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

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmiss
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Man
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Man
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Man
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Man
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Autom

```
df.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual
1	1	Hyundai Creta 1.6 CRDi SX	Pune	2015	41000	Diesel	Manual

```
df.tail()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmissi
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manu
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manu

```
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
```

```
a=df['Location'].unique()
a
```

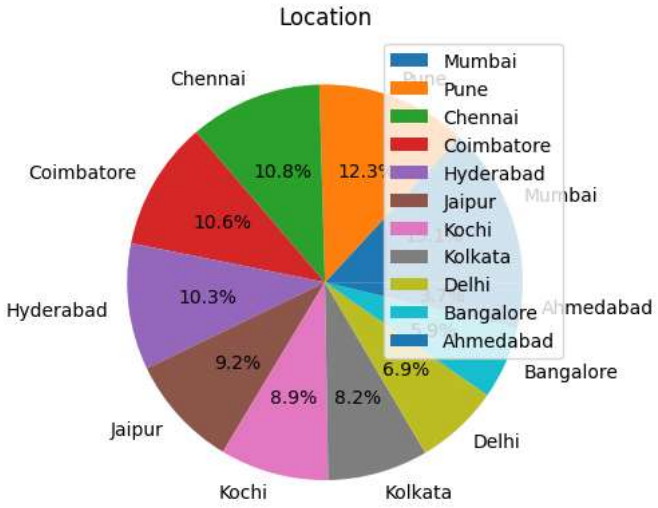
```
array(['Mumbai', 'Pune', 'Chennai', 'Coimbatore', 'Hyderabad', 'Jaipur',
       'Kochi', 'Kolkata', 'Delhi', 'Bangalore', 'Ahmedabad'],
      dtype=object)
```

```
a1=df['Location'].value_counts()
a1
```

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

```
plt.pie(a1,labels=a,autopct='%1.1f%%')
plt.legend(loc='upper right')
plt.title("Location")
```

```
Text(0.5, 1.0, 'Location')
```




```
b=df['Fuel_Type'].unique()
b
```

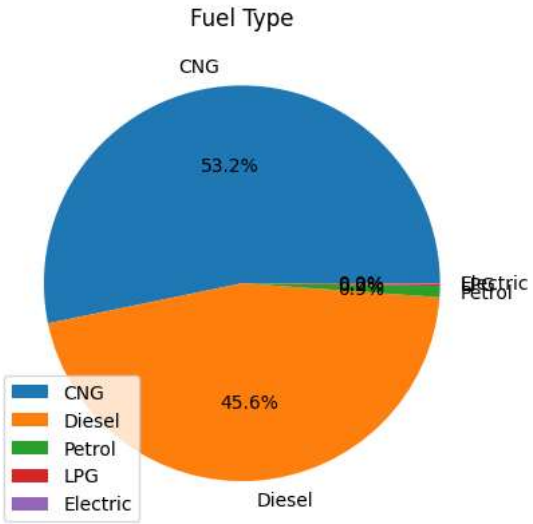
```
array(['CNG', 'Diesel', 'Petrol', 'LPG', 'Electric'], dtype=object)
```

```
b1=df['Fuel_Type'].value_counts()
b1
```

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

```
plt.pie(b1,labels=b,autopct='%1.1f%%')
plt.legend(loc='lower left')
plt.title("Fuel Type")
```

```
Text(0.5, 1.0, 'Fuel Type')
```




```
c=df['Transmission'].unique()
c
```

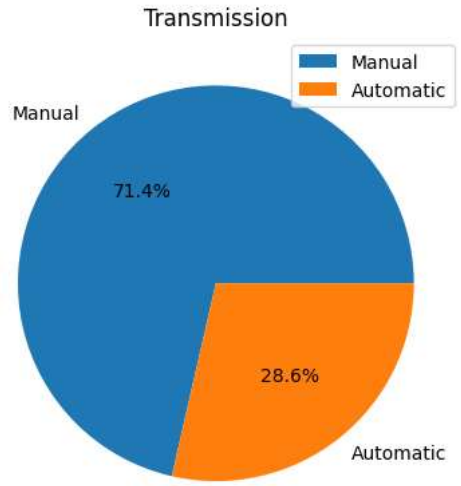
```
array(['Manual', 'Automatic'], dtype=object)
```

```
c1=df['Transmission'].value_counts()
c1
```

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

```
plt.pie(c1,labels=c,autopct='%1.1f%%')
plt.legend(loc='upper right')
plt.title("Transmission")
```

```
Text(0.5, 1.0, 'Transmission')
```




```
d=df['Owner_Type'].unique()
d
```

