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

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mil
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	k
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	1
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	2
4	4	Audi A4 New 2.0 TDI Multitronic	Coimbatore	2013	40670	Diesel	Automatic	Second	
...	...	...	...	...	...	...	...	...	
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	
6016	6016	Mahindra Xylo D4 BSIV	Jaipur	2012	55000	Diesel	Manual	Second	
6017	6017	Maruti Wagon R VXI	Kolkata	2013	46000	Petrol	Manual	First	
6018	6018	Chevrolet Beat Diesel	Hyderabad	2011	47000	Diesel	Manual	First	2

6019 rows × 14 columns

```
train.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
```

```
train.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
```

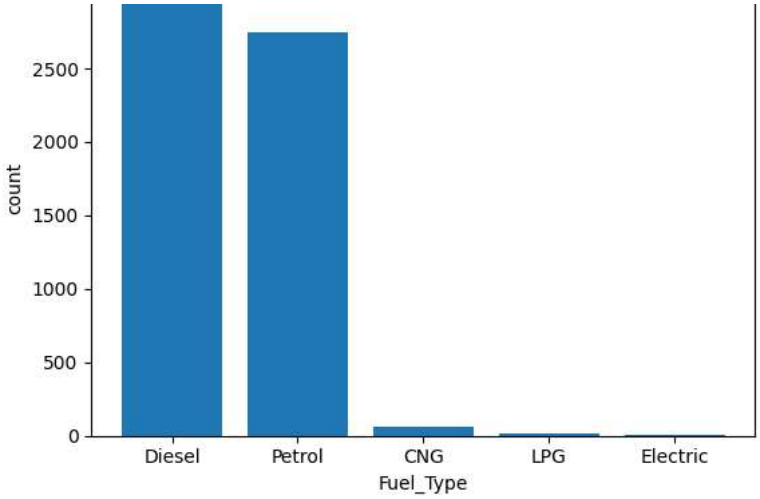
```
train.describe()
```

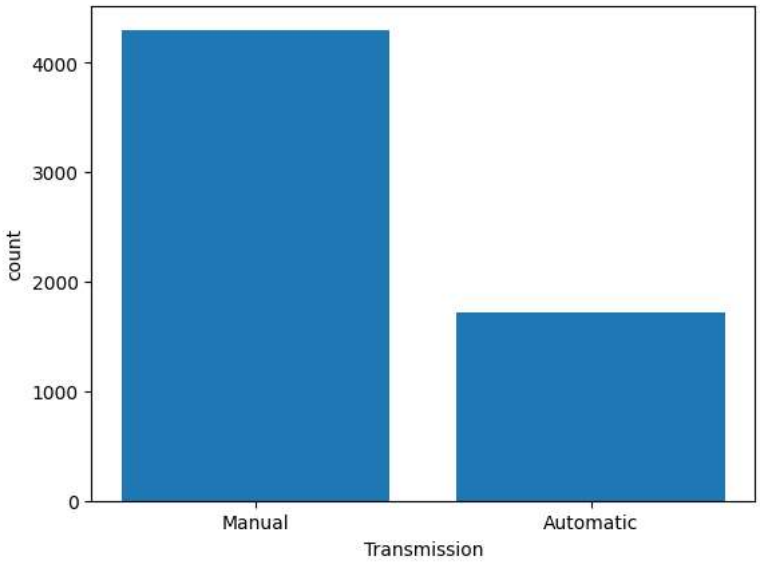
	Unnamed: 0	Year	Kilometers_Driven	Seats	Price
count	6019.000000	6019.000000	6.019000e+03	5977.000000	6019.000000
mean	3009.000000	2013.358199	5.873838e+04	5.278735	9.479468
std	1737.679967	3.269742	9.126884e+04	0.808840	11.187917
min	0.000000	1998.000000	1.710000e+02	0.000000	0.440000
25%	1504.500000	2011.000000	3.400000e+04	5.000000	3.500000
50%	3009.000000	2014.000000	5.300000e+04	5.000000	5.640000
75%	4513.500000	2016.000000	7.300000e+04	5.000000	9.950000
max	6018.000000	2019.000000	6.500000e+06	10.000000	160.000000

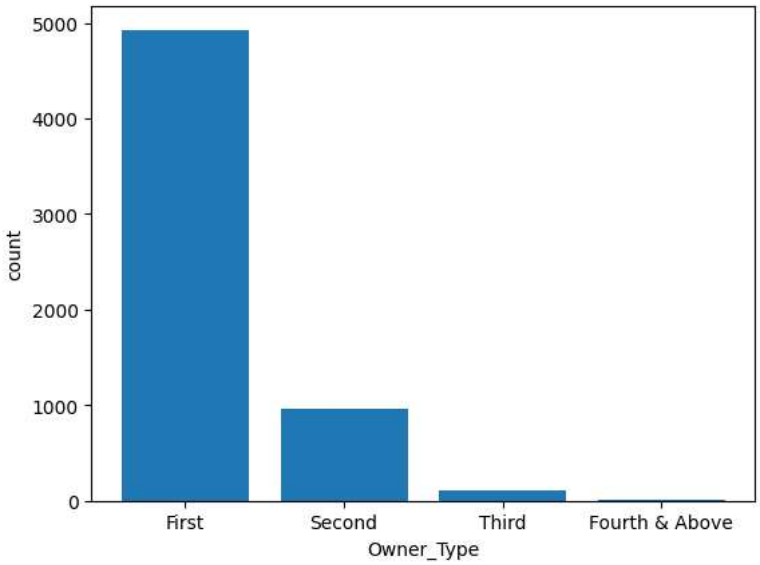
```
lst=['Name','Location','Fuel_Type','Transmission','Owner_Type']
for i in lst:
  count=train[i].value_counts()
  print('column',i,'have',len(count),'unique values')
  print(count.index)
  print('*'*100)
```

```
column Name have 1878 unique values
Index(['Mahindra XUV500 W8 2WD', 'Maruti Swift VDI', 'Honda City 1.5 S MT',
       'Maruti Swift Dzire VDI', 'Maruti Swift VDI BSIV', 'Maruti Ritz VDi',
       'Hyundai i10 Sportz', 'Toyota Fortuner 3.0 Diesel',
       'Honda Amaze S i-Dtech', 'Hyundai Grand i10 Sportz',
       ...
       'Mahindra Scorpio SLE BSIII', 'Land Rover Discovery HSE Luxury 3.0 TD6',
       'Hyundai Tucson 2.0 Dual VTVT 2WD AT GL', 'Audi A4 2.0 TFSI',
       'Volvo S60 D4 SUMMUM', 'Ford Fiesta Titanium 1.5 TDCi',
       'Mahindra Scorpio S10 AT 4WD', 'Hyundai i20 1.2 Era',
       'Toyota Camry W4 (AT)', 'Mahindra Xylo D4 BSIV'],
      dtype='object', length=1878)
****************************************************************************************************
column Location have 11 unique values
Index(['Mumbai', 'Hyderabad', 'Kochi', 'Coimbatore', 'Pune', 'Delhi',
       'Kolkata', 'Chennai', 'Jaipur', 'Bangalore', 'Ahmedabad'],
      dtype='object')
****************************************************************************************************
column Fuel_Type have 5 unique values
Index(['Diesel', 'Petrol', 'CNG', 'LPG', 'Electric'], dtype='object')
****************************************************************************************************
column Transmission have 2 unique values
Index(['Manual', 'Automatic'], dtype='object')
****************************************************************************************************
column Owner_Type have 4 unique values
Index(['First', 'Second', 'Third', 'Fourth & Above'], dtype='object')
****************************************************************************************************
```

```
lst=['Location','Fuel_Type','Transmission','Owner_Type']
for i in lst:
  count1=train[i].value_counts()
  plt.bar(count1.index,count1)
  plt.xlabel(i)
  plt.ylabel('count')
  plt.show()
```


2500
2000
1500
1000
500
0
count
Diesel
Petrol
CNG
LPG
Electric
Fuel_Type


4000
3000
2000
1000
0
count
Manual
Automatic
Transmission


5000
4000
3000
2000
1000
0
count
First
Second
Third
Fourth & Above
Owner_Type

```
#we should drop columns named - unnamed,newprize,name
#newprize column have large missing value and name have large set of unique values
#get dummies encoding
df1=pd.get_dummies(train[['Location','Fuel_Type','Transmission','Owner_Type']],drop_first=True)
df1
```

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

6019 rows × 18 columns

```
df2=pd.concat([train,df1],axis=1)
df2
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mil
0	0	Maruti Wagon R LXI CNG	Mumbai	2010	72000	CNG	Manual	First	k
1	1	Hyundai Creta 1.6 CRDi SX Option	Pune	2015	41000	Diesel	Manual	First	1
2	2	Honda Jazz V	Chennai	2011	46000	Petrol	Manual	First	
3	3	Maruti Ertiga VDI	Chennai	2012	87000	Diesel	Manual	First	2
4	4	Audi A4 New 2.0	Coimbatore	2013	40670	Diesel	Automatic	Second	

```
df2.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')
```

BSIV

```
#test data file dont have a column named 'Fuel_Type_Electric' therefore we should drop it
df3=df2.drop(['Unnamed: 0','Name','Location','Fuel_Type','Transmission','Owner_Type','New_Price'],axis=1)
```

VXI

```
#replace unit from mileage,engine,power
df3['Mileage']=df3['Mileage'].str.replace('km/kg','')
df3['Mileage']=df3['Mileage'].str.replace('kmpl','')
df3['Engine']=df3['Engine'].str.replace('CC','')
df3['Power']=df3['Engine'].str.replace('bhp','')
#there is 'null' in engine,power,mileage given in description
df3['Mileage']=df3['Mileage'].str.replace('null','0')
df3['Engine']=df3['Engine'].str.replace('null','0')
df3['Power']=df3['Power'].str.replace('null','0')
df3
```

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

6019 rows × 25 columns

```
df3.dtypes
#uint8 un directional integer
```

```
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 datatype of object into int
df3['Engine']=df3['Engine'].astype(float)
df3['Mileage']=df3['Mileage'].astype(float)
df3['Power']=df3['Power'].astype(float)
df3.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
```

```
df3.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
```

```
#consider the '0' value we give instead of 'null' as a missing value and replace with NaN
df3.loc[df3.Engine==0,'Engine']=np.NaN
df3.loc[df3.Mileage==0,'Mileage']=np.NaN
df3.loc[df3.Power==0,'Power']=np.NaN
```

```
df3.isna().sum()
```

```
Year                         0
Kilometers_Driven            0
Mileage                     70
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
```

```
#filling missing values
df3['Mileage']=df3['Mileage'].fillna(df3['Mileage'].mean())
df3['Engine']=df3['Engine'].fillna(df3['Engine'].mean())
df3['Power']=df3['Power'].fillna(df3['Power'].mean())
df3['Seats']=df3['Seats'].fillna(df3['Seats'].mode()[0])
```

```
df3.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=df3.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	998.0	5.0	0	0	0	0
1	2015	41000	19.67	1582.0	1582.0	5.0	0	0	0	0
2	2011	46000	18.20	1199.0	1199.0	5.0	0	1	0	0
3	2012	87000	20.77	1248.0	1248.0	7.0	0	1	0	0
4	2013	40670	15.20	1968.0	1968.0	5.0	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...
6014	2014	27365	28.40	1248.0	1248.0	5.0	0	0	0	1
6015	2015	100000	24.40	1120.0	1120.0	5.0	0	0	0	0
6016	2012	55000	14.00	2498.0	2498.0	8.0	0	0	0	0
6017	2013	46000	18.90	998.0	998.0	5.0	0	0	0	0
6018	2011	47000	25.44	936.0	936.0	5.0	0	0	0	0

6019 rows × 23 columns

```
y=df3['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
```

