

Car price prediction

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

```
In [2]: data=pd.read_csv("D:/csvfiles/CarPrice.csv")
```

```
In [3]: data.head()
```

```
Out[3]:
```

| | car_ID | symboling | CarName | fueltype | aspiration | doornumber | carbody | drivewheel | enginelocation | wheelbase | ... | engine | enginesize | fuelsystem | bor |
|---|--------|-----------|-----------------------------|----------|------------|------------|-------------|------------|----------------|-----------|-----|--------|------------|------------|-----|
| 0 | 1 | 3 | alfa-romero giulia | gas | std | two | convertible | rwd | front | 88.6 | ... | 130 | mpfi | | |
| 1 | 2 | 3 | alfa-romero stelvio | gas | std | two | convertible | rwd | front | 88.6 | ... | 130 | mpfi | | |
| 2 | 3 | 1 | alfa-romero Quadrifoglio | gas | std | two | hatchback | rwd | front | 94.5 | ... | 152 | mpfi | | |
| 3 | 4 | 2 | audi 100 ls | gas | std | four | sedan | fwd | front | 99.8 | ... | 109 | mpfi | | |
| 4 | 5 | 2 | audi 100ls | gas | std | four | sedan | 4wd | front | 99.4 | ... | 136 | mpfi | | |

5 rows × 26 columns

```
In [4]: data['enginelocation'].unique()
```

```
Out[4]: array(['front', 'rear'], dtype=object)
```

```
In [5]: data['fuelsystem'].unique()
```

```
Out[5]: array(['mpfi', '2bbl', 'mfi', '1bbl', 'spfi', '4bbl', 'idi', 'spdi'],
dtype=object)
```

```
In [6]: data['drivewheel'].unique()
```

```
Out[6]: array(['rwd', 'fwd', '4wd'], dtype=object)
```

```
In [7]: data['drivewheel']=data['drivewheel'].replace("4wd","fwd")
```

```
In [8]: data['drivewheel'].unique()
```

```
Out[8]: array(['rwd', 'fwd'], dtype=object)
```

```
In [9]: data['fueltype'].unique()
```

```
Out[9]: array(['gas', 'diesel'], dtype=object)
```

```
In [10]: data['aspiration'].unique()
```

```
Out[10]: array(['std', 'turbo'], dtype=object)
```

```
In [11]: data['carbody'].unique()
```

```
Out[11]: array(['convertible', 'hatchback', 'sedan', 'wagon', 'hardtop'],  
              dtype=object)
```

```
In [12]: data['doornumber'].unique()
```

```
Out[12]: array(['two', 'four'], dtype=object)
```

```
In [13]: data.corr()
```

Out[13]:

| | car_ID | symboling | wheelbase | carlength | carwidth | carheight | curbweight | enginesize | boreratio | stroke | compressionratio | horsepower | peakrpm | citympg | highwaympg | price |
|------------------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|-----------|-----------|------------------|------------|-----------|-----------|------------|-----------|
| car_ID | 1.000000 | -0.151621 | 0.129729 | 0.170636 | 0.052387 | 0.255960 | 0.071962 | -0.033930 | 0.260064 | -0.160824 | 0.150276 | -0.015006 | -0.203789 | 0.015940 | 0.011255 | -0.109093 |
| symboling | -0.151621 | 1.000000 | -0.531954 | -0.357612 | -0.232919 | -0.541038 | -0.227691 | -0.105790 | -0.130051 | -0.008735 | -0.178515 | 0.070873 | 0.273606 | -0.035823 | 0.034606 | -0.079978 |
| wheelbase | 0.129729 | -0.531954 | 1.000000 | 0.874587 | 0.795144 | 0.589435 | 0.776386 | 0.569329 | 0.488750 | 0.160959 | 0.249786 | 0.353294 | -0.360469 | -0.470414 | -0.544082 | 0.577816 |
| carlength | 0.170636 | -0.357612 | 0.874587 | 1.000000 | 0.841118 | 0.491029 | 0.877728 | 0.683360 | 0.606454 | 0.129533 | 0.158414 | 0.552623 | -0.287242 | -0.670909 | -0.704662 | 0.682920 |
| carwidth | 0.052387 | -0.232919 | 0.795144 | 0.841118 | 1.000000 | 0.279210 | 0.867032 | 0.735433 | 0.559150 | 0.182942 | 0.181129 | 0.640732 | -0.220012 | -0.642704 | -0.677218 | 0.759325 |
| carheight | 0.255960 | -0.541038 | 0.589435 | 0.491029 | 0.279210 | 1.000000 | 0.295572 | 0.067149 | 0.171071 | -0.055307 | 0.261214 | -0.108802 | -0.320411 | -0.048640 | -0.107358 | 0.119336 |
| curbweight | 0.071962 | -0.227691 | 0.776386 | 0.877728 | 0.867032 | 0.295572 | 1.000000 | 0.850594 | 0.648480 | 0.168790 | 0.151362 | 0.750739 | -0.266243 | -0.757414 | -0.797465 | 0.835305 |
| enginesize | -0.033930 | -0.105790 | 0.569329 | 0.683360 | 0.735433 | 0.067149 | 0.850594 | 1.000000 | 0.583774 | 0.203129 | 0.028971 | 0.809769 | -0.244660 | -0.653658 | -0.677470 | 0.874145 |
| boreratio | 0.260064 | -0.130051 | 0.488750 | 0.606454 | 0.559150 | 0.171071 | 0.648480 | 0.583774 | 1.000000 | -0.055909 | 0.005197 | 0.573677 | -0.254976 | -0.584532 | -0.587012 | 0.553173 |
| stroke | -0.160824 | -0.008735 | 0.160959 | 0.129533 | 0.182942 | -0.055307 | 0.168790 | 0.203129 | -0.055909 | 1.000000 | 0.186110 | 0.080940 | -0.067964 | -0.042145 | -0.043931 | 0.079443 |
| compressionratio | 0.150276 | -0.178515 | 0.249786 | 0.158414 | 0.181129 | 0.261214 | 0.151362 | 0.028971 | 0.005197 | 0.186110 | 1.000000 | -0.204326 | -0.435741 | 0.324701 | 0.265201 | 0.067984 |
| horsepower | -0.015006 | 0.070873 | 0.353294 | 0.552623 | 0.640732 | -0.108802 | 0.750739 | 0.809769 | 0.573677 | 0.080940 | -0.204326 | 1.000000 | -0.435741 | 0.324701 | 0.265201 | 0.067984 |
| peakrpm | -0.203789 | 0.273606 | -0.360469 | -0.287242 | -0.220012 | -0.320411 | -0.266243 | -0.244660 | -0.254976 | -0.067964 | -0.435741 | -0.435741 | 1.000000 | -0.435741 | -0.435741 | -0.435741 |
| citympg | 0.015940 | -0.035823 | -0.470414 | -0.670909 | -0.642704 | -0.048640 | -0.757414 | -0.653658 | -0.584532 | -0.042145 | 0.324701 | 0.324701 | -0.435741 | 1.000000 | 0.324701 | 0.324701 |
| highwaympg | 0.011255 | 0.034606 | -0.544082 | -0.704662 | -0.677218 | -0.107358 | -0.797465 | -0.677470 | -0.587012 | -0.043931 | 0.265201 | 0.265201 | -0.435741 | 0.324701 | 1.000000 | 0.265201 |
| price | -0.109093 | -0.079978 | 0.577816 | 0.682920 | 0.759325 | 0.119336 | 0.835305 | 0.874145 | 0.553173 | 0.079443 | 0.067984 | 0.067984 | -0.435741 | 0.324701 | 0.265201 | 1.000000 |

```
In [14]: data['car_ID']
```

Out[14]:

| | |
|-----|-----|
| 0 | 1 |
| 1 | 2 |
| 2 | 3 |
| 3 | 4 |
| 4 | 5 |
| ... | |
| 200 | 201 |
| 201 | 202 |
| 202 | 203 |
| 203 | 204 |
| 204 | 205 |

Name: car_ID, Length: 205, dtype: int64

```
In [15]: data=data.iloc[:,1:]
```

```
In [16]: data
```

