Name: yeswanth Vallabhaneni
Student Id: 16359049

```
load_dataset.isnull().sum()
```

	0
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	38
New_Price	5847
Price	0

dtype: int64

	Unnamed: 0	Year	Kilometers_Driven	Mileage	Engine	Power	Seats	New_Price	Price
count	5847.000000	5847.000000	5.847000e+03	5845.000000	5811.000000	5811.000000	5809.000000	0.0	5847.000000
mean	3013.181461	2013.448435	5.841013e+04	18.158496	1631.552573	113.803144	5.286452	NaN	9.653742
std	1736.398890	3.194949	9.237971e+04	4.358246	601.972587	53.896719	0.806668	NaN	11.275966
min	1.000000	1998.000000	1.710000e+02	0.000000	72.000000	34.200000	2.000000	NaN	0.440000
25%	1509.500000	2012.000000	3.346750e+04	15.260000	1198.000000	78.000000	5.000000	NaN	3.550000
50%	3015.000000	2014.000000	5.257600e+04	18.190000	1497.000000	98.600000	5.000000	NaN	5.750000
75%	4517.500000	2016.000000	7.249050e+04	21.100000	1991.000000	139.010000	5.000000	NaN	10.250000
max	6018.000000	2019.000000	6.500000e+06	28.400000	5998.000000	560.000000	10.000000	NaN	160.000000

Mileage:
- The mean (18.16) and median (18.19) for Mileage are very close, that shows minimal skew and a likely symmetric distribution.

- A small standard deviation (4.36) suggests low variability, but minor outliers might still be present.

- Median for imputation is chosen to handle any potential outliers robustly, ensuring typical values remain unaffected.

Engine:

- The mean (1631.55) is higher than the median (1497), indicates a right-skewed distribution.

- A high standard deviation (601.97) indicates significant variability, likely due to extreme values.

- So used the median for imputation to avoid the influence of large engine sizes, giving a more typical, representative value.

Power:

- Mean (113.80) is higher than the median (98.6), indicating potential right skew.

- High standard deviation (53.89) indicates variability due to high-performance vehicles.

- Imputing with the median avoids skew from outliers, gives a more accurate central tendency for typical power values.

Seats:

- Used mean as an imputation ,it maintains the central tendency of the variable and ensures that the imputed values align well with the overall distribution of the data.

New_Price:

- We dropped this column

- The column New_Price has almost all of its values missing.