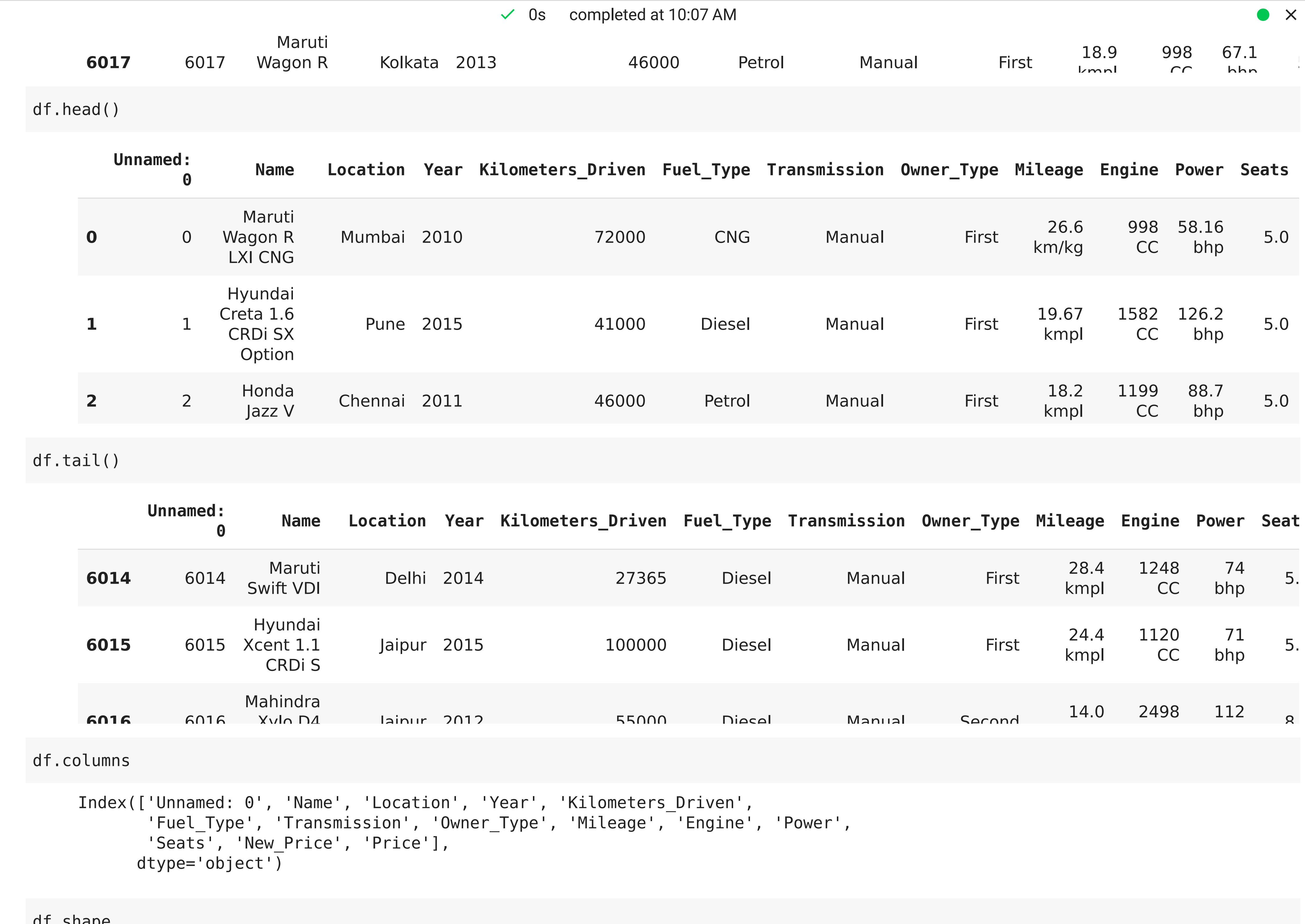
```
# car economy price and test_data_csv
```

```
# car ecnomy used as training and test_data used as testing
```

```
import numpy as np
import pandas as pd
df=pd.read_csv('/content/car economy price.csv')
df
```

