In [94]:

```
import pandas as pd
import numpy as np
import matplotlib as mp
import seaborn as sns
```

In [95]:

```
df = pd.read_csv('imports-85.data', sep=",")
headers = ["symboling","normalized-losses","make","fuel-type","asp
iration", "num-of-doors","body-style",
         "drive-wheels","engine-location","wheel-base", "length","
width","height","curb-weight","engine-type",
         "num-of-cylinders", "engine-size","fuel-system","bore","s
troke","compression-ratio","horsepower",
         "peak-rpm","city-mpg","highway-mpg","price"]
df.columns = headers
df
```

Out[95]:

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front
1	1	?	alfa-romero	gas	std	two	hatchback	rwd	front
2	2	164	audi	gas	std	four	sedan	fwd	front
3	2	164	audi	gas	std	four	sedan	4wd	front
4	2	?	audi	gas	std	two	sedan	fwd	front
...	...	...	...	...	...	...	...	...	...
199	-1	95	volvo	gas	std	four	sedan	rwd	front
200	-1	95	volvo	gas	turbo	four	sedan	rwd	front
201	-1	95	volvo	gas	std	four	sedan	rwd	front
202	-1	95	volvo	diesel	turbo	four	sedan	rwd	front
203	-1	95	volvo	gas	turbo	four	sedan	rwd	front

204 rows × 26 columns

In [96]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   symboling          204 non-null    int64  
 1   normalized-losses  204 non-null    object 
 2   make               204 non-null    object 
 3   fuel-type          204 non-null    object 
 4   aspiration         204 non-null    object 
 5   num-of-doors       204 non-null    object 
 6   body-style         204 non-null    object 
 7   drive-wheels       204 non-null    object 
 8   engine-location    204 non-null    object 
 9   wheel-base         204 non-null    float64
 10  length             204 non-null    float64
 11  width              204 non-null    float64
 12  height             204 non-null    float64
 13  curb-weight        204 non-null    int64  
 14  engine-type        204 non-null    object 
 15  num-of-cylinders   204 non-null    object 
 16  engine-size        204 non-null    int64  
 17  fuel-system        204 non-null    object 
 18  bore               204 non-null    object 
 19  stroke             204 non-null    object 
 20  compression-ratio  204 non-null    float64
 21  horsepower         204 non-null    object 
 22  peak-rpm           204 non-null    object 
 23  city-mpg           204 non-null    int64  
 24  highway-mpg        204 non-null    int64  
 25  price              204 non-null    object 
dtypes: float64(5), int64(5), object(16)
memory usage: 41.6+ KB
```

In [97]:

```
df= df.replace('?',np.nan)
```

In [98]:

```
# Replacing normalized_losses
df['normalized-losses'] = df['normalized-losses'].astype(float)
mean_of_nl =  df['normalized-losses'].mean()
df['normalized-losses'] = df['normalized-losses'].replace(np.nan,m
ean_of_nl)
```

In [99]:

```
# Replacing num_of_doors
num_of_door = df['num-of-doors'].mode()
df['num-of-doors'] = df['num-of-doors'].replace(np.nan,num_of_door
[0])
```

In [100]:

```
# Replacing bore
df['bore'] = df['bore'].astype(float)
mean_of_bore =  df['bore'].mean()
df['bore'] = df['bore'].replace(np.nan,mean_of_bore)
```

In [101]:

```
# Replacing stroke
df['stroke'] = df['stroke'].astype(float)
mean_of_stroke =  df['stroke'].mean()
df['stroke'] = df['stroke'].replace(np.nan,mean_of_stroke)
```

In [102]:

```
# Replacing horsepower
df['horsepower'] = df['horsepower'].astype(float)
mean_of_horsepower =  df['horsepower'].mean()
df['horsepower'] = df['horsepower'].replace(np.nan,mean_of_horsepo
wer)
```

In [103]:

```
# Replacing peak-rpm
df['peak-rpm'] = df['peak-rpm'].astype(float)
mean_of_peak_rpm =  df['peak-rpm'].mean()
df['peak-rpm'] = df['peak-rpm'].replace(np.nan,mean_of_peak_rpm)
```

In [104]:

```
# Replacing price
df['price'] = df['price'].astype(float)
mean_of_price =  df['price'].mean()
df['price'] = df['price'].replace(np.nan,mean_of_price)
```

In [105]:

```
df.isna().sum()
```

Out[105]:

```
symboling            0
normalized-losses    0
make                 0
fuel-type            0
aspiration           0
num-of-doors         0
body-style           0
drive-wheels         0
engine-location      0
wheel-base           0
length               0
width                0
height               0
curb-weight          0
engine-type          0
num-of-cylinders     0
engine-size          0
fuel-system          0
bore                 0
stroke               0
compression-ratio    0
horsepower           0
peak-rpm             0
city-mpg             0
highway-mpg          0
price                0
dtype: int64
```

In [106]:

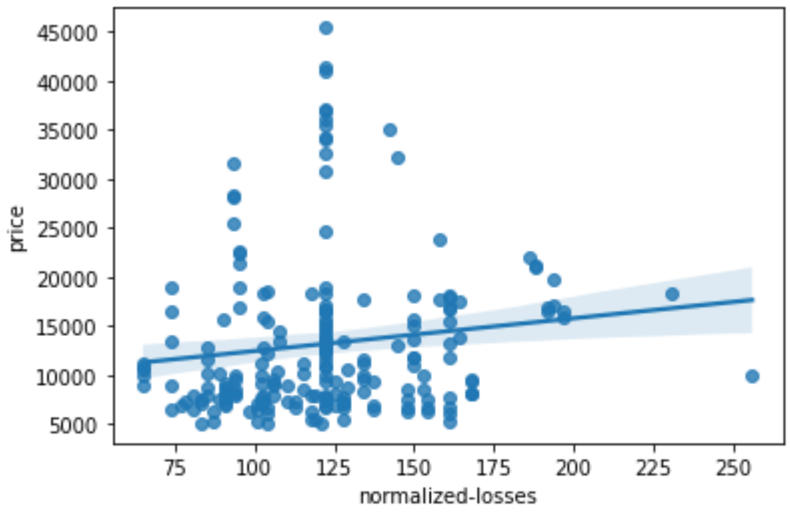
```
df.to_csv('cars_data.csv')
```

```
In [107]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   symboling          204 non-null    int64  
 1   normalized-losses  204 non-null    float64
 2   make               204 non-null    object 
 3   fuel-type          204 non-null    object 
 4   aspiration         204 non-null    object 
 5   num-of-doors       204 non-null    object 
 6   body-style         204 non-null    object 
 7   drive-wheels       204 non-null    object 
 8   engine-location    204 non-null    object 
 9   wheel-base         204 non-null    float64
 10  length             204 non-null    float64
 11  width              204 non-null    float64
 12  height             204 non-null    float64
 13  curb-weight        204 non-null    int64  
 14  engine-type        204 non-null    object 
 15  num-of-cylinders   204 non-null    object 
 16  engine-size        204 non-null    int64  
 17  fuel-system        204 non-null    object 
 18  bore               204 non-null    float64
 19  stroke             204 non-null    float64
 20  compression-ratio  204 non-null    float64
 21  horsepower         204 non-null    float64
 22  peak-rpm           204 non-null    float64
 23  city-mpg           204 non-null    int64  
 24  highway-mpg        204 non-null    int64  
 25  price              204 non-null    float64
dtypes: float64(11), int64(5), object(10)
memory usage: 41.6+ KB
```

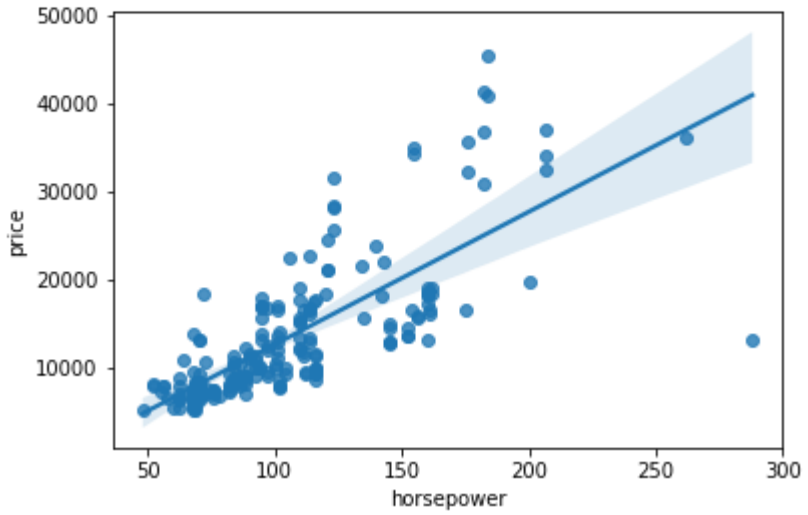
```
In [108]: sns.regplot(x="normalized-losses",y="price", data=df)
```

