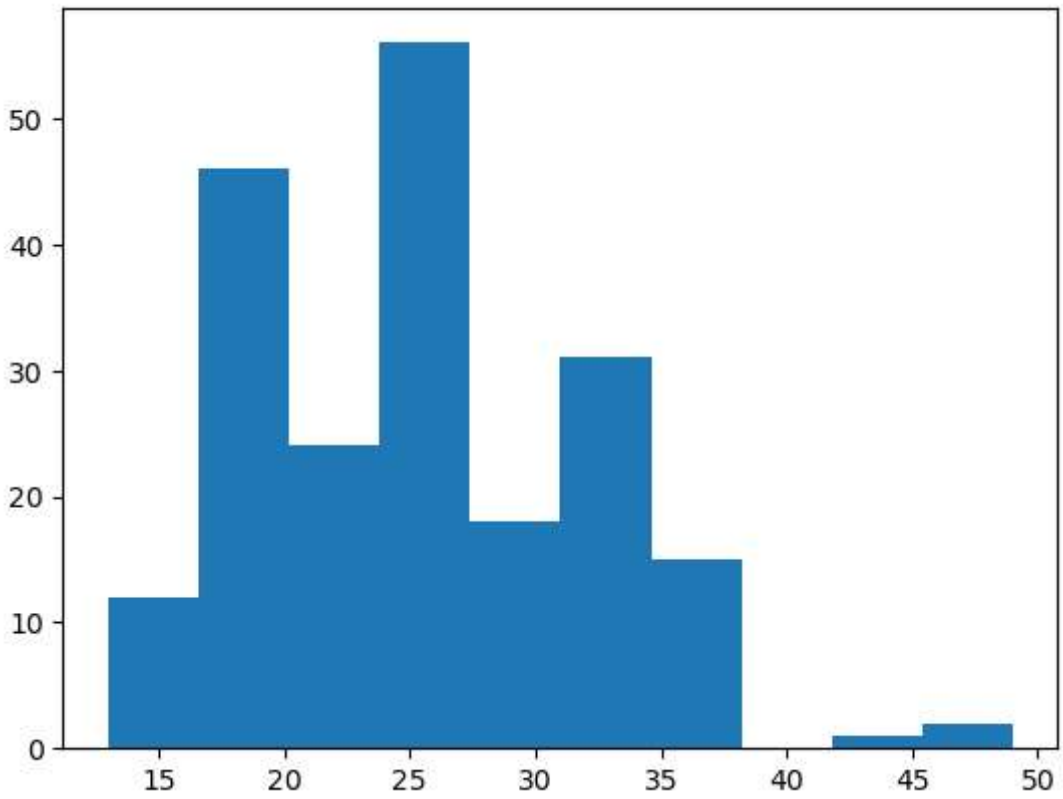
```

plt.hist(dataset['peakrpm'])
```

Out[107]:
```
(array([17., 10., 41., 33., 31., 51.,  8., 12.,  0.,  2.]),
 array([4150., 4395., 4640., 4885., 5130., 5375., 5620., 5865., 6110.,
        6355., 6600.]),
 <BarContainer object of 10 artists>)
```



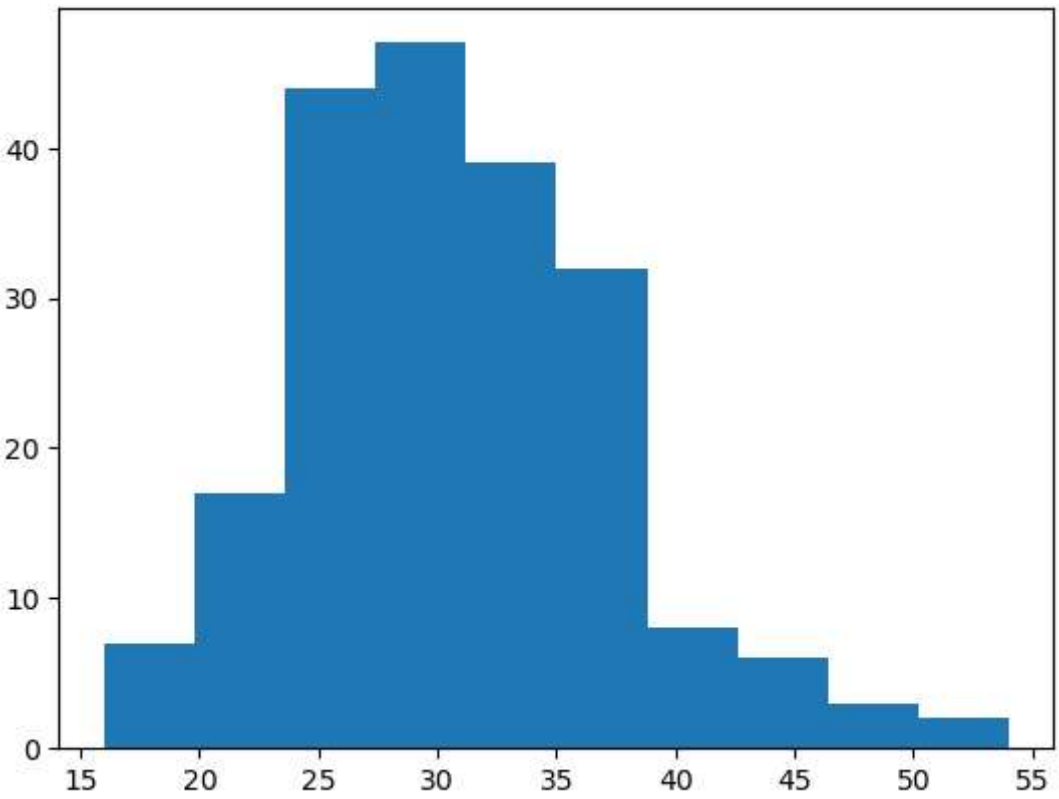
In [108]:
```
#8. Create Histograms for rest of the variables (excluding the ones that h


plt.hist(dataset['citympg'])
```

Out[108]:
```
(array([12., 46., 24., 56., 18., 31., 15.,  0.,  1.,  2.]),
 array([13. , 16.6, 20.2, 23.8, 27.4, 31. , 34.6, 38.2, 41.8, 45.4, 49. ]),
 <BarContainer object of 10 artists>)
```