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



✓ 0s completed at 10:07 AM

6017	6017	Maruti Wagon R	Kolkata	2013	46000	Petrol	Manual	First	18.9 kmpl	998 CC	67.1 bhp

```
df.head()
```

	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

```
df.tail()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seat
6014	6014	Maruti Swift VDI	Delhi	2014	27365	Diesel	Manual	First	28.4 kmpl	1248 CC	74 bhp	5.
6015	6015	Hyundai Xcent 1.1 CRDi S	Jaipur	2015	100000	Diesel	Manual	First	24.4 kmpl	1120 CC	71 bhp	5.
6016	6016	Mahindra Xylo D4	Jaipur	2012	55000	Diesel	Manual	Second	14.0	2498	112	8.

```
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
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
```

```
# new_price having 90% above missing value,hence drop the column
# and drop unnamed:0
# alse drp name :because there are 2000 different names
```

```
df['Name'].value_counts()
# also drp name :because there are 2000 different names
```

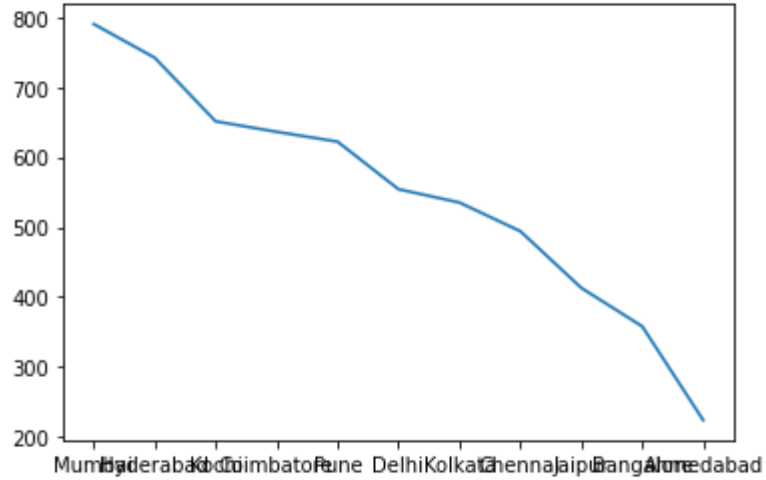
```
Mahindra XUV500 W8 2WD           49
Maruti Swift VDI                 45
Honda City 1.5 S MT              34
Maruti Swift Dzire VDI           34
Maruti Swift VDI BSIV            31
                                 ..
Ford Fiesta Titanium 1.5 TDCi     1
Mahindra Scorpio S10 AT 4WD       1
Hyundai i20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

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

```
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
```

```
# graphically represent the count of a location
import matplotlib.pyplot as plt
plt.plot(loc)
```

```
[<matplotlib.lines.Line2D at 0x7ff29f4adbb0>]
```



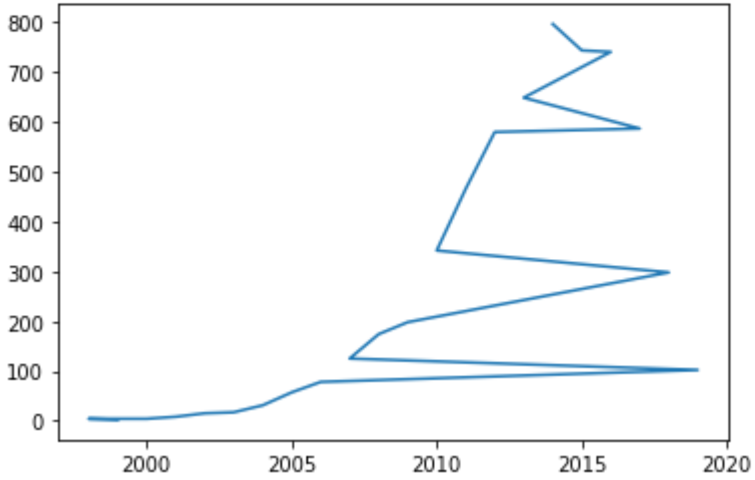
```
year=df['Year'].value_counts()
year
```

```
2014    797
2015    744
```

```
2016    741
2013    649
2017    587
2012    580
2011    466
2010    342
2018    298
2009    198
2008    174
2007    125
2019    102
2006     78
2005     57
2004     31
2003     17
2002     15
2001      8
2000      4
1998      4
1999      2
Name: Year, dtype: int64
```

```
plt.plot(year)
```

```
[<matplotlib.lines.Line2D at 0x7ff29f40d1c0>]
```