Out[16]:

| | symboling | CarName | fueltype | aspiration | doornumber | carbody | drivewheel | enginelocation | wheelbase | carlength | ... | enginesize | fuelsystem |
|-----|-----------|--------------------------|----------|------------|------------|-------------|------------|----------------|-----------|-----------|-----|------------|------------|
| 0 | 3 | alfa-romero giulia | gas | std | two | convertible | rwd | front | 88.6 | 168.8 | ... | 130 | mpfi |
| 1 | 3 | alfa-romero stelvio | gas | std | two | convertible | rwd | front | 88.6 | 168.8 | ... | 130 | mpfi |
| 2 | 1 | alfa-romero Quadrifoglio | gas | std | two | hatchback | rwd | front | 94.5 | 171.2 | ... | 152 | mpfi |
| 3 | 2 | audi 100 ls | gas | std | four | sedan | fwd | front | 99.8 | 176.6 | ... | 109 | mpfi |
| 4 | 2 | audi 100ls | gas | std | four | sedan | fwd | front | 99.4 | 176.6 | ... | 136 | mpfi |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 200 | -1 | volvo 145e (sw) | gas | std | four | sedan | rwd | front | 109.1 | 188.8 | ... | 141 | mpfi |
| 201 | -1 | volvo 144ea | gas | turbo | four | sedan | rwd | front | 109.1 | 188.8 | ... | 141 | mpfi |
| 202 | -1 | volvo 244dl | gas | std | four | sedan | rwd | front | 109.1 | 188.8 | ... | 173 | mpfi |
| 203 | -1 | volvo 246 | diesel | turbo | four | sedan | rwd | front | 109.1 | 188.8 | ... | 145 | idi |
| 204 | -1 | volvo 264gl | gas | turbo | four | sedan | rwd | front | 109.1 | 188.8 | ... | 141 | mpfi |

