```
import pandas as pd
import numpy as np
df=pd.read_csv('/content/train-data.csv')
df
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Eng
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968
...	...	...	...	...	...	...	...	...	...	
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	24.4 kmpl	1120
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Second	14.0 kmpl	2498
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual	First	18.9 kmpl	998
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual	First	25.44 kmpl	936

6019 rows × 14 columns

```
df.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC

```
df.tail()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248 CC	7
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	24.4 kmpl	1120 CC	7
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	11
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual	First	18.9 kmpl	998 CC	67.
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual	First	25.44 kmpl	936 CC	57.

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6019 entries, 0 to 6018
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Unnamed: 0         6019 non-null   int64  
 1   Name               6019 non-null   object 
 2   Location           6019 non-null   object 
 3   Year               6019 non-null   int64  
 4   Kilometers_Driven  6019 non-null   int64  
 5   Fuel_Type          6019 non-null   object 
 6   Transmission       6019 non-null   object 
 7   Owner_Type         6019 non-null   object 
 8   Mileage            6017 non-null   object 
 9   Engine             5983 non-null   object 
 10  Power              5983 non-null   object 
 11  Seats              5977 non-null   float64
 12  New_Price          824 non-null    object 
 13  Price              6019 non-null   float64
dtypes: float64(2), int64(3), object(9)
memory usage: 658.5+ KB
```

```
df.columns
```

```
Index(['Unnamed: 0', 'Name', 'Location', 'Year', 'Kilometers_Driven',
       'Fuel_Type', 'Transmission', 'Owner_Type', 'Mileage', 'Engine', 'Power',
       'Seats', 'New_Price', 'Price'],
      dtype='object')
```

```
df.shape
```

```
(6019, 14)
```

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

```
Unnamed: 0            0
Name                  0
Location              0
Year                  0
Kilometers_Driven     0
Fuel_Type             0
Transmission          0
Owner_Type            0
Mileage               2
Engine               36
Power                36
Seats                42
New_Price          5195
Price                 0
dtype: int64
```

```
df.dtypes
```

```
Unnamed: 0           int64
Name                object
Location            object
Year                 int64
Kilometers_Driven    int64
```

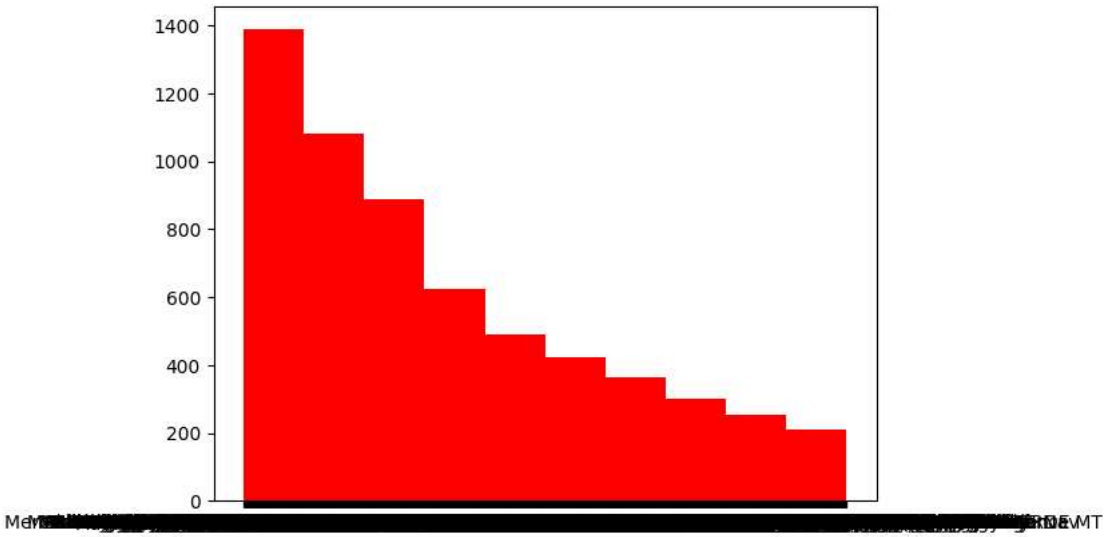
```
Fuel_Type            object
Transmission         object
Owner_Type           object
Mileage              object
Engine               object
Power                object
Seats               float64
New_Price            object
Price               float64
dtype: object
```

```
df['Name'].value_counts()
```

```
Mahindra XUV500 W8 2WD          49
Maruti Swift VDI                45
Honda City 1.5 S MT             34
Maruti Swift Dzire VDI          34
Maruti Swift VDI BSIV           31
                                ..
Ford Fiesta Titanium 1.5 TDCi    1
Mahindra Scorpio S10 AT 4WD      1
Hyundai i20 1.2 Era              1
Toyota Camry W4 (AT)             1
Mahindra Xylo D4 BSIV            1
Name: Name, Length: 1878, dtype: int64
```

```
import matplotlib.pyplot as plt
plt.hist(df['Name'],color="red")
```

```
(array([1388., 1080.,  889.,  623.,  490.,  424.,  363.,  300.,  252.,
         210.]),
 array([   0. ,  187.7,  375.4,  563.1,  750.8,  938.5, 1126.2, 1313.9,
        1501.6, 1689.3, 1877. ]),
 <BarContainer object of 10 artists>)
```




```
df['Location'].value_counts()
```

```
Mumbai        790
Hyderabad     742
Kochi         651
Coimbatore    636
Pune          622
Delhi         554
Kolkata       535
Chennai       494
Jaipur        413
Bangalore     358
Ahmedabad     224
Name: Location, dtype: int64
```

```
plt.hist(df['Location'])
```