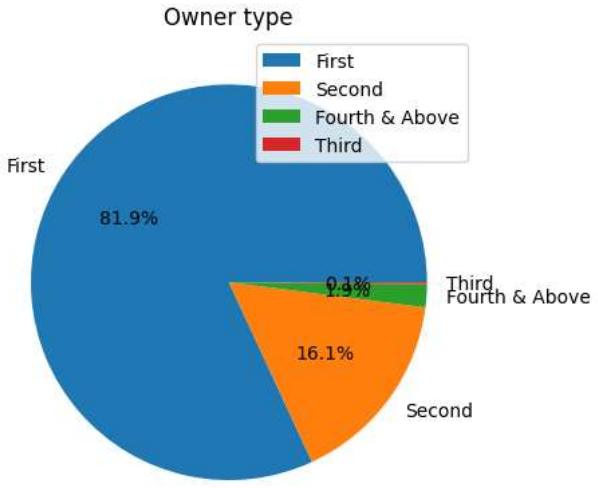
```
array(['First', 'Second', 'Fourth & Above', 'Third'], dtype=object)
```

```
d1=df['Owner_Type'].value_counts()
d1
```

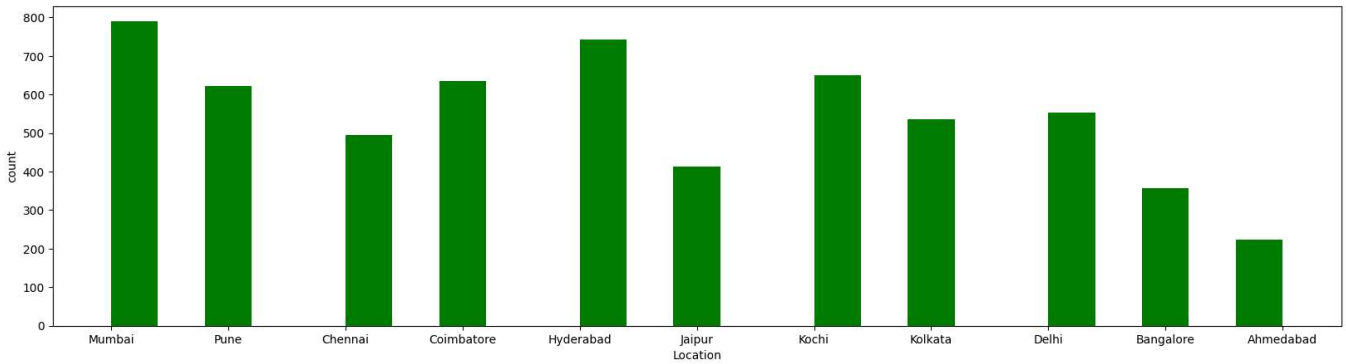
```
Owner_Type
First             4929
Second             968
Third              113
Fourth & Above       9
Name: count, dtype: int64
```

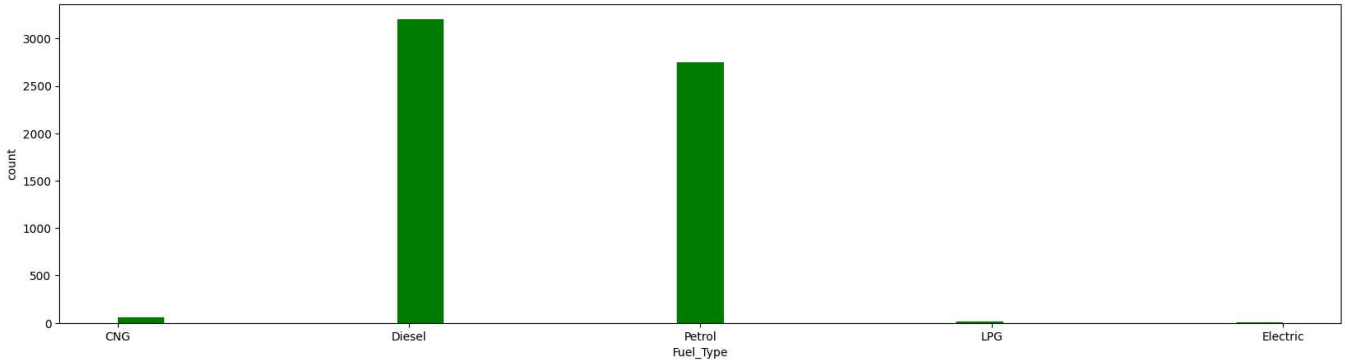
```
plt.pie(d1,labels=d,autopct='%1.1f%%')
plt.legend(loc='upper right')
plt.title("Owner type")
```

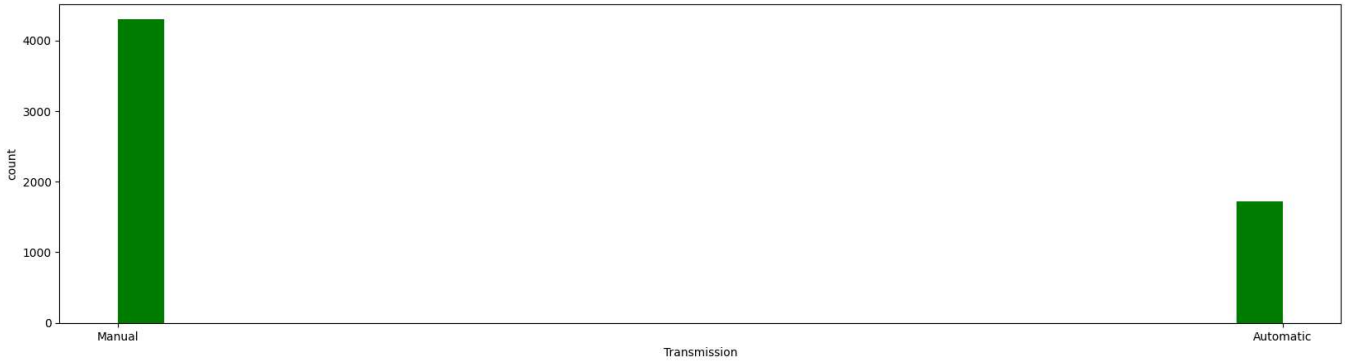
```
Text(0.5, 1.0, 'Owner type')
```




```
col=['Location','Fuel_Type',"Transmission",'Owner_Type']
for i in col:
  plt.figure(figsize=(20,5))
  plt.hist(df[i],bins=25,color='g')
  plt.xlabel(i)
  plt.ylabel('count')
```







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

```
array(['Maruti Wagon R LXI CNG', 'Hyundai Creta 1.6 CRDi SX Option',
       'Honda Jazz V', ..., 'Tata Bolt Revotron XT', 'Honda',
       'Mahindra Xylo D4 BSIV'], dtype=object)
```

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

```
Name
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: count, Length: 1878, dtype: int64
```

```
# collect only the brand name because it is difficult to encode the name column and we can't drop that it is importan
df['Brand']=df['Name'].apply(lambda x:x.split()[0])
df
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	26.6 km/kg	998 CC	58.16 bhp	5.0
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	19.67 kmpl	1582 CC	126.2 bhp	5.0
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	18.2 kmpl	1199 CC	88.7 bhp	5.0
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	20.77 kmpl	1248 CC	88.76 bhp	7.0
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	15.2 kmpl	1968 CC	140.8 bhp	5.0
...	...	...	...	...	...	...	...	...	...	...	...	...
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.0
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	24.4 kmpl	1120 CC	71 bhp	5.0
6016	6016	Mahindra Xylo D4	Jaipur	2012	55000	Diesel	Manual	Second	14.0 kmpl	2498 CC	112 bhp	8.0

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

```
Brand
Maruti           1211
Hyundai          1107
Honda             608
Toyota            411
Mercedes-Benz     318
Volkswagen        315
Ford              300
Mahindra          272
BMW               267
Audi              236
Tata              186
Skoda             173
Renault           145
Chevrolet         121
Nissan             91
Land               60
Jaguar             40
Fiat               28
Mitsubishi         27
Mini               26
Volvo              21
Porsche            18
Jeep               15
Datsun             13
Force               3
ISUZU               2
Smart               1
Ambassador          1
Isuzu               1
Bentley             1
Lamborghini         1
Name: count, dtype: int64
```

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

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Hyderabad	Location_Jaipur	Location_
0	0	0	0	0	0	0	
1	0	0	0	0	0	0	
2	0	1	0	0	0	0	
3	0	1	0	0	0	0	
4	0	0	1	0	0	0	
...	...	...	...	...	...	...	
6014	0	0	0	1	0	0	
6015	0	0	0	0	0	1	
6016	0	0	0	0	0	1	
6017	0	0	0	0	0	0	
6018	0	0	0	0	1	0	

6019 rows × 48 columns

```
# concatnate both dataframes
df2=pd.concat([df,df1],axis=1)
df2
```

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

6019 rows × 63 columns

```
df2.drop(['Unnamed: 0','Name','Location','Fuel_Type','Transmission','Owner_Type','New_Price','Brand','Brand_Audi','Brand_Force','Br
df2
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Location_Bangalore	Location_Chennai	Location_Coimbatore
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

