

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

	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmi
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	Au

```
df.drop(["Unnamed: 0"],axis=1, inplace=True)
df.head()
```

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

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

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

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

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

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

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

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

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

```
df.isnull().sum()
```

```
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.shape

(6019, 13)
```

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

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

6019 rows × 18 columns



```
dfe = pd.concat([df, dummy],axis=1)
dfe=dfe.drop(['Location','Fuel_Type','Transmission','Owner_Type','New_Price','Name
dfe
```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Locatio
0	2010	72000	26.6 km/kg	998 CC	58.16 bhp	5.0	1.75	
1	2015	41000	19.67 kmpl	1582 CC	126.2 bhp	5.0	12.50	
2	2011	46000	18.2 kmpl	1199 CC	88.7 bhp	5.0	4.50	
3	2012	87000	20.77 kmpl	1248 CC	88.76 bhp	7.0	6.00	
4	2013	40670	15.2 kmpl	1968 CC	140.8 bhp	5.0	17.74	

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

```

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

```

```
dfe.dtypes
```

```

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

```

```

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

```

#To replace object such as unit to number

```

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

```

```

dfe['Power']=dfe['Power'].str.replace('null','0')

```

```

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

```

dfe.dtypes

```

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

```

```

Owner_Type_Second      uint8
Owner_Type_Third       uint8
dtype: object

```

dfe.dtypes

```

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

```

dfe.isna().sum()

```

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

```

```

Owner_Type_Second      0
Owner_Type_Third       0
dtype: int64

```

```

import numpy as np
dfe.loc[dfe.Engine==0, 'Engine']=np.NaN
dfe.loc[dfe.Power==0, 'Power']=np.NaN

```

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

```

Year      0
Kilometers_Driven  0
Mileage    2
Engine     36
Power     143
Seats      42
Price      0
Location_Bangalore  0
Location_Chennai  0
Location_Coimbatore  0
Location_Delhi  0
Location_Hyderabad  0
Location_Jaipur  0
Location_Kochi  0
Location_Kolkata  0
Location_Mumbai  0
Location_Pune  0
Fuel_Type_Diesel  0
Fuel_Type_Electric  0
Fuel_Type_LPG  0
Fuel_Type_Petrol  0
Transmission_Manual  0
Owner_Type_Fourth & Above  0
Owner_Type_Second  0
Owner_Type_Third  0
dtype: int64

```

```

dfe=dfe.drop('Fuel_Type_Electric',axis=1)
dfe['Engine']=dfe['Engine'].fillna(dfe['Engine'].mean())
dfe['Power']=dfe['Power'].fillna(dfe['Power'].mean())
dfe['Seats']=dfe['Seats'].fillna(dfe['Seats'].mode()[0])
dfe['Mileage']=dfe['Mileage'].fillna(dfe['Mileage'].mean())
dfe

```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Price	Locatio
0	2010	72000	26.60	998.0	58.16	5.0	1.75	
1	2015	41000	19.67	1582.0	126.20	5.0	12.50	
2	2011	46000	18.20	1199.0	88.70	5.0	4.50	
3	2012	87000	20.77	1248.0	88.76	7.0	6.00	
4	2013	40670	15.20	1968.0	140.80	5.0	17.74	
...

```
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

```

```

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

```

```

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

```

```
LinearRegression()
```

```

dft=pd.read_csv('test-data.csv')
dft

```


	Unnamed: 0	Name	Location	Year	Kilometers_Driven	Fuel_Type	Træ
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 × 13 columns

```
dft.drop(["Unnamed: 0"],axis=1, inplace=True)
dft.head()
```

	Name	Location	Year	Kilometers_Driven	Fuel_Type	Transmission	Owner_Type
0	Maruti Alto K10 LXI CNG	Delhi	2014	40929	CNG	Manual	
1	Maruti Alto 800 2016- 2019 LXI	Coimbatore	2013	54493	Petrol	Manual	
2	Toyota Innova Crysta Touring Sport 2.4 MT	Mumbai	2017	34000	Diesel	Manual	
3	Toyota Etios Live CD	Hyderabad	2012	139000	Diesel	Manual	

```
dft['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
```

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

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

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

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

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

```
First      1023
Second     184
Third       24
```

Fourth & Above 3
Name: Owner_Type, dtype: int64

dft.isnull().sum()

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

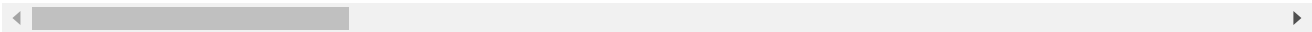
dft.shape

(1234, 12)

dummy1=pd.get_dummies(dft[['Location','Fuel_Type','Transmission','Owner_Type']], d
dummy1

