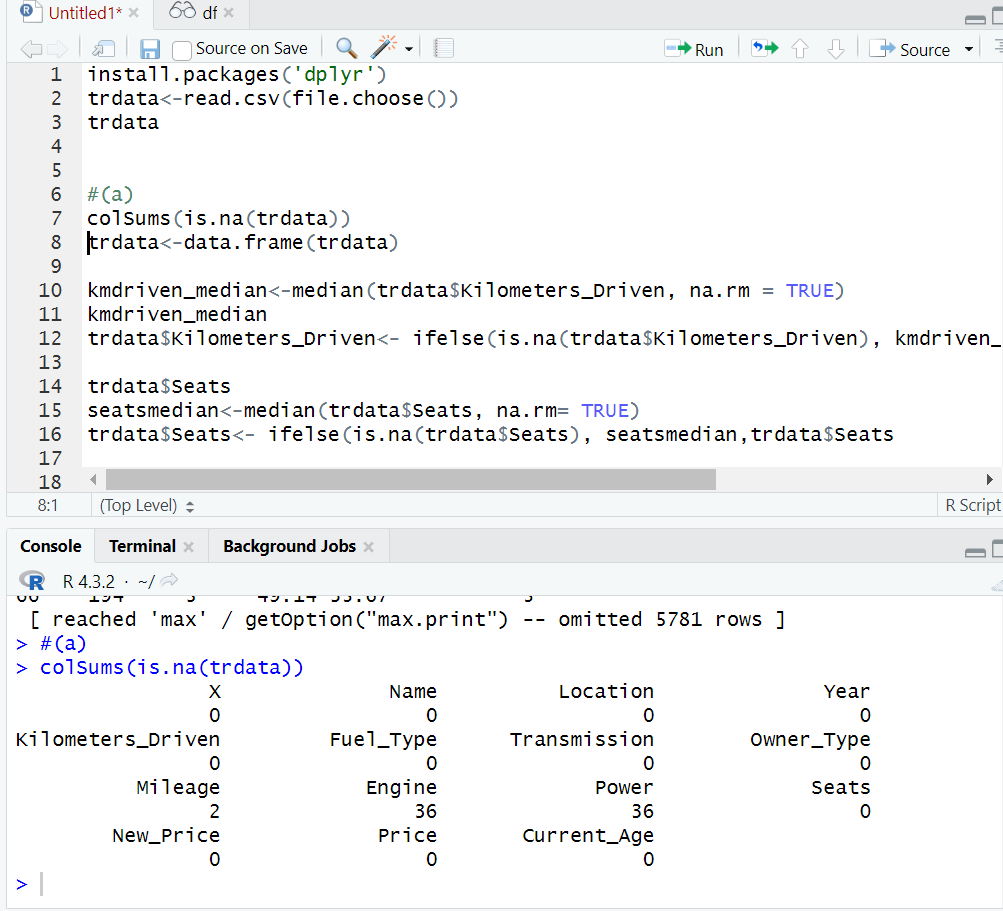
a) Look for the missing values in all the columns and either impute them (replace with mean,