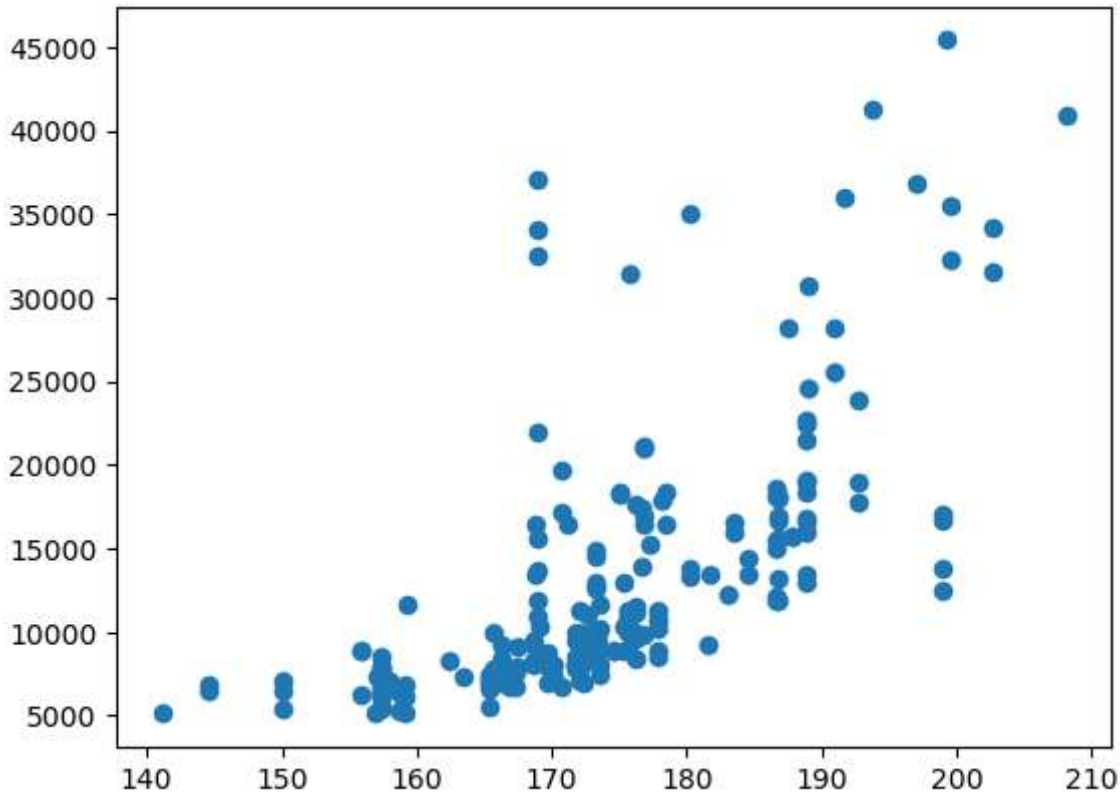
In [109]:
```
#8. Create Histograms for rest of the variables (excluding the ones that 


plt.hist(dataset['highwaympg'])
```

Out[109]:
```
(array([ 7., 17., 44., 47., 39., 32.,  8.,  6.,  3.,  2.]),
 array([16. , 19.8, 23.6, 27.4, 31.2, 35. , 38.8, 42.6, 46.4, 50.2, 54. ]),
 <BarContainer object of 10 artists>)
```



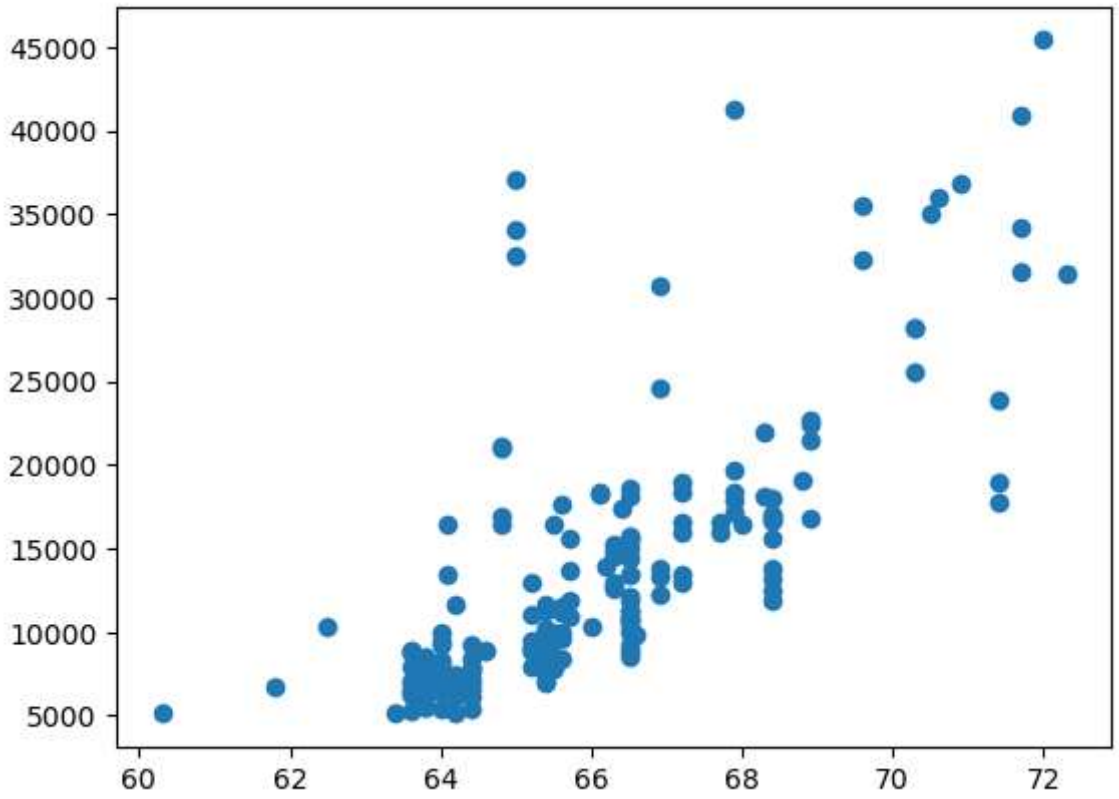
In [110]:
```
#9. Create scatter plot and provide insights for the following:
#a. Car length vs price


plt.scatter(dataset['carlength'], dataset['price'])
```

