In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
other_path = "https://s3-api.us-geo.objectstorage.softlayer.net/cf-courses-data/CognitiveClass/DA0101EN/auto.csv"
```

In [3]:

```
df = pd.read_csv(other_path,header= None)
```

In [4]:

```
headers = ["symboling","normalized-losses","make","fuel-type","aspiration", "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","stroke","compression-ratio","horsepower",
         "peak-rpm","city-mpg","highway-mpg","price"]
```

In [5]:

```
df.columns = headers
```

In [6]:

```
df.head()
```

Out[6]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | ? | alfa-romero | gas | std | two | convertible | rwd | front | |
| **1** | 3 | ? | alfa-romero | gas | std | two | convertible | rwd | front | |
| **2** | 1 | ? | alfa-romero | gas | std | two | hatchback | rwd | front | |
| **3** | 2 | 164 | audi | gas | std | four | sedan | fwd | front | |
| **4** | 2 | 164 | audi | gas | std | four | sedan | 4wd | front | |

5 rows × 26 columns

In [7]:

```
pd.options.display.max_columns = None
pd.options.display.max_rows = None
```

In [8]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 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   aspiration         205 non-null    object 
 5   num-of-doors       205 non-null    object 
 6   body-style         205 non-null    object 
 7   drive-wheels       205 non-null    object 
 8   engine-location    205 non-null    object 
 9   wheel-base         205 non-null    float64
 10  length             205 non-null    float64
 11  width              205 non-null    float64
 12  height             205 non-null    float64
 13  curb-weight        205 non-null    int64  
 14  engine-type        205 non-null    object 
 15  num-of-cylinders   205 non-null    object 
 16  engine-size        205 non-null    int64  
 17  fuel-system        205 non-null    object 
 18  bore               205 non-null    object 
 19  stroke             205 non-null    object 
 20  compression-ratio  205 non-null    float64
 21  horsepower         205 non-null    object 
 22  peak-rpm           205 non-null    object 
 23  city-mpg           205 non-null    int64  
 24  highway-mpg        205 non-null    int64  
 25  price              205 non-null    object 
dtypes: float64(5), int64(5), object(16)
memory usage: 41.8+ KB
```

In [9]:

```
df.describe(include="all")
```

Out[9]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 205.000000 | 205 | 205 | 205 | 205 | 205 | 205 | 205 | 205 | 20 |
| **unique** | NaN | 52 | 22 | 2 | 2 | 3 | 5 | 3 | 2 | |
| **top** | NaN | ? | toyota | gas | std | four | sedan | fwd | front | |
| **freq** | NaN | 41 | 32 | 185 | 168 | 114 | 96 | 120 | 202 | |
| **mean** | 0.834146 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 9 |
| **std** | 1.245307 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | |
| **min** | -2.000000 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 8 |
| **25%** | 0.000000 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 9 |
| **50%** | 1.000000 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 9 |
| **75%** | 2.000000 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 10 |
| **max** | 3.000000 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | 12 |

In [10]:

```
df.head()
```

Out[10]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | ? | alfa-romero | gas | std | two | convertible | rwd | front | |
| **1** | 3 | ? | alfa-romero | gas | std | two | convertible | rwd | front | |
| **2** | 1 | ? | alfa-romero | gas | std | two | hatchback | rwd | front | |
| **3** | 2 | 164 | audi | gas | std | four | sedan | fwd | front | |
| **4** | 2 | 164 | audi | gas | std | four | sedan | 4wd | front | |

# Replacing ? with Nan

In [11]:

```
df.replace("?",np.nan,inplace= True)
```

In [12]:

```
df.head()
```

Out[12]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | NaN | alfa-romero | gas | std | two | convertible | rwd | front | |
| **1** | 3 | NaN | alfa-romero | gas | std | two | convertible | rwd | front | |
| **2** | 1 | NaN | alfa-romero | gas | std | two | hatchback | rwd | front | |
| **3** | 2 | 164 | audi | gas | std | four | sedan | fwd | front | |
| **4** | 2 | 164 | audi | gas | std | four | sedan | 4wd | front | |

# Converting Object to float

In [13]:

```
df.select_dtypes(include=["int","float"])
```

Out[13]:

| | wheel-base | length | width | height | compression-ratio |
|---|---|---|---|---|---|
| **0** | 88.6 | 168.8 | 64.1 | 48.8 | 9.00 |
| **1** | 88.6 | 168.8 | 64.1 | 48.8 | 9.00 |
| **2** | 94.5 | 171.2 | 65.5 | 52.4 | 9.00 |
| **3** | 99.8 | 176.6 | 66.2 | 54.3 | 10.00 |
| **4** | 99.4 | 176.6 | 66.4 | 54.3 | 8.00 |
| **5** | 99.8 | 177.3 | 66.3 | 53.1 | 8.50 |
| **6** | 105.8 | 192.7 | 71.4 | 55.7 | 8.50 |
| **7** | 105.8 | 192.7 | 71.4 | 55.7 | 8.50 |
| **8** | 105.8 | 192.7 | 71.4 | 55.9 | 8.30 |
| **9** | 99.5 | 178.2 | 67.9 | 52.0 | 7.00 |
| **10** | 101.2 | 176.8 | 64.8 | 54.3 | 8.80 |
| **11** | 101.2 | 176.8 | 64.8 | 54.3 | 8.80 |
| **12** | 101.2 | 176.8 | 64.8 | 54.3 | 9.00 |
| **13** | 101.2 | 176.8 | 64.8 | 54.3 | 9.00 |
| **14** | 103.5 | 189.0 | 66.9 | 55.7 | 9.00 |
| **15** | 103.5 | 189.0 | 66.9 | 55.7 | 8.00 |
| **16** | 103.5 | 193.8 | 67.9 | 53.7 | 8.00 |
| **17** | 110.0 | 197.0 | 70.9 | 56.3 | 8.00 |
| **18** | 88.4 | 141.1 | 60.3 | 53.2 | 9.50 |
| **19** | 94.5 | 155.9 | 63.6 | 52.0 | 9.60 |
| **20** | 94.5 | 158.8 | 63.6 | 52.0 | 9.60 |
| **21** | 93.7 | 157.3 | 63.8 | 50.8 | 9.41 |
| **22** | 93.7 | 157.3 | 63.8 | 50.8 | 9.40 |
| **23** | 93.7 | 157.3 | 63.8 | 50.8 | 7.60 |
| **24** | 93.7 | 157.3 | 63.8 | 50.6 | 9.40 |
| **25** | 93.7 | 157.3 | 63.8 | 50.6 | 9.40 |
| **26** | 93.7 | 157.3 | 63.8 | 50.6 | 9.40 |
| **27** | 93.7 | 157.3 | 63.8 | 50.6 | 7.60 |
| **28** | 103.3 | 174.6 | 64.6 | 59.8 | 8.50 |
| **29** | 95.9 | 173.2 | 66.3 | 50.2 | 7.00 |
| **30** | 86.6 | 144.6 | 63.9 | 50.8 | 9.60 |
| **31** | 86.6 | 144.6 | 63.9 | 50.8 | 9.20 |
| **32** | 93.7 | 150.0 | 64.0 | 52.6 | 10.10 |
| **33** | 93.7 | 150.0 | 64.0 | 52.6 | 9.20 |
| **34** | 93.7 | 150.0 | 64.0 | 52.6 | 9.20 |
| **35** | 96.5 | 163.4 | 64.0 | 54.5 | 9.20 |
| **36** | 96.5 | 157.1 | 63.9 | 58.3 | 9.20 |

| | wheel-base | length | width | height | compression-ratio |
|---|---|---|---|---|---|
| **37** | 96.5 | 167.5 | 65.2 | 53.3 | 9.00 |
| **38** | 96.5 | 167.5 | 65.2 | 53.3 | 9.00 |
| **39** | 96.5 | 175.4 | 65.2 | 54.1 | 9.00 |
| **40** | 96.5 | 175.4 | 62.5 | 54.1 | 9.00 |
| **41** | 96.5 | 175.4 | 65.2 | 54.1 | 9.00 |
| **42** | 96.5 | 169.1 | 66.0 | 51.0 | 9.10 |
| **43** | 94.3 | 170.7 | 61.8 | 53.5 | 8.50 |
| **44** | 94.5 | 155.9 | 63.6 | 52.0 | 9.60 |
| **45** | 94.5 | 155.9 | 63.6 | 52.0 | 9.60 |
| **46** | 96.0 | 172.6 | 65.2 | 51.4 | 9.20 |
| **47** | 113.0 | 199.6 | 69.6 | 52.8 | 8.10 |
| **48** | 113.0 | 199.6 | 69.6 | 52.8 | 8.10 |
| **49** | 102.0 | 191.7 | 70.6 | 47.8 | 11.50 |
| **50** | 93.1 | 159.1 | 64.2 | 54.1 | 9.00 |
| **51** | 93.1 | 159.1 | 64.2 | 54.1 | 9.00 |
| **52** | 93.1 | 159.1 | 64.2 | 54.1 | 9.00 |
| **53** | 93.1 | 166.8 | 64.2 | 54.1 | 9.00 |
| **54** | 93.1 | 166.8 | 64.2 | 54.1 | 9.00 |
| **55** | 95.3 | 169.0 | 65.7 | 49.6 | 9.40 |
| **56** | 95.3 | 169.0 | 65.7 | 49.6 | 9.40 |
| **57** | 95.3 | 169.0 | 65.7 | 49.6 | 9.40 |
| **58** | 95.3 | 169.0 | 65.7 | 49.6 | 9.40 |
| **59** | 98.8 | 177.8 | 66.5 | 53.7 | 8.60 |
| **60** | 98.8 | 177.8 | 66.5 | 55.5 | 8.60 |
| **61** | 98.8 | 177.8 | 66.5 | 53.7 | 8.60 |
| **62** | 98.8 | 177.8 | 66.5 | 55.5 | 8.60 |
| **63** | 98.8 | 177.8 | 66.5 | 55.5 | 22.70 |
| **64** | 98.8 | 177.8 | 66.5 | 55.5 | 8.60 |
| **65** | 104.9 | 175.0 | 66.1 | 54.4 | 8.00 |
| **66** | 104.9 | 175.0 | 66.1 | 54.4 | 22.00 |
| **67** | 110.0 | 190.9 | 70.3 | 56.5 | 21.50 |
| **68** | 110.0 | 190.9 | 70.3 | 58.7 | 21.50 |
| **69** | 106.7 | 187.5 | 70.3 | 54.9 | 21.50 |
| **70** | 115.6 | 202.6 | 71.7 | 56.3 | 21.50 |
| **71** | 115.6 | 202.6 | 71.7 | 56.5 | 8.30 |
| **72** | 96.6 | 180.3 | 70.5 | 50.8 | 8.30 |
| **73** | 120.9 | 208.1 | 71.7 | 56.7 | 8.00 |
| **74** | 112.0 | 199.2 | 72.0 | 55.4 | 8.00 |
| **75** | 102.7 | 178.4 | 68.0 | 54.8 | 8.00 |