6019 rows × 50 columns

```
df2.dtypes
```

```
Year                          int64
Kilometers_Driven             int64
Mileage                      object
Engine                       object
Power                        object
Seats                       float64
Price                       float64
Location_Bangalore            int64
Location_Chennai              int64
Location_Coimbatore           int64
Location_Delhi                int64
Location_Hyderabad            int64
Location_Jaipur               int64
Location_Kochi                int64
Location_Kolkata              int64
Location_Mumbai               int64
Location_Pune                 int64
Fuel_Type_Diesel              int64
Fuel_Type_LPG                 int64
Fuel_Type_Petrol              int64
Transmission_Manual           int64
Owner_Type_Fourth & Above     int64
Owner_Type_Second             int64
Owner_Type_Third              int64
Brand_BMW                     int64
Brand_Bentley                 int64
Brand_Chevrolet               int64
Brand_Datsun                  int64
Brand_Fiat                    int64
Brand_Ford                    int64
Brand_Honda                   int64
Brand_Hyundai                 int64
Brand_ISUZU                   int64
Brand_Isuzu                   int64
Brand_Jaguar                  int64
Brand_Jeep                    int64
Brand_Land                    int64
Brand_Mahindra                int64
Brand_Maruti                  int64
Brand_Mercedes-Benz           int64
Brand_Mini                    int64
Brand_Mitsubishi              int64
Brand_Nissan                  int64
Brand_Porsche                 int64
Brand_Renault                 int64
Brand_Skoda                   int64
Brand_Tata                    int64
Brand_Toyota                  int64
Brand_Volkswagen              int64
Brand_Volvo                   int64
dtype: object
```

```
#  remove the units in mileage,engine and power column

df2['Mileage']=df2['Mileage'].str.replace('km/kg','')
df2['Mileage']=df2['Mileage'].str.replace('kmpl','')
df2['Engine']=df2['Engine'].str.replace('CC','')
df2['Power']=df2['Power'].str.replace('bhp','')
df2['Mileage']=df2['Mileage'].str.replace('null','0')
df2['Engine']=df2['Engine'].str.replace('null','0')
df2['Power']=df2['Power'].str.replace('null','0')

df2
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Location_Bangalore	Location_Chennai	Location_Coimbatore
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 × 50 columns

```
# datatype conversion of mileage,engine and power column
df2['Mileage']=df2['Mileage'].astype(float)
df2['Engine']=df2['Engine'].astype(float)
df2['Power']=df2['Power'].astype(float)


df2.dtypes
```

```
Year                           int64
Kilometers_Driven              int64
Mileage                      float64
Engine                       float64
Power                        float64
Seats                        float64
Price                        float64
Location_Bangalore             int64
Location_Chennai               int64
Location_Coimbatore            int64
Location_Delhi                 int64
Location_Hyderabad             int64
Location_Jaipur                int64
Location_Kochi                 int64
Location_Kolkata               int64
Location_Mumbai                int64
Location_Pune                  int64
Fuel_Type_Diesel               int64
Fuel_Type_LPG                  int64
Fuel_Type_Petrol               int64
Transmission_Manual            int64
Owner_Type_Fourth & Above      int64
Owner_Type_Second              int64
Owner_Type_Third               int64
Brand_BMW                      int64
Brand_Bentley                  int64
Brand_Chevrolet                int64
Brand_Datsun                   int64
Brand_Fiat                     int64
Brand_Ford                     int64
Brand_Honda                    int64
Brand_Hyundai                  int64
Brand_ISUZU                    int64
Brand_Isuzu                    int64
Brand_Jaguar                   int64
Brand_Jeep                     int64
Brand_Land                     int64
Brand_Mahindra                 int64
```

```
Brand_Maruti                  int64
Brand_Mercedes-Benz           int64
Brand_Mini                    int64
Brand_Mitsubishi              int64
Brand_Nissan                  int64
Brand_Porsche                 int64
Brand_Renault                 int64
Brand_Skoda                   int64
Brand_Tata                    int64
Brand_Toyota                  int64
Brand_Volkswagen              int64
Brand_Volvo                   int64
dtype: object
```

```
# mileage,engine power ===========>null======>0
# treat the 0 as missing values
df2.loc[df2.Engine==0,'Engine']=np.NaN
df2.loc[df2.Mileage==0,'Mileage']=np.NaN
df2.loc[df2.Power==0,'Power']=np.NaN
```