Out[110]: <matplotlib.collections.PathCollection at 0x209554b61d0>



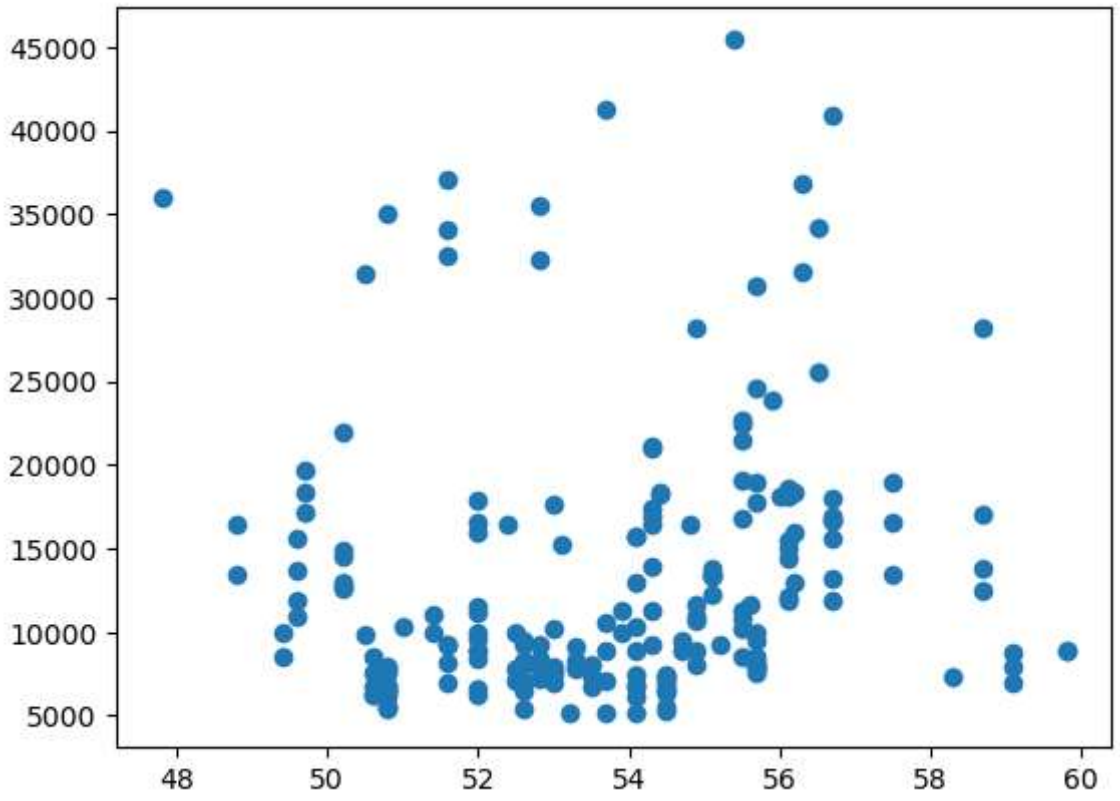
In [111]:
```
#b. Car width vs price


plt.scatter(dataset['carwidth'], dataset['price'])
```

Out[111]: <matplotlib.collections.PathCollection at 0x20953610910>



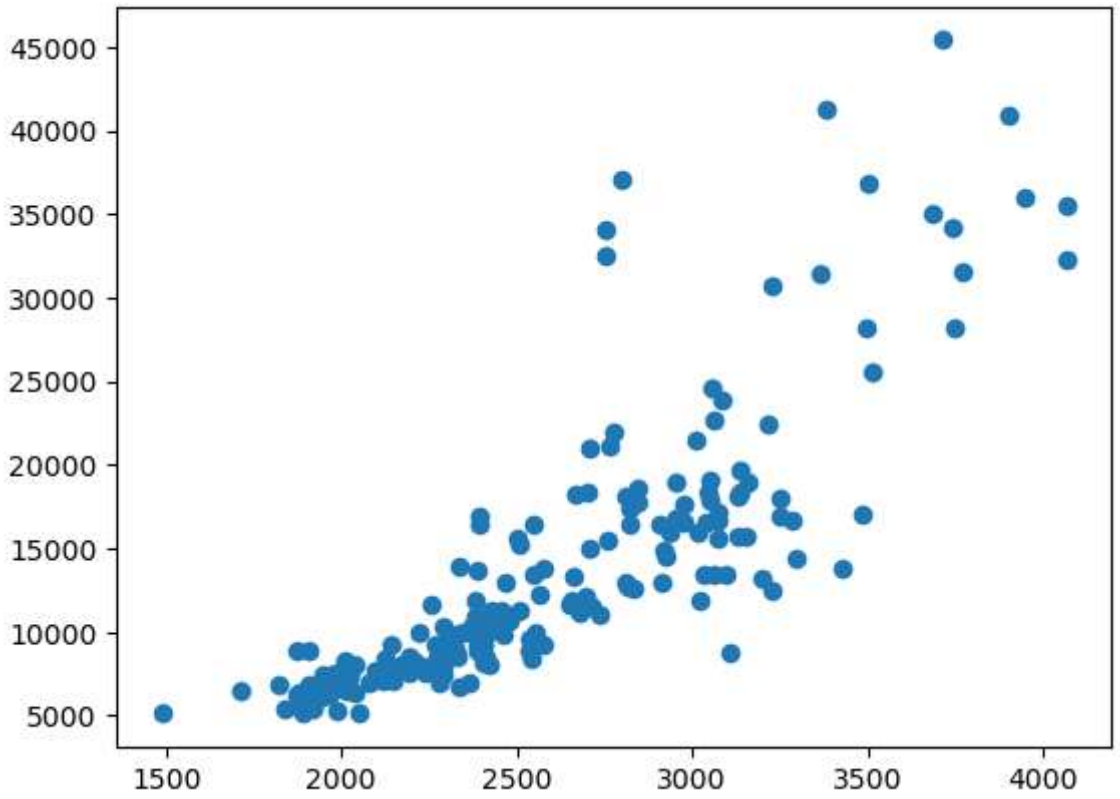
In [112]:
```
#c. Car height vs price


plt.scatter(dataset['carheight'], dataset['price'])
```