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

	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	
1	1	Maruti Alto 800 2016-	Coimbatore	2013	54493	Petrol	Manual	Second	24.7 kmpl	796 CC	47.3 bhp	5.0	

```
test.info
```

```
<bound method DataFrame.info of       Unnamed: 0                                               Name  \
0              0                            Maruti Alto K10 LXI CNG
1              1                           Maruti Alto 800 2016-2019 LXI
2              2             Toyota Innova Crysta Touring Sport 2.4 MT
3              3                                Toyota Etios Liva GD
4              4                                   Hyundai i20 Magna
...          ...                                                 ...
1229        1229                     Volkswagen Vento Diesel Trendline
1230        1230                                Volkswagen Polo GT TSI
1231        1231                                Nissan Micra Diesel XV
1232        1232                                Volkswagen Polo GT TSI
1233        1233  Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...

        Location  Year  Kilometers_Driven Fuel_Type Transmission Owner_Type  \
0          Delhi  2014              40929       CNG       Manual      First
1     Coimbatore  2013              54493    Petrol       Manual     Second
2         Mumbai  2017              34000    Diesel       Manual      First
3      Hyderabad  2012             139000    Diesel       Manual      First
4         Mumbai  2014              29000    Petrol       Manual      First
...          ...   ...                ...       ...          ...        ...
1229   Hyderabad  2011              89411    Diesel       Manual      First
1230      Mumbai  2015              59000    Petrol    Automatic      First
1231     Kolkata  2012              28000    Diesel       Manual      First
1232        Pune  2013              52262    Petrol    Automatic      Third
1233       Kochi  2014              72443    Diesel    Automatic      First

          Mileage   Engine      Power  Seats   New_Price
0     32.26 km/kg   998 CC   58.2 bhp    4.0         NaN
1       24.7 kmpl   796 CC   47.3 bhp    5.0         NaN
2       13.68 kmpl  2393 CC  147.8 bhp    7.0  25.27 Lakh
3       23.59 kmpl  1364 CC   null bhp    5.0         NaN
4       18.5 kmpl   1197 CC  82.85 bhp    5.0         NaN
...           ...      ...        ...    ...         ...
1229    20.54 kmpl  1598 CC  103.6 bhp    5.0         NaN
1230    17.21 kmpl  1197 CC  103.6 bhp    5.0         NaN
1231    23.08 kmpl  1461 CC   63.1 bhp    5.0         NaN
1232     17.2 kmpl  1197 CC  103.6 bhp    5.0         NaN
1233     10.0 kmpl  2148 CC    170 bhp    5.0         NaN

[1234 rows x 13 columns]>
```

```
test.describe()
```

	Unnamed: 0	Year	Kilometers_Driven	Seats
count	1234.000000	1234.000000	1234.000000	1223.000000
mean	616.500000	2013.400324	58507.288493	5.284546
std	356.369424	3.179700	35598.702098	0.825622
min	0.000000	1996.000000	1000.000000	2.000000
25%	308.250000	2011.000000	34000.000000	5.000000
50%	616.500000	2014.000000	54572.500000	5.000000
75%	924.750000	2016.000000	75000.000000	5.000000
max	1233.000000	2019.000000	350000.000000	10.000000

```
test.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
```

```
test.columns
```

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

```
test.isna().sum()
```

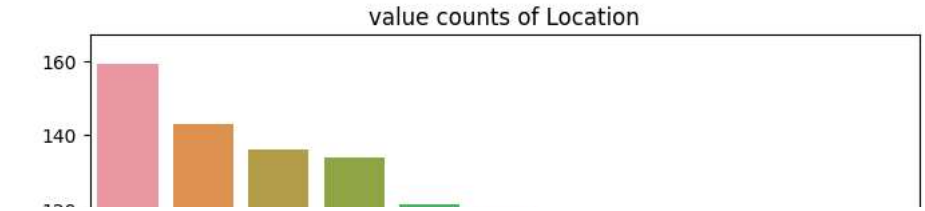
```
Unnamed: 0             0
Name                   0
Location               0
Year                   0
Kilometers_Driven      0
Fuel_Type              0
Transmission           0
Owner_Type             0
Mileage                0
Engine                10
Power                 10
Seats                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

```
test['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
```

```
column='Location'
plt.figure(figsize=(8,6))
sns.countplot(data=test,x=column,order=test[column].value_counts().index)
plt.title('value counts of Location')
plt.xticks(rotation=45)
plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```

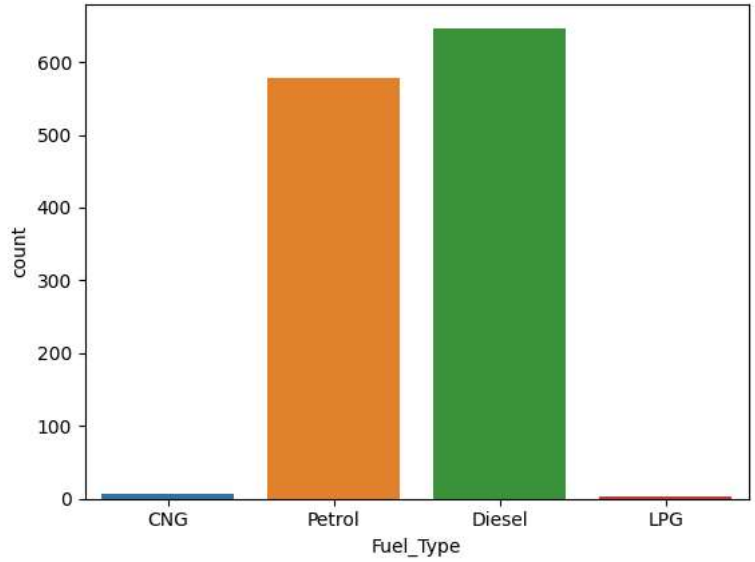



```
test['Fuel_Type'].value_counts()
```