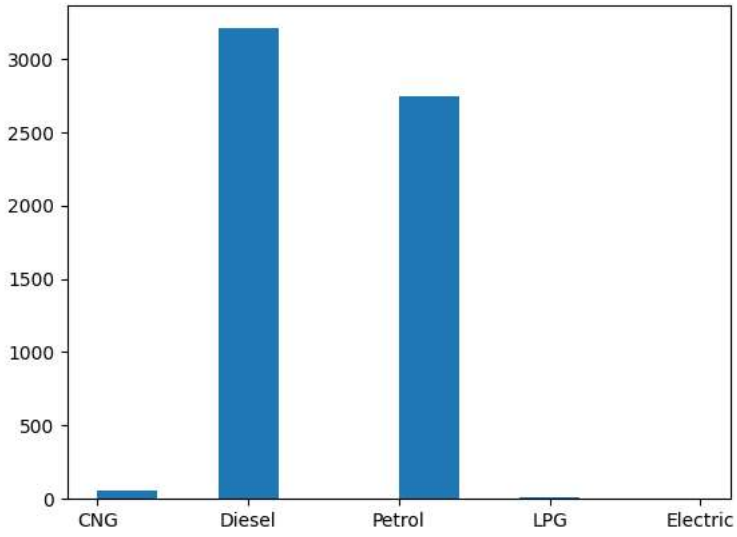
```
(array([790., 622., 494., 636., 742., 413., 651., 535., 554., 582.]),
 array([ 0.,  1.,  2.,  3.,  4.,  5.,  6.,  7.,  8.,  9., 10.]),
 <BarContainer object of 10 artists>)
```


```
df['Fuel_Type'].value_counts()
```

```
Diesel      3205
Petrol      2746
CNG           56
LPG           10
Electric       2
Name: Fuel_Type, dtype: int64
```

```
plt.hist(df['Fuel_Type'])
```

```
(array([5.600e+01, 0.000e+00, 3.205e+03, 0.000e+00, 0.000e+00, 2.746e+03,
        0.000e+00, 1.000e+01, 0.000e+00, 2.000e+00]),
 array([0. , 0.4, 0.8, 1.2, 1.6, 2. , 2.4, 2.8, 3.2, 3.6, 4. ]),
 <BarContainer object of 10 artists>)
```




```
df['Transmission'].value_counts()
```

```
Manual       4299
Automatic    1720
Name: Transmission, dtype: int64
```

```
plt.hist(df['Transmission'])
```

```
(array([4299.,    0.,    0.,    0.,    0.,    0.,    0.,    0.,    0.,
        1720.]),
 array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <BarContainer object of 10 artists>)
```