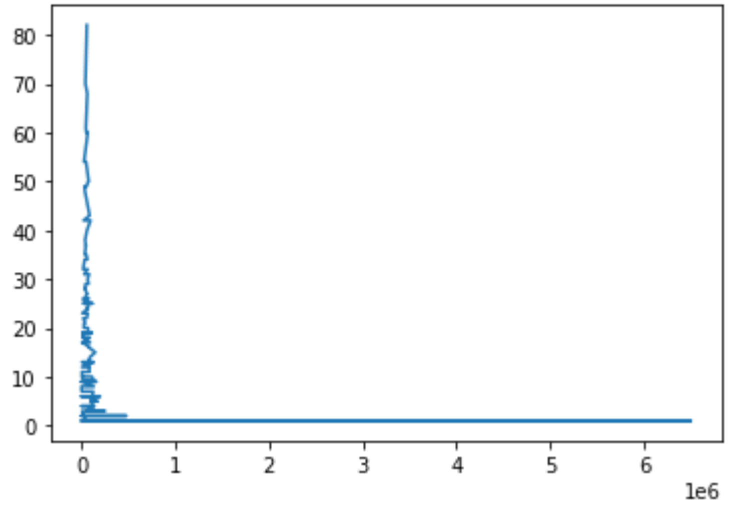
```
kd=df['Kilometers_Driven'].value_counts()
kd
```

```
60000    82
```

```
60000     82
45000     70
65000     68
50000     61
55000     60
          ..
28937      1
82085      1
68465      1
63854      1
27365      1
Name: Kilometers_Driven, Length: 3093, dtype: int64
```

```
plt.plot(kd)
```

```
[<matplotlib.lines.Line2D at 0x7ff29f3de310>]
```

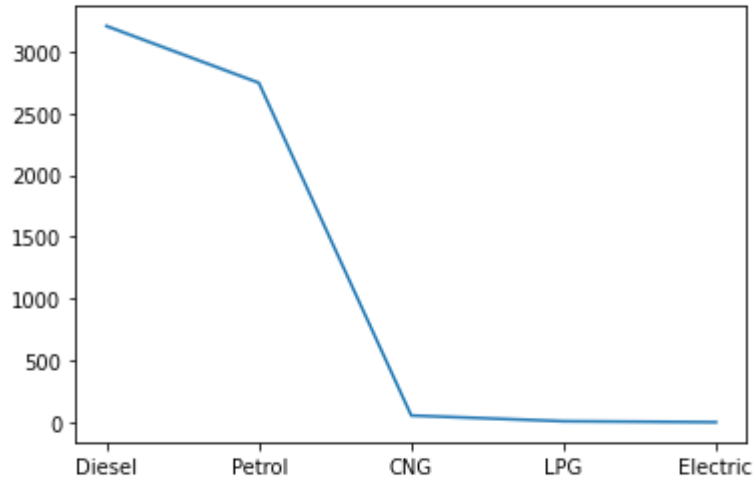


```
ft=df['Fuel_Type'].value_counts()
ft
```

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

```
plt.plot(ft)
```

```
[<matplotlib.lines.Line2D at 0x7ff29f3aefd0>]
```