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

```
sns.countplot(x=test['Fuel_Type'])
```

```
<Axes: xlabel='Fuel_Type', ylabel='count'>
```




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

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

```
sns.countplot(x=test['Transmission'])
```

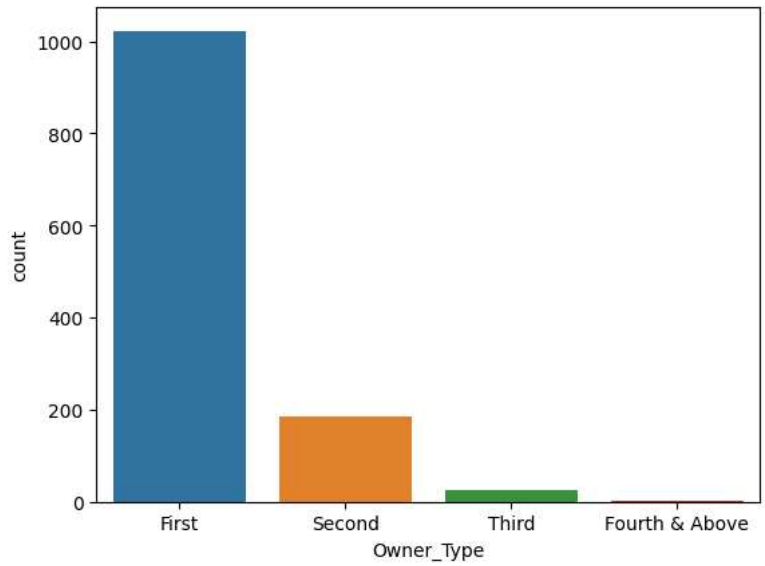
```
<Axes: xlabel='Transmission', ylabel='count'>
```

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

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

```
sns.countplot(x=test['Owner_Type'])
```

```
<Axes: xlabel='Owner_Type', ylabel='count'>
```




```
test.columns
```

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

```
test1=pd.get_dummies(test[['Location','Fuel_Type','Transmission','Owner_Type']],drop_first=True)
test1
```

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Hyderabad	Location_Jaipur	Location_Kochi	L
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

```
test2=pd.concat([test,test1],axis=1)
test2
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	...	Location_Kolk
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 × 30 columns

```
test2=test2.drop(['Unnamed: 0','Name','Location','Fuel_Type','Transmission','Owner_Type','New_Price'],axis=1)
```

```
test2.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
```

```
test2['Mileage']=test2['Mileage'].str.replace('km/kg','')
test2['Mileage']=test2['Mileage'].str.replace('kmpl','')
test2['Engine']=test2['Engine'].str.replace('CC','')
test2['Power']=test2['Power'].str.replace('bhp','')
test2['Mileage']=test2['Mileage'].str.replace('null','0')
test2['Engine']=test2['Engine'].str.replace('null','0')
test2['Power']=test2['Power'].str.replace('null','0')
test2
```

	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
...	...	...	...	...	...	...	...	...	...	...
1229	2011	89411	20.54	1598	103.6	5.0	0	0	0	0
1230	2015	59000	17.21	1197	103.6	5.0	0	0	0	0
1231	2012	28000	23.08	1461	63.1	5.0	0	0	0	0
1232	2013	52262	17.2	1197	103.6	5.0	0	0	0	0
1233	2014	72443	10.0	2148	170	5.0	0	0	0	0

1234 rows × 23 columns

```
test2['Engine']=test2['Engine'].astype(float)
test2['Power']=test2['Power'].astype(float)
test2['Mileage']=test2['Mileage'].astype(float)
test2
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Lo
0	2014	40929	32.26	998.0	58.20	4.0	0	0	
1	2013	54493	24.70	796.0	47.30	5.0	0	0	
2	2017	34000	13.68	2393.0	147.80	7.0	0	0	
3	2012	139000	23.59	1364.0	0.00	5.0	0	0	
4	2014	29000	18.50	1197.0	82.85	5.0	0	0	
...	...	...	...	...	...	...	...	...	
1229	2011	89411	20.54	1598.0	103.60	5.0	0	0	
1230	2015	59000	17.21	1197.0	103.60	5.0	0	0	
1231	2012	28000	23.08	1461.0	63.10	5.0	0	0	
1232	2013	52262	17.20	1197.0	103.60	5.0	0	0	
1233	2014	72443	10.00	2148.0	170.00	5.0	0	0	