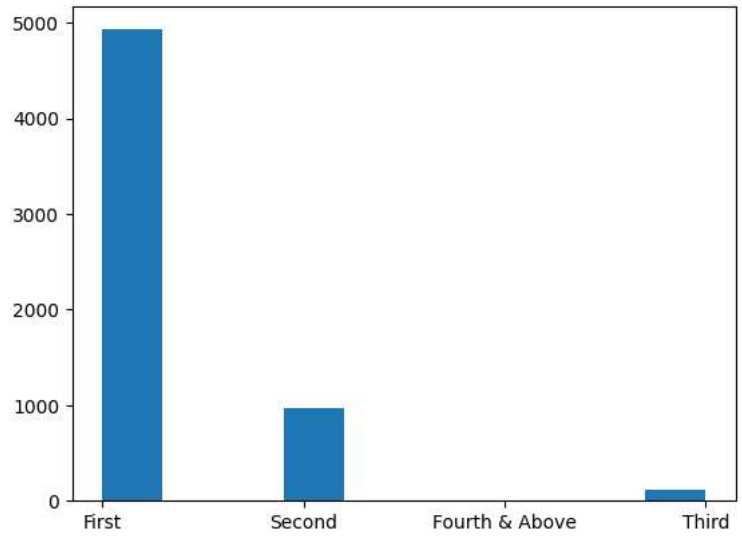

```
df['Owner_Type'].value_counts()
```

```
First             4929
Second             968
Third              113
Fourth & Above       9
Name: Owner_Type, dtype: int64
```

Manual Automatic

```
plt.hist(df['Owner_Type'])
```

```
(array([4929.,    0.,    0.,  968.,    0.,    0.,    9.,    0.,    0.,
         113.]),
 array([0. , 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, 2.4, 2.7, 3. ]),
 <BarContainer object of 10 artists>)
```




```
#drop==>unnamed:0,name,new_price
#encoding==>Location,Fuel_Type,Transmission,Owner_Type


df1=pd.get_dummies(df[['Location','Fuel_Type','Transmission','Owner_Type']],drop_first=True)
df1
```

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Hyderabad	Location_Jaipur	Location_Kochi	Lo
0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	
2	0	1	0	0	0	0	0	
3	0	1	0	0	0	0	0	
4	0	0	1	0	0	0	0	

```
#concat==>to compaign 2 data frames===df & df1
dfe=pd.concat([df,df1],axis=1)
dfe
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	...	Location_Mumbai
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	...	1
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	...	0
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	...	0
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	...	0
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248 CC	...	0
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	24.4 kmpl	1120 CC	...	0
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	...	0
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual	First	18.9 kmpl	998 CC	...	0
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual	First	25.44 kmpl	936 CC	...	0

6019 rows × 32 columns

```
dfe.columns
```

```
Index(['Unnamed: 0', 'Name', 'Location', 'Year', 'Kilometers_Driven',
       'Fuel_Type', 'Transmission', 'Owner_Type', 'Mileage', 'Engine', 'Power',
       'Seats', 'New_Price', 'Price', 'Location_Bangalore', 'Location_Chennai',
       'Location_Coimbatore', 'Location_Delhi', 'Location_Hyderabad',
       'Location_Jaipur', 'Location_Kochi', 'Location_Kolkata',
       'Location_Mumbai', 'Location_Pune', 'Fuel_Type_Diesel',
       'Fuel_Type_Electric', 'Fuel_Type_LPG', 'Fuel_Type_Petrol',
       'Transmission_Manual', 'Owner_Type_Fourth & Above', 'Owner_Type_Second',
       'Owner_Type_Third'],
      dtype='object')
```