205 rows × 25 columns

```
In [17]: correlation=data.corr()
```

```
In [18]: correlation['price']
```

```
Out[18]: symboling      -0.079978  
         wheelbase     0.577816  
         carlength     0.682920  
         carwidth      0.759325  
         carheight     0.119336  
         curbweight    0.835305  
         enginesize     0.874145  
         boreratio     0.553173  
         stroke        0.079443  
         compressionratio 0.067984  
         horsepower    0.808139  
         peakrpm      -0.085267  
         citympg       -0.685751  
         highwaympg    -0.697599  
         price         1.000000  
         Name: price, dtype: float64
```

```
In [19]: ##by seeing the correlation between dependent and independent variables we can conclude that symboling,peakrpm,compressionratio,stroke,car height doesnot contribute more to price so we can remove by to increase accuracy
```

```
In [20]: data=data.drop(['symboling','peakrpm','compressionratio','carheight','stroke'],axis=1)
```

```
In [21]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 20 columns):
 #   Column              Non-Null Count  Dtype  
---  -
 0   CarName             205 non-null   object 
 1   fueltype            205 non-null   object 
 2   aspiration          205 non-null   object 
 3   doornumber          205 non-null   object 
 4   carbody             205 non-null   object 
 5   drivewheel          205 non-null   object 
 6   enginelocation      205 non-null   object 
 7   wheelbase           205 non-null   float64 
 8   carlength           205 non-null   float64 
 9   carwidth            205 non-null   float64 
10  curbweight          205 non-null   int64   
11  enginetype          205 non-null   object 
12  cylindernumber      205 non-null   object 
13  enginesize          205 non-null   int64   
14  fuelsystem          205 non-null   object 
15  boreratio           205 non-null   float64 
16  horsepower          205 non-null   int64   
17  citympg             205 non-null   int64   
18  highwaympg          205 non-null   int64   
19  price               205 non-null   float64 
dtypes: float64(5), int64(5), object(10)
memory usage: 32.2+ KB
```

```
In [22]: correlation=data.corr()
```

```
In [23]: correlation['price']
```

```
Out[23]: wheelbase    0.577816
carlength    0.682920
carwidth     0.759325
curbweight   0.835305
enginesize   0.874145
boreratio    0.553173
horsepower   0.808139
citympg      -0.685751
highwaympg   -0.697599
price        1.000000
Name: price, dtype: float64
```

```
In [24]: data['wheelbase'].unique()
```

```
Out[24]: array([ 88.6,  94.5,  99.8,  99.4, 105.8,  99.5, 101.2, 103.5, 110. ,
                88.4,  93.7, 103.3,  95.9,  86.6,  96.5,  94.3,  96. , 113. ,
                102. ,  93.1,  95.3,  98.8, 104.9, 106.7, 115.6,  96.6, 120.9,
                112. , 102.7,  93. ,  96.3,  95.1,  97.2, 100.4,  91.3,  99.2,
                107.9, 114.2, 108. ,  89.5,  98.4,  96.1,  99.1,  93.3,  97. ,
                96.9,  95.7, 102.4, 102.9, 104.5,  97.3, 104.3, 109.1])
```

```
In [25]: data['enginesize'].unique()
```

```
Out[25]: array([130, 152, 109, 136, 131, 108, 164, 209,  61,  90,  98, 122, 156,
                92,  79, 110, 111, 119, 258, 326,  91,  70,  80, 140, 134, 183,
                234, 308, 304,  97, 103, 120, 181, 151, 194, 203, 132, 121, 146,
                171, 161, 141, 173, 145], dtype=int64)
```

```
In [26]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   CarName                205 non-null   object
1   fueltype               205 non-null   object
2   aspiration             205 non-null   object
3   doornumber             205 non-null   object
4   carbody                205 non-null   object
5   drivewheel            205 non-null   object
6   enginelocation         205 non-null   object
7   wheelbase              205 non-null   float64
8   carlength              205 non-null   float64
9   carwidth               205 non-null   float64
10  curbweight             205 non-null   int64
11  enginetype             205 non-null   object
12  cylindernumber         205 non-null   object
13  enginesize             205 non-null   int64
14  fuelsystem             205 non-null   object
15  boreratio              205 non-null   float64
16  horsepower             205 non-null   int64
17  citympg                205 non-null   int64
18  highwaympg            205 non-null   int64
19  price                  205 non-null   float64
dtypes: float64(5), int64(5), object(10)
memory usage: 32.2+ KB
```

```
In [27]: data['CarName'].unique().size
```

```
Out[27]: 147
```



```
In [28]: data['CarName'].unique()
```

```
Out[28]: array(['alfa-romero giulia', 'alfa-romero stelvio',  
               'alfa-romero Quadrifoglio', 'audi 100 ls', 'audi 100ls',  
               'audi fox', 'audi 5000', 'audi 4000', 'audi 5000s (diesel)',  
               'bmw 320i', 'bmw x1', 'bmw x3', 'bmw z4', 'bmw x4', 'bmw x5',  
               'chevrolet impala', 'chevrolet monte carlo', 'chevrolet vega 2300',  
               'dodge rampage', 'dodge challenger se', 'dodge d200',  
               'dodge monaco (sw)', 'dodge colt hardtop', 'dodge colt (sw)',  
               'dodge coronet custom', 'dodge dart custom',  
               'dodge coronet custom (sw)', 'honda civic', 'honda civic cvcc',  
               'honda accord cvcc', 'honda accord lx', 'honda civic 1500 gl',  
               'honda accord', 'honda civic 1300', 'honda prelude',  
               'honda civic (auto)', 'isuzu MU-X', 'isuzu D-Max ',  
               'isuzu D-Max V-Cross', 'jaguar xj', 'jaguar xf', 'jaguar xk',  
               'maxda rx3', 'maxda glc deluxe', 'mazda rx2 coupe', 'mazda rx-4',  
               'mazda glc deluxe', 'mazda 626', 'mazda glc', 'mazda rx-7 gs',  
               'mazda glc 4', 'mazda glc custom l', 'mazda glc custom',  
               'buick electra 225 custom', 'buick century luxus (sw)',  
               'buick century', 'buick skyhawk', 'buick opel isuzu deluxe',  
               'buick skylark', 'buick century special',  
               'buick regal sport coupe (turbo)', 'mercury cougar',  
               'mitsubishi mirage', 'mitsubishi lancer', 'mitsubishi outlander',  
               'mitsubishi g4', 'mitsubishi mirage g4', 'mitsubishi montero',  
               'mitsubishi pajero', 'Nissan versa', 'nissan gt-r', 'nissan rogue',  
               'nissan latio', 'nissan titan', 'nissan leaf', 'nissan juke',  
               'nissan note', 'nissan clipper', 'nissan nv200', 'nissan dayz',  
               'nissan fuga', 'nissan otti', 'nissan teana', 'nissan kicks',  
               'peugeot 504', 'peugeot 304', 'peugeot 504 (sw)', 'peugeot 604sl',  
               'peugeot 505s turbo diesel', 'plymouth fury iii',  
               'plymouth cricket', 'plymouth satellite custom (sw)',  
               'plymouth fury gran sedan', 'plymouth valiant', 'plymouth duster',  
               'porsche macan', 'porcshce panamera', 'porsche cayenne',  
               'porsche boxter', 'renault 12tl', 'renault 5 gtl', 'saab 99e',  
               'saab 99le', 'saab 99gle', 'subaru', 'subaru dl', 'subaru brz',  
               'subaru baja', 'subaru r1', 'subaru r2', 'subaru trezia',  
               'subaru tribeca', 'toyota corona mark ii', 'toyota corona',  
               'toyota corolla 1200', 'toyota corona hardtop',  
               'toyota corolla 1600 (sw)', 'toyota carina', 'toyota mark ii',  
               'toyota corolla', 'toyota corolla liftback',  
               'toyota celica gt liftback', 'toyota corolla tercel',  
               'toyota corona liftback', 'toyota starlet', 'toyota tercel',  
               'toyota cressida', 'toyota celica gt', 'toyouta tercel',  
               'volkswagen rabbit', 'volkswagen 1131 deluxe sedan',  
               'volkswagen model 111', 'volkswagen type 3', 'volkswagen 411 (sw)',  
               'volkswagen super beetle', 'volkswagen dasher', 'vw dasher',  
               'vw rabbit', 'volkswagen rabbit', 'volkswagen rabbit custom',
```

```
'volvo 145e (sw)', 'volvo 144ea', 'volvo 244dl', 'volvo 245',  
'volvo 264gl', 'volvo diesel', 'volvo 246'], dtype=object)
```

```
In [29]: data['fueltype'].unique()
```

```
Out[29]: array(['gas', 'diesel'], dtype=object)
```

```
In [30]: from sklearn.preprocessing import LabelEncoder
```

```
In [31]: encoder=LabelEncoder()
```

```
In [32]: data['fueltype']=encoder.fit_transform((data['fueltype']))
```

```
In [33]: data['fueltype'].unique()  ##gas=1 diesel=0
```

```
Out[33]: array([1, 0])
```

```
In [34]: data['aspiration'].unique()
```

```
Out[34]: array(['std', 'turbo'], dtype=object)
```

```
In [35]: data['aspiration']=encoder.fit_transform((data['aspiration']))
```

```
In [36]: data['doornumber'].unique()
```

```
Out[36]: array(['two', 'four'], dtype=object)
```

```
In [37]: data['doornumber']=encoder.fit_transform((data['doornumber']))
```

```
In [38]: data['carbody'].unique()
```

```
Out[38]: array(['convertible', 'hatchback', 'sedan', 'wagon', 'hardtop'],  
              dtype=object)
```

```
In [39]: data.head()
```

```
Out[39]:
```