```
df2.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_LPG                 0
Fuel_Type_Petrol              0
Transmission_Manual           0
Owner_Type_Fourth & Above     0
Owner_Type_Second             0
Owner_Type_Third              0
Brand_BMW                     0
Brand_Bentley                 0
Brand_Chevrolet               0
Brand_Datsun                  0
Brand_Fiat                    0
Brand_Ford                    0
Brand_Honda                   0
Brand_Hyundai                 0
Brand_ISUZU                   0
Brand_Isuzu                   0
Brand_Jaguar                  0
Brand_Jeep                    0
Brand_Land                    0
Brand_Mahindra                0
Brand_Maruti                  0
Brand_Mercedes-Benz           0
Brand_Mini                    0
Brand_Mitsubishi              0
Brand_Nissan                  0
Brand_Porsche                 0
Brand_Renault                 0
Brand_Skoda                   0
Brand_Tata                    0
Brand_Toyota                  0
Brand_Volkswagen              0
Brand_Volvo                   0
dtype: int64
```

```
df2['Mileage'].fillna(df2['Mileage'].mean(), inplace=True)
df2['Engine'].fillna(df2['Engine'].mean(), inplace=True)
df2['Power'].fillna(df2['Power'].mean(), inplace=True)
df2['Seats'].fillna(df2['Seats'].mode()[0], inplace=True)
```

```
df2.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_LPG               0
Fuel_Type_Petrol            0
Transmission_Manual         0
Owner_Type_Fourth & Above   0
Owner_Type_Second           0
Owner_Type_Third            0
Brand_BMW                   0
Brand_Bentley               0
Brand_Chevrolet             0
Brand_Datsun                0
Brand_Fiat                  0
Brand_Ford                  0
Brand_Honda                 0
Brand_Hyundai               0
Brand_ISUZU                 0
Brand_Isuzu                 0
Brand_Jaguar                0
Brand_Jeep                  0
Brand_Land                  0
Brand_Mahindra              0
Brand_Maruti                0
Brand_Mercedes-Benz         0
Brand_Mini                  0
Brand_Mitsubishi            0
Brand_Nissan                0
Brand_Porsche               0
Brand_Renault               0
Brand_Skoda                 0
Brand_Tata                  0
Brand_Toyota                0
Brand_Volkswagen            0
Brand_Volvo                 0
dtype: int64
```

```
x=df2.drop(['Price'],axis=1)
x
```

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

6019 rows × 49 columns

```
y=df2['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
```

```
# model.fit(x,y)
# ypred=fit with z

# TESTING DATA
df_tst=pd.read_csv('/content/test-data.csv')
df_tst
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seat
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual	First	32.26 km/kg	998 CC	58.2 bhp	4.
1	1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	Manual	Second	24.7 kmpl	796 CC	47.3 bhp	5.
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual	First	13.68 kmpl	2393 CC	147.8 bhp	7.
3	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual	First	23.59 kmpl	1364 CC	null bhp	5.
4	4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual	First	18.5 kmpl	1197 CC	82.85 bhp	5.
...	...	...	...	...	...	...	...	...	...	...	...	.
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	First	20.54 kmpl	1598 CC	103.6 bhp	5.
1230	1230	Volkswagen Polo GT	Mumbai	2015	59000	Petrol	Automatic	First	17.21	1197	103.6	5.

```
df_tst.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats	New
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual	First	32.26 km/kg	998 CC	58.2 bhp	4.0	
		Maruti Alto											

```
df_tst.tail()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	First	20.54 kmpl	1598 CC	103.6 bhp	5.0
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	First	17.21 kmpl	1197 CC	103.6 bhp	5.0
1231	1231	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	5.0
		Volkswagen										

```
df_tst.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
```

```
df_tst.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
dtype: object
```

```
aa=df_tst['Location'].unique()
aa
```

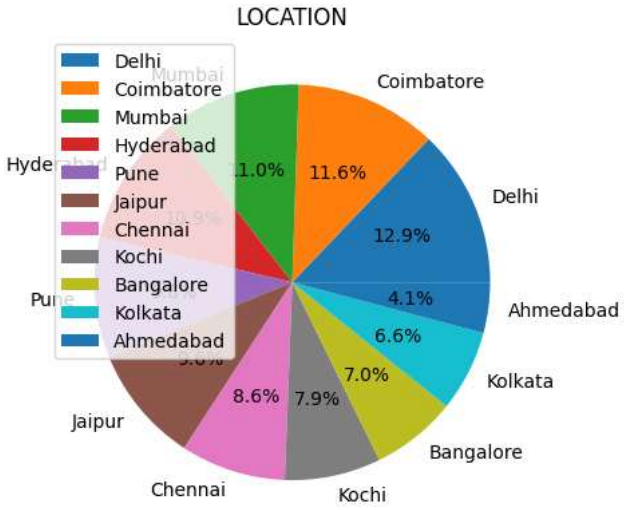
```
array(['Delhi', 'Coimbatore', 'Mumbai', 'Hyderabad', 'Pune', 'Jaipur',
       'Chennai', 'Kochi', 'Bangalore', 'Kolkata', 'Ahmedabad'],
      dtype=object)
```

```
aa1=df_tst['Location'].value_counts()
aa1
```