| | wheel-base | length | width | height | compression-ratio |
|---|---|---|---|---|---|
| **76** | 93.7 | 157.3 | 64.4 | 50.8 | 9.40 |
| **77** | 93.7 | 157.3 | 64.4 | 50.8 | 9.40 |
| **78** | 93.7 | 157.3 | 64.4 | 50.8 | 9.40 |
| **79** | 93.0 | 157.3 | 63.8 | 50.8 | 7.60 |
| **80** | 96.3 | 173.0 | 65.4 | 49.4 | 7.50 |
| **81** | 96.3 | 173.0 | 65.4 | 49.4 | 8.50 |
| **82** | 95.9 | 173.2 | 66.3 | 50.2 | 7.00 |
| **83** | 95.9 | 173.2 | 66.3 | 50.2 | 7.00 |
| **84** | 95.9 | 173.2 | 66.3 | 50.2 | 7.00 |
| **85** | 96.3 | 172.4 | 65.4 | 51.6 | 8.50 |
| **86** | 96.3 | 172.4 | 65.4 | 51.6 | 8.50 |
| **87** | 96.3 | 172.4 | 65.4 | 51.6 | 7.50 |
| **88** | 96.3 | 172.4 | 65.4 | 51.6 | 7.50 |
| **89** | 94.5 | 165.3 | 63.8 | 54.5 | 9.40 |
| **90** | 94.5 | 165.3 | 63.8 | 54.5 | 21.90 |
| **91** | 94.5 | 165.3 | 63.8 | 54.5 | 9.40 |
| **92** | 94.5 | 165.3 | 63.8 | 54.5 | 9.40 |
| **93** | 94.5 | 170.2 | 63.8 | 53.5 | 9.40 |
| **94** | 94.5 | 165.3 | 63.8 | 54.5 | 9.40 |
| **95** | 94.5 | 165.6 | 63.8 | 53.3 | 9.40 |
| **96** | 94.5 | 165.3 | 63.8 | 54.5 | 9.40 |
| **97** | 94.5 | 170.2 | 63.8 | 53.5 | 9.40 |
| **98** | 95.1 | 162.4 | 63.8 | 53.3 | 9.40 |
| **99** | 97.2 | 173.4 | 65.2 | 54.7 | 8.50 |
| **100** | 97.2 | 173.4 | 65.2 | 54.7 | 8.50 |
| **101** | 100.4 | 181.7 | 66.5 | 55.1 | 9.00 |
| **102** | 100.4 | 184.6 | 66.5 | 56.1 | 9.00 |
| **103** | 100.4 | 184.6 | 66.5 | 55.1 | 9.00 |
| **104** | 91.3 | 170.7 | 67.9 | 49.7 | 9.00 |
| **105** | 91.3 | 170.7 | 67.9 | 49.7 | 7.80 |
| **106** | 99.2 | 178.5 | 67.9 | 49.7 | 9.00 |
| **107** | 107.9 | 186.7 | 68.4 | 56.7 | 8.40 |
| **108** | 107.9 | 186.7 | 68.4 | 56.7 | 21.00 |
| **109** | 114.2 | 198.9 | 68.4 | 58.7 | 8.40 |
| **110** | 114.2 | 198.9 | 68.4 | 58.7 | 21.00 |
| **111** | 107.9 | 186.7 | 68.4 | 56.7 | 8.40 |
| **112** | 107.9 | 186.7 | 68.4 | 56.7 | 21.00 |
| **113** | 114.2 | 198.9 | 68.4 | 56.7 | 8.40 |
| **114** | 114.2 | 198.9 | 68.4 | 58.7 | 21.00 |

| | wheel-base | length | width | height | compression-ratio |
|---|---|---|---|---|---|
| **115** | 107.9 | 186.7 | 68.4 | 56.7 | 8.40 |
| **116** | 107.9 | 186.7 | 68.4 | 56.7 | 21.00 |
| **117** | 108.0 | 186.7 | 68.3 | 56.0 | 7.00 |
| **118** | 93.7 | 157.3 | 63.8 | 50.8 | 9.40 |
| **119** | 93.7 | 157.3 | 63.8 | 50.8 | 7.60 |
| **120** | 93.7 | 157.3 | 63.8 | 50.6 | 9.40 |
| **121** | 93.7 | 167.3 | 63.8 | 50.8 | 9.40 |
| **122** | 93.7 | 167.3 | 63.8 | 50.8 | 9.40 |
| **123** | 103.3 | 174.6 | 64.6 | 59.8 | 8.50 |
| **124** | 95.9 | 173.2 | 66.3 | 50.2 | 7.00 |
| **125** | 94.5 | 168.9 | 68.3 | 50.2 | 9.50 |
| **126** | 89.5 | 168.9 | 65.0 | 51.6 | 9.50 |
| **127** | 89.5 | 168.9 | 65.0 | 51.6 | 9.50 |
| **128** | 89.5 | 168.9 | 65.0 | 51.6 | 9.50 |
| **129** | 98.4 | 175.7 | 72.3 | 50.5 | 10.00 |
| **130** | 96.1 | 181.5 | 66.5 | 55.2 | 8.70 |
| **131** | 96.1 | 176.8 | 66.6 | 50.5 | 8.70 |
| **132** | 99.1 | 186.6 | 66.5 | 56.1 | 9.31 |
| **133** | 99.1 | 186.6 | 66.5 | 56.1 | 9.30 |
| **134** | 99.1 | 186.6 | 66.5 | 56.1 | 9.30 |
| **135** | 99.1 | 186.6 | 66.5 | 56.1 | 9.30 |
| **136** | 99.1 | 186.6 | 66.5 | 56.1 | 9.00 |
| **137** | 99.1 | 186.6 | 66.5 | 56.1 | 9.00 |
| **138** | 93.7 | 156.9 | 63.4 | 53.7 | 9.00 |
| **139** | 93.7 | 157.9 | 63.6 | 53.7 | 8.70 |
| **140** | 93.3 | 157.3 | 63.8 | 55.7 | 8.70 |
| **141** | 97.2 | 172.0 | 65.4 | 52.5 | 9.50 |
| **142** | 97.2 | 172.0 | 65.4 | 52.5 | 9.50 |
| **143** | 97.2 | 172.0 | 65.4 | 52.5 | 9.00 |
| **144** | 97.0 | 172.0 | 65.4 | 54.3 | 9.00 |
| **145** | 97.0 | 172.0 | 65.4 | 54.3 | 7.70 |
| **146** | 97.0 | 173.5 | 65.4 | 53.0 | 9.00 |
| **147** | 97.0 | 173.5 | 65.4 | 53.0 | 9.00 |
| **148** | 96.9 | 173.6 | 65.4 | 54.9 | 9.00 |
| **149** | 96.9 | 173.6 | 65.4 | 54.9 | 7.70 |
| **150** | 95.7 | 158.7 | 63.6 | 54.5 | 9.00 |
| **151** | 95.7 | 158.7 | 63.6 | 54.5 | 9.00 |
| **152** | 95.7 | 158.7 | 63.6 | 54.5 | 9.00 |
| **153** | 95.7 | 169.7 | 63.6 | 59.1 | 9.00 |

| | wheel-base | length | width | height | compression-ratio |
|---|---|---|---|---|---|
| **154** | 95.7 | 169.7 | 63.6 | 59.1 | 9.00 |
| **155** | 95.7 | 169.7 | 63.6 | 59.1 | 9.00 |
| **156** | 95.7 | 166.3 | 64.4 | 53.0 | 9.00 |
| **157** | 95.7 | 166.3 | 64.4 | 52.8 | 9.00 |
| **158** | 95.7 | 166.3 | 64.4 | 53.0 | 22.50 |
| **159** | 95.7 | 166.3 | 64.4 | 52.8 | 22.50 |
| **160** | 95.7 | 166.3 | 64.4 | 53.0 | 9.00 |
| **161** | 95.7 | 166.3 | 64.4 | 52.8 | 9.00 |
| **162** | 95.7 | 166.3 | 64.4 | 52.8 | 9.00 |
| **163** | 94.5 | 168.7 | 64.0 | 52.6 | 9.00 |
| **164** | 94.5 | 168.7 | 64.0 | 52.6 | 9.00 |
| **165** | 94.5 | 168.7 | 64.0 | 52.6 | 9.40 |
| **166** | 94.5 | 168.7 | 64.0 | 52.6 | 9.40 |
| **167** | 98.4 | 176.2 | 65.6 | 52.0 | 9.30 |
| **168** | 98.4 | 176.2 | 65.6 | 52.0 | 9.30 |
| **169** | 98.4 | 176.2 | 65.6 | 52.0 | 9.30 |
| **170** | 98.4 | 176.2 | 65.6 | 52.0 | 9.30 |
| **171** | 98.4 | 176.2 | 65.6 | 52.0 | 9.30 |
| **172** | 98.4 | 176.2 | 65.6 | 53.0 | 9.30 |
| **173** | 102.4 | 175.6 | 66.5 | 54.9 | 8.70 |
| **174** | 102.4 | 175.6 | 66.5 | 54.9 | 22.50 |
| **175** | 102.4 | 175.6 | 66.5 | 53.9 | 8.70 |
| **176** | 102.4 | 175.6 | 66.5 | 54.9 | 8.70 |
| **177** | 102.4 | 175.6 | 66.5 | 53.9 | 8.70 |
| **178** | 102.9 | 183.5 | 67.7 | 52.0 | 9.30 |
| **179** | 102.9 | 183.5 | 67.7 | 52.0 | 9.30 |
| **180** | 104.5 | 187.8 | 66.5 | 54.1 | 9.20 |
| **181** | 104.5 | 187.8 | 66.5 | 54.1 | 9.20 |
| **182** | 97.3 | 171.7 | 65.5 | 55.7 | 23.00 |
| **183** | 97.3 | 171.7 | 65.5 | 55.7 | 9.00 |
| **184** | 97.3 | 171.7 | 65.5 | 55.7 | 23.00 |
| **185** | 97.3 | 171.7 | 65.5 | 55.7 | 9.00 |
| **186** | 97.3 | 171.7 | 65.5 | 55.7 | 9.00 |
| **187** | 97.3 | 171.7 | 65.5 | 55.7 | 23.00 |
| **188** | 97.3 | 171.7 | 65.5 | 55.7 | 10.00 |
| **189** | 94.5 | 159.3 | 64.2 | 55.6 | 8.50 |
| **190** | 94.5 | 165.7 | 64.0 | 51.4 | 8.50 |
| **191** | 100.4 | 180.2 | 66.9 | 55.1 | 8.50 |
| **192** | 100.4 | 180.2 | 66.9 | 55.1 | 23.00 |