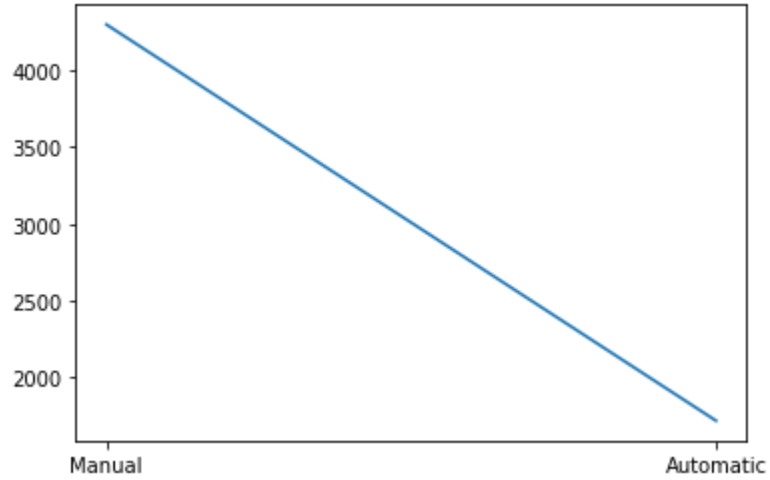


```
trans=df['Transmission'].value_counts()
trans
```

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

```
plt.plot(trans)
```

```
[<matplotlib.lines.Line2D at 0x7ff29f306430>]
```



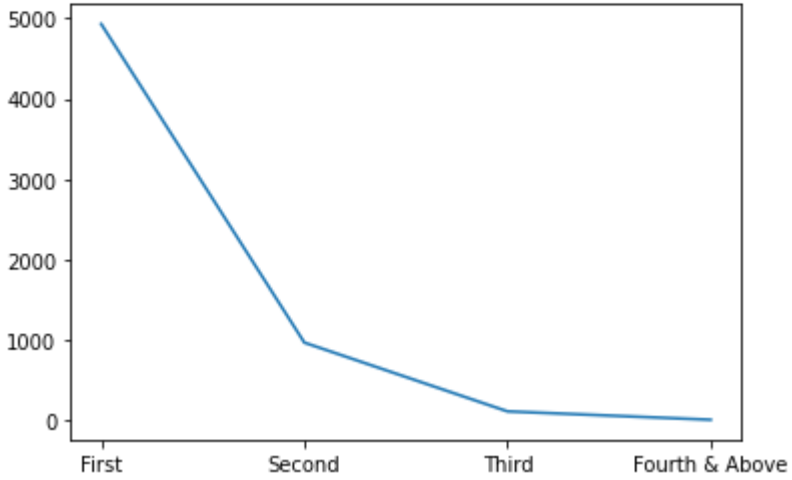
```
ot=df['Owner_Type'].value_counts()
```

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

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

```
plt.plot(ot)
```

```
[<matplotlib.lines.Line2D at 0x7ff29f2c2d90>]
```

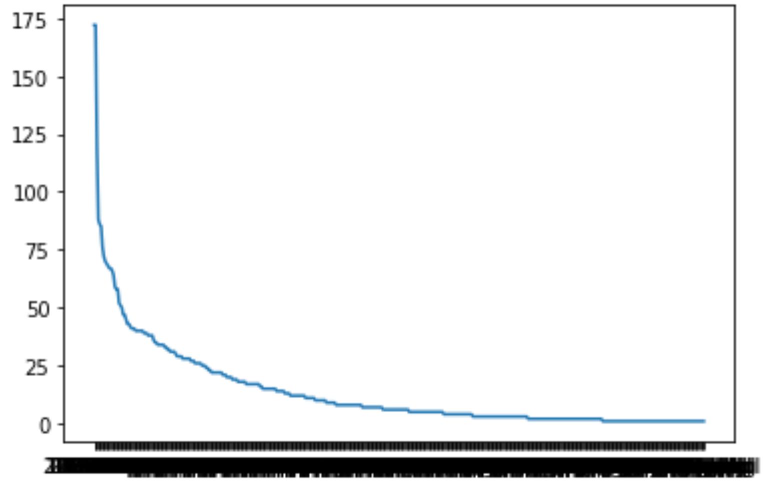


```
milage=df['Mileage'].value_counts()
milage
```

```
18.9 kmpl     172
17.0 kmpl     172
18.6 kmpl     119
20.36 kmpl     88
21.1 kmpl      86
             ...
27.28 kmpl      1
14.57 kmpl      1
22.8 km/kg      1
8.0 kmpl        1
17.24 kmpl      1
Name: Mileage, Length: 442, dtype: int64
```

```
plt.plot(milage)
```

```
[<matplotlib.lines.Line2D at 0x7ff29f298d90>]
```

