PDS-Assignment-2

A) Look for the missing values in all the columns and either impute them (replace with mean,

median, or mode) or drop them. Justify your action for this task.

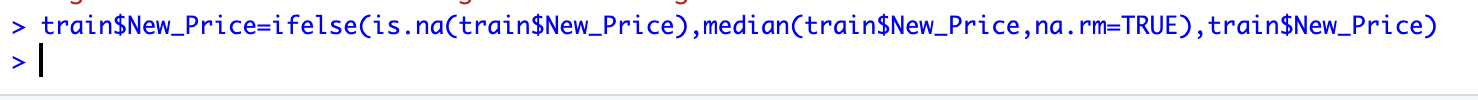
A screenshot of a computer error

Description automatically generatedA screenshot of a data

Description automatically generatedA screenshot of a computer

Description automatically generatedA close-up of a wall

Description automatically generated

A screenshot of a table

Description automatically generated

I have replaced NA values with median as the values have more variance. So, replacing with mean is not ideal and dataset has many NA values in new\_price so dropping the rows results in removing lost of data.

B) Remove the units from some of the attributes and only keep the numerical values (for

example remove kmpl from “Mileage”, CC from “Engine”, bhp from “Power”, and lakh from

“New\_price”).

A computer code with blue and red text

Description automatically generatedA screenshot of a table

Description automatically generated

C) Change the categorical variables (“Fuel\_Type” and “Transmission”) into numerical one hot

encoded value

A screenshot of a computer

Description automatically generated

D) Create one more feature and add this column to the dataset (you can use mutate function in

R for this). For example, you can calculate the current age of the car by subtracting “Year” value

from the current yea

A close-up of a computer code

Description automatically generatedA screenshot of a table

Description automatically generated