

```
import numpy as np
import pandas as pd
df=pd.read_csv("/content/car_economy_price.csv")
print(df)
```

```

   Unnamed: 0      Name      Location      Year \
0           0  Maruti Wagon R LXI CNG      Mumbai  2010
1           1  Hyundai Creta 1.6 CRDi SX Option      Pune  2015
2           2      Honda Jazz V      Chennai  2011
3           3  Maruti Ertiga VDI      Chennai  2012
4           4  Audi A4 New 2.0 TDI Multitronic  Coimbatore  2013
...         ...      ...      ...      ...
6014        6014  Maruti Swift VDI      Delhi  2014
6015        6015  Hyundai Xcent 1.1 CRDi S      Jaipur  2015
6016        6016  Mahindra Xylo D4 BSIV      Jaipur  2012
6017        6017  Maruti Wagon R VXI      Kolkata  2013
6018        6018  Chevrolet Beat Diesel      Hyderabad  2011
```

```

      Kilometers_Driven  Fuel_Type  Transmission  Owner_Type      Mileage \
0           72000      CNG      Manual      First  26.6 km/kg
1           41000      Diesel      Manual      First  19.67 kmpl
2           46000      Petrol      Manual      First  18.2 kmpl
3           87000      Diesel      Manual      First  20.77 kmpl
4           40670      Diesel      Automatic      Second  15.2 kmpl
...         ...      ...      ...      ...
6014          27365      Diesel      Manual      First  28.4 kmpl
6015         100000      Diesel      Manual      First  24.4 kmpl
6016          55000      Diesel      Manual      Second  14.0 kmpl
6017          46000      Petrol      Manual      First  18.9 kmpl
6018          47000      Diesel      Manual      First  25.44 kmpl
```

```

      Engine      Power  Seats  New_Price  Price
0      998 CC  58.16 bhp   5.0      NaN   1.75
1     1582 CC  126.2 bhp   5.0      NaN  12.50
2     1199 CC   88.7 bhp   5.0  8.61 Lakh   4.50
3     1248 CC   88.76 bhp   7.0      NaN   6.00
4     1968 CC  140.8 bhp   5.0      NaN  17.74
...     ...      ...      ...      ...
6014    1248 CC   74 bhp   5.0  7.88 Lakh   4.75
6015    1120 CC   71 bhp   5.0      NaN   4.00
6016    2498 CC  112 bhp   8.0      NaN   2.90
6017     998 CC   67.1 bhp   5.0      NaN   2.65
6018     936 CC   57.6 bhp   5.0      NaN   2.50
```

[6019 rows x 14 columns]

```
print(df.head())
print(df.tail())
```

```

   Unnamed: 0      Name      Location      Year \
0           0  Maruti Wagon R LXI CNG      Mumbai  2010
1           1  Hyundai Creta 1.6 CRDi SX Option      Pune  2015
2           2      Honda Jazz V      Chennai  2011
3           3  Maruti Ertiga VDI      Chennai  2012
4           4  Audi A4 New 2.0 TDI Multitronic  Coimbatore  2013

      Kilometers_Driven  Fuel_Type  Transmission  Owner_Type      Mileage      Engine \
0           72000      CNG      Manual      First  26.6 km/kg      998 CC
1           41000      Diesel      Manual      First  19.67 kmpl     1582 CC
2           46000      Petrol      Manual      First  18.2 kmpl     1199 CC
3           87000      Diesel      Manual      First  20.77 kmpl     1248 CC
4           40670      Diesel      Automatic      Second  15.2 kmpl     1968 CC

      Power  Seats  New_Price  Price
0  58.16 bhp   5.0      NaN   1.75
1  126.2 bhp   5.0      NaN  12.50
2   88.7 bhp   5.0  8.61 Lakh   4.50
3  88.76 bhp   7.0      NaN   6.00
4  140.8 bhp   5.0      NaN  17.74

   Unnamed: 0      Name      Location      Year \
6014        6014  Maruti Swift VDI      Delhi  2014
6015        6015  Hyundai Xcent 1.1 CRDi S      Jaipur  2015
6016        6016  Mahindra Xylo D4 BSIV      Jaipur  2012
6017        6017  Maruti Wagon R VXI      Kolkata  2013
6018        6018  Chevrolet Beat Diesel      Hyderabad  2011

      Kilometers_Driven  Fuel_Type  Transmission  Owner_Type      Mileage \
6014          27365      Diesel      Manual      First  28.4 kmpl
6015         100000      Diesel      Manual      First  24.4 kmpl
6016          55000      Diesel      Manual      Second  14.0 kmpl
6017          46000      Petrol      Manual      First  18.9 kmpl
6018          47000      Diesel      Manual      First  25.44 kmpl

      Engine      Power  Seats  New_Price  Price
6014    1248 CC   74 bhp   5.0  7.88 Lakh   4.75
6015    1120 CC   71 bhp   5.0      NaN   4.00
6016    2498 CC  112 bhp   8.0      NaN   2.90
```

```
6017  998 CC  67.1 bhp   5.0      NaN  2.65
6018  936 CC  57.6 bhp   5.0      NaN  2.50
```

```
df.columns
```