| | wheel-base | length | width | height | compression-ratio |
|---|---|---|---|---|---|
| **193** | 100.4 | 183.1 | 66.9 | 55.1 | 9.00 |
| **194** | 104.3 | 188.8 | 67.2 | 56.2 | 9.50 |
| **195** | 104.3 | 188.8 | 67.2 | 57.5 | 9.50 |
| **196** | 104.3 | 188.8 | 67.2 | 56.2 | 9.50 |
| **197** | 104.3 | 188.8 | 67.2 | 57.5 | 9.50 |
| **198** | 104.3 | 188.8 | 67.2 | 56.2 | 7.50 |
| **199** | 104.3 | 188.8 | 67.2 | 57.5 | 7.50 |
| **200** | 109.1 | 188.8 | 68.9 | 55.5 | 9.50 |
| **201** | 109.1 | 188.8 | 68.8 | 55.5 | 8.70 |
| **202** | 109.1 | 188.8 | 68.9 | 55.5 | 8.80 |
| **203** | 109.1 | 188.8 | 68.9 | 55.5 | 23.00 |
| **204** | 109.1 | 188.8 | 68.9 | 55.5 | 9.50 |

In [14]:

```
cat_var = df.select_dtypes(include=["object"])
```

In [15]:

```
cat_var.columns
```

Out[15]:

```
Index(['normalized-losses', 'make', 'fuel-type', 'aspiration', 'num-of-doo
rs',
       'body-style', 'drive-wheels', 'engine-location', 'engine-type',
       'num-of-cylinders', 'fuel-system', 'bore', 'stroke', 'horsepower',
       'peak-rpm', 'price'],
      dtype='object')
```

In [16]:

```
convert_type_float ={'normalized-losses':float,'bore':float, 'stroke':float, 'horsepowe
r':float,'peak-rpm':float, 'price':float}
```

In [17]:

```
d = df.astype(convert_type_float)
```

In [18]:

```
d.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   symboling          205 non-null    int64  
 1   normalized-losses  164 non-null    float64
 2   make               205 non-null    object 
 3   fuel-type          205 non-null    object 
 4   aspiration         205 non-null    object 
 5   num-of-doors       203 non-null    object 
 6   body-style         205 non-null    object 
 7   drive-wheels       205 non-null    object 
 8   engine-location    205 non-null    object 
 9   wheel-base         205 non-null    float64
 10  length             205 non-null    float64
 11  width              205 non-null    float64
 12  height             205 non-null    float64
 13  curb-weight        205 non-null    int64  
 14  engine-type        205 non-null    object 
 15  num-of-cylinders   205 non-null    object 
 16  engine-size        205 non-null    int64  
 17  fuel-system        205 non-null    object 
 18  bore               201 non-null    float64
 19  stroke             201 non-null    float64
 20  compression-ratio  205 non-null    float64
 21  horsepower         203 non-null    float64
 22  peak-rpm           203 non-null    float64
 23  city-mpg           205 non-null    int64  
 24  highway-mpg        205 non-null    int64  
 25  price              201 non-null    float64
dtypes: float64(11), int64(5), object(10)
memory usage: 41.8+ KB
```

# Replacing null value with mean for integers and mode for objects

In [19]:

```
num_var1 = d.select_dtypes(include=["int","float"])
```

In [20]:

```
cat_var1= d.select_dtypes(include=["object"])
```

In [21]:

```
num_var1.isnull().sum()
```

Out[21]:

```
normalized-losses    41
wheel-base            0
length                0
width                 0
height                0
bore                  4
stroke                4
compression-ratio     0
horsepower            2
peak-rpm              2
price                 4
dtype: int64
```

