🡪 **Action for Question – (a)**

Here I’m performing data cleaning on a given dataset, which contains car’s related data, here my goal is to handle missing values, below screenshot shows the missing values in the dataset.

A screen shot of a computer

Description automatically generated

Also, this dataset contains both numerical and categorical type of data.

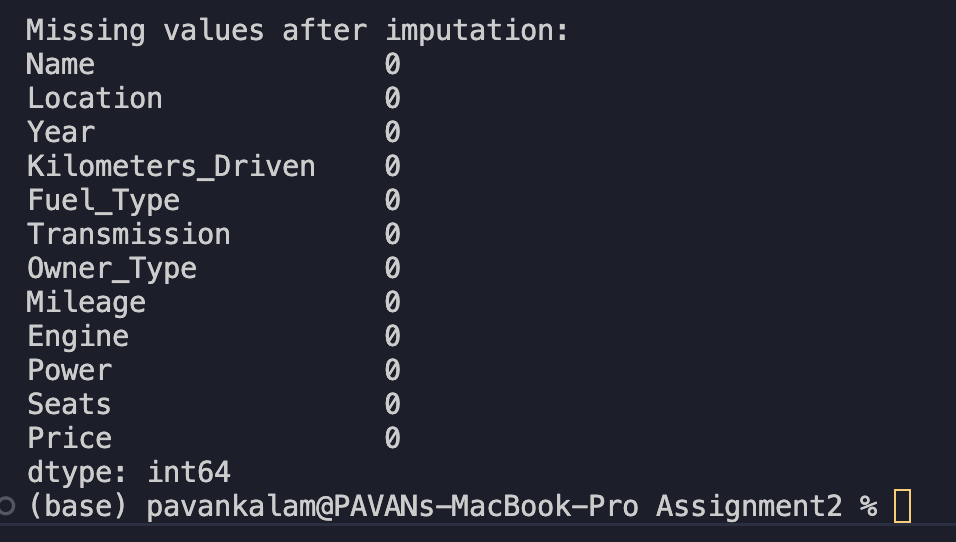
Also, by observing the above screenshot only the numerical columns have missing values.

Moreover, ‘New\_Price’ column has more missing values.

To handle this approach first I tried to handle the missing values other than ‘New\_Price’ column, for ‘categorical columns’ I choose to impute with ‘mode’, and for ‘numerical columns’ I choose to impute with ‘median’.

Once this completes, now started working with the ‘New\_Price’ column. Here, I observed that, for the given dataset total number of rows were ‘5847’, from these many rows ‘5032’ rows were missing for this ‘New\_Price’ column (i.e., around 86% of the rows are missing). So, in this case I choose to drop that column for further analysis.

After all this, the missing values count of the dataset is as below,



This is the action which I took for this task (Question-a)