PRE-PROCESS THE DATA

1.Import Required Libraries

In [30]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
```

2.Read the Datasets

In [8]:
```
data=pd.read_csv(r"C:\Users\admin\OneDrive\Desktop\IBM DATASET\car data.csv")
```

In [9]:
```
data.head()
```

Out[9]:

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	Petrol	Dealer	Manual	0
1	sx4	2013	4.75	9.54	43000	Diesel	Dealer	Manual	0
2	ciaz	2017	7.25	9.85	6900	Petrol	Dealer	Manual	0
3	wagon r	2011	2.85	4.15	5200	Petrol	Dealer	Manual	0
4	swift	2014	4.60	6.87	42450	Diesel	Dealer	Manual	0

In [10]:
```
data.info
```

Out[10]:
```
<bound method DataFrame.info of     Car_Name  Year  Selling_Price  Present_Price  Kms_Driven Fuel_Type  \
0       ritz  2014           3.35           5.59       27000    Petrol
1        sx4  2013           4.75           9.54       43000    Diesel
2       ciaz  2017           7.25           9.85        6900    Petrol
3    wagon r  2011           2.85           4.15        5200    Petrol
4      swift  2014           4.60           6.87       42450    Diesel
..       ...   ...            ...            ...         ...       ...
296     city  2016           9.50          11.60       33988    Diesel
297     brio  2015           4.00           5.90       60000    Petrol
298     city  2009           3.35          11.00       87934    Petrol
299     city  2017          11.50          12.50        9000    Diesel
300     brio  2016           5.30           5.90        5464    Petrol

    Seller_Type Transmission  Owner
0        Dealer       Manual      0
1        Dealer       Manual      0
2        Dealer       Manual      0
3        Dealer       Manual      0
4        Dealer       Manual      0
..          ...          ...    ...
296      Dealer       Manual      0
297      Dealer       Manual      0
298      Dealer       Manual      0
299      Dealer       Manual      0
300      Dealer       Manual      0

[301 rows x 9 columns]>
```

In [6]:
```
data.shape
```

Out[6]:
```
(301, 9)
```

In [8]:
```
data.isnull().sum()
```

Out[8]:
```
Car_Name         0
Year             0
Selling_Price    0
Present_Price    0
Kms_Driven       0
Fuel_Type        0
Seller_Type      0
Transmission     0
Owner            0
dtype: int64
```

3.Cleaning the Dataset

Encoding the categorial values

In [15]:
```
data.replace({'Fuel_Type':{'Petrol':0,'Diesel':1,'CNG':2}},inplace=True)
data.replace({'Seller_Type':{'Dealer':0,'Individual':1}},inplace=True)
data.replace({'Transmission':{'Manual':0,'Automatic':1}},inplace=True)
```

In [16]:
```
data.head()
```

Out[16]:

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	ritz	2014	3.35	5.59	27000	0	0	0	0
1	sx4	2013	4.75	9.54	43000	1	0	0	0
2	ciaz	2017	7.25	9.85	6900	0	0	0	0
3	wagon r	2011	2.85	4.15	5200	0	0	0	0
4	swift	2014	4.60	6.87	42450	1	0	0	0

4.Splitting Data into Independent And Dependent Variable

In [18]:
```
x=data.drop(['Car_Name','Selling_Price'],axis=1)
x
```

Out[18]:

	Year	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
0	2014	5.59	27000	0	0	0	0
1	2013	9.54	43000	1	0	0	0
2	2017	9.85	6900	0	0	0	0
3	2011	4.15	5200	0	0	0	0
4	2014	6.87	42450	1	0	0	0
...	...	...	...	...	...	...	...
296	2016	11.60	33988	1	0	0	0
297	2015	5.90	60000	0	0	0	0
298	2009	11.00	87934	0	0	0	0
299	2017	12.50	9000	1	0	0	0
300	2016	5.90	5464	0	0	0	0

301 rows × 7 columns

In [20]:
```
y=data['Selling_Price']
y
```

Out[20]:
```
0       3.35
1       4.75
2       7.25
3       2.85
4       4.60
       ...
296     9.50
297     4.00
298     3.35
299    11.50
300     5.30
Name: Selling_Price, Length: 301, dtype: float64
```

In [23]:
```
X_Train, X_Test, Y_Train, Y_Test = train_test_split(x, y, test_size=0.3, random_state=0)
```

In [24]:
```
X_Train.shape,X_Test.shape
```

Out[24]:
```
((210, 7), (91, 7))
```

In [25]:
```
Y_Train.shape,Y_Test.shape
```

Out[25]:
```
((210,), (91,))
```

In [26]:
```
X_Train
```

Out[26]:

	Year	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
222	2014	7.60	77632	1	0	0	0
283	2016	11.80	9010	0	0	0	0
44	2012	2.69	50000	0	0	0	0
245	2012	9.40	71000	1	0	0	0
191	2012	0.57	25000	0	1	0	1
...	...	...	...	...	...	...	...
251	2013	9.90	56701	0	0	0	0
192	2007	0.75	49000	0	1	0	1
117	2015	1.90	14000	0	1	0	0
47	2006	4.15	65000	0	0	0	0
172	2014	0.64	13700	0	1	0	0

210 rows × 7 columns

In [27]:
```
X_Test
```

Out[27]:

	Year	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission	Owner
223	2015	9.400	61381	1	0	0	0
150	2011	0.826	6000	0	1	0	0
226	2015	5.700	24678	0	0	0	0
296	2016	11.600	33988	1	0	0	0
52	2017	19.770	15000	1	0	1	0
...	...	...	...	...	...	...	...
240	2012	9.400	32322	1	0	0	0
76	2013	14.680	72000	0	0	0	0
145	2012	0.810	19000	0	1	0	0
300	2016	5.900	5464	0	0	0	0
135	2015	0.740	5000	0	1	0	0

91 rows × 7 columns

In [28]:
```
Y_Train
```

Out[28]:
```
222    6.00
283    8.99
44     1.25
245    5.20
191    0.20
       ...
251    5.00
192    0.20
117    1.10
47     1.05
172    0.40
Name: Selling_Price, Length: 210, dtype: float64
```

In [29]:
```
Y_Test
```

Out[29]:
```
223     8.25
150     0.50
226     5.25
296     9.50
52     18.00
       ...
240     5.35
76      5.50
145     0.60
300     5.30
135     0.65
Name: Selling_Price, Length: 91, dtype: float64
```