```
Out[108]: <AxesSubplot:xlabel='normalized-losses', ylabel='price'>
```



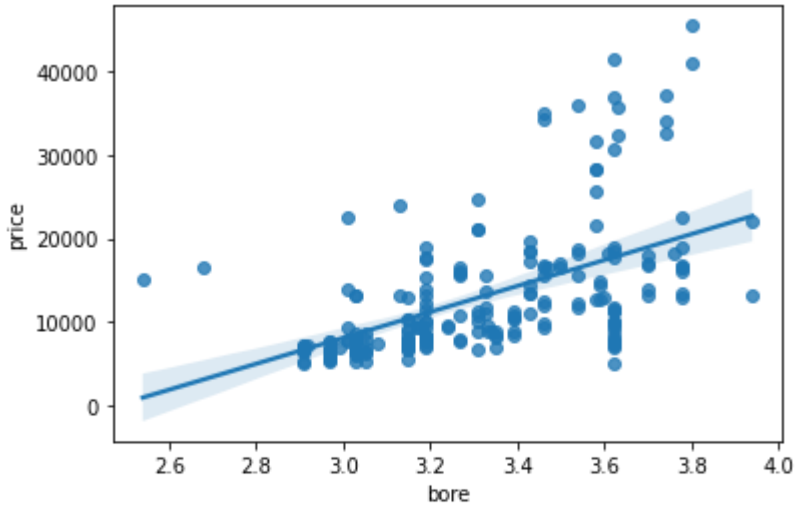
```
In [109]: sns.regplot(x="horsepower",y="price", data=df)
```

```
Out[109]: <AxesSubplot:xlabel='horsepower', ylabel='price'>
```



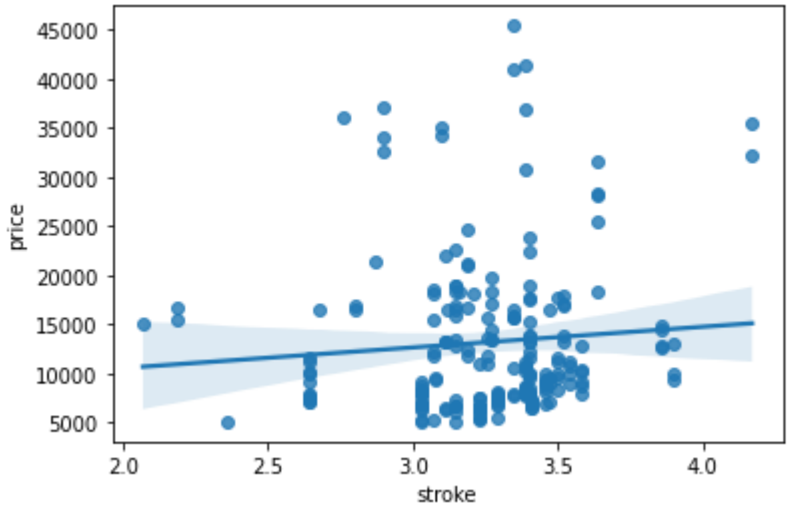
```
In [110]: sns.regplot(x="bore",y="price", data=df)
```

```
Out[110]: <AxesSubplot:xlabel='bore', ylabel='price'>
```



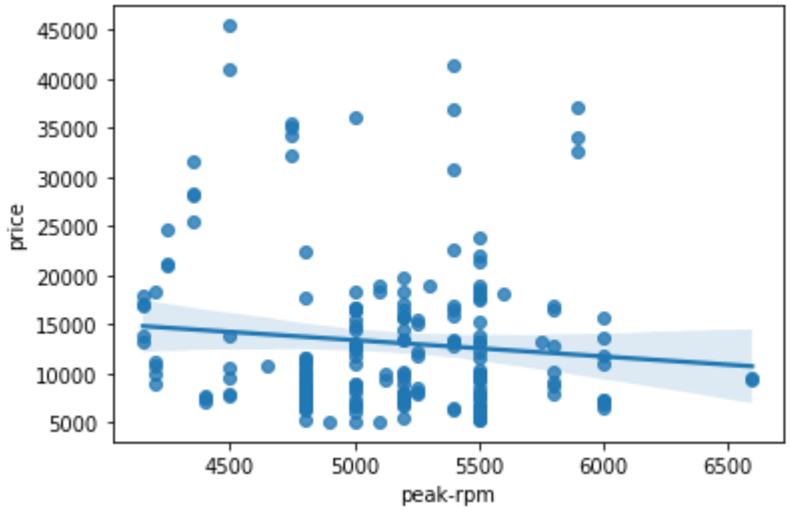
```
In [111]: sns.regplot(x="stroke",y="price", data=df)
```

```
Out[111]: <AxesSubplot:xlabel='stroke', ylabel='price'>
```



```
In [112]: sns.regplot(x="peak-rpm",y="price", data=df)
```

```
Out[112]: <AxesSubplot:xlabel='peak-rpm', ylabel='price'>
```



```
In [ ]: '''
        Conclusion:-

        SATVIK:- Looking at all the graphs, i guess 'bore' & 'horsepower'
        will be the attributes affecting the price.

        '''
```

```
In [ ]: # Practical 1 End
```

```
In [113]: # Practical 2 Start
```

In [114]:

```
#categorical attribute
numerics = ['int16', 'int32', 'int64', 'float16', 'float32', 'floa
t64']
dataframe = df.select_dtypes(exclude=numerics)
dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 10 columns):
 #   Column             Non-Null Count  Dtype 
---  ------             --------------  ----- 
 0   make               204 non-null    object
 1   fuel-type          204 non-null    object
 2   aspiration         204 non-null    object
 3   num-of-doors       204 non-null    object
 4   body-style         204 non-null    object
 5   drive-wheels       204 non-null    object
 6   engine-location    204 non-null    object
 7   engine-type        204 non-null    object
 8   num-of-cylinders   204 non-null    object
 9   fuel-system        204 non-null    object
dtypes: object(10)
memory usage: 16.1+ KB
```

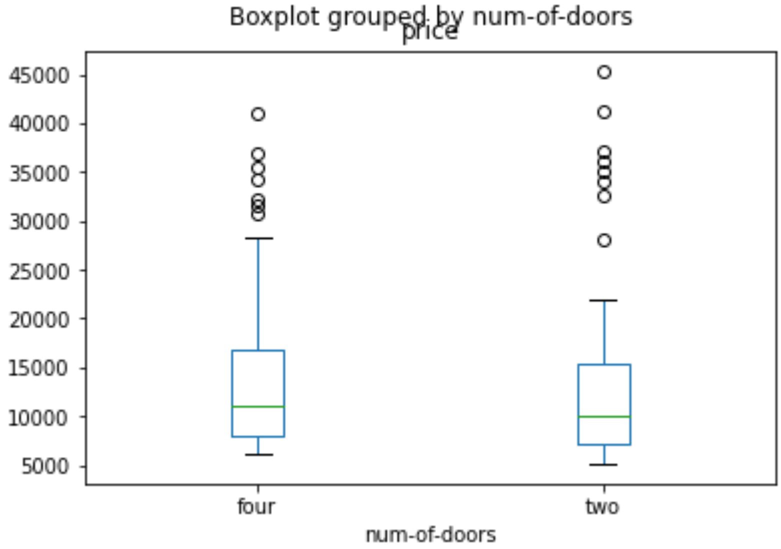
In [115]:

```
#contionous attributes
dataframe = df.select_dtypes(include=numerics)
dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 16 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   symboling          204 non-null    int64  
 1   normalized-losses  204 non-null    float64
 2   wheel-base         204 non-null    float64
 3   length             204 non-null    float64
 4   width              204 non-null    float64
 5   height             204 non-null    float64
 6   curb-weight        204 non-null    int64  
 7   engine-size        204 non-null    int64  
 8   bore               204 non-null    float64
 9   stroke             204 non-null    float64
 10  compression-ratio  204 non-null    float64
 11  horsepower         204 non-null    float64
 12  peak-rpm           204 non-null    float64
 13  city-mpg           204 non-null    int64  
 14  highway-mpg        204 non-null    int64  
 15  price              204 non-null    float64
dtypes: float64(11), int64(5)
memory usage: 25.6 KB
```

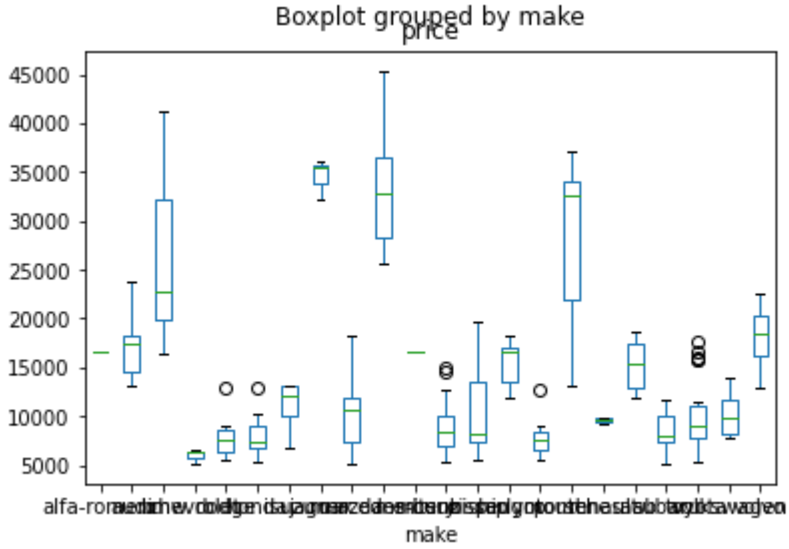
```
In [116]: df.boxplot(by ='num-of-doors', column =['price'], grid = False) #N
          ot Important
          #df.boxplot(by ='day', column =['price'], grid = False)
```

```
Out[116]: <AxesSubplot:title={'center':'price'}, xlabel='num-of-doors'>
```



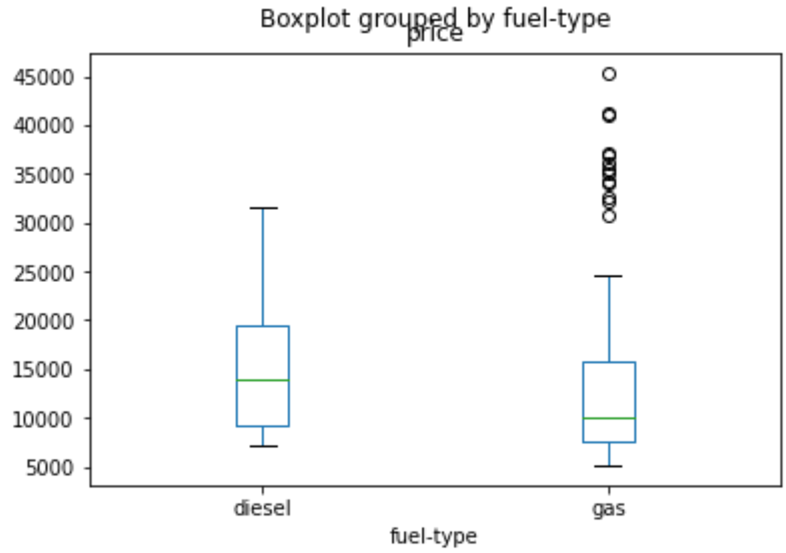
```
In [117]: df.boxplot(by ='make', column =['price'], grid = False) #Not Impor
          tant
```

```
Out[117]: <AxesSubplot:title={'center':'price'}, xlabel='make'>
```



```
In [118]: df.boxplot(by ='fuel-type', column =['price'], grid = False)#Not Important

Out[118]: <AxesSubplot:title={'center':'price'}, xlabel='fuel-type'>
```


Boxplot grouped by fuel-type
price
diesel
gas
fuel-type