| | CarName | fueltype | aspiration | doornumber | carbody | drivewheel | enginelocation | wheelbase | carlength | carwidth | curbweight | enginetype | cylind |
|---|-----------------------------|----------|------------|------------|-------------|------------|----------------|-----------|-----------|----------|------------|------------|--------|
| 0 | alfa-romero giulia | 1 | 0 | 1 | convertible | rwd | front | 88.6 | 168.8 | 64.1 | 2548 | dohc | |
| 1 | alfa-romero stelvio | 1 | 0 | 1 | convertible | rwd | front | 88.6 | 168.8 | 64.1 | 2548 | dohc | |
| 2 | alfa-romero Quadrifoglio | 1 | 0 | 1 | hatchback | rwd | front | 94.5 | 171.2 | 65.5 | 2823 | ohcv | |
| 3 | audi 100 ls | 1 | 0 | 0 | sedan | fwd | front | 99.8 | 176.6 | 66.2 | 2337 | ohc | |
| 4 | audi 100ls | 1 | 0 | 0 | sedan | fwd | front | 99.4 | 176.6 | 66.4 | 2824 | ohc | |

```
In [40]: data.shape
```

```
Out[40]: (205, 20)
```

```
In [41]: data['drivewheel'].unique()
```

```
Out[41]: array(['rwd', 'fwd'], dtype=object)
```

```
In [42]: data['drivewheel']=encoder.fit_transform((data['drivewheel']))
```

```
In [43]: data['enginelocation'].unique()
```

```
Out[43]: array(['front', 'rear'], dtype=object)
```

```
In [44]: data['enginelocation']=encoder.fit_transform((data['enginelocation']))
```

```
In [45]: data['enginetype'].unique()
```

```
Out[45]: array(['dohc', 'ohcv', 'ohc', 'l', 'rotor', 'ohcf', 'dohcv'], dtype=object)
```

```
In [46]: data['cylindernumber'].unique()
```

```
Out[46]: array(['four', 'six', 'five', 'three', 'twelve', 'two', 'eight'],  
              dtype=object)
```

```
In [47]: data['fuelsystem'].unique()
```

```
Out[47]: array(['mpfi', '2bbl', 'mfi', '1bbl', 'spfi', '4bbl', 'idi', 'spdi'],  
              dtype=object)
```

```
In [48]: data.head()
```

