In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```python
df=pd.read_csv("cars.csv")
```

In [3]:
```python
df.head()
```

Out[3]:

	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine-type
0	3	?	alfa-romero	gas	convertible	rwd	front	64.1	48.8	dohc
1	3	?	alfa-romero	gas	convertible	rwd	front	64.1	48.8	dohc
2	1	?	alfa-romero	gas	hatchback	rwd	front	65.5	52.4	ohcv
3	2	164	audi	gas	sedan	fwd	front	66.2	54.3	ohc
4	2	164	audi	gas	sedan	4wd	front	66.4	54.3	ohc

```
1  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 15 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   symboling          205 non-null    int64
 1   normalized-losses  205 non-null    object
 2   make               205 non-null    object
 3   fuel-type          205 non-null    object
 4   body-style         205 non-null    object
 5   drive-wheels       205 non-null    object
 6   engine-location    205 non-null    object
 7   width              205 non-null    float64
 8   height             205 non-null    float64
 9   engine-type        205 non-null    object
 10  engine-size        205 non-null    int64
 11  horsepower         205 non-null    object
 12  city-mpg           205 non-null    int64
 13  highway-mpg        205 non-null    int64
 14  price              205 non-null    int64
dtypes: float64(2), int64(5), object(8)
memory usage: 24.1+ KB
```

In [5]:

```
1  # total no of records :205, column 15
2  # normalized-losses and horsepower contains numeric data but they are in object .
```

```
1  df['normalized-losses'].value_counts()
```

```
?       41
161     11
91       8
150      7
134      6
128      6
104      6
85       5
94       5
65       5
102      5
74       5
168      5
103      5
95       5
106      4
93       4
118      4
148      4
122      4
83       3
125      3
154      3
115      3
137      3
101      3
119      2
87       2
89       2
192      2
197      2
158      2
81       2
188      2
194      2
153      2
129      2
108      2
110      2
164      2
145      2
113      2
256      1
107      1
90       1
231      1
142      1
121      1
78       1
98       1
186      1
77       1
Name: normalized-losses, dtype: int64
```

```
1  df['horsepower'].value_counts()
```

```
68     19
70     11
69     10
116     9
110     8
95      7
88      6
62      6
101     6
160     6
114     6
84      5
97      5
102     5
145     5
82      5
76      5
111     4
92      4
123     4
86      4
90      3
73      3
85      3
207     3
182     3
121     3
152     3
112     2
56      2
161     2
156     2
94      2
52      2
?       2
162     2
155     2
184     2
100     2
176     2
55      1
262     1
134     1
115     1
140     1
48      1
58      1
60      1
78      1
135     1
200     1
64      1
120     1
72      1
```

```
154    1
288    1
143    1
142    1
175    1
106    1
Name: horsepower, dtype: int64
```

```python
1  df['normalized-losses'].replace("?",np.nan,inplace=True)
```

```python
1  df.isnull().sum()
```

```
symboling             0
normalized-losses    41
make                  0
fuel-type             0
body-style            0
drive-wheels          0
engine-location       0
width                 0
height                0
engine-type           0
engine-size           0
horsepower            0
city-mpg              0
highway-mpg           0
price                 0
dtype: int64
```

In [10]:

```python
1  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 15 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   symboling          205 non-null    int64
 1   normalized-losses  164 non-null    object
 2   make               205 non-null    object
 3   fuel-type          205 non-null    object
 4   body-style         205 non-null    object
 5   drive-wheels       205 non-null    object
 6   engine-location    205 non-null    object
 7   width              205 non-null    float64
 8   height             205 non-null    float64
 9   engine-type        205 non-null    object
 10  engine-size        205 non-null    int64
 11  horsepower         205 non-null    object
 12  city-mpg           205 non-null    int64
 13  highway-mpg        205 non-null    int64
 14  price              205 non-null    int64
dtypes: float64(2), int64(5), object(8)
memory usage: 24.1+ KB
```

In [11]:

```python
1  df['normalized-losses']=df['normalized-losses'].astype('float64')
```

In [12]:

```python
1  n_mean=df['normalized-losses'].mean()
```

In [13]:

```python
1  n_mean
```

Out[13]:

```
122.0
```

In [14]:

```python
1  df['normalized-losses'].fillna(n_mean,inplace=True)
```

```
1  df.isnull().sum()
```

```
symboling           0
normalized-losses   0
make                0
fuel-type           0
body-style          0
drive-wheels        0
engine-location     0
width               0
height              0
engine-type         0
engine-size         0
horsepower          0
city-mpg            0
highway-mpg         0
price               0
dtype: int64
```

```
1  df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 15 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   symboling          205 non-null    int64
 1   normalized-losses  205 non-null    float64
 2   make               205 non-null    object
 3   fuel-type          205 non-null    object
 4   body-style         205 non-null    object
 5   drive-wheels       205 non-null    object
 6   engine-location    205 non-null    object
 7   width              205 non-null    float64
 8   height             205 non-null    float64
 9   engine-type        205 non-null    object
 10  engine-size        205 non-null    int64
 11  horsepower         205 non-null    object
 12  city-mpg           205 non-null    int64
 13  highway-mpg        205 non-null    int64
 14  price              205 non-null    int64
dtypes: float64(3), int64(5), object(7)
memory usage: 24.1+ KB
```

```
1  df.head(10)
```

	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine-type
0	3	122.0	alfa-romero	gas	convertible	rwd	front	64.1	48.8	dohc
1	3	122.0	alfa-romero	gas	convertible	rwd	front	64.1	48.8	dohc
2	1	122.0	alfa-romero	gas	hatchback	rwd	front	65.5	52.4	ohcv
3	2	164.0	audi	gas	sedan	fwd	front	66.2	54.3	ohc
4	2	164.0	audi	gas	sedan	4wd	front	66.4	54.3	ohc
5	2	122.0	audi	gas	sedan	fwd	front	66.3	53.1	ohc
6	1	158.0	audi	gas	sedan	fwd	front	71.4	55.7	ohc
7	1	122.0	audi	gas	wagon	fwd	front	71.4	55.7	ohc
8	1	158.0	audi	gas	sedan	fwd	front	71.4	55.9	ohc
9	0	122.0	audi	gas	hatchback	4wd	front	67.9	52.0	ohc

```
1  df['horsepower'].replace("?",np.nan,inplace=True)
```

```
1  df.isnull().sum()
```

```
symboling            0
normalized-losses    0
make                 0
fuel-type            0
body-style           0
drive-wheels         0
engine-location      0
width                0
height               0
engine-type          0
engine-size          0
horsepower           2
city-mpg             0
highway-mpg          0
price                0
dtype: int64
```

In [20]:

```python
1  df.dropna(inplace=True)
```

In [21]:

```python
1  df.isnull().sum()
```

Out[21]:

```
symboling            0
normalized-losses    0
make                 0
fuel-type            0
body-style           0
drive-wheels         0
engine-location      0
width                0
height               0
engine-type          0
engine-size          0
horsepower           0
city-mpg             0
highway-mpg          0
price                0
dtype: int64
```

In [22]:

```python
1  df['horsepower'] = df['horsepower'].astype('float64')
```

In [23]:

```python
1  df.dtypes
```

Out[23]:

```
symboling              int64
normalized-losses    float64
make                  object
fuel-type             object
body-style            object
drive-wheels          object
engine-location       object
width                float64
height               float64
engine-type           object
engine-size            int64
horsepower           float64
city-mpg               int64
highway-mpg            int64
price                  int64
dtype: object
```
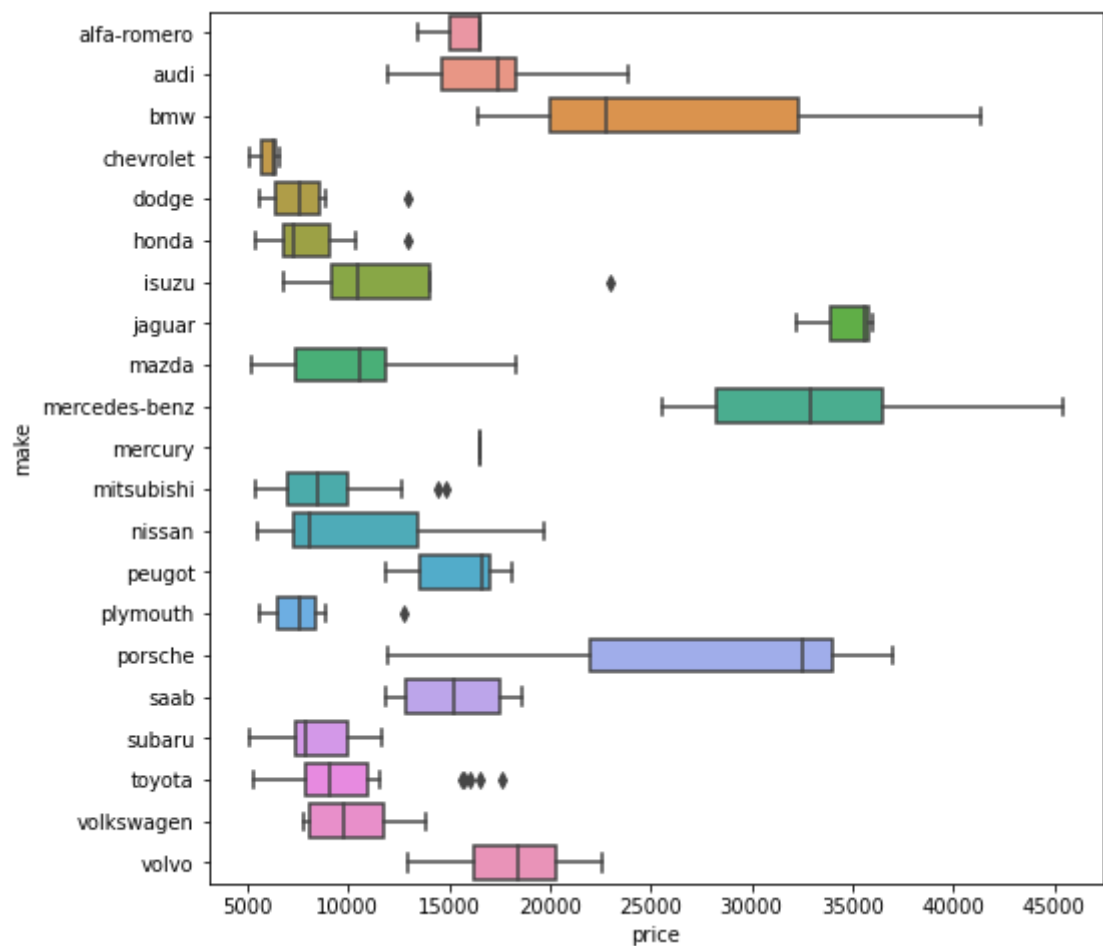
In [24]:

```python
1  #handling outlier
```

```
1  plt.figure(figsize=(8,8))
2  sns.boxplot(data=df,x="price",y="make")
```

Out[25]:

```
<AxesSubplot:xlabel='price', ylabel='make'>
```

```python
1  df.loc[(df['fuel-type']=="gas")]
```

Out[26]:

	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine-type
0	3	122.0	alfa-romero	gas	convertible	rwd	front	64.1	48.8	dohc
1	3	122.0	alfa-romero	gas	convertible	rwd	front	64.1	48.8	dohc
2	1	122.0	alfa-romero	gas	hatchback	rwd	front	65.5	52.4	ohcv
3	2	164.0	audi	gas	sedan	fwd	front	66.2	54.3	ohc
4	2	164.0	audi	gas	sedan	4wd	front	66.4	54.3	ohc
...	...	...	...	...	...	...	...	...	...	...
199	-1	74.0	volvo	gas	wagon	rwd	front	67.2	57.5	ohc
200	-1	95.0	volvo	gas	sedan	rwd	front	68.9	55.5	ohc
201	-1	95.0	volvo	gas	sedan	rwd	front	68.8	55.5	ohc
202	-1	95.0	volvo	gas	sedan	rwd	front	68.9	55.5	ohcv
204	-1	95.0	volvo	gas	sedan	rwd	front	68.9	55.5	ohc

183 rows × 15 columns

In [27]:

```python
1  df.loc[(df['make']=='toyota')& (df['price']>14000)]
```

Out[27]:

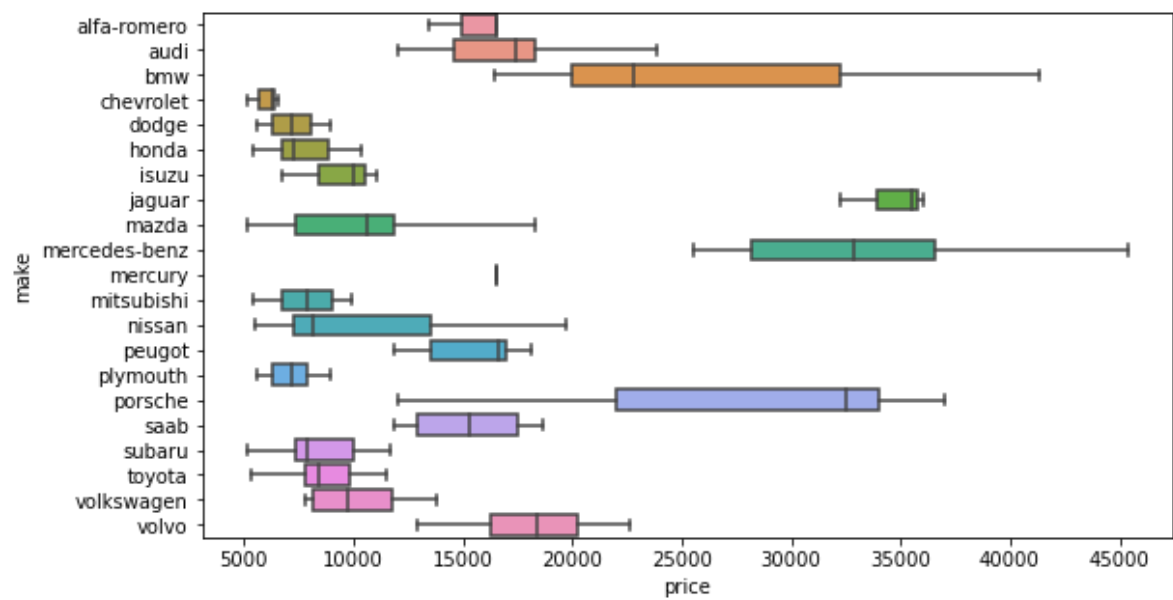|	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine-type
172	2	134.0	toyota	gas	convertible	rwd	front	65.6	53.0	ohc
178	3	197.0	toyota	gas	hatchback	rwd	front	67.7	52.0	dohc
179	3	197.0	toyota	gas	hatchback	rwd	front	67.7	52.0	dohc
180	-1	90.0	toyota	gas	sedan	rwd	front	66.5	54.1	dohc
181	-1	122.0	toyota	gas	wagon	rwd	front	66.5	54.1	dohc

In [28]:

```python
1  df.drop([172,178,179,180,181],axis=0,inplace=True)
```

```
1  plt.figure(figsize=(9,5))
2  sns.boxplot(data=df,x="price",y="make")
```

Out[40]:

```
<AxesSubplot:xlabel='price', ylabel='make'>
```


In [30]:

```
1  df.loc[(df['make']=='mitsubishi') & (df['price']>11000)]
```

Out[30]:

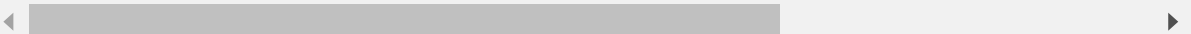|	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine type
82	3	122.0	mitsubishi	gas	hatchback	fwd	front	66.3	50.2	oh
83	3	122.0	mitsubishi	gas	hatchback	fwd	front	66.3	50.2	oh
84	3	122.0	mitsubishi	gas	hatchback	fwd	front	66.3	50.2	oh

In [31]:

```
1  df.drop([82,83,84],axis=0,inplace=True)
```

```
1  df.loc[(df['make']=='plymouth')& (df['price']>10000)]
```

Out[32]:

	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine-type
124	3	122.0	plymouth	gas	hatchback	rwd	front	66.3	50.2	oh

In [33]:

```
1  df.loc[(df['make']=='isuzu')& (df['price']>20000)]
```

Out[33]:

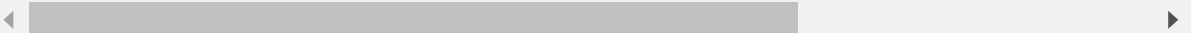|	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine-type	engine-s
45	0	122.0	isuzu	gas	sedan	fwd	front	63.6	52.0	ohc	

In [34]:

```
1  df.loc[(df['make']=='honda')& (df['price']>11000)]
```

Out[34]:

	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine-type	engi s
41	0	85.0	honda	gas	sedan	fwd	front	65.2	54.1	ohc	

In [35]:

```
1  df.loc[(df['make']=='dodge')& (df['price']>10000)]
```

Out[35]:

	symboling	normalized-losses	make	fuel-type	body-style	drive-wheels	engine-location	width	height	engine-type	
29	3	145.0	dodge	gas	hatchback	fwd	front	66.3	50.2	ohc	

In [36]:

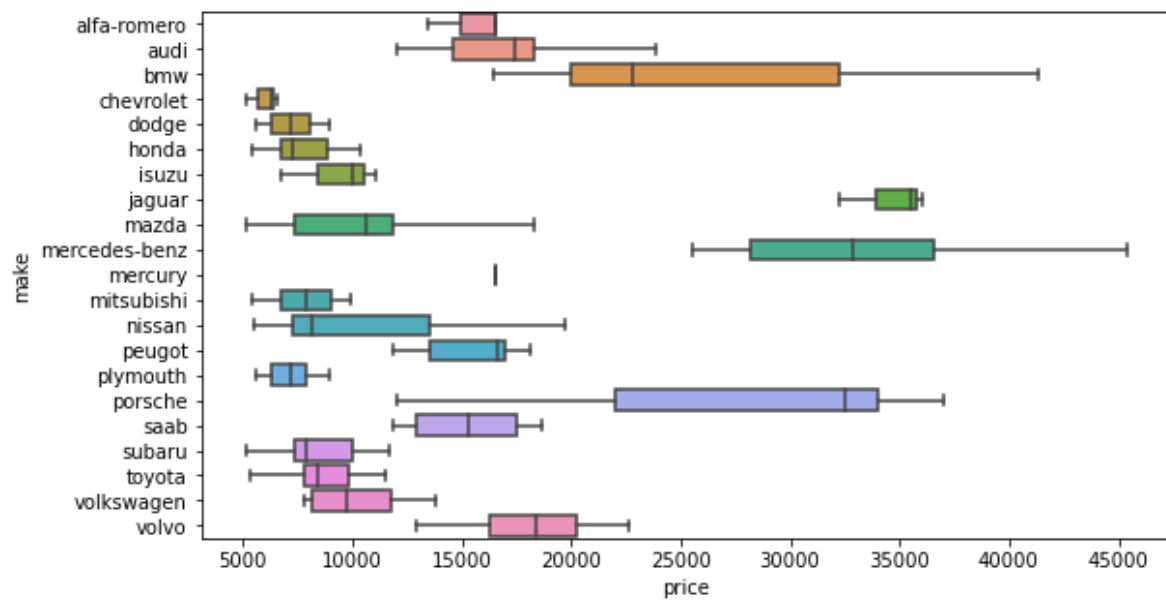```
1 df.drop([124,45,41,29],axis=0,inplace=True)
```

```
1  plt.figure(figsize=(9,5))
2  sns.boxplot(data=df,x="price",y="make")
```

Out[42]:

```
<AxesSubplot:xlabel='price', ylabel='make'>
```


In [ ]:

```
1
```