PDS ASSIGNMENT

* Checking for the null values and checking for the how many null values

data.isna().sum()

Unnamed: 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 5032

Price 0

dtype: int64

* We have missing values in Mileage, Engine, Power, Seats, and New\_Price. So, those missing values can be replaced by using mean, median, and mode.

df1['Mileage'].fillna(df1['Mileage'].median(),inplace=True)

df1['Engine'].fillna(df1['Engine'].mean(),inplace=True)

df1['Power'].fillna(df1['Power'].median(),inplace=True)

df1['Seats'].fillna(df1['Seats'].mode()[0],inplace=True)

df1.drop('New\_Price', axis=1, inplace=True)

* The reason for this is that outliers have little effect on the median, which often reflects the majority of results.
* Since the mode is the most often occurring number of seats, we must use it to replace any missing values in the Seat column.
* We are eliminating the New Price column as it contains more than 50% missing values, which might introduce inaccurate data and lead to wrong analysis.