ASSIGNMENT-2

16352972

KAVYA BALINENI

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.

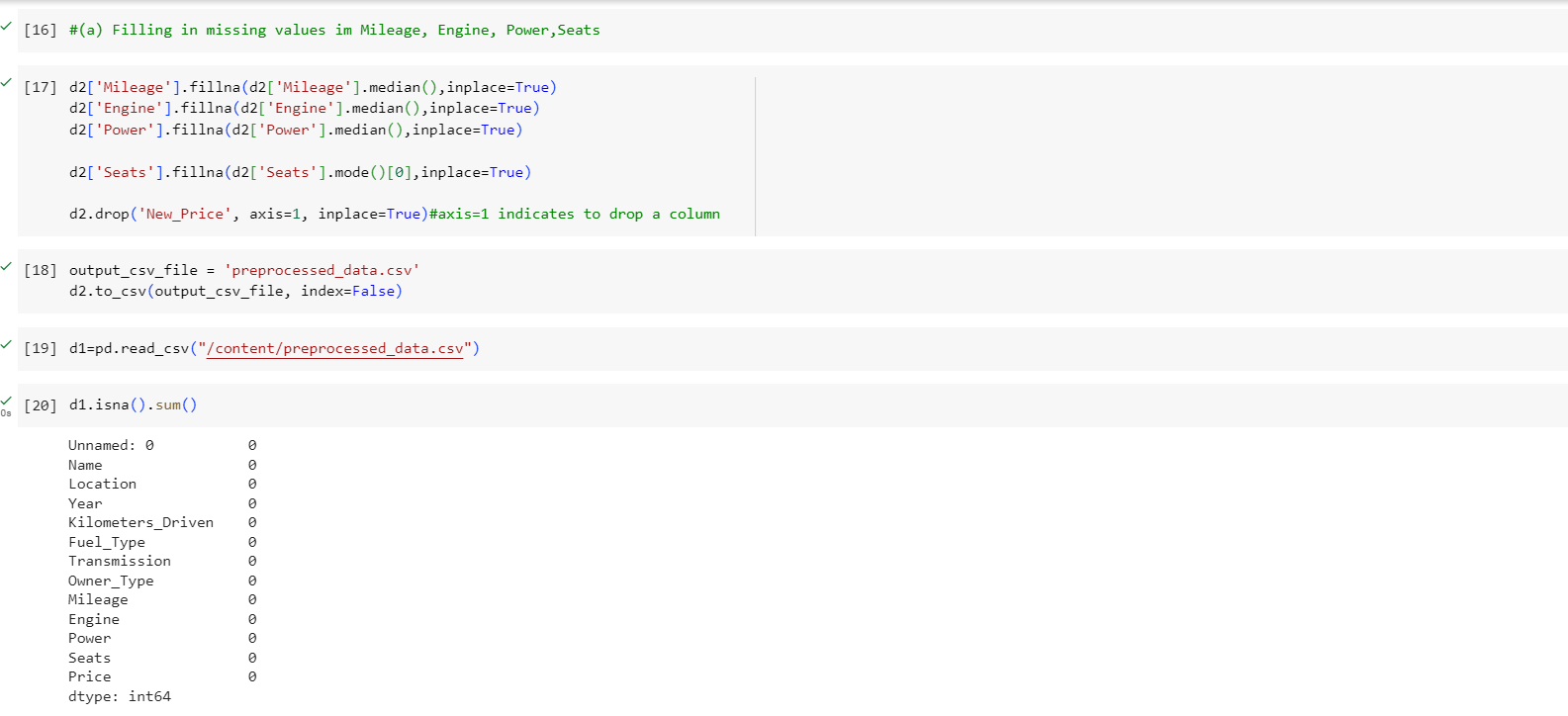
A screenshot of a computer

Description automatically generated

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 screenshot of a computer

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 year.

A screenshot of a computer

Description automatically generated

e) Perform select, filter, rename, mutate, arrange, and summarize with group by operations (or their equivalent operations in python) on this dataset.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated