In [31]:

```
import numpy as np
import pandas as pd
from sklearn import preprocessing
from sklearn import linear_model
```

In [4]:

```
df = pd.read_csv("CarPrice_Assignment.csv")
df
```

Out[4]:

	car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel
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 [7]:

```
df.info
```

Out[7]:

```
<bound method DataFrame.info of      car_ID  symboling                   C
arName fueltype aspiration  \
0         1          3        alfa-romero giulia      gas        std
1         2          3       alfa-romero stelvio      gas        std
2         3          1  alfa-romero Quadrifoglio      gas        std
3         4          2               audi 100 ls      gas        std
4         5          2                audi 100ls      gas        std
..      ...        ...                       ...      ...        ...
200     201         -1           volvo 145e (sw)      gas        std
201     202         -1               volvo 144ea      gas      turbo
202     203         -1               volvo 244dl      gas        std
203     204         -1                 volvo 246   diesel      turbo
204     205         -1               volvo 264gl      gas      turbo

    doornumber      carbody drivewheel enginelocation  wheelbase  ...  \
0          two  convertible        rwd          front       88.6  ...
1          two  convertible        rwd          front       88.6  ...
2          two    hatchback        rwd          front       94.5  ...
3         four        sedan        fwd          front       99.8  ...
4         four        sedan        4wd          front       99.4  ...
..         ...          ...        ...            ...        ...  ...
200       four        sedan        rwd          front      109.1  ...
201       four        sedan        rwd          front      109.1  ...
202       four        sedan        rwd          front      109.1  ...
203       four        sedan        rwd          front      109.1  ...
204       four        sedan        rwd          front      109.1  ...

     enginesize  fuelsystem  boreratio  stroke compressionratio horsepower
\
0           130        mpfi       3.47    2.68              9.0        111
1           130        mpfi       3.47    2.68              9.0        111
2           152        mpfi       2.68    3.47              9.0        154
3           109        mpfi       3.19    3.40             10.0        102
4           136        mpfi       3.19    3.40              8.0        115
..          ...         ...        ...     ...              ...        ...
200         141        mpfi       3.78    3.15              9.5        114
201         141        mpfi       3.78    3.15              8.7        160
202         173        mpfi       3.58    2.87              8.8        134
203         145         idi       3.01    3.40             23.0        106
204         141        mpfi       3.78    3.15              9.5        114

     peakrpm citympg  highwaympg    price
0       5000      21          27  13495.0
1       5000      21          27  16500.0
2       5000      19          26  16500.0
3       5500      24          30  13950.0
4       5500      18          22  17450.0
..       ...     ...         ...      ...
200     5400      23          28  16845.0
201     5300      19          25  19045.0
202     5500      18          23  21485.0
203     4800      26          27  22470.0
204     5400      19          25  22625.0

[205 rows x 26 columns]>
```

In [9]:

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

Out[9]:

```
car_ID              0
symboling           0
CarName             0
fueltype            0
aspiration          0
doornumber          0
carbody             0
drivewheel          0
enginelocation      0
wheelbase           0
carlength           0
carwidth            0
carheight           0
curbweight          0
enginetype          0
cylindernumber      0
enginesize          0
fuelsystem          0
boreratio           0
stroke              0
compressionratio    0
horsepower          0
peakrpm             0
citympg             0
highwaympg          0
price               0
dtype: int64
```

In [ ]:

```

```

Data Preprocessing

In [13]:

```
lab_en = preprocessing.LabelEncoder()
lab_en.fit(df.fueltype)
df.fueltype = lab_en.transform(df.fueltype)
```

In [14]:

```
lab_en.fit(df.doornumber)
df.doornumber = lab_en.transform(df.doornumber)
```

In [15]:

```
lab_en.fit(df.aspiration)
df.aspiration = lab_en.transform(df.aspiration)
```

In [16]:

```
lab_en.fit(df.drivewheel)
df.drivewheel = lab_en.transform(df.drivewheel)
```

In [17]:

```
lab_en.fit(df.carbody)
df.carbody = lab_en.transform(df.carbody)
```

In [18]:

```
lab_en.fit(df.enginelocation)
df.enginelocation = lab_en.transform(df.enginelocation)
```

In [19]:

```
lab_en.fit(df.enginetype)
df.enginetype = lab_en.transform(df.enginetype)
```

In [20]:

```
lab_en.fit(df.cylindernumber)
df.cylindernumber = lab_en.transform(df.cylindernumber)
```

In [21]:

```
lab_en.fit(df.fuelsystem)
df.fuelsystem = lab_en.transform(df.fuelsystem)
```

In [39]:

```
df
```

Out[39]:

e	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	...	engine
ɔ a	1	0	1	0	2	0	88.6	...	
ɔ ɔ	1	0	1	0	2	0	88.6	...	
ɔ ɔ	1	0	1	2	2	0	94.5	...	
s	1	0	0	3	1	0	99.8	...	
s	1	0	0	3	0	0	99.4	...	
.	...	...	...	...	...	...	...	...	
e)	1	0	0	3	2	0	109.1	...	
a	1	1	0	3	2	0	109.1	...	
ll	1	0	0	3	2	0	109.1	...	
ɔ	0	1	0	3	2	0	109.1	...	
ll	1	1	0	3	2	0	109.1	...	

In [34]:

```
lreg = linear_model.LinearRegression()

indipendent_var = df[[ 'fueltype', 'aspiration',
       'doornumber', 'carbody', 'drivewheel', 'enginelocation', 'wheelbase',
       'carlength', 'carwidth', 'carheight', 'curbweight', 'enginetype',
       'cylindernumber', 'enginesize', 'fuelsystem', 'boreratio', 'stroke',
       'compressionratio', 'horsepower', 'peakrpm', 'citympg', 'highwaympg']]
dependent_var = df.price

lreg.fit(indipendent_var,dependent_var )
```

Out[34]:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=F
alse)
```

In [35]:

```
lreg.coef_
```

Out[35]:

```
array([-9.21677014e+03,  2.42352286e+02, -5.13930613e+02, -8.19139857e+02,
        1.16616708e+03,  1.03974427e+04,  1.10593867e+02, -2.91102195e+01,
        7.12602036e+02,  1.84232716e+02,  2.35662037e+00,  1.90948576e+02,
        1.28166071e+02,  1.01622195e+02, -1.48294905e+02, -2.74450799e+03,
       -2.82460445e+03, -5.48092812e+02,  2.06295041e+01,  2.10440991e+00,
       -1.45461967e+02,  1.35183217e+02])
```

In [37]:

```
lreg.intercept_
```

Out[37]:

```
-49313.25002503426
```

In [44]:

```
lreg.predict([[1,0,1,0,2,0,88.6,168.8,64.1,48.8,2548,0,0,130,0,3.47,2.68,9,111,5000,21,27
```

Out[44]:

```
array([13438.04815291])
```

In [38]:

```
df['price'][0]
```

Out[38]:

```
13495.0
```

In [ ]:

```
predicted price is similar to the actual value
```