	Location_Bangalore	Location_Chennai	Location_Coimbatore	Location_Delhi
0	0	0	0	0
1	0	0	1	0
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

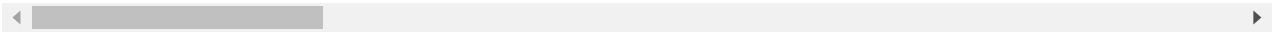


dft1 = pd.concat([dft, dummy1],axis=1)
dft1=dft1.drop(['Location','Fuel_Type','Transmission','Owner_Type','New_Price','Nai

dft1

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

1234 rows × 23 columns



dft1.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

```

```

dft1['Mileage']=dft1['Mileage'].str.replace('kmpl','')
dft1['Mileage']=dft1['Mileage'].str.replace('km/kg','')
dft1['Engine']=dft1['Engine'].str.replace('CC','')
dft1['Power']=dft1['Power'].str.replace('bhp','')

```

```
dft1['Power']=dft1['Power'].str.replace('null','0')
```

```

dft1['Engine']=dft1['Engine'].astype(float)
dft1['Power']=dft1['Power'].astype(float)
dft1['Mileage']=dft1['Mileage'].astype(float)

```

```

import numpy as np
dft1.loc[dft1.Engine==0,'Engine']=np.NaN
dft1.loc[dft1.Power==0,'Power']=np.NaN

```

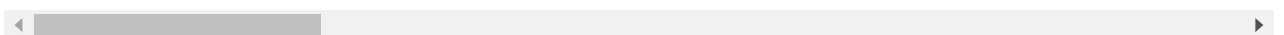
```

dft1['Engine']=dft1['Engine'].fillna(dft1['Engine'].mean())
dft1['Power']=dft1['Power'].fillna(dft1['Power'].mean())
dft1['Seats']=dft1['Seats'].fillna(dft1['Seats'].mode()[0])
dft1

```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Ba
0	2014	40929	32.26	998.0	58.20000	4.0	
1	2013	54493	24.70	796.0	47.30000	5.0	
2	2017	34000	13.68	2393.0	147.80000	7.0	
3	2012	139000	23.59	1364.0	110.38042	5.0	
4	2014	29000	18.50	1197.0	82.85000	5.0	
...	
1229	2011	89411	20.54	1598.0	103.60000	5.0	
1230	2015	59000	17.21	1197.0	103.60000	5.0	
1231	2012	28000	23.08	1461.0	63.10000	5.0	
1232	2013	52262	17.20	1197.0	103.60000	5.0	
1233	2014	72443	10.00	2148.0	170.00000	5.0	

1234 rows × 23 columns



```
dft1.isna().sum()
```

```

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

```

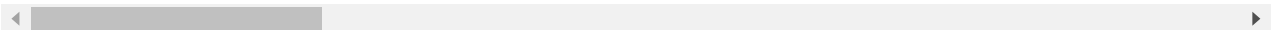
```

x_test=dft1
x_test.head()

```

	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	Location_Banga
0	2014	40929	32.26	998.0	58.20000	4.0	
1	2013	54493	24.70	796.0	47.30000	5.0	
2	2017	34000	13.68	2393.0	147.80000	7.0	
3	2012	139000	23.59	1364.0	110.38042	5.0	
4	2014	29000	18.50	1197.0	82.85000	5.0	

5 rows × 23 columns



```
x_test.columns
```

```

Index(['Year', 'Kilometers_Driven', 'Mileage', 'Engine', 'Power', 'Seats',
      'Location_Bangalore', 'Location_Chennai', 'Location_Coimbatore',
      'Location_Delhi', 'Location_Hyderabad', 'Location_Jaipur',
      'Location_Kochi', 'Location_Kolkata', 'Location_Mumbai',
      'Location_Pune', 'Fuel_Type_Diesel', 'Fuel_Type_LPG',
      'Fuel_Type_Petrol', 'Transmission_Manual', 'Owner_Type_Fourth &
      Above',
      'Owner_Type_Second', 'Owner_Type_Third'],
      dtype='object')

```

```
y_pred=reg.predict(x_test)  
y_pred
```

```
array([ 2.59595345, -1.60908461, 16.27901151, ...,  0.02500755,  
        9.24070839, 21.8250085 ])
```

[Colab paid products](#) - [Cancel contracts here](#)