```
dfe.drop(["Unnamed: 0","Name","Location","Fuel_Type","Transmission","Owner_Type","New_Price"],axis=1,inplace=True)
dfe
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Location_Bangalore	Location_Chennai	Location_Coimbatore	...	Loca
0	2010	72000	26.6 km/kg	998 CC	58.16 bhp	5.0	1.75	0	0	0	...	
1	2015	41000	19.67 kmpl	1582 CC	126.2 bhp	5.0	12.50	0	0	0	...	
2	2011	46000	18.2 kmpl	1199 CC	88.7 bhp	5.0	4.50	0	1	0	...	
3	2012	87000	20.77 kmpl	1248 CC	88.76 bhp	7.0	6.00	0	1	0	...	
4	2013	40670	15.2 kmpl	1968 CC	140.8 bhp	5.0	17.74	0	0	1	...	
...	...	...	...	...	...	...	...	...	...	...	...	
6014	2014	27365	28.4 kmpl	1248 CC	74 bhp	5.0	4.75	0	0	0	...	
6015	2015	100000	24.4 kmpl	1120 CC	71 bhp	5.0	4.00	0	0	0	...	
6016	2012	55000	14.0 kmpl	2498 CC	112 bhp	8.0	2.90	0	0	0	...	
6017	2013	46000	18.9 kmpl	998 CC	67.1 bhp	5.0	2.65	0	0	0	...	
6018	2011	47000	25.44 kmpl	936 CC	57.6 bhp	5.0	2.50	0	0	0	...	

```
#string replace

dfe['Mileage']=dfe['Mileage'].str.replace('km/kg','')
dfe['Mileage']=dfe['Mileage'].str.replace('kmpl','')
dfe['Engine']=dfe['Engine'].str.replace('CC','')
dfe['Power']=dfe['Power'].str.replace('bhp','')
#null
dfe['Mileage']=dfe['Mileage'].str.replace('null','0')
dfe['Engine']=dfe['Engine'].str.replace('null','0')
dfe['Power']=dfe['Power'].str.replace('null','0')
dfe
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Location_Bangalore	Location_Chennai	Location_Coimbatore	...	Loca
0	2010	72000	26.6	998	58.16	5.0	1.75	0	0	0	...	
1	2015	41000	19.67	1582	126.2	5.0	12.50	0	0	0	...	
2	2011	46000	18.2	1199	88.7	5.0	4.50	0	1	0	...	
3	2012	87000	20.77	1248	88.76	7.0	6.00	0	1	0	...	
4	2013	40670	15.2	1968	140.8	5.0	17.74	0	0	1	...	
...	...	...	...	...	...	...	...	...	...	...	...	
6014	2014	27365	28.4	1248	74	5.0	4.75	0	0	0	...	
6015	2015	100000	24.4	1120	71	5.0	4.00	0	0	0	...	
6016	2012	55000	14.0	2498	112	8.0	2.90	0	0	0	...	
6017	2013	46000	18.9	998	67.1	5.0	2.65	0	0	0	...	
6018	2011	47000	25.44	936	57.6	5.0	2.50	0	0	0	...	

6019 rows × 25 columns

```
dfe.dtypes
```

```
Year                    int64
Kilometers_Driven       int64
Mileage                object
Engine                 object
Power                  object
Seats                 float64
Price                 float64
Location_Bangalore      uint8
```

```
Location_Chennai              uint8
Location_Coimbatore           uint8
Location_Delhi                uint8
Location_Hyderabad            uint8
Location_Jaipur               uint8
Location_Kochi                uint8
Location_Kolkata              uint8
Location_Mumbai               uint8
Location_Pune                 uint8
Fuel_Type_Diesel              uint8
Fuel_Type_Electric            uint8
Fuel_Type_LPG                 uint8
Fuel_Type_Petrol              uint8
Transmission_Manual           uint8
Owner_Type_Fourth & Above     uint8
Owner_Type_Second             uint8
Owner_Type_Third              uint8
dtype: object
```

```
#convert into float
dfe['Engine']=dfe['Engine'].astype(float)
dfe['Power']=dfe['Power'].astype(float)
dfe['Mileage']=dfe['Mileage'].astype(float)
dfe.dtypes
```