Out[48]:

| | CarName | fueltype | aspiration | doornumber | carbbody | drivewheel | enginelocation | wheelbase | carlength | carwidth | curbweight | enginetype | cylind |
|---|-----------------------------|----------|------------|------------|-------------|------------|----------------|-----------|-----------|----------|------------|------------|--------|
| 0 | alfa-romero giulia | 1 | 0 | 1 | convertible | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | dohc | |
| 1 | alfa-romero stelvio | 1 | 0 | 1 | convertible | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | dohc | |
| 2 | alfa-romero Quadrifoglio | 1 | 0 | 1 | hatchback | 1 | 0 | 94.5 | 171.2 | 65.5 | 2823 | ohcv | |
| 3 | audi 100 ls | 1 | 0 | 0 | sedan | 0 | 0 | 99.8 | 176.6 | 66.2 | 2337 | ohc | |
| 4 | audi 100ls | 1 | 0 | 0 | sedan | 0 | 0 | 99.4 | 176.6 | 66.4 | 2824 | ohc | |

```
In [49]: data=data.iloc[:,1:]    ###as carname contains 147 different values and length of dataset is 205 we can remove the carname
```

```
In [50]: data.head()
```

Out[50]:

| | fueltype | aspiration | doornumber | carbbody | drivewheel | enginelocation | wheelbase | carlength | carwidth | curbweight | enginetype | cylindernumber | e |
|---|----------|------------|------------|-------------|------------|----------------|-----------|-----------|----------|------------|------------|----------------|---|
| 0 | 1 | 0 | 1 | convertible | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | dohc | four | |
| 1 | 1 | 0 | 1 | convertible | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | dohc | four | |
| 2 | 1 | 0 | 1 | hatchback | 1 | 0 | 94.5 | 171.2 | 65.5 | 2823 | ohcv | six | |
| 3 | 1 | 0 | 0 | sedan | 0 | 0 | 99.8 | 176.6 | 66.2 | 2337 | ohc | four | |
| 4 | 1 | 0 | 0 | sedan | 0 | 0 | 99.4 | 176.6 | 66.4 | 2824 | ohc | five | |

```
In [51]: y=data['price']
```

```
In [52]: print(y)
```

```
0      13495.0
1      16500.0
2      16500.0
3      13950.0
4      17450.0
...
200     16845.0
201     19045.0
202     21485.0
203     22470.0
204     22625.0
Name: price, Length: 205, dtype: float64
```

```
In [53]: data=data.drop(['price'],axis=1)
```

```
In [54]: data.head()
```

```
Out[54]:
```

| | fueltype | aspiration | doornumber | carbody | drivewheel | enginelocation | wheelbase | carlength | carwidth | curbweight | enginetype | cylindernumber | e |
|---|----------|------------|------------|-------------|------------|----------------|-----------|-----------|----------|------------|------------|----------------|---|
| 0 | 1 | 0 | 1 | convertible | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | dohc | four | |
| 1 | 1 | 0 | 1 | convertible | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | dohc | four | |
| 2 | 1 | 0 | 1 | hatchback | 1 | 0 | 94.5 | 171.2 | 65.5 | 2823 | ohcv | six | |
| 3 | 1 | 0 | 0 | sedan | 0 | 0 | 99.8 | 176.6 | 66.2 | 2337 | ohc | four | |
| 4 | 1 | 0 | 0 | sedan | 0 | 0 | 99.4 | 176.6 | 66.4 | 2824 | ohc | five | |

```
In [55]: data=pd.get_dummies(data,['carbody','enginetype','cylindernumber','fuelsystem'],drop_first=True)
```