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

```
plt.pie(aa1,labels=aa,autopct='%1.1f%%')
plt.legend(loc='upper left')
plt.title("LOCATION")
```

```
Text(0.5, 1.0, 'LOCATION')
```




```
bb=df_tst['Fuel_Type'].unique()
bb
```

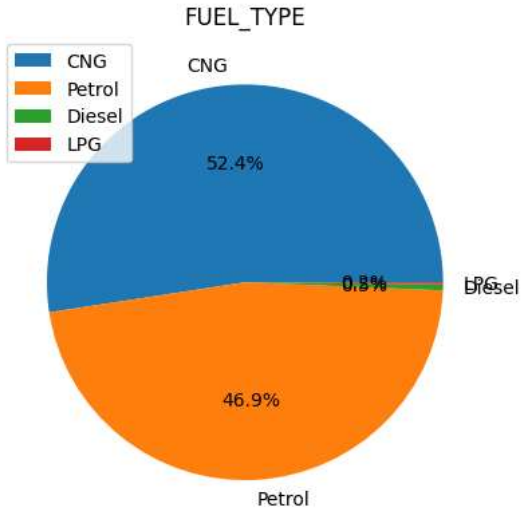
```
array(['CNG', 'Petrol', 'Diesel', 'LPG'], dtype=object)
```

```
bb1=df_tst['Fuel_Type'].value_counts()
bb1
```

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

```
plt.pie(bb1,labels=bb,autopct='%1.1f%%')
plt.legend(loc='upper left')
plt.title("FUEL_TYPE")
```

```
Text(0.5, 1.0, 'FUEL_TYPE')
```




```
cc=df_tst['Transmission'].unique()
cc
```

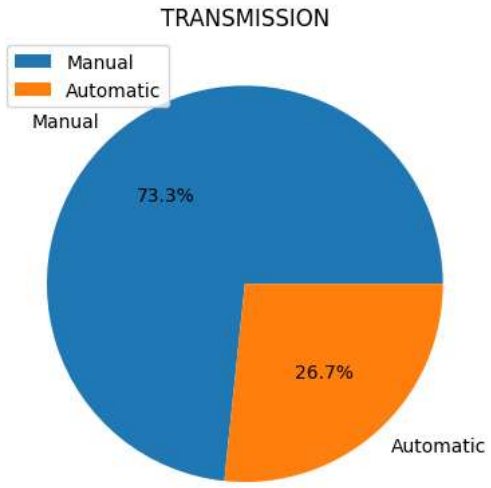
```
array(['Manual', 'Automatic'], dtype=object)
```

```
cc1=df_tst['Transmission'].value_counts()
cc1
```

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

```
plt.pie(cc1,labels=cc,autopct='%1.1f%%')
plt.legend(loc='upper left')
plt.title("TRANSMISSION")
```

```
Text(0.5, 1.0, 'TRANSMISSION')
```




```
dd=df_tst['Owner_Type'].unique()
dd
```

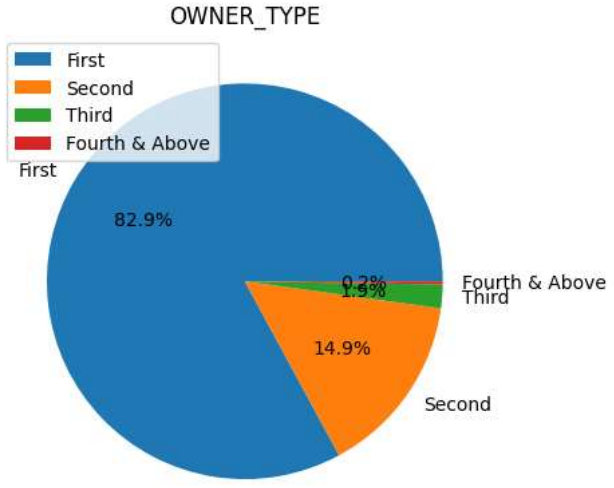
```
array(['First', 'Second', 'Third', 'Fourth & Above'], dtype=object)
```

```
dd1=df_tst['Owner_Type'].value_counts()
dd1
```