```
Year                           int64
Kilometers_Driven              int64
Mileage                      float64
Engine                       float64
Power                        float64
Seats                        float64
Price                        float64
Location_Bangalore             uint8
Location_Chennai               uint8
Location_Coimbatore            uint8
Location_Delhi                 uint8
Location_Hyderabad             uint8
Location_Jaipur                uint8
Location_Kochi                 uint8
Location_Kolkata               uint8
Location_Mumbai                uint8
Location_Pune                  uint8
Fuel_Type_Diesel               uint8
Fuel_Type_Electric             uint8
Fuel_Type_LPG                  uint8
Fuel_Type_Petrol               uint8
Transmission_Manual            uint8
Owner_Type_Fourth & Above      uint8
Owner_Type_Second              uint8
Owner_Type_Third               uint8
dtype: object
```

```
dfe.isna().sum()
```

```
Year                           0
Kilometers_Driven              0
Mileage                        2
Engine                        36
Power                         36
Seats                         42
Price                          0
Location_Bangalore             0
Location_Chennai               0
Location_Coimbatore            0
Location_Delhi                 0
Location_Hyderabad             0
Location_Jaipur                0
Location_Kochi                 0
Location_Kolkata               0
Location_Mumbai                0
Location_Pune                  0
Fuel_Type_Diesel               0
Fuel_Type_Electric             0
Fuel_Type_LPG                  0
Fuel_Type_Petrol               0
Transmission_Manual            0
Owner_Type_Fourth & Above      0
Owner_Type_Second              0
Owner_Type_Third               0
dtype: int64
```

```
#setting missing value==converting int or float to NaN


dfe.loc[dfe.Engine==0,'Engine']=np.NaN
dfe.loc[dfe.Power==0,'Power']=np.NaN
dfe.loc[dfe.Mileage==0,'Mileage']=np.NaN




dfe.isna().sum()
```

```
Year                           0
Kilometers_Driven              0
Mileage                       70
Engine                        36
Power                        143
Seats                         42
Price                          0
Location_Bangalore             0
Location_Chennai               0
Location_Coimbatore            0
Location_Delhi                 0
Location_Hyderabad             0
Location_Jaipur                0
Location_Kochi                 0
Location_Kolkata               0
Location_Mumbai                0
Location_Pune                  0
Fuel_Type_Diesel               0
Fuel_Type_Electric             0
Fuel_Type_LPG                  0
Fuel_Type_Petrol               0
Transmission_Manual            0
Owner_Type_Fourth & Above      0
Owner_Type_Second              0
Owner_Type_Third               0
dtype: int64
```

```
#handling missing value
#mileage,engine,power==>mean
#seat==>mode

dfe['Mileage']=dfe['Mileage'].fillna(dfe['Mileage'].mean())
dfe['Engine']=dfe['Engine'].fillna(dfe['Engine'].mean())
dfe['Power']=dfe['Power'].fillna(dfe['Power'].mean())
dfe['Seats']=dfe['Seats'].fillna(dfe['Seats'].mode()[0])


dfe.isna().sum()
```

```
Year                          0
Kilometers_Driven             0
Mileage                       0
Engine                        0
Power                         0
Seats                         0
Price                         0
Location_Bangalore            0
Location_Chennai              0
Location_Coimbatore           0
Location_Delhi                0
Location_Hyderabad            0
Location_Jaipur               0
Location_Kochi                0
Location_Kolkata              0
Location_Mumbai               0
Location_Pune                 0
Fuel_Type_Diesel              0
Fuel_Type_Electric            0
Fuel_Type_LPG                 0
Fuel_Type_Petrol              0
Transmission_Manual           0
Owner_Type_Fourth & Above     0
Owner_Type_Second             0
Owner_Type_Third              0
dtype: int64
```

```
x=dfe.drop(['Price','Fuel_Type_Electric'],axis=1)
x
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi
0	2010	72000	26.60	998.0	58.16	5.0	0	0	0	0
1	2015	41000	19.67	1582.0	126.20	5.0	0	0	0	0
2	2011	46000	18.20	1199.0	88.70	5.0	0	1	0	0
3	2012	87000	20.77	1248.0	88.76	7.0	0	1	0	0
4	2013	40670	15.20	1968.0	140.80	5.0	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...
6014	2014	27365	28.40	1248.0	74.00	5.0	0	0	0	1
6015	2015	100000	24.40	1120.0	71.00	5.0	0	0	0	0
6016	2012	55000	14.00	2498.0	112.00	8.0	0	0	0	0
6017	2013	46000	18.90	998.0	67.10	5.0	0	0	0	0
6018	2011	47000	25.44	936.0	57.60	5.0	0	0	0	0

6019 rows × 23 columns

```
y=dfe['Price']
y
```

```
0        1.75
1       12.50
2        4.50
3        6.00
4       17.74
        ...
6014     4.75
6015     4.00
6016     2.90
6017     2.65
6018     2.50
Name: Price, Length: 6019, dtype: float64
```

```
df2=pd.read_csv('/content/test-data.csv')
df2
```

Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type

```
df2.columns
```

```
Index(['Unnamed: 0', 'Name', 'Location', 'Year', 'Kilometers_Driven',
       'Fuel_Type', 'Transmission', 'Owner_Type', 'Mileage', 'Engine', 'Power',
       'Seats', 'New_Price'],
      dtype='object')
```

```
df2.isna().sum()
```

```
Unnamed: 0              0
Name                    0
Location                0
Year                    0
Kilometers_Driven       0
Fuel_Type               0
Transmission            0
Owner_Type              0
Mileage                 0
Engine                 10
Power                  10
Seats                  11
New_Price            1052
dtype: int64
```

```
df2['Name'].value_counts()
```

```
Maruti Alto LXi                                     9
Honda City 1.5 V MT                                 8
Maruti Swift Dzire VDI                              8
Volkswagen Polo 1.2 MPI Highline                    8
Hyundai i10 Magna                                   7
                                                   ..
Hyundai Santro GLS I - Euro II                      1
Honda City i DTec VX Option BL                      1
Land Rover Discovery 4 SDV6 SE                      1
Hyundai Verna CRDi 1.6 SX Option                    1
Mercedes-Benz E-Class 2009-2013 E 220 CDI Avantgarde    1
Name: Name, Length: 769, dtype: int64
```

```
df2['Location'].value_counts()
```

```
Mumbai        159
Pune          143
Coimbatore    136
Hyderabad     134
Kochi         121
Kolkata       119
Delhi         106
Chennai        97
Jaipur         86
Bangalore      82
Ahmedabad      51
Name: Location, dtype: int64
```

```
df2['Fuel_Type'].value_counts()
```

```
Diesel    647
Petrol    579
CNG         6
LPG         2
Name: Fuel_Type, dtype: int64
```

```
df2['Transmission'].value_counts()
```

```
Manual       905
Automatic    329
Name: Transmission, dtype: int64
```

```
df2['Owner_Type'].value_counts()
```

```
First             1023
Second             184
Third               24
Fourth & Above       3
Name: Owner_Type, dtype: int64
```

```
df3=pd.get_dummies(df2[['Location','Fuel_Type','Transmission','Owner_Type']],drop_first=True)
df3
```

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Hyderabad	Location_Jaipur	Location_Kochi	Lo
0	0	0	0	1	0	0	0	
1	0	0	1	0	0	0	0	
2	0	0	0	0	0	0	0	
3	0	0	0	0	1	0	0	
4	0	0	0	0	0	0	0	
...	...	...	...	...	...	...	...	
1229	0	0	0	0	1	0	0	
1230	0	0	0	0	0	0	0	
1231	0	0	0	0	0	0	0	
1232	0	0	0	0	0	0	0	
1233	0	0	0	0	0	0	1	

1234 rows × 17 columns

```
dfg=pd.concat([df2,df3],axis=1)
dfg
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	...	Location_Kolka

```
dfg.columns
```

```
Index(['Unnamed: 0', 'Name', 'Location', 'Year', 'Kilometers_Driven',
       'Fuel_Type', 'Transmission', 'Owner_Type', 'Mileage', 'Engine', 'Power',
       'Seats', 'New_Price', 'Location_Bangalore', 'Location_Chennai',
       'Location_Coimbatore', 'Location_Delhi', 'Location_Hyderabad',
       'Location_Jaipur', 'Location_Kochi', 'Location_Kolkata',
       'Location_Mumbai', 'Location_Pune', 'Fuel_Type_Diesel', 'Fuel_Type_LPG',
       'Fuel_Type_Petrol', 'Transmission_Manual', 'Owner_Type_Fourth & Above',
       'Owner_Type_Second', 'Owner_Type_Third'],
      dtype='object')
```

```
dfg.drop(["Unnamed: 0","Name","Location","Fuel_Type","Transmission","Owner_Type","New_Price"],axis=1,inplace=True)
dfg
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi
0	2014	40929	32.26 km/kg	998 CC	58.2 bhp	4.0	0	0	0	1
1	2013	54493	24.7 kmpl	796 CC	47.3 bhp	5.0	0	0	1	0
2	2017	34000	13.68 kmpl	2393 CC	147.8 bhp	7.0	0	0	0	0
3	2012	139000	23.59 kmpl	1364 CC	null bhp	5.0	0	0	0	0
4	2014	29000	18.5 kmpl	1197 CC	82.85 bhp	5.0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...
1229	2011	89411	20.54 kmpl	1598 CC	103.6 bhp	5.0	0	0	0	0
1230	2015	59000	17.21 kmpl	1197 CC	103.6 bhp	5.0	0	0	0	0
1231	2012	28000	23.08 kmpl	1461 CC	63.1 bhp	5.0	0	0	0	0
1232	2013	52262	17.2 kmpl	1197 CC	103.6 bhp	5.0	0	0	0	0
1233	2014	72443	10.0 kmpl	2148 CC	170 bhp	5.0	0	0	0	0

1234 rows × 23 columns

```
#string replace

dfg['Mileage']=dfg['Mileage'].str.replace('km/kg','')
dfg['Mileage']=dfg['Mileage'].str.replace('kmpl','')
dfg['Engine']=dfg['Engine'].str.replace('CC','')
dfg['Power']=dfg['Power'].str.replace('bhp','')
#null
dfg['Mileage']=dfg['Mileage'].str.replace('null','0')
dfg['Engine']=dfg['Engine'].str.replace('null','0')
dfg['Power']=dfg['Power'].str.replace('null','0')
dfg
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi
0	2014	40929	32.26	998	58.2	4.0	0	0	0	1
1	2013	54493	24.7	796	47.3	5.0	0	0	1	0
2	2017	34000	13.68	2393	147.8	7.0	0	0	0	0
3	2012	139000	23.59	1364	0	5.0	0	0	0	0
4	2014	29000	18.5	1197	82.85	5.0	0	0	0	0

```
dfg.dtypes
```

```
Year                          int64
Kilometers_Driven             int64
Mileage                      object
Engine                       object
Power                        object
Seats                       float64
Location_Bangalore            uint8
Location_Chennai              uint8
Location_Coimbatore           uint8
Location_Delhi                uint8
Location_Hyderabad            uint8
Location_Jaipur               uint8
Location_Kochi                uint8
Location_Kolkata              uint8
Location_Mumbai               uint8
Location_Pune                 uint8
Fuel_Type_Diesel              uint8
Fuel_Type_LPG                 uint8
Fuel_Type_Petrol              uint8
Transmission_Manual           uint8
Owner_Type_Fourth & Above     uint8
Owner_Type_Second             uint8
Owner_Type_Third              uint8
dtype: object
```

```
#convert into float
dfg['Engine']=dfg['Engine'].astype(float)
dfg['Power']=dfg['Power'].astype(float)
dfg['Mileage']=dfg['Mileage'].astype(float)
dfg.dtypes
```

```
Year                          int64
Kilometers_Driven             int64
Mileage                     float64
Engine                      float64
Power                       float64
Seats                       float64
Location_Bangalore            uint8
Location_Chennai              uint8
Location_Coimbatore           uint8
Location_Delhi                uint8
Location_Hyderabad            uint8
Location_Jaipur               uint8
Location_Kochi                uint8
Location_Kolkata              uint8
Location_Mumbai               uint8
Location_Pune                 uint8
Fuel_Type_Diesel              uint8
Fuel_Type_LPG                 uint8
Fuel_Type_Petrol              uint8
Transmission_Manual           uint8
Owner_Type_Fourth & Above     uint8
Owner_Type_Second             uint8
Owner_Type_Third              uint8
dtype: object
```

```
dfg.isna().sum()
```

```
Year                          0
Kilometers_Driven             0
Mileage                       0
Engine                       10
Power                        10
Seats                        11
Location_Bangalore            0
Location_Chennai              0
Location_Coimbatore           0
```

```
Location_Delhi               0
Location_Hyderabad           0
Location_Jaipur              0
Location_Kochi               0
Location_Kolkata             0
Location_Mumbai              0
Location_Pune                0
Fuel_Type_Diesel             0
Fuel_Type_LPG                0
Fuel_Type_Petrol             0
Transmission_Manual          0
Owner_Type_Fourth & Above    0
Owner_Type_Second            0
Owner_Type_Third             0
dtype: int64
```

```
#setting missing value==converting int or float to NaN


dfg.loc[dfg.Engine==0,'Engine']=np.NaN
dfg.loc[dfg.Power==0,'Power']=np.NaN
dfg.loc[dfg.Mileage==0,'Mileage']=np.NaN


dfg.isna().sum()
```

```
Year                         0
Kilometers_Driven            0
Mileage                     13
Engine                      10
Power                       32
Seats                       11
Location_Bangalore           0
Location_Chennai             0
Location_Coimbatore          0
Location_Delhi               0
Location_Hyderabad           0
Location_Jaipur              0
Location_Kochi               0
Location_Kolkata             0
Location_Mumbai              0
Location_Pune                0
Fuel_Type_Diesel             0
Fuel_Type_LPG                0
Fuel_Type_Petrol             0
Transmission_Manual          0
Owner_Type_Fourth & Above    0
Owner_Type_Second            0
Owner_Type_Third             0
dtype: int64
```

```
#handling missing value
#mileage,engine,power==>mean
#seat==>mode

dfg['Mileage']=dfg['Mileage'].fillna(dfg['Mileage'].mean())
dfg['Engine']=dfg['Engine'].fillna(dfg['Engine'].mean())
dfg['Power']=dfg['Power'].fillna(dfg['Power'].mean())
dfg['Seats']=dfg['Seats'].fillna(dfg['Seats'].mode()[0])


dfg.isna().sum()
```

```
Year                         0
Kilometers_Driven            0
Mileage                      0
Engine                       0
Power                        0
Seats                        0
Location_Bangalore           0
Location_Chennai             0
Location_Coimbatore          0
Location_Delhi               0
Location_Hyderabad           0
Location_Jaipur              0
Location_Kochi               0
Location_Kolkata             0
Location_Mumbai              0
Location_Pune                0
Fuel_Type_Diesel             0
Fuel_Type_LPG                0
```

```
Fuel_Type_Petrol            0
Transmission_Manual         0
Owner_Type_Fourth & Above   0
Owner_Type_Second           0
Owner_Type_Third            0
dtype: int64
```

```
x_test=dfg
```

```
x_test.describe()
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats
count	1234.000000	1234.000000	1234.000000	1234.000000	1234.000000	1234.000000
mean	2013.400324	58507.288493	18.367355	1593.584967	110.380420	5.282010
std	3.179700	35598.702098	4.051338	562.054549	50.870564	0.822366
min	1996.000000	1000.000000	7.940000	624.000000	34.200000	2.000000
25%	2011.000000	34000.000000	15.400000	1198.000000	76.200000	5.000000
50%	2014.000000	54572.500000	18.367355	1461.000000	98.600000	5.000000
75%	2016.000000	75000.000000	21.100000	1968.000000	126.300000	5.000000
max	2019.000000	350000.000000	32.260000	5998.000000	616.000000	10.000000

8 rows × 23 columns

```
x_test.shape
```

```
(1234, 23)
```

```
x.shape
```

```
(6019, 23)
```

```
from sklearn.linear_model import LinearRegression
model=LinearRegression()
model.fit(x,y)
y_pred=model.predict(x_test)
y_pred
```

```
array([ 2.87588492, -1.29344912, 16.1069494 , ...,  0.1378514 ,
        9.27293255, 21.48043251])
```