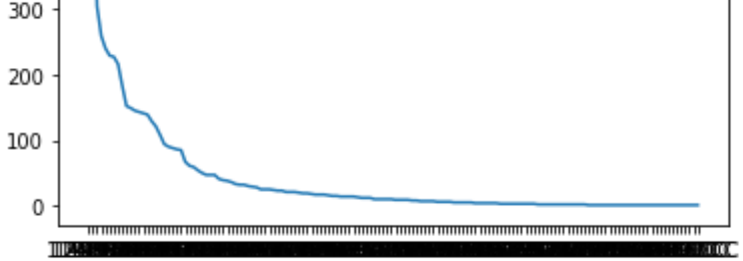


```
engine=df['Engine'].value_counts()
engine
```

```
1197 CC    606
1248 CC    512
1498 CC    304
998 CC     259
2179 CC    240
          ...
2999 CC      1
2147 CC      1
2495 CC      1
3200 CC      1
1797 CC      1
Name: Engine, Length: 146, dtype: int64
```

```
plt.plot(engine)
```

```
[<matplotlib.lines.Line2D at 0x7ff29efc1af0>]
```

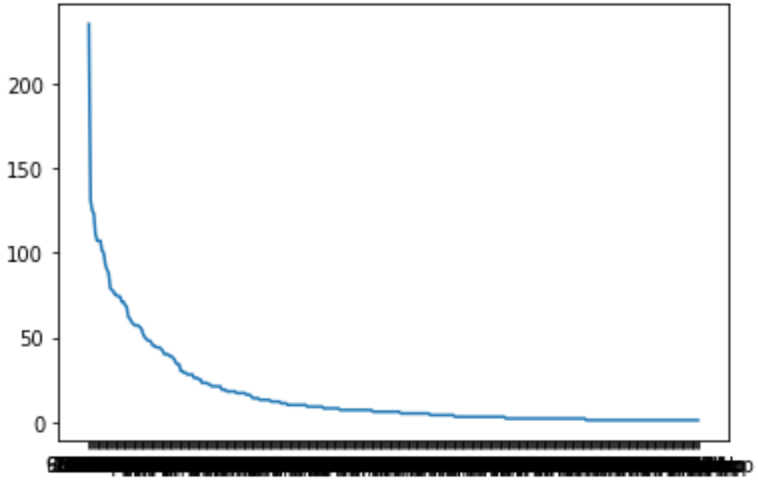


```
power=df['Power'].value_counts()
power
```

```
74 bhp         235
98.6 bhp       131
73.9 bhp       125
140 bhp        123
78.9 bhp       111
              ...
76.9 bhp         1
201 bhp          1
199.3 bhp        1
95 bhp           1
181.04 bhp       1
Name: Power, Length: 372, dtype: int64
```

```
plt.plot(power)
```

```
[<matplotlib.lines.Line2D at 0x7ff29ed5c250>]
```