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

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

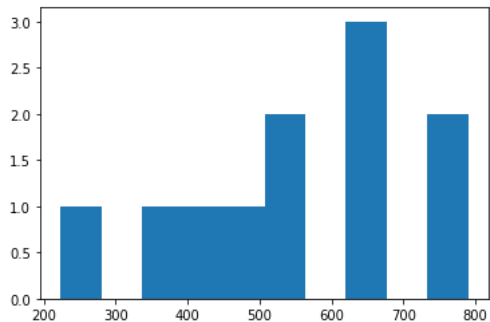
```
#count and graphical rep
#location
#fuel_type
#owner type
```

```
#location
```

```
a=df["Location"].value_counts()
a
```

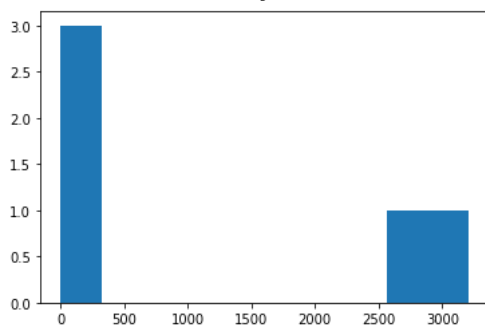
```
import matplotlib.pyplot as plt
plt.hist(a)
```

```
(array([1., 0., 1., 1., 1., 2., 0., 3., 0., 2.]),
 array([224. , 280.6, 337.2, 393.8, 450.4, 507. , 563.6, 620.2, 676.8,
       733.4, 790. ]),
 <a list of 10 Patch objects>)
```

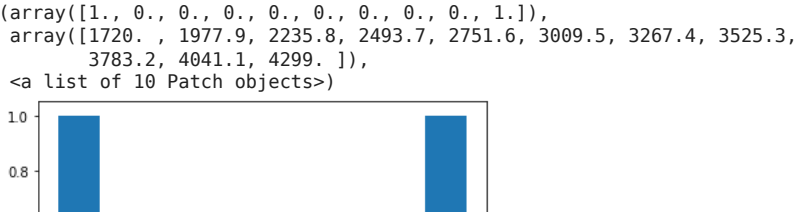


```
#fuel_type
b=df["Fuel_Type"].value_counts()
plt.hist(b)
```