In [56]: data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 37 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   fueltype                             205 non-null    int32
1   aspiration                           205 non-null    int32
2   doornumber                           205 non-null    int32
3   drivewheel                           205 non-null    int32
4   enginelocation                       205 non-null    int32
5   wheelbase                           205 non-null    float64
6   carlength                           205 non-null    float64
7   carwidth                            205 non-null    float64
8   curbweight                           205 non-null    int64
9   enginesize                           205 non-null    int64
10  boreratio                            205 non-null    float64
11  horsepower                           205 non-null    int64
12  citympg                              205 non-null    int64
13  highwaympg                           205 non-null    int64
14  carbody_hardtop                       205 non-null    uint8
15  carbody_hatchback                     205 non-null    uint8
16  carbody_sedan                         205 non-null    uint8
17  carbody_wagon                         205 non-null    uint8
18  enginetype_dohcv                      205 non-null    uint8
19  enginetype_l                          205 non-null    uint8
20  enginetype_ohc                       205 non-null    uint8
21  enginetype_ohcf                       205 non-null    uint8
22  enginetype_ohcv                       205 non-null    uint8
23  enginetype_rotor                      205 non-null    uint8
24  cylindernumber_five                   205 non-null    uint8
25  cylindernumber_four                   205 non-null    uint8
26  cylindernumber_six                    205 non-null    uint8
27  cylindernumber_three                  205 non-null    uint8
28  cylindernumber_twelve                 205 non-null    uint8
29  cylindernumber_two                    205 non-null    uint8
30  fuelsystem_2bbl                       205 non-null    uint8
31  fuelsystem_4bbl                       205 non-null    uint8
32  fuelsystem_idi                        205 non-null    uint8
33  fuelsystem_mfi                        205 non-null    uint8
34  fuelsystem_mphi                       205 non-null    uint8
35  fuelsystem_spdi                       205 non-null    uint8
36  fuelsystem_spfi                       205 non-null    uint8
dtypes: float64(4), int32(5), int64(5), uint8(23)
memory usage: 23.1 KB
```

```
In [57]: data.head(5)
```

Out[57]:

| | fueltype | aspiration | doornumber | drivewheel | enginelocation | wheelbase | carlength | carwidth | curbweight | enginesize | ... | cylindernumber_three | cyli |
|---|----------|------------|------------|------------|----------------|-----------|-----------|----------|------------|------------|-----|----------------------|------|
| 0 | 1 | 0 | 1 | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | 130 | ... | 0 | |
| 1 | 1 | 0 | 1 | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | 130 | ... | 0 | |
| 2 | 1 | 0 | 1 | 1 | 0 | 94.5 | 171.2 | 65.5 | 2823 | 152 | ... | 0 | |
| 3 | 1 | 0 | 0 | 0 | 0 | 99.8 | 176.6 | 66.2 | 2337 | 109 | ... | 0 | |
| 4 | 1 | 0 | 0 | 0 | 0 | 99.4 | 176.6 | 66.4 | 2824 | 136 | ... | 0 | |

5 rows × 37 columns

```
In [58]: data['price']=y
```

```
In [59]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 38 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   fueltype                               205 non-null    int32
1   aspiration                             205 non-null    int32
2   doornumber                             205 non-null    int32
3   drivewheel                             205 non-null    int32
4   enginelocation                         205 non-null    int32
5   wheelbase                             205 non-null    float64
6   carlength                             205 non-null    float64
7   carwidth                              205 non-null    float64
8   curbweight                             205 non-null    int64
9   enginesize                             205 non-null    int64
10  boreratio                              205 non-null    float64
11  horsepower                             205 non-null    int64
12  citympg                                205 non-null    int64
13  highwaympg                             205 non-null    int64
14  carbody_hardtop                         205 non-null    uint8
15  carbody_hatchback                       205 non-null    uint8
16  carbody_sedan                           205 non-null    uint8
17  carbody_wagon                           205 non-null    uint8
18  enginetype_dohcv                        205 non-null    uint8
19  enginetype_l                            205 non-null    uint8
20  enginetype_ohc                          205 non-null    uint8
21  enginetype_ohcf                         205 non-null    uint8
22  enginetype_ohcv                         205 non-null    uint8
23  enginetype_rotor                        205 non-null    uint8
24  cylindernumber_five                     205 non-null    uint8
25  cylindernumber_four                     205 non-null    uint8
26  cylindernumber_six                      205 non-null    uint8
27  cylindernumber_three                    205 non-null    uint8
28  cylindernumber_twelve                   205 non-null    uint8
29  cylindernumber_two                      205 non-null    uint8
30  fuelsystem_2bbl                         205 non-null    uint8
31  fuelsystem_4bbl                         205 non-null    uint8
32  fuelsystem_idi                          205 non-null    uint8
33  fuelsystem_mfi                          205 non-null    uint8
34  fuelsystem_mphi                         205 non-null    uint8
35  fuelsystem_spdi                         205 non-null    uint8
36  fuelsystem_spfi                         205 non-null    uint8
37  price                                  205 non-null    float64
dtypes: float64(5), int32(5), int64(5), uint8(23)
memory usage: 24.7 KB
```


Regression M0del