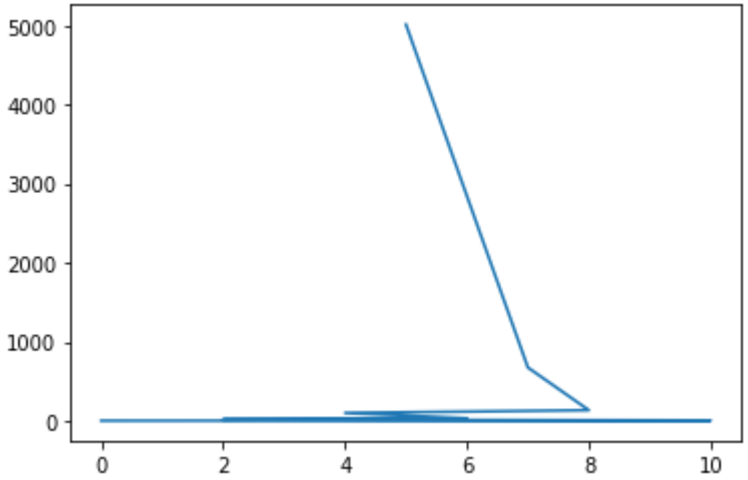


```
seats=df['Seats'].value_counts()
seats
```

```
5.0     5014
7.0      674
8.0      134
4.0       99
6.0       31
2.0       16
10.0       5
9.0        3
0.0        1
Name: Seats, dtype: int64
```

```
plt.plot(seats)
```

```
[<matplotlib.lines.Line2D at 0x7ff29eb028e0>]
```



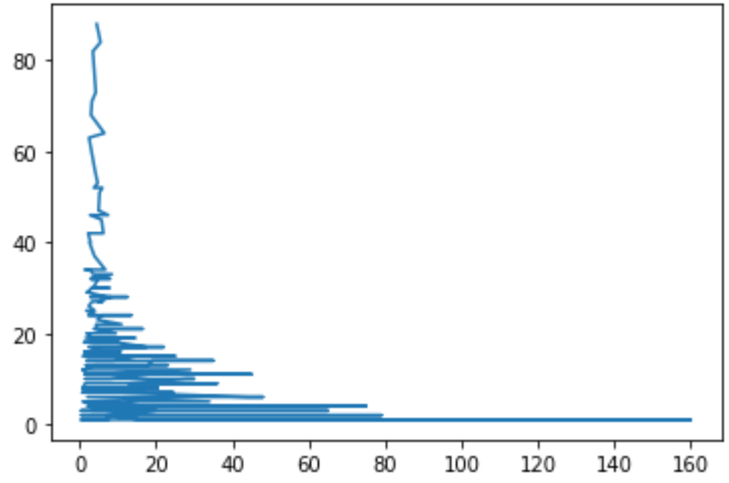
```
price=df['Price'].value_counts()
price
```

```
4.50     88
5.50     84
3.50     82
4.25     73
3.25     71
         ..
11.62     1
```

```
43.60     1
19.05     1
3.94      1
7.43      1
Name: Price, Length: 1373, dtype: int64
```

```
plt.plot(price)
```

```
[<matplotlib.lines.Line2D at 0x7ff29e9e7550>]
```



encoding ==>get_dummy

```
# encoding technique inside pandas  (other are inside machine learning)
```

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

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi	Location_Hyderabad	Location_Jaipur	Locatio
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 × 18 columns

```
# combine encoding dummy data frme and others
df1=pd.concat([df,dummy],axis=1)
df1
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	...	Locat
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 × 32 columns

```
dfe=df1.drop(['Unnamed: 0','Name','Location','Fuel_Type','Transmission','Owner_Type','New_Price','Fuel_Type_Electric'],axis=1)
dfe
```

	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	1968	140.8	5.0	17.74	0	0	1

4	2013	40670	kmpl	CC	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 × 24 columns

```
# replace the string in  te column
dfe['Mileage']=dfe['Mileage'].str.replace('km/kg','')
dfe['Mileage']=dfe['Mileage'].str.replace('kmpl','')
dfe['Power']=dfe['Power'].str.replace('bhp','')
dfe['Engine']=dfe['Engine'].str.replace('CC','')
dfe
```

	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 × 24 columns

```
# null value can be replaced by 0(string)
dfe['Mileage']=dfe['Mileage'].str.replace('null','0')
dfe['Power']=dfe['Power'].str.replace('null','0')
dfe['Engine']=dfe['Engine'].str.replace('null','0')
dfe
```

	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

6018	2011	47000	25.44	936	57.6	5.0	2.50	0	0	0

6019 rows × 24 columns