Out[112]: <matplotlib.collections.PathCollection at 0x2095397fd50>



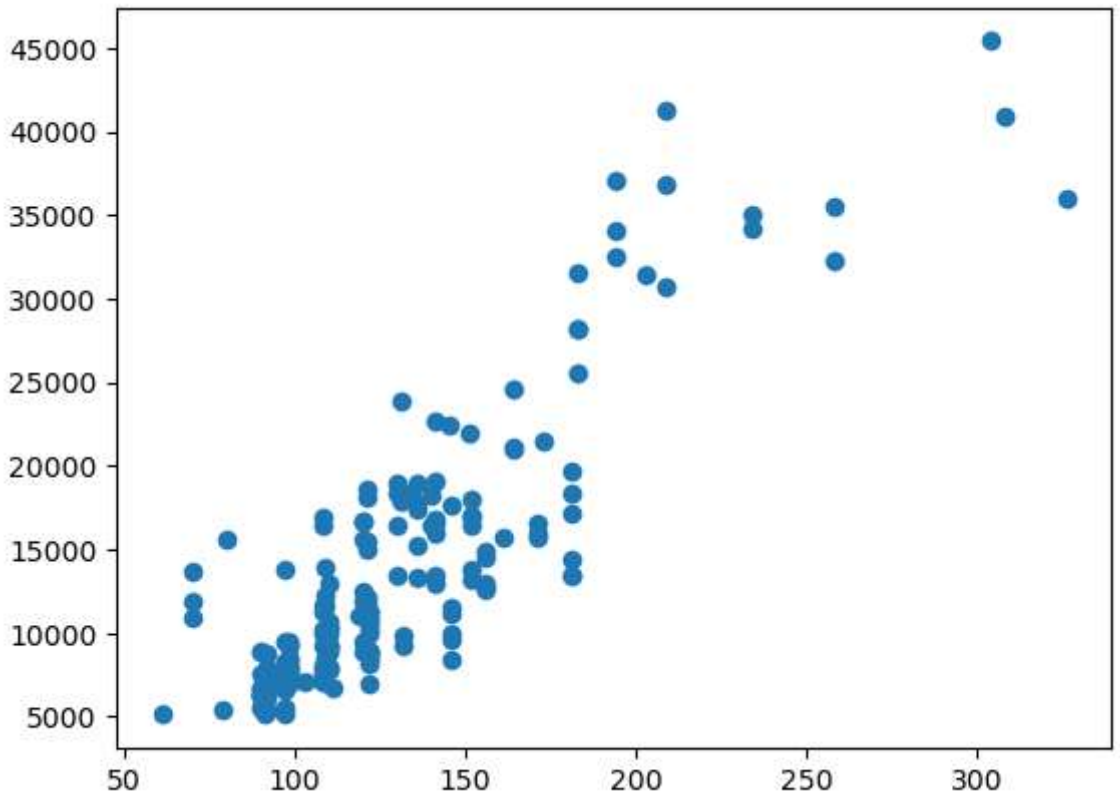
In [113]:
```
#d. Car weight vs price


plt.scatter(dataset['curbweight'], dataset['price'])
```

Out[113]: <matplotlib.collections.PathCollection at 0x20953635850>



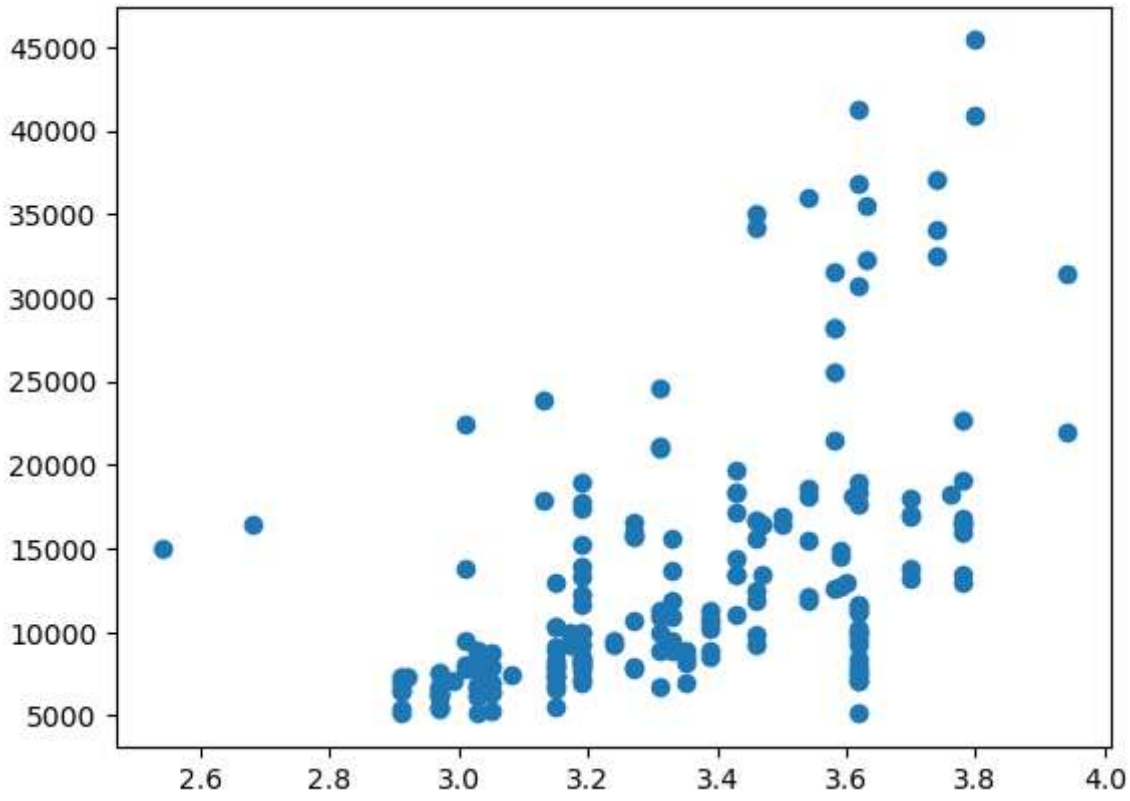
In [114]:
```
#10. Create scatter plot for the following variables w.r.t. price as prov

#✓ Engine size


plt.scatter(dataset['enginesize'], dataset['price'])
```