```
In [60]: from sklearn.linear_model import LinearRegression
```

```
In [61]: model=LinearRegression()
```

```
In [62]: from sklearn.model_selection import train_test_split
```

```
In [63]: x=data.iloc[:, :-1]
```

```
In [64]: x
```

```
Out[64]:
```

| | fueltype | aspiration | doornumber | drivewheel | enginelocation | wheelbase | carlength | carwidth | curbweight | enginesize | ... | cylindernumber_three | cylindernumber_four |
|-----|----------|------------|------------|------------|----------------|-----------|-----------|----------|------------|------------|-----|----------------------|---------------------|
| 0 | 1 | 0 | 1 | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | 130 | ... | 0 | 0 |
| 1 | 1 | 0 | 1 | 1 | 0 | 88.6 | 168.8 | 64.1 | 2548 | 130 | ... | 0 | 0 |
| 2 | 1 | 0 | 1 | 1 | 0 | 94.5 | 171.2 | 65.5 | 2823 | 152 | ... | 0 | 0 |
| 3 | 1 | 0 | 0 | 0 | 0 | 99.8 | 176.6 | 66.2 | 2337 | 109 | ... | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 | 0 | 99.4 | 176.6 | 66.4 | 2824 | 136 | ... | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 200 | 1 | 0 | 0 | 1 | 0 | 109.1 | 188.8 | 68.9 | 2952 | 141 | ... | 0 | 0 |
| 201 | 1 | 1 | 0 | 1 | 0 | 109.1 | 188.8 | 68.8 | 3049 | 141 | ... | 0 | 0 |
| 202 | 1 | 0 | 0 | 1 | 0 | 109.1 | 188.8 | 68.9 | 3012 | 173 | ... | 0 | 0 |
| 203 | 0 | 1 | 0 | 1 | 0 | 109.1 | 188.8 | 68.9 | 3217 | 145 | ... | 0 | 0 |
| 204 | 1 | 1 | 0 | 1 | 0 | 109.1 | 188.8 | 68.9 | 3062 | 141 | ... | 0 | 0 |

205 rows × 37 columns

```
In [65]: y=data['price']
```

```
In [66]: x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=42,test_size=0.2)
```

```
In [67]: model=LinearRegression()
```

```
In [68]: model.fit(x_train,y_train)
```

```
Out[68]: LinearRegression()
```

```
In [69]: y_pred=model.predict(x_test)
```

```
In [70]: from sklearn.metrics import r2_score
```

```
In [71]: score=r2_score(y_test,y_pred)
```

```
In [72]: print(score*100)
```

87.2243957721253

Random Forest Model

```
In [73]: from sklearn.ensemble import RandomForestRegressor
```

```
In [74]: x=data.iloc[:, :-1]
```

```
In [75]: y=data['price']
```

```
In [76]: from sklearn.datasets import make_regression
```

```
In [77]: x,y=make_regression(n_features=4, n_informative=2,random_state=0, shuffle=False)
```

```
In [78]: model=RandomForestRegressor(max_depth=2,random_state=42)
```

```
In [79]: model.fit(x,y)
```

```
Out[79]: RandomForestRegressor(max_depth=2, random_state=42)
```

```
In [80]: y_pred=model.predict(x)
```

```
In [81]: score=r2_score(y,y_pred)
```

```
In [82]: score=score*100
```

```
In [83]: print(score)
```

84.75417782248233