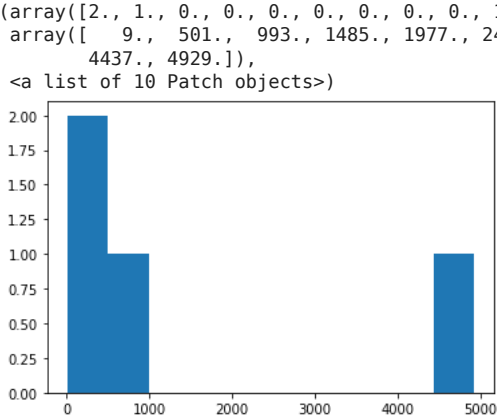
```
(array([3., 0., 0., 0., 0., 0., 0., 0., 1., 1.]),
 array([2.0000e+00, 3.2230e+02, 6.4260e+02, 9.6290e+02, 1.2832e+03,
       1.6035e+03, 1.9238e+03, 2.2441e+03, 2.5644e+03, 2.8847e+03,
       3.2050e+03]),
 <a list of 10 Patch objects>)
```



```
#transmission
c=df["Transmission"].value_counts()
plt.hist(c)
```



```
#owner type
d=df["Owner_Type"].value_counts()
plt.hist(d)
```



```
#string==> loation, fuel_type, #transmission , owner_type

#get_dummy

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

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_De
0	0	0	0	0
1	0	0	0	0
2	0	1	0	0
3	0	1	0	0
4	0	0	1	0
...
6014	0	0	0	0
6015	0	0	0	0
6016	0	0	0	0
6017	0	0	0	0
6018	0	0	0	0

6019 rows × 18 columns

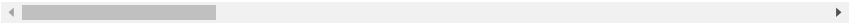
```
#combain two data frame
dfe=pd.concat([df,dummy],axis=1)
dfe
```

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

```
dfe=dfe.drop(['Unnamed: 0', 'Location', 'Fuel_Type', 'Power', 'Fuel_Type_Electric', 'Transmission', 'Owner_Type', 'Name', 'New_Price'])
dfe
```

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

6019 rows × 23 columns



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

Year	0
Kilometers_Driven	0
Mileage	2
Engine	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_LPG           0
Fuel_Type_Petrol        0
Transmission_Manual     0
Owner_Type_Fourth & Above 0
Owner_Type_Second       0
Owner_Type_Third        0
dtype: int64

#string replace(eg: milage 22 kmpl==>22)
dfe['Mileage']=dfe['Mileage'].str.replace('km/kg','')
dfe['Mileage']=dfe['Mileage'].str.replace('kmpl','')
dfe['Engine']=dfe['Engine'].str.replace('CC','')
#null
dfe['Mileage']=dfe['Mileage'].str.replace('null','0')
dfe['Engine']=dfe['Engine'].str.replace('null','0')
dfe
```

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

6019 rows × 23 columns

```
dfe.dtypes

Year                int64
Kilometers_Driven   int64
Mileage             object
Engine              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_LPG         uint8
Fuel_Type_Petrol      uint8
Transmission_Manual   uint8
Owner_Type_Fourth & Above uint8
Owner_Type_Second     uint8
Owner_Type_Third      uint8
dtype: object

#covert object to int or flot
dfe['Engine']=dfe['Engine'].astype(float)
dfe['Mileage']=dfe['Mileage'].astype(float)
```

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

```
Year          0
Kilometers_Driven  0
Mileage        2
Engine        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_LPG  0
Fuel_Type_Petrol  0
Transmission_Manual  0
Owner_Type_Fourth & Above  0
Owner_Type_Second  0
Owner_Type_Third  0
dtype: int64
```

```
#replace zero value (mileage,power,engine)
```

```
import numpy as np
```

```
dfe.loc[dfe.Engine==0,'Engine']=np.NaN #not a number #read as missing value
```

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

```
dfe['Engine']=dfe['Engine'].fillna(dfe['Engine'].mean())
```

```
dfe['Mileage']=dfe['Mileage'].fillna(dfe['Mileage'].mean())
```

```
dfe['Seats']=dfe['Seats'].fillna(dfe['Seats'].mode()[0])
```

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

```
Year          0
Kilometers_Driven  0
Mileage        0
Engine        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
dtype: int64
```

```
x=dfe['Price']
```

```
x
```

```
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
```