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

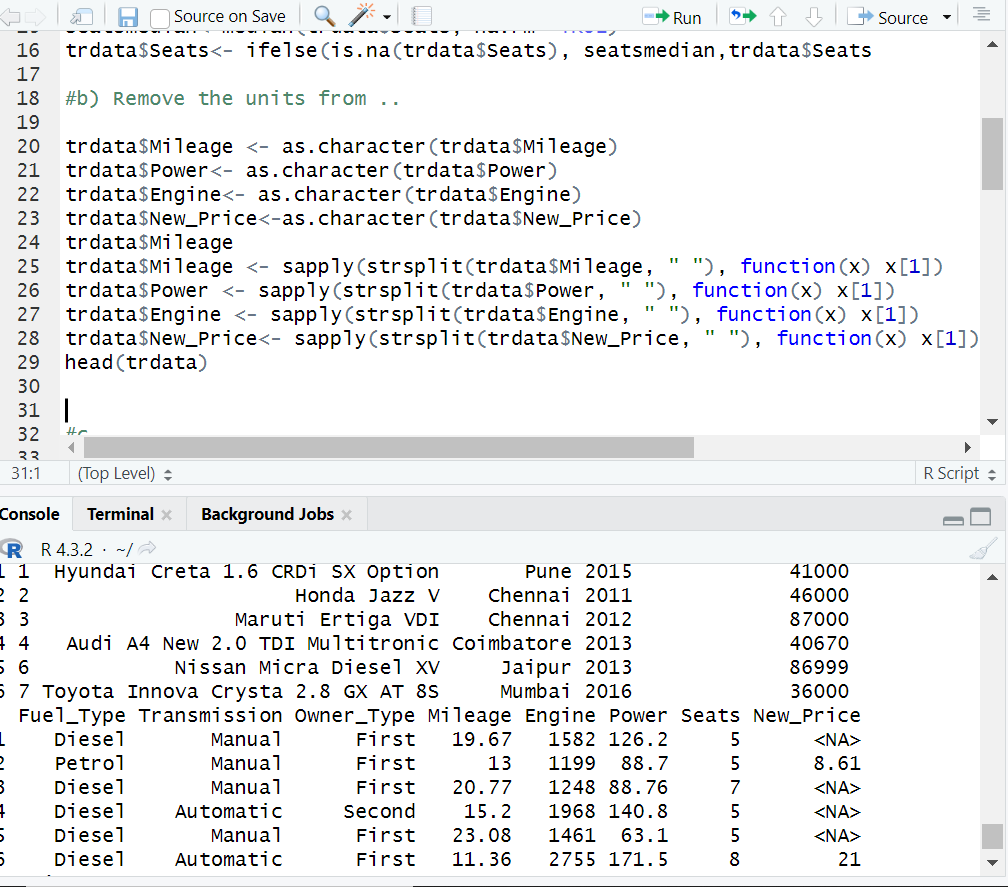
* I selected median because it gave more accurate average value ( since there might be some cars which are used very less or too much so they show extreme values )



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

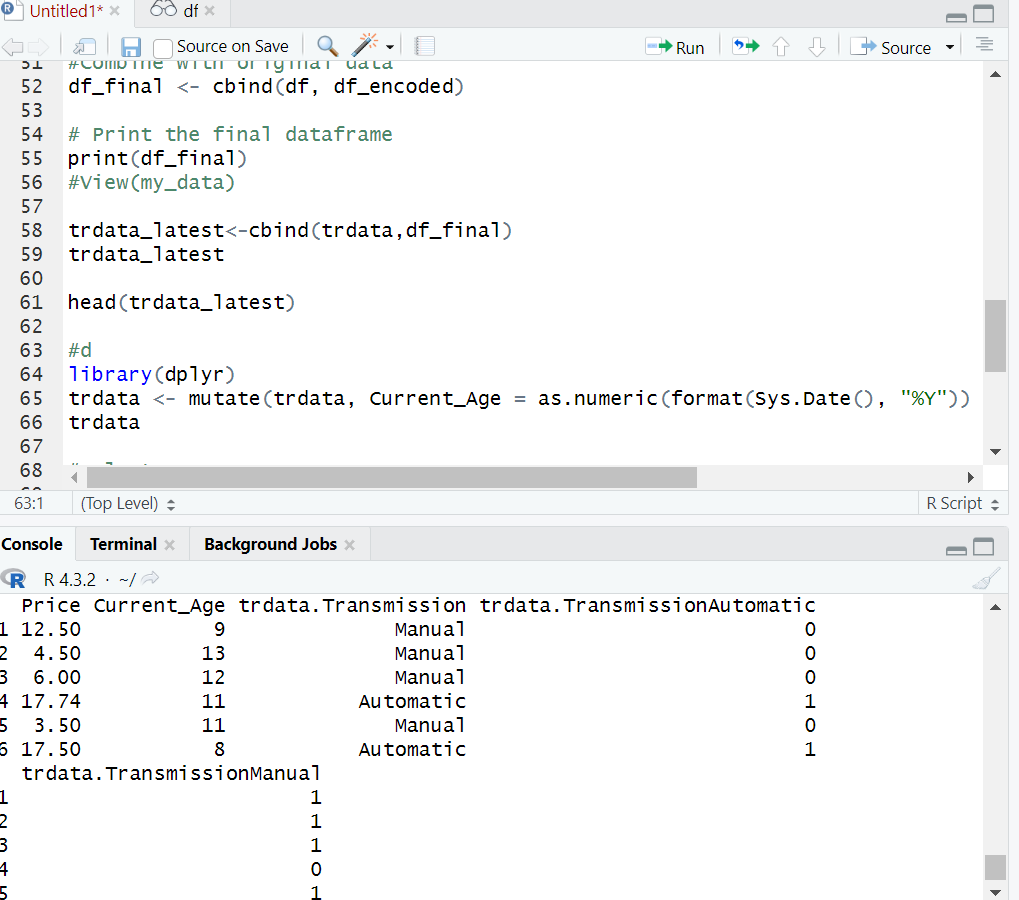
example remove kmpl from “Mileage”, CC from “Engine”, bhp from “Power”, and lakh from