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

```
plt.pie(dd1,labels=dd,autopct='%1.1f%%')
plt.legend(loc='upper left')
plt.title("OWNER_TYPE")
```

```
Text(0.5, 1.0, 'OWNER_TYPE')
```




```
df_tst['Brand']=df_tst['Name'].apply(lambda x:x.split()[0])
df_tst
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seat
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual	First	32.26 km/kg	998 CC	58.2 bhp	4.
1	1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	Manual	Second	24.7 kmpl	796 CC	47.3 bhp	5.
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual	First	13.68 kmpl	2393 CC	147.8 bhp	7.
3	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual	First	23.59 kmpl	1364 CC	null bhp	5.
4	4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual	First	18.5 kmpl	1197 CC	82.85 bhp	5.
...	...	...	...	...	...	...	...	...	...	...	...	.
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	First	20.54 kmpl	1598 CC	103.6 bhp	5.
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	First	17.21 kmpl	1197 CC	103.6 bhp	5.

```
df_tst['Brand'].value_counts()
```

```
Brand
Maruti           233
Hyundai          233
Honda            135
Toyota            96
Mercedes-Benz     62
Mahindra          59
Volkswagen        59
Ford              51
Audi              49
BMW               45
Tata              42
Chevrolet         30
Skoda             29
Nissan            26
Renault           25
Fiat              10
Mitsubishi         9
Jaguar             8
Volvo              7
Land               7
Mini               5
Datsun             4
Jeep               4
Bentley            1
Hindustan          1
Isuzu              1
Porsche            1
ISUZU              1
OpelCorsa          1
Name: count, dtype: int64
```

```
df1_tst=pd.get_dummies(df_tst[['Location','Fuel_Type','Transmission','Owner_Type','Brand']],dtype=int,drop_first=True)
df1_tst
```

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

1234 rows × 45 columns

```
df2_tst=pd.concat([df_tst,df1_tst],axis=1)
df2_tst
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	...	Brand_
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual	First	32.26 km/kg	998 CC	...	
1	1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	Manual	Second	24.7 kmpl	796 CC	...	
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual	First	13.68 kmpl	2393 CC	...	
3	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	Manual	First	23.59 kmpl	1364 CC	...	
4	4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	Manual	First	18.5 kmpl	1197 CC	...	
...	...	...	...	...	...	...	...	...	...	...	...	
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	First	20.54 kmpl	1598 CC	...	
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	First	17.21 kmpl	1197 CC	...	
1231	1231	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual	First	23.08 kmpl	1461 CC	...	
1232	1232	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	Automatic	Third	17.2 kmpl	1197 CC	...	
1233	1233	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...	Kochi	2014	72443	Diesel	Automatic	First	10.0 kmpl	2148 CC	...	

1234 rows × 59 columns

```
df2_tst.drop(['Unnamed: 0','Name','Location','Fuel_Type','Transmission','Owner_Type','New_Price','Brand'],axis=1,inplace=True)
df2_tst
```

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

1234 rows × 51 columns

```
df2_tst.dtypes
```

```
Year                         int64
Kilometers_Driven            int64
Mileage                     object
Engine                      object
Power                       object
Seats                      float64
Location_Bangalore           int64
Location_Chennai             int64
Location_Coimbatore          int64
Location_Delhi               int64
Location_Hyderabad           int64
Location_Jaipur              int64
Location_Kochi               int64
Location_Kolkata             int64
Location_Mumbai              int64
Location_Pune                int64
Fuel_Type_Diesel             int64
Fuel_Type_LPG                int64
Fuel_Type_Petrol             int64
Transmission_Manual          int64
Owner_Type_Fourth & Above    int64
Owner_Type_Second            int64
Owner_Type_Third             int64
Brand_BMW                    int64
Brand_Bentley                int64
Brand_Chevrolet              int64
Brand_Datsun                 int64
Brand_Fiat                   int64
Brand_Ford                   int64
Brand_Hindustan              int64
Brand_Honda                  int64
Brand_Hyundai                int64
Brand_ISUZU                  int64
Brand_Isuzu                  int64
Brand_Jaguar                 int64
Brand_Jeep                   int64
Brand_Land                   int64
Brand_Mahindra               int64
Brand_Maruti                 int64
Brand_Mercedes-Benz          int64
Brand_Mini                   int64
Brand_Mitsubishi             int64
Brand_Nissan                 int64
Brand_OpelCorsa              int64
Brand_Porsche                int64
Brand_Renault                int64
Brand_Skoda                  int64
Brand_Tata                   int64
Brand_Toyota                 int64
Brand_Volkswagen             int64
```

```
Brand_Volvo                    int64
dtype: object
```

```
df2_tst['Mileage']=df2_tst['Mileage'].str.replace('km/kg','')
df2_tst['Mileage']=df2_tst['Mileage'].str.replace('kmpl','')
df2_tst['Engine']=df2_tst['Engine'].str.replace('CC','')
df2_tst['Power']=df2_tst['Power'].str.replace('bhp','')
df2_tst['Mileage']=df2_tst['Mileage'].str.replace('null','0')
df2_tst['Engine']=df2_tst['Engine'].str.replace('null','0')
df2_tst['Power']=df2_tst['Power'].str.replace('null','0')
```

```
df2_tst
```

