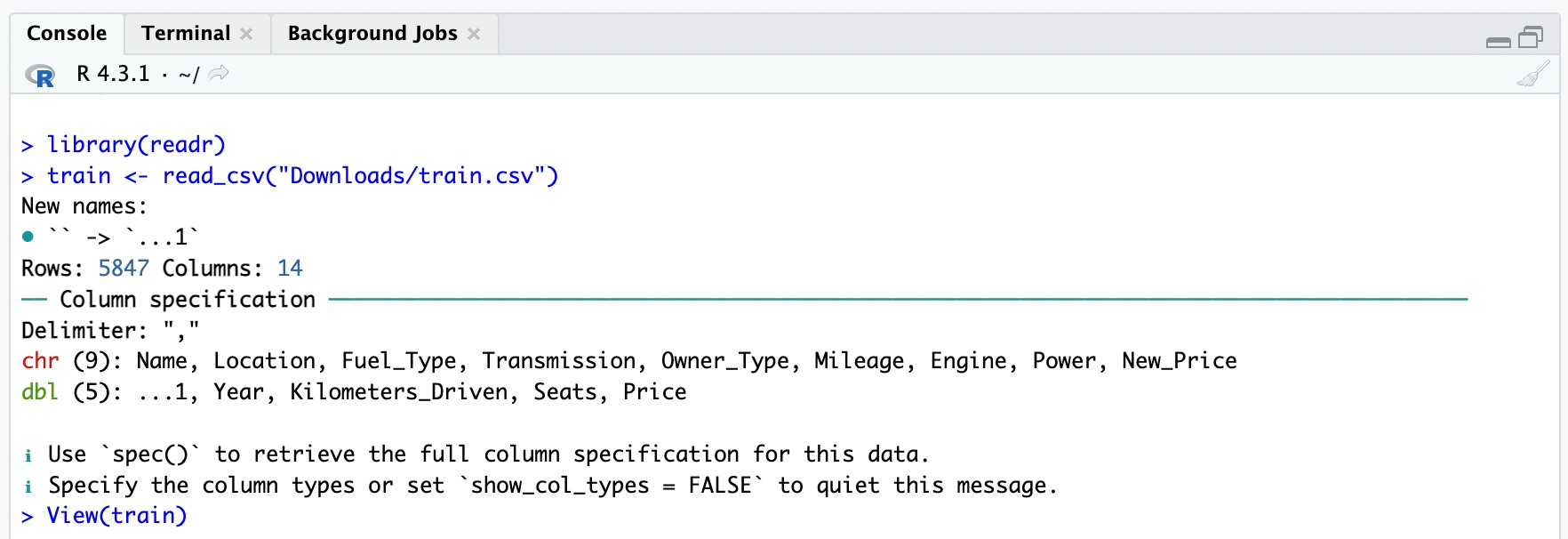