Out[114]: <matplotlib.collections.PathCollection at 0x20953631e10>



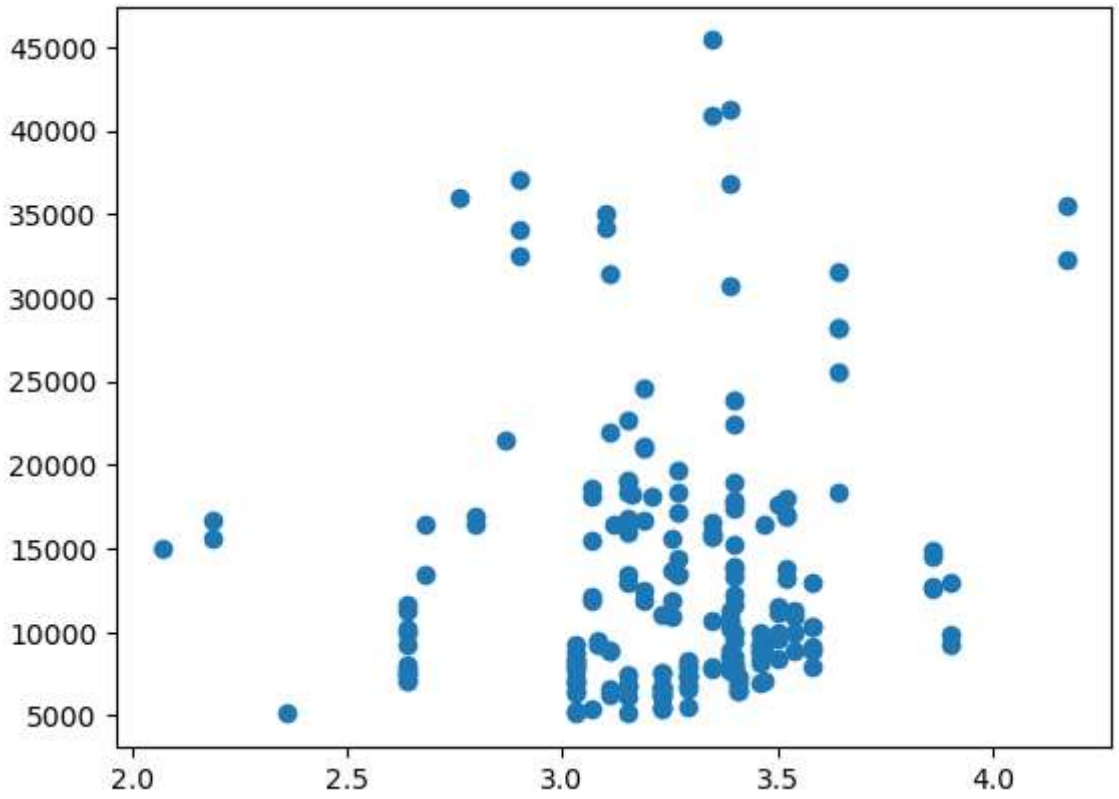
In [115]:
```
#✓ bore ratio


plt.scatter(dataset['boreratio'], dataset['price'])
```

Out[115]: <matplotlib.collections.PathCollection at 0x20953847d10>



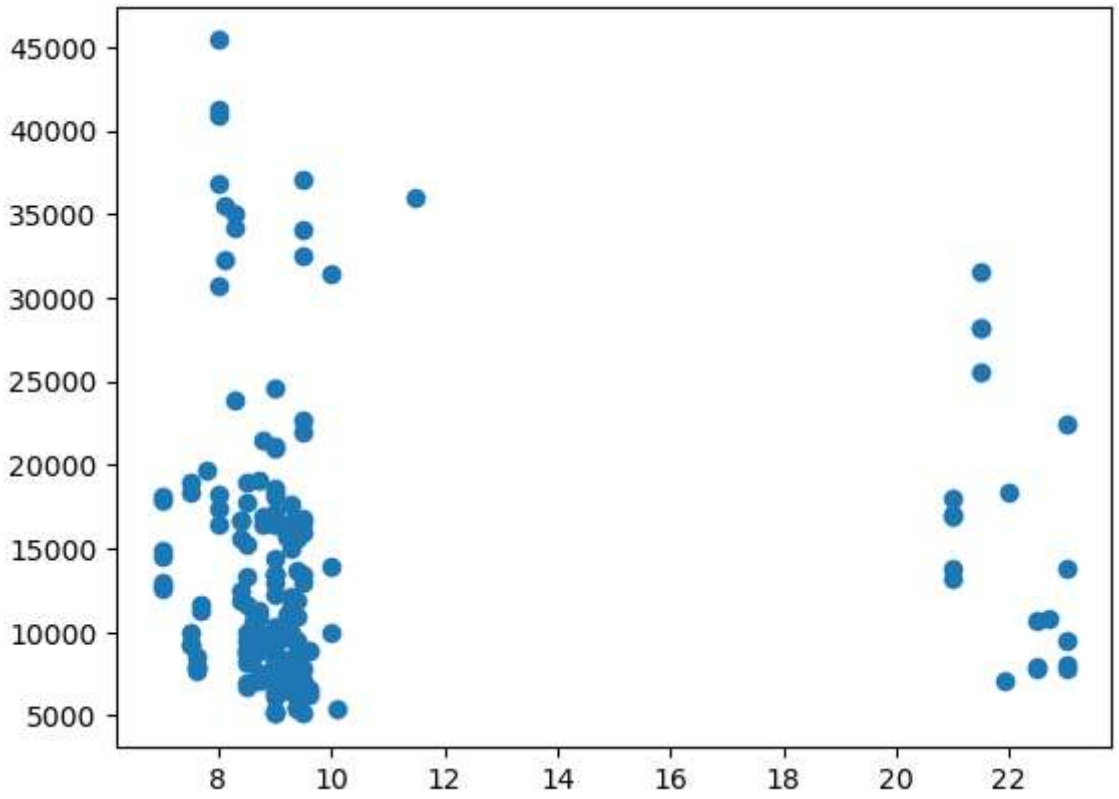
In [116]:
```
#✓ stroke


plt.scatter(dataset['stroke'], dataset['price'])
```