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

1234 rows × 51 columns

```
df2_tst['Mileage']=df2_tst['Mileage'].astype(float)
df2_tst['Engine']=df2_tst['Engine'].astype(float)
df2_tst['Power']=df2_tst['Power'].astype(float)
```

```
df2_tst.loc[df2_tst.Engine==0,'Engine']=np.NaN
df2_tst.loc[df2_tst.Mileage==0,'Mileage']=np.NaN
df2_tst.loc[df2_tst.Power==0,'Power']=np.NaN
```

```
df2_tst.dtypes
```

```
Year                         int64
Kilometers_Driven            int64
Mileage                    float64
Engine                     float64
Power                      float64
Seats                      float64
Location_Bangalore           int64
Location_Chennai             int64
Location_Coimbatore          int64
Location_Delhi               int64
Location_Hyderabad           int64
Location_Jaipur              int64
Location_Kochi               int64
Location_Kolkata             int64
Location_Mumbai              int64
Location_Pune                int64
Fuel_Type_Diesel             int64
Fuel_Type_LPG                int64
Fuel_Type_Petrol             int64
Transmission_Manual          int64
Owner_Type_Fourth & Above    int64
Owner_Type_Second            int64
Owner_Type_Third             int64
Brand_BMW                    int64
Brand_Bentley                int64
Brand_Chevrolet              int64
Brand_Datsun                 int64
Brand_Fiat                   int64
Brand_Ford                   int64
Brand_Hindustan              int64
Brand_Honda                  int64
Brand_Hyundai                int64
Brand_ISUZU                  int64
Brand_Isuzu                  int64
Brand_Jaguar                 int64
Brand_Jeep                   int64
```

```
Brand_Land                    int64
Brand_Mahindra                int64
Brand_Maruti                  int64
Brand_Mercedes-Benz           int64
Brand_Mini                    int64
Brand_Mitsubishi              int64
Brand_Nissan                  int64
Brand_OpelCorsa               int64
Brand_Porsche                 int64
Brand_Renault                 int64
Brand_Skoda                   int64
Brand_Tata                    int64
Brand_Toyota                  int64
Brand_Volkswagen              int64
Brand_Volvo                   int64
dtype: object
```

```
df2_tst.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
Brand_BMW                    0
Brand_Bentley                0
Brand_Chevrolet              0
Brand_Datsun                 0
Brand_Fiat                   0
Brand_Ford                   0
Brand_Hindustan              0
Brand_Honda                  0
Brand_Hyundai                0
Brand_ISUZU                  0
Brand_Isuzu                  0
Brand_Jaguar                 0
Brand_Jeep                   0
Brand_Land                   0
Brand_Mahindra               0
Brand_Maruti                 0
Brand_Mercedes-Benz          0
Brand_Mini                   0
Brand_Mitsubishi             0
Brand_Nissan                 0
Brand_OpelCorsa              0
Brand_Porsche                0
Brand_Renault                0
Brand_Skoda                  0
Brand_Tata                   0
Brand_Toyota                 0
Brand_Volkswagen             0
Brand_Volvo                  0
dtype: int64
```

```
df2_tst['Mileage'].fillna(df2_tst['Mileage'].mean(), inplace=True)
df2_tst['Engine'].fillna(df2_tst['Engine'].mean(), inplace=True)
df2_tst['Power'].fillna(df2_tst['Power'].mean(), inplace=True)
df2_tst['Seats'].fillna(df2_tst['Seats'].mode()[0], inplace=True)
```

```
df2_tst.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
Brand_BMW                    0
Brand_Bentley                0
Brand_Chevrolet              0
Brand_Datsun                 0
Brand_Fiat                   0
Brand_Ford                   0
Brand_Hindustan              0
Brand_Honda                  0
Brand_Hyundai                0
Brand_ISUZU                  0
Brand_Isuzu                  0
```