A.Finding the missing values in all the columns

A screenshot of a computer

Description automatically generated

* From the above output, we can see that the columns Mileage, Engine, Power, Seats, New\_Price have the count of 2,36,36,38,5032 missing values respectively.
* Before imputing, cleaning the data is essential in order perform mean, median, mode. Unit values should be removed from Mileage, Engine, Power as the mean operation can be done only on numerical values.

A screenshot of a computer

Description automatically generated

* Dropping the New\_Price column and first column as the New\_Price column have more number of missing values which doesn’t add any meaning to our analysis. Dropping the column will be the good decision for analysis. Similarly First column shows nothing but row number. Performing mean and mode operations for numerical and character columns respectively.

A screenshot of a computer code

Description automatically generated

Checking the missing values after cleaning the data and performing imputation:

A screen shot of a computer

Description automatically generated

B) Below is the snapshot of clean dataset showing after the removal of unit values from Mileage, Engine, Power and New\_Price columns

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

C. Changing the Categorical columns like Fuel\_Type and Transmission into Numerical columns:

A screenshot of a computer

Description automatically generated

Output:

A screenshot of a computer

Description automatically generated

D. Creating a new variable Current Age where Current Age= Year – Current Year

A screenshot of a computer

Description automatically generated

Output:

A screenshot of a table

Description automatically generated