In [22]:

```
cat_var1.isnull().sum()
```

Out[22]:

```
make               0
fuel-type          0
aspiration         0
num-of-doors       2
body-style         0
drive-wheels       0
engine-location    0
engine-type        0
num-of-cylinders   0
fuel-system        0
dtype: int64
```

In [23]:

```
num_var1.fillna(num_var1.mean()[0],inplace= True)
num_var1.isnull().sum()
```

```
C:\Users\shubh\anaconda3\lib\site-packages\pandas\core\frame.py:4153: Sett
ingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  downcast=downcast,
```

Out[23]:

```
normalized-losses    0
wheel-base           0
length               0
width                0
height               0
bore                 0
stroke               0
compression-ratio    0
horsepower           0
peak-rpm             0
price                0
dtype: int64
```

In [24]:

```
cat_var1['num-of-doors'].fillna(cat_var1['num-of-doors'].mode()[0],inplace= True)
cat_var1.isnull().sum()
```

```
C:\Users\shubh\anaconda3\lib\site-packages\pandas\core\generic.py:6245: Se
ttingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  self._update_inplace(new_data)
```

Out[24]:

```
make               0
fuel-type          0
aspiration         0
num-of-doors       0
body-style         0
drive-wheels       0
engine-location    0
engine-type        0
num-of-cylinders   0
fuel-system        0
dtype: int64
```

In [25]:

```
d.update(num_var1)
```

In [26]:

```
d.update(cat_var1)
```

In [27]:

```
d.isnull().sum().sum()
```

Out[27]:

```
0
```

In [28]:

```
d.head()
```

Out[28]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | 122.0 | alfa-romero | gas | std | two | convertible | rwd | front | |
| **1** | 3 | 122.0 | alfa-romero | gas | std | two | convertible | rwd | front | |
| **2** | 1 | 122.0 | alfa-romero | gas | std | two | hatchback | rwd | front | |
| **3** | 2 | 164.0 | audi | gas | std | four | sedan | fwd | front | |
| **4** | 2 | 164.0 | audi | gas | std | four | sedan | 4wd | front | |

# converting number to interger

In [29]:

```
d["num-of-cylinders"].value_counts()
```

Out[29]:

```
four      159
six        24
five       11
eight       5
two         4
three       1
twelve      1
Name: num-of-cylinders, dtype: int64
```

In [30]:

```
word_number ={"four":4,"six":6,"five":5,"eight":8,"two":2,"three":3,"twelve":12}
```

In [31]:

```
d.replace({"num-of-cylinders":word_number},inplace=True)
```

In [32]:

```
d.replace({"num-of-doors":word_number},inplace=True)
```

In [33]:

```
d.head()
```

Out[33]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **1** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **2** | 1 | 122.0 | alfa-romero | gas | std | 2 | hatchback | rwd | front | |
| **3** | 2 | 164.0 | audi | gas | std | 4 | sedan | fwd | front | |
| **4** | 2 | 164.0 | audi | gas | std | 4 | sedan | 4wd | front | |

# Bining

In [34]:

```
bins = np.linspace(min(d["horsepower"]),max(d["horsepower"]),4)
```

In [35]:

```
group_names = ["Low","Medium","High"]
```

In [36]:

```
d["horsepower_binned"] = pd.cut(d['horsepower'],bins,labels=group_names,include_lowest=True)
```

In [37]:

```
d.head()
```

Out[37]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **1** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **2** | 1 | 122.0 | alfa-romero | gas | std | 2 | hatchback | rwd | front | |
| **3** | 2 | 164.0 | audi | gas | std | 4 | sedan | fwd | front | |
| **4** | 2 | 164.0 | audi | gas | std | 4 | sedan | 4wd | front | |

# Standardisation of columns "highway mpg","citympg" by converting it from L/100km = 235 / mpg

In [38]:

```
d["city/235mpg"]=235/d["city-mpg"]
```

In [39]:

```
d["highway/235mpg"]=235/d["highway-mpg"]
```

In [40]:

```
d.head()
```

Out[40]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **1** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **2** | 1 | 122.0 | alfa-romero | gas | std | 2 | hatchback | rwd | front | |
| **3** | 2 | 164.0 | audi | gas | std | 4 | sedan | fwd | front | |
| **4** | 2 | 164.0 | audi | gas | std | 4 | sedan | 4wd | front | |

# Normalization of columns"height","weigth","length"

In [41]:

```
d["length"] = d["length"]/d["length"].max()
```

In [42]:

```
d["width"] = d["width"]/d["width"].max()
```

In [43]:

```
d["height"] = d["height"]/d["height"].max()
```

In [44]:

```
d.head()
```

Out[44]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **1** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **2** | 1 | 122.0 | alfa-romero | gas | std | 2 | hatchback | rwd | front | |
| **3** | 2 | 164.0 | audi | gas | std | 4 | sedan | fwd | front | |
| **4** | 2 | 164.0 | audi | gas | std | 4 | sedan | 4wd | front | |