```
# check the columns having intiger
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_LPG                 uint8
Fuel_Type_Petrol              uint8
Transmission_Manual           uint8
Owner_Type_Fourth & Above     uint8
Owner_Type_Second             uint8
Owner_Type_Third              uint8
dtype: object
```

```
# here milage ,engine,power are not intiger
# these are converted into float
dfe['Mileage']=dfe['Mileage'].astype(float)
dfe['Engine']=dfe['Engine'].astype(float)
dfe['Power']=dfe['Power'].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_LPG                  uint8
Fuel_Type_Petrol               uint8
Transmission_Manual            uint8
Owner_Type_Fourth & Above      uint8
Owner_Type_Second              uint8
Owner_Type_Third               uint8
dtype: object
```

```
# milage power and engine  =====>null=====>0
# it is not posiible
# 0 replaced by Nan (ACT AS MISSING VALUE)
# intiger / float can be replaced
# missing value increased
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
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Location_Bangalore	Location_Chennai	Location_Coimbatore
0	2010	72000	26.60	998.0	58.16	5.0	1.75	0	0	0
1	2015	41000	19.67	1582.0	126.20	5.0	12.50	0	0	0
2	2011	46000	18.20	1199.0	88.70	5.0	4.50	0	1	0

3	2012	87000	20.77	1248.0	88.76	7.0	6.00	0	1	0
4	2013	40670	15.20	1968.0	140.80	5.0	17.74	0	0	1
...	...	...	...	...	...	...	...	...	...	..
6014	2014	27365	28.40	1248.0	74.00	5.0	4.75	0	0	0
6015	2015	100000	24.40	1120.0	71.00	5.0	4.00	0	0	0
6016	2012	55000	14.00	2498.0	112.00	8.0	2.90	0	0	0
6017	2013	46000	18.90	998.0	67.10	5.0	2.65	0	0	0
6018	2011	47000	25.44	936.0	57.60	5.0	2.50	0	0	0

6019 rows × 24 columns

```
# find the miising value
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_LPG                0
Fuel_Type_Petrol             0
Transmission_Manual          0
```

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

```
# milage engine power can be replaced by mean
# seat will be filled by mode
dfe['Engine']=dfe['Engine'].fillna(dfe['Engine'].mean())
dfe['Power']=dfe['Power'].fillna(dfe['Power'].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
    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
    dtype: int64
```

```
# seperate x and y
# x is except price
# y is price
# y is price
x=dfe.drop(['Price'],axis=1)
y=dfe['Price']
```

```
# we need the testing data test-data.csv
```

test-data.csv

```
import numpy as np
import pandas as pd
dfx=pd.read_csv("/content/test-data.csv")
dfx
```

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

		Diesel Trendline							kmpl	CC	bhp
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	First	17.21 kmpl	1197 CC	103.6 bhp
1231	1231	Nissan Micra	Kolkata	2012	28000	Diesel	Manual	First	23.08	1461	63.1

```
dfx.head()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Seats
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 2016-2019 LXI	Coimbatore	2013	54493	Petrol	Manual	Second	24.7 kmpl	796 CC	47.3 bhp	5.0
		Toyota Innova										

```
dfx.tail()
```

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type	Mileage	Engine	Power	Se
1229	1229	Volkswagen Vento Diesel Trendline	Hyderabad	2011	89411	Diesel	Manual	First	20.54 kmpl	1598 CC	103.6 bhp	
1230	1230	Volkswagen Polo GT TSI	Mumbai	2015	59000	Petrol	Automatic	First	17.21 kmpl	1197 CC	103.6 bhp	
1231	1231	Nissan Micra Diesel XV	Kolkata	2012	28000	Diesel	Manual	First	23.08 kmpl	1461 CC	63.1 bhp	
1232	1232	Volkswagen Polo GT TSI	Pune	2013	52262	Petrol	Automatic	Third	17.2 kmpl	1197 CC	103.6 bhp	
		Mercedes-										

```
                    Mercedes
                        Benz
                     E Class                                                       10.0     2148     170
```