1234 rows × 23 columns

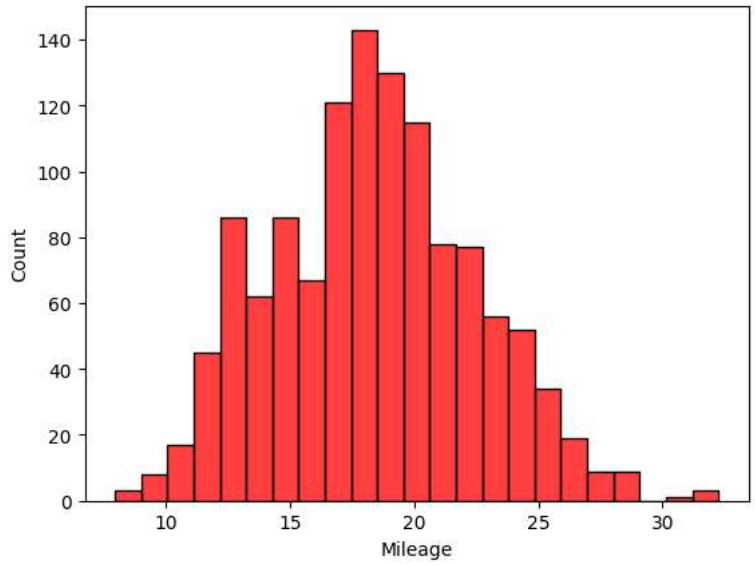
```
test2.loc[test2.Engine==0,'Engine']=np.NaN
test2.loc[test2.Mileage==0,'Mileage']=np.NaN
test2.loc[test2.Power==0,'Power']=np.NaN
```

```
test2.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
```

```
sns.histplot(x=test2['Mileage'],color='red')
```

```
<Axes: xlabel='Mileage', ylabel='Count'>
```

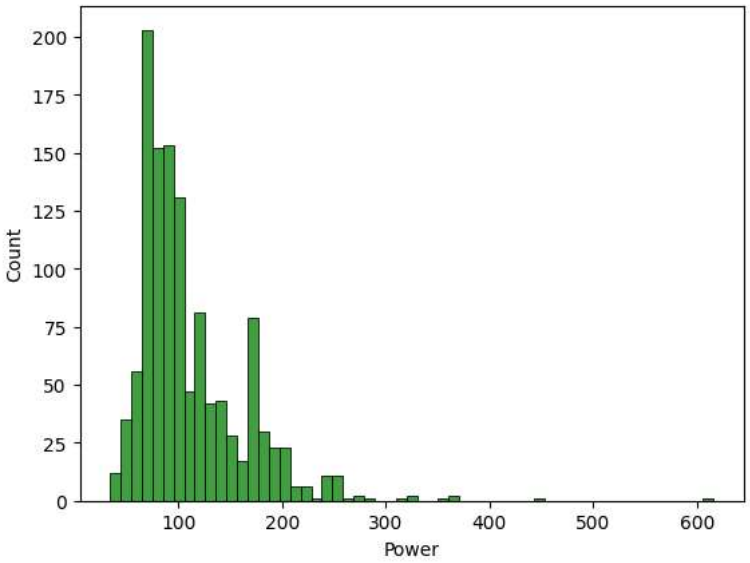



```
sns.histplot(x=test2['Engine'])
```

```
<Axes: xlabel='Engine', ylabel='Count'>
```


```
sns.histplot(x=test2['Power'],color='green')
```

```
<Axes: xlabel='Power', ylabel='Count'>
```




```
a=test2['Mileage'].mean()
test2['Mileage'].fillna(a,inplace=True)
b=test2['Engine'].median()
test2['Engine'].fillna(b,inplace=True)
c=test2['Power'].median()
test2['Power'].fillna(c,inplace=True)
d=test2['Seats'].mode()[0]
test2['Seats'].fillna(d,inplace=True)
```

```
test2.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
```

```
from sklearn.ensemble import RandomForestRegressor
rf=RandomForestRegressor(n_estimators=100)
rf.fit(x,y)
ypred=rf.predict(test2)
ypred
```