In [45]:

```
d["drive-wheels"].value_counts()
```

Out[45]:

```
fwd    120
rwd     76
4wd      9
Name: drive-wheels, dtype: int64
```

In [46]:

```
fuel_type_dummies = pd.get_dummies(d["fuel-type"])
fuel_type_dummies
```

Out[46]:

| | diesel | gas |
|---|---|---|
| **0** | 0 | 1 |
| **1** | 0 | 1 |
| **2** | 0 | 1 |
| **3** | 0 | 1 |
| **4** | 0 | 1 |
| **5** | 0 | 1 |
| **6** | 0 | 1 |
| **7** | 0 | 1 |
| **8** | 0 | 1 |
| **9** | 0 | 1 |
| **10** | 0 | 1 |
| **11** | 0 | 1 |
| **12** | 0 | 1 |
| **13** | 0 | 1 |
| **14** | 0 | 1 |
| **15** | 0 | 1 |
| **16** | 0 | 1 |
| **17** | 0 | 1 |
| **18** | 0 | 1 |
| **19** | 0 | 1 |
| **20** | 0 | 1 |
| **21** | 0 | 1 |
| **22** | 0 | 1 |
| **23** | 0 | 1 |
| **24** | 0 | 1 |
| **25** | 0 | 1 |
| **26** | 0 | 1 |
| **27** | 0 | 1 |
| **28** | 0 | 1 |
| **29** | 0 | 1 |
| **30** | 0 | 1 |
| **31** | 0 | 1 |
| **32** | 0 | 1 |
| **33** | 0 | 1 |
| **34** | 0 | 1 |
| **35** | 0 | 1 |
| **36** | 0 | 1 |

| | diesel | gas |
|---|---|---|
| **37** | 0 | 1 |
| **38** | 0 | 1 |
| **39** | 0 | 1 |
| **40** | 0 | 1 |
| **41** | 0 | 1 |
| **42** | 0 | 1 |
| **43** | 0 | 1 |
| **44** | 0 | 1 |
| **45** | 0 | 1 |
| **46** | 0 | 1 |
| **47** | 0 | 1 |
| **48** | 0 | 1 |
| **49** | 0 | 1 |
| **50** | 0 | 1 |
| **51** | 0 | 1 |
| **52** | 0 | 1 |
| **53** | 0 | 1 |
| **54** | 0 | 1 |
| **55** | 0 | 1 |
| **56** | 0 | 1 |
| **57** | 0 | 1 |
| **58** | 0 | 1 |
| **59** | 0 | 1 |
| **60** | 0 | 1 |
| **61** | 0 | 1 |
| **62** | 0 | 1 |
| **63** | 1 | 0 |
| **64** | 0 | 1 |
| **65** | 0 | 1 |
| **66** | 1 | 0 |
| **67** | 1 | 0 |
| **68** | 1 | 0 |
| **69** | 1 | 0 |
| **70** | 1 | 0 |
| **71** | 0 | 1 |
| **72** | 0 | 1 |
| **73** | 0 | 1 |
| **74** | 0 | 1 |
| **75** | 0 | 1 |

| | diesel | gas |
|---|---|---|
| **76** | 0 | 1 |
| **77** | 0 | 1 |
| **78** | 0 | 1 |
| **79** | 0 | 1 |
| **80** | 0 | 1 |
| **81** | 0 | 1 |
| **82** | 0 | 1 |
| **83** | 0 | 1 |
| **84** | 0 | 1 |
| **85** | 0 | 1 |
| **86** | 0 | 1 |
| **87** | 0 | 1 |
| **88** | 0 | 1 |
| **89** | 0 | 1 |
| **90** | 1 | 0 |
| **91** | 0 | 1 |
| **92** | 0 | 1 |
| **93** | 0 | 1 |
| **94** | 0 | 1 |
| **95** | 0 | 1 |
| **96** | 0 | 1 |
| **97** | 0 | 1 |
| **98** | 0 | 1 |
| **99** | 0 | 1 |
| **100** | 0 | 1 |
| **101** | 0 | 1 |
| **102** | 0 | 1 |
| **103** | 0 | 1 |
| **104** | 0 | 1 |
| **105** | 0 | 1 |
| **106** | 0 | 1 |
| **107** | 0 | 1 |
| **108** | 1 | 0 |
| **109** | 0 | 1 |
| **110** | 1 | 0 |
| **111** | 0 | 1 |
| **112** | 1 | 0 |
| **113** | 0 | 1 |
| **114** | 1 | 0 |