Out[116]: <matplotlib.collections.PathCollection at 0x2095398f750>



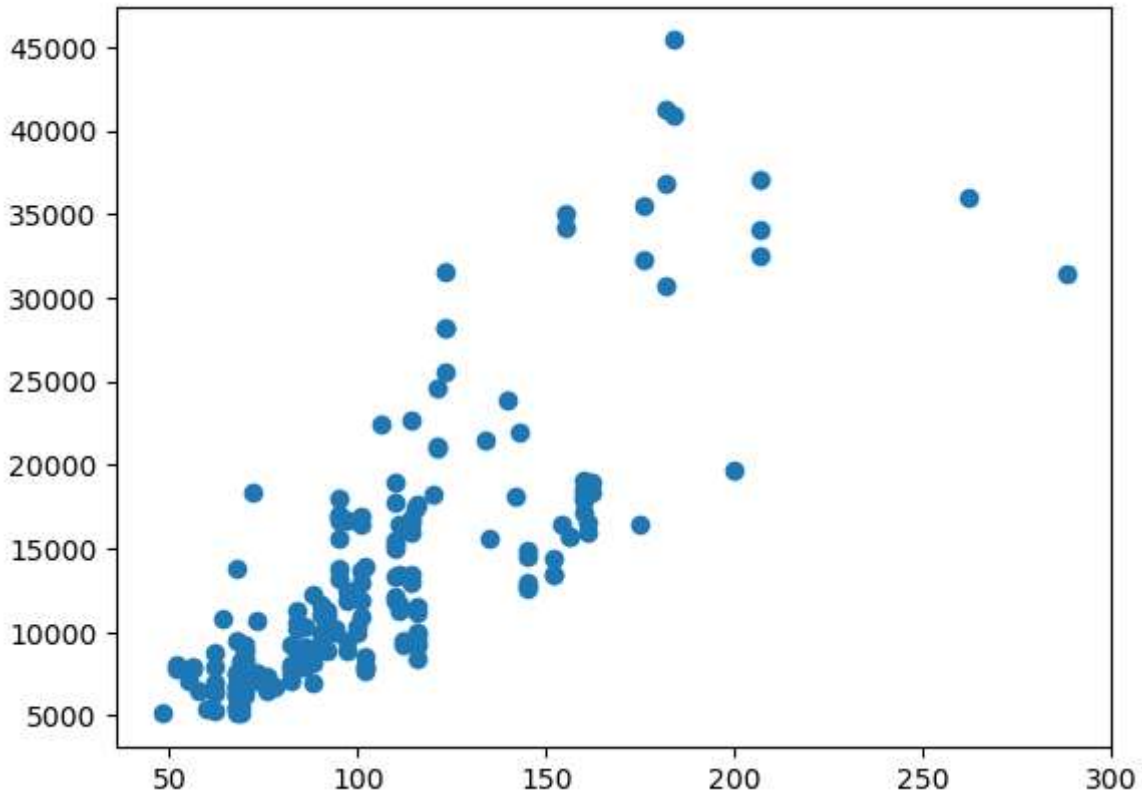
In [117]:
```
#✓ compression ratio


plt.scatter(dataset['compressionratio'], dataset['price'])
```

Out[117]: <matplotlib.collections.PathCollection at 0x2095398e0d0>



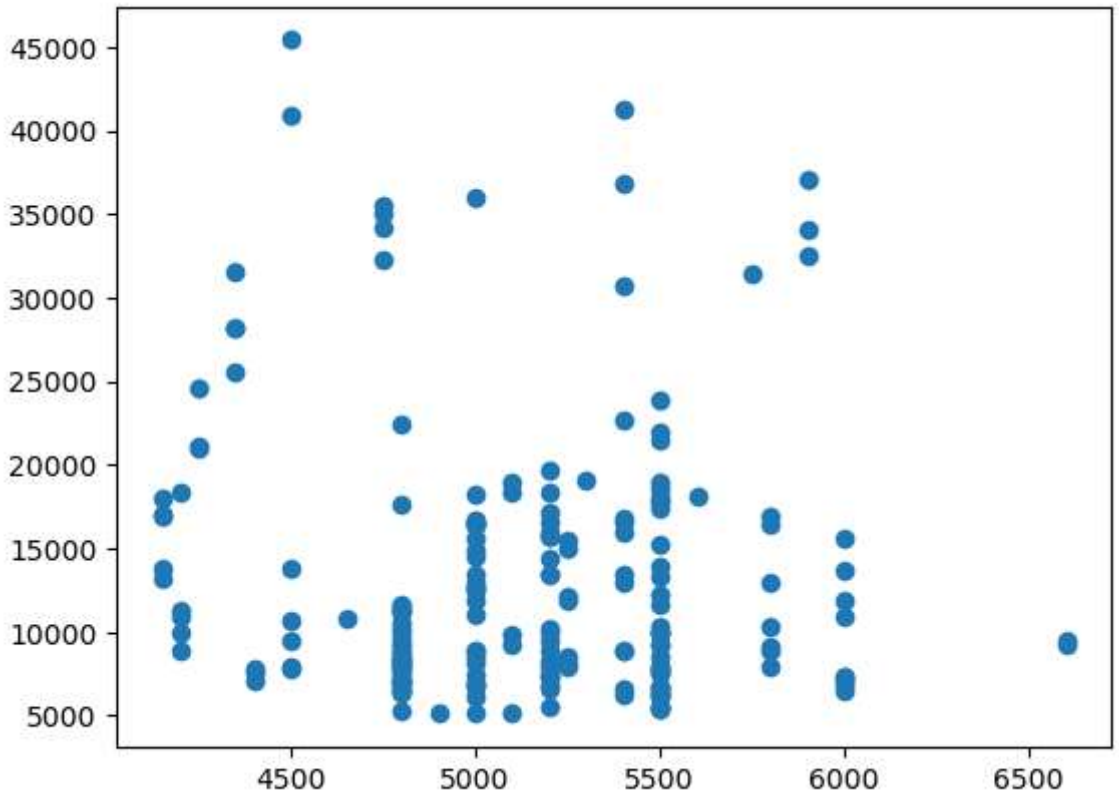
In [118]:
```
#✓ horse power


plt.scatter(dataset['horsepower'], dataset['price'])
```

Out[118]: <matplotlib.collections.PathCollection at 0x209538f7510>



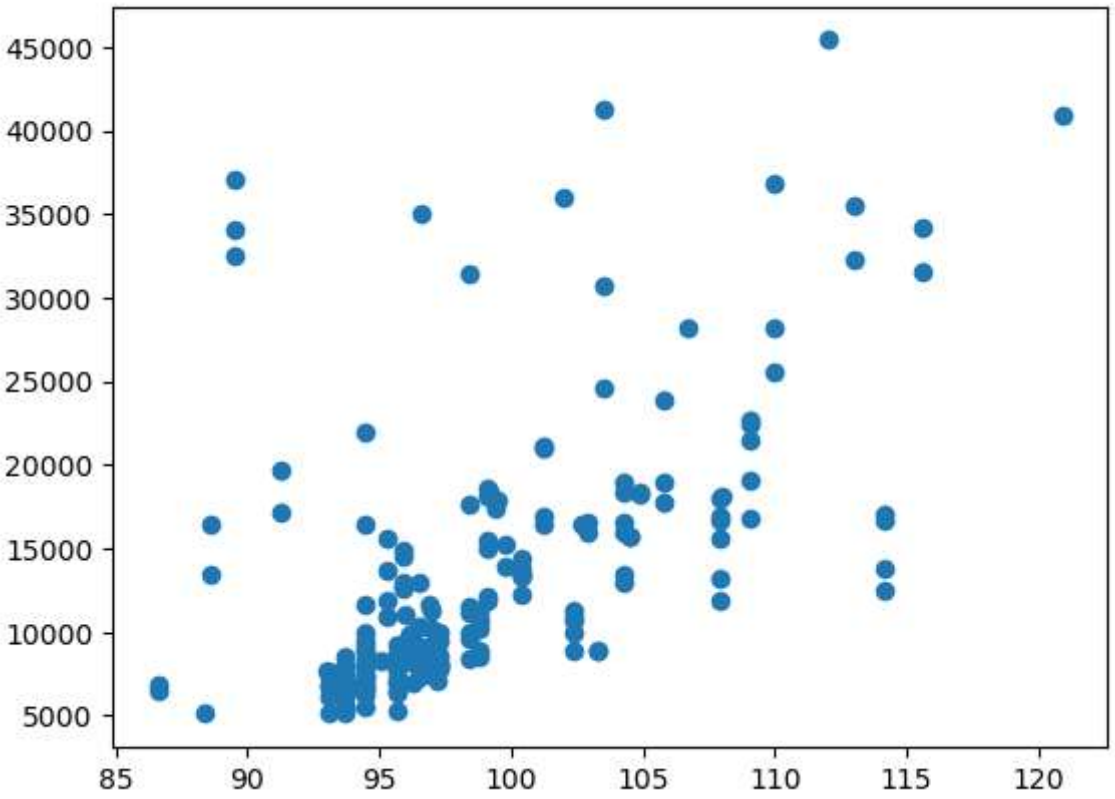
In [119]:
```
#✓ peak rpm


plt.scatter(dataset['peakrpm'], dataset['price'])
```

Out[119]: <matplotlib.collections.PathCollection at 0x2095372e1d0>



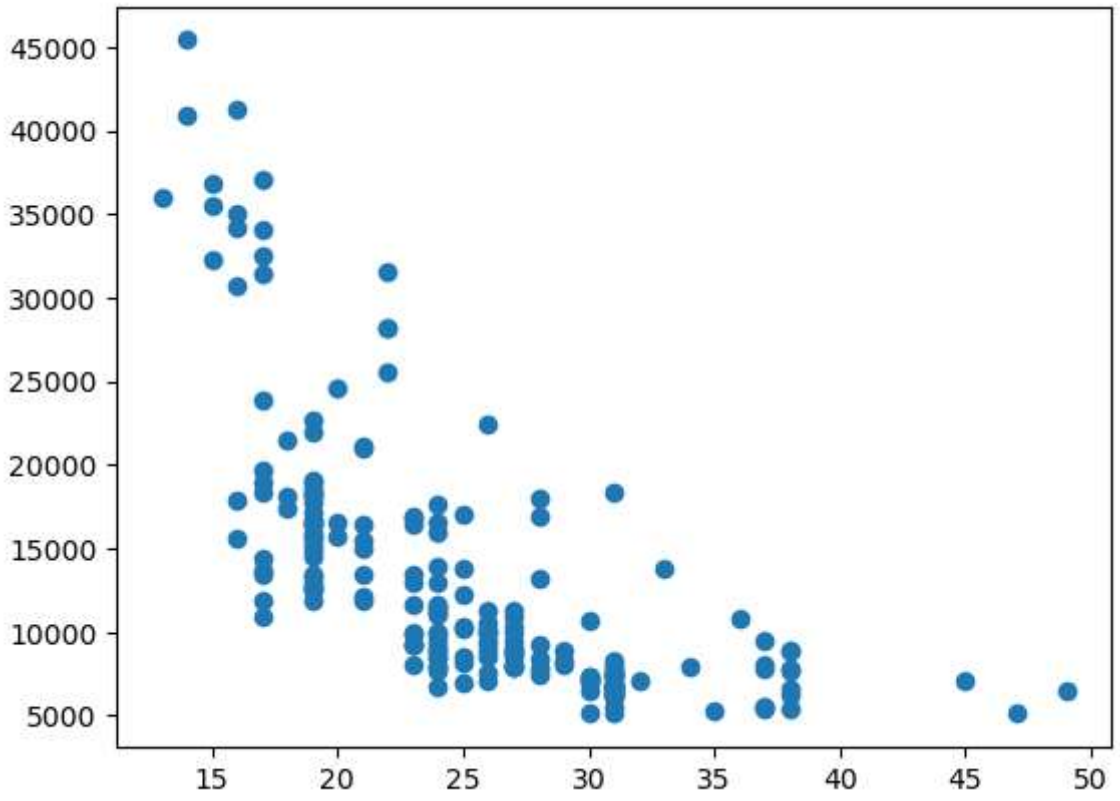
In [120]:
```
#✓ wheel base


plt.scatter(dataset['wheelbase'], dataset['price'])
```