“New\_price”).



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

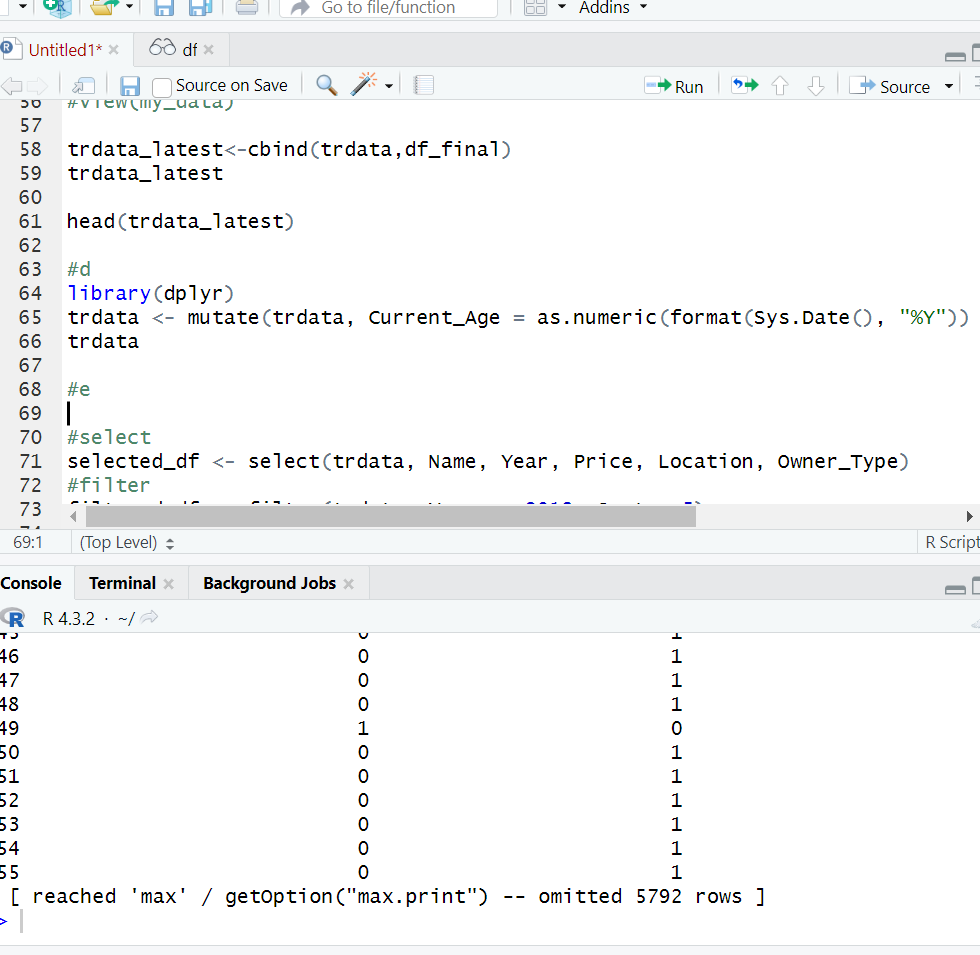
encoded value.



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.



e) Perform select, filter, rename, mutate, arrange and summarize with group by operations (or

their equivalent operations in python) on this dataset.