| | diesel | gas |
|---|---|---|
| **115** | 0 | 1 |
| **116** | 1 | 0 |
| **117** | 0 | 1 |
| **118** | 0 | 1 |
| **119** | 0 | 1 |
| **120** | 0 | 1 |
| **121** | 0 | 1 |
| **122** | 0 | 1 |
| **123** | 0 | 1 |
| **124** | 0 | 1 |
| **125** | 0 | 1 |
| **126** | 0 | 1 |
| **127** | 0 | 1 |
| **128** | 0 | 1 |
| **129** | 0 | 1 |
| **130** | 0 | 1 |
| **131** | 0 | 1 |
| **132** | 0 | 1 |
| **133** | 0 | 1 |
| **134** | 0 | 1 |
| **135** | 0 | 1 |
| **136** | 0 | 1 |
| **137** | 0 | 1 |
| **138** | 0 | 1 |
| **139** | 0 | 1 |
| **140** | 0 | 1 |
| **141** | 0 | 1 |
| **142** | 0 | 1 |
| **143** | 0 | 1 |
| **144** | 0 | 1 |
| **145** | 0 | 1 |
| **146** | 0 | 1 |
| **147** | 0 | 1 |
| **148** | 0 | 1 |
| **149** | 0 | 1 |
| **150** | 0 | 1 |
| **151** | 0 | 1 |
| **152** | 0 | 1 |
| **153** | 0 | 1 |

| | diesel | gas |
|---|---|---|
| **154** | 0 | 1 |
| **155** | 0 | 1 |
| **156** | 0 | 1 |
| **157** | 0 | 1 |
| **158** | 1 | 0 |
| **159** | 1 | 0 |
| **160** | 0 | 1 |
| **161** | 0 | 1 |
| **162** | 0 | 1 |
| **163** | 0 | 1 |
| **164** | 0 | 1 |
| **165** | 0 | 1 |
| **166** | 0 | 1 |
| **167** | 0 | 1 |
| **168** | 0 | 1 |
| **169** | 0 | 1 |
| **170** | 0 | 1 |
| **171** | 0 | 1 |
| **172** | 0 | 1 |
| **173** | 0 | 1 |
| **174** | 1 | 0 |
| **175** | 0 | 1 |
| **176** | 0 | 1 |
| **177** | 0 | 1 |
| **178** | 0 | 1 |
| **179** | 0 | 1 |
| **180** | 0 | 1 |
| **181** | 0 | 1 |
| **182** | 1 | 0 |
| **183** | 0 | 1 |
| **184** | 1 | 0 |
| **185** | 0 | 1 |
| **186** | 0 | 1 |
| **187** | 1 | 0 |
| **188** | 0 | 1 |
| **189** | 0 | 1 |
| **190** | 0 | 1 |
| **191** | 0 | 1 |
| **192** | 1 | 0 |

| | diesel | gas |
|---|---|---|
| **193** | 0 | 1 |
| **194** | 0 | 1 |
| **195** | 0 | 1 |
| **196** | 0 | 1 |
| **197** | 0 | 1 |
| **198** | 0 | 1 |
| **199** | 0 | 1 |
| **200** | 0 | 1 |
| **201** | 0 | 1 |
| **202** | 0 | 1 |
| **203** | 1 | 0 |
| **204** | 0 | 1 |

In [47]:

```
drive_wheels_dummies = pd.get_dummies(d["drive-wheels"])
```

In [48]:

```
engine_type_dummies = pd.get_dummies(d["engine-type"])
```

In [49]:

```
fuel_system_dummies = pd.get_dummies(d["fuel-system"])
```

In [50]:

```
horsepower_binned_dummies = pd.get_dummies(["horsepower_binned"])
horsepower_binned_dummies
```

Out[50]:

| | horsepower_binned |
|---|---|
| **0** | 1 |

In [51]:

```
d2 = pd.concat([d,fuel_type_dummies,drive_wheels_dummies,engine_type_dummies,fuel_syste
m_dummies,horsepower_binned_dummies],axis=1)
```

In [52]:

```
d2.head()
```

Out[52]:

| | symboling | normalized-losses | make | fuel-type | aspiration | num-of-doors | body-style | drive-wheels | engine-location | wh b |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **1** | 3 | 122.0 | alfa-romero | gas | std | 2 | convertible | rwd | front | |
| **2** | 1 | 122.0 | alfa-romero | gas | std | 2 | hatchback | rwd | front | |
| **3** | 2 | 164.0 | audi | gas | std | 4 | sedan | fwd | front | |
| **4** | 2 | 164.0 | audi | gas | std | 4 | sedan | 4wd | front | |

In [53]:

```
d2.drop(columns=["fuel-type","drive-wheels","engine-type","fuel-system","horsepower_binned"],axis=1,inplace= True)
```

In [54]:

```
d2.head()
```

Out[54]:

| | symboling | normalized-losses | make | aspiration | num-of-doors | body-style | engine-location | wheel-base | length |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 3 | 122.0 | alfa-romero | std | 2 | convertible | front | 88.6 | 0.811148 |
| **1** | 3 | 122.0 | alfa-romero | std | 2 | convertible | front | 88.6 | 0.811148 |
| **2** | 1 | 122.0 | alfa-romero | std | 2 | hatchback | front | 94.5 | 0.822681 |
| **3** | 2 | 164.0 | audi | std | 4 | sedan | front | 99.8 | 0.848630 |
| **4** | 2 | 164.0 | audi | std | 4 | sedan | front | 99.4 | 0.848630 |

In [ ]:

```

```

In [ ]:

```

```