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

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

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Tra
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	
1	1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	
3	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	
4	4	Hyundai i20 Nios	Mumbai	2014	29000	Petrol	

```
print(test_df.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_df.columns
```

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

```
#get dummy
dummy2=pd.get_dummies(test_df[['Location','Fuel_Type','Transmission','Owner_Type']],drop_first=True)
dummy2
```

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_De
0	0	0	0	0
1	0	0	0	1
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
...
1229	0	0	0	0
1230	0	0	0	0
1231	0	0	0	0
1232	0	0	0	0
1233	0	0	0	0

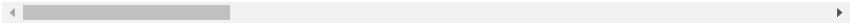
1234 rows × 17 columns



```
#combai
dfe1=pd.concat([test_df,dummy2],axis=1)
dfe1
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Tra
0	0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	
1	1	Maruti Alto 800 2016-2019 LXI	Coimbatore	2013	54493	Petrol	
2	2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	
3	3	Toyota Etios Liva GD	Hyderabad	2012	139000	Diesel	
4	4	Hyundai i20 Magna	Mumbai	2014	29000	Petrol	
...	
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	
1231	1231	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	
1232	1232	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	
1233	1233	Mercedes-Benz E-Class 2009-2013 E 220 CDI Avan...	Kochi	2014	72443	Diesel	

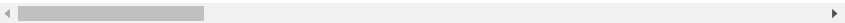
1234 rows × 30 columns




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

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

1234 rows × 22 columns



```
dfel.isna().sum()
```

Year	0
Kilometers_Driven	0
Mileage	0
Engine	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	

```
#string replace
dfel['Mileage']=dfel['Mileage'].str.replace('km/kg','')
dfel['Mileage']=dfel['Mileage'].str.replace('kmpl','')
dfel['Engine']=dfel['Engine'].str.replace('CC','')
dfel['Mileage']=dfel['Mileage'].str.replace('null','')
dfel['Engine']=dfel['Engine'].str.replace('null','')
```

```
dfel
```

	Year	Kilometers_Driven	Mileage	Engine	Seats	Location_Bangalore	Lc
0	2014	40929	32.26	998	4.0		0
1	2013	54493	24.7	796	5.0		0
2	2017	34000	13.68	2393	7.0		0
3	2012	139000	23.59	1364	5.0		0
4	2014	29000	18.5	1197	5.0		0
...
1229	2011	89411	20.54	1598	5.0		0
1230	2015	59000	17.21	1197	5.0		0
1231	2012	28000	23.08	1461	5.0		0

```
dfel.dtypes

Year                int64
Kilometers_Driven  int64
Mileage            object
Engine            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 object to int
import numpy as np
dfel['Engine']=dfel['Engine'].astype(float)
dfel['Mileage']=dfel['Mileage'].astype(float)
dfel.loc[dfel.Seats==0, 'Seats']=np.NaN
dfel['Seats']=dfel['Seats'].astype(float)
```

```
dfel.dtypes

Year                int64
Kilometers_Driven  int64
Mileage            float64
Engine            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

dfel.isna().sum()

Year                0
Kilometers_Driven  0
```

```

Mileage          0
Engine           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

```

```

dfel.loc[dfel.Seats==0, 'Seats']=np.NaN
dfel.loc[dfel.Engine==0, 'Engine']=np.NaN

```

```

dfel['Engine']=dfel['Engine'].fillna(dfel['Engine'].mean())
dfel['Seats']=dfel['Seats'].fillna(dfel['Seats'].mode()[0])

```

```
dfel.isna().sum()
```

```

Year          0
Kilometers_Driven 0
Mileage        0
Engine         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

```

```
z=dfel
```

```

from sklearn.linear_model import LinearRegression
model=LinearRegression()
model.fit(y,x)

```

```
LinearRegression()
```

```

y_pred=model.predict(z)
y_pred

```

```

array([ 4.66918815, -0.79356764, 15.66061843, ...,  4.23076859,
        9.13334362, 21.86596393])

```