```
dfx.shape
```

```
(1234, 13)
```

```
dfx.columns
```

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

```
dfx.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
```

```
dfx['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
```

```
    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
```

```
dfx['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
```

```
dfx['Year'].value_counts()
```

```
    2015    185
    2016    145
    2013    142
    2014    128
    2017    122
    2011    113
    2012    110
    2010     65
    2018     63
    2009     54
    2008     33
    2007     23
    2019     17
    2006     11
    2005     11
    2004      4
    2003      3
    2002      3
    1996      1
    2000      1
    Name: Year, dtype: int64
```

```
dfx['Kilometers_Driven'].value_counts()
```

```
dfx['Kilometers_Driven'].value_counts()
```

```
65000    18
70000    17
45000    16
46000    14
60000    14
         ..
45250     1
89190     1
48184     1
42125     1
72443     1
Name: Kilometers_Driven, Length: 755, dtype: int64
```

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

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

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

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

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

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

```
dfx['Mileage'].value_counts()
```

```
17.0 kmpl     35
18.9 kmpl     29
18.6 kmpl     25
21.1 kmpl     20
```

```
    21.1 kmpl     20
    18.0 kmpl     20
                  ..
    23.5 kmpl      1
    17.19 kmpl     1
    21.02 kmpl     1
    19.33 kmpl     1
    27.28 kmpl     1
    Name: Mileage, Length: 301, dtype: int64
```

```
dfx['Engine'].value_counts()
```

```
    1197 CC    126
    1248 CC     98
    1498 CC     66
    1198 CC     54
    998 CC      50
              ...
    1948 CC      1
    1299 CC      1
    5998 CC      1
    2362 CC      1
    1047 CC      1
    Name: Engine, Length: 104, dtype: int64
```

```
dfx['Power'].value_counts()
```

```
    74 bhp        45
    98.6 bhp      35
    73.9 bhp      27
    82 bhp        23
    null bhp      22
                  ..
    97.7 bhp       1
    161 bhp        1
    167.7 bhp      1
    245.41 bhp     1
    92.7 bhp       1
    Name: Power, Length: 249, dtype: int64
```

```
dfx['Seats'].value_counts()
```

```
5.0     1033
7.0      122
8.0       36
4.0       20
6.0        7
10.0       3
2.0        2
Name: Seats, dtype: int64
```

```
# get dummy===>(split columns)
```

```
dummy1=pd.get_dummies(dfx[['Location','Fuel_Type','Transmission','Owner_Type']],drop_first=True)
dummy1
```

	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 × 17 columns

```
# combine encoding dummy data frme and others
df2=pd.concat([dfx,dummy1],axis=1)
df2
```

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

```
dfa=df2.drop(['Unnamed: 0','Name','Location','Fuel_Type','Transmission','Owner_Type','New_Price'],axis=1)
dfa
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Locat
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	

			kmpl	CC	bhp				
1233	2014	72443	10.0 kmpl	2148 CC	170 bhp	5.0	0	0	0

1234 rows × 23 columns

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

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Locat
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	null	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 × 23 columns

```
dfa['Mileage']=dfa['Mileage'].str.replace('null','0')
dfa['Power']=dfa['Power'].str.replace('null','0')
dfa['Engine']=dfa['Engine'].str.replace('null','0')
dfa
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Bangalore	Location_Chennai	Location_Coimbatore	Locat
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 × 23 columns

```
dfa.dtypes
```

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

```
    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
```

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

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

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

1234 rows × 23 columns

```
dfa.isna().sum()
```

```
Year                            0
Kilometers_Driven               0
Mileage                        12
```

```
    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
```

```
dfa['Engine']=dfa['Engine'].fillna(dfe['Engine'].mean())
dfa['Power']=dfa['Power'].fillna(dfe['Power'].mean())
dfa['Mileage']=dfa['Mileage'].fillna(dfe['Mileage'].mean())
dfa['Seats']=dfa['Seats'].fillna(dfe['Seats'].mode()[0])
dfa.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.linear_model import LinearRegression
model=LinearRegression()
model.fit(x,y)
y_pred=model.predict(dfa)
y_pred
```

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