Out[120]: <matplotlib.collections.PathCollection at 0x20955595f50>



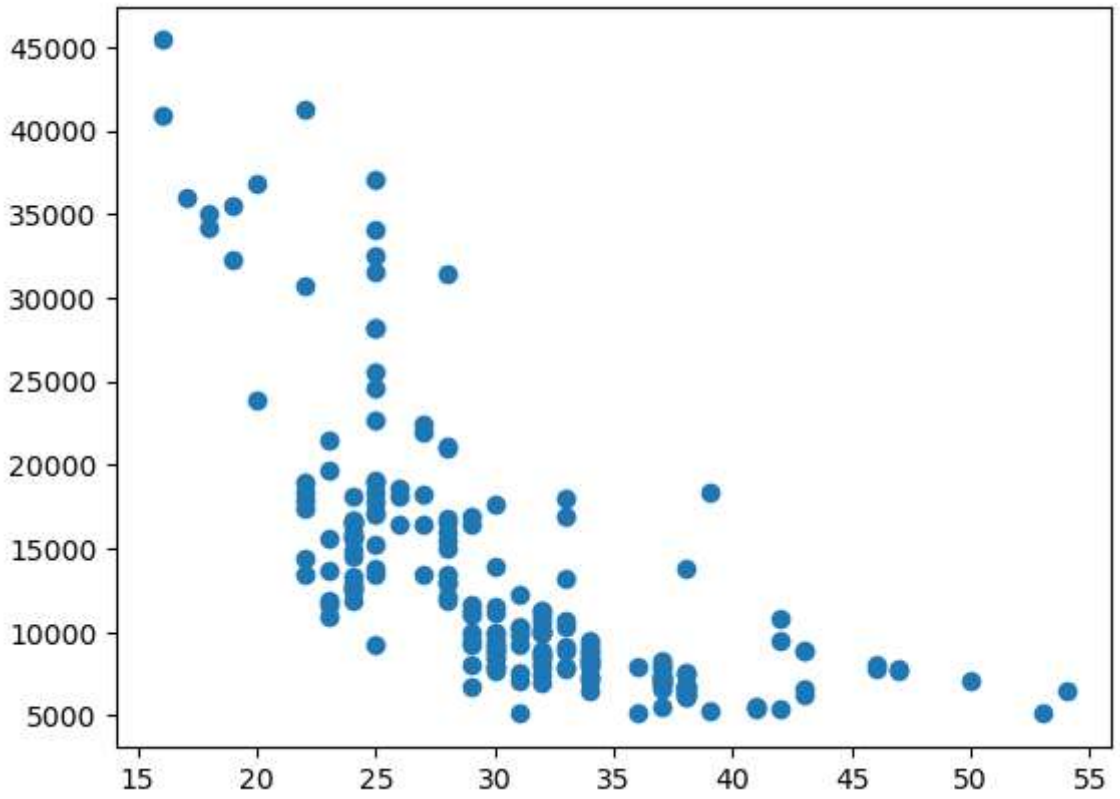
In [121]:
```
#✓ city mpg


plt.scatter(dataset['citympg'], dataset['price'])
```

Out[121]: <matplotlib.collections.PathCollection at 0x209555ae1d0>



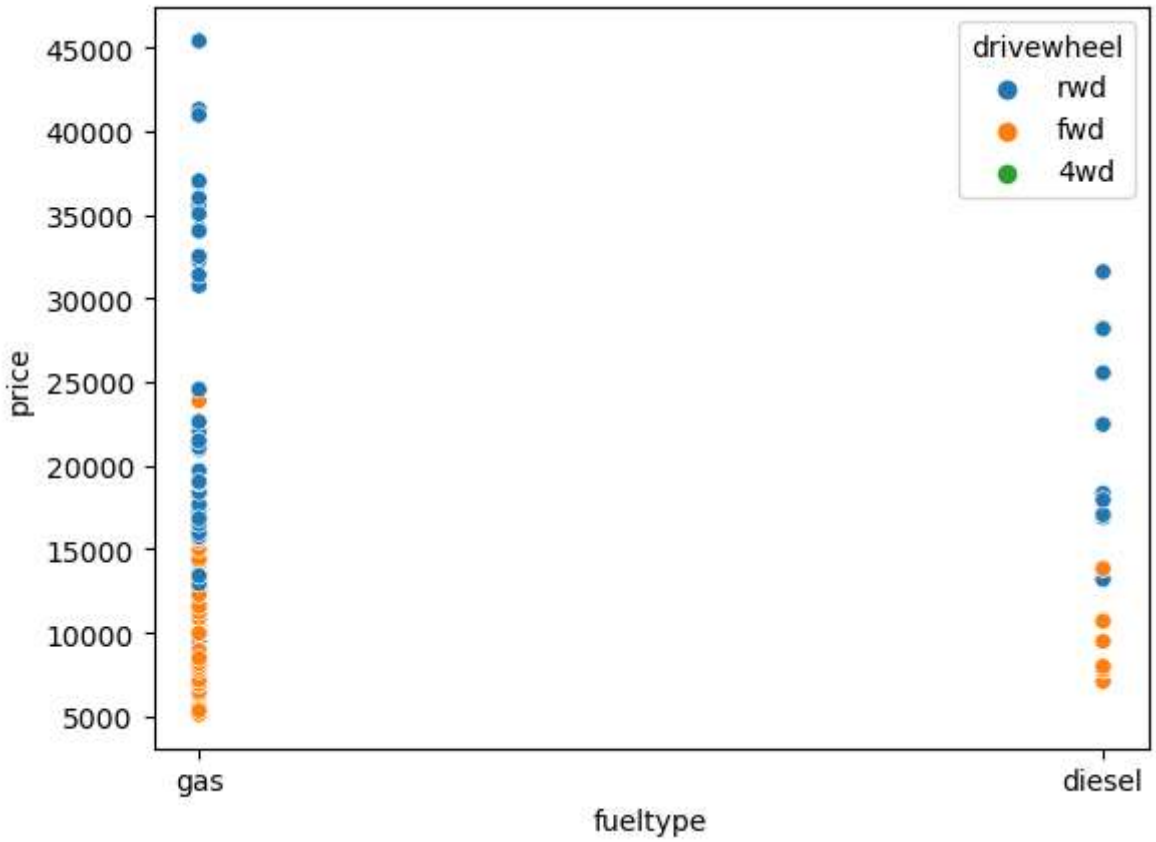
In [122]:
```
#✓ highway mpg


plt.scatter(dataset['highwaympg'], dataset['price'])
```

Out[122]: <matplotlib.collections.PathCollection at 0x20951f79cd0>



In [123]:
```
1 sns.scatterplot(x=dataset['fueltype'],y=dataset['price'],hue=dataset['dri
```

Out[123]: <Axes: xlabel='fueltype', ylabel='price'>



In [ ]:
```
1 #12. After above exercise for visualization, can you list down important v
2
3
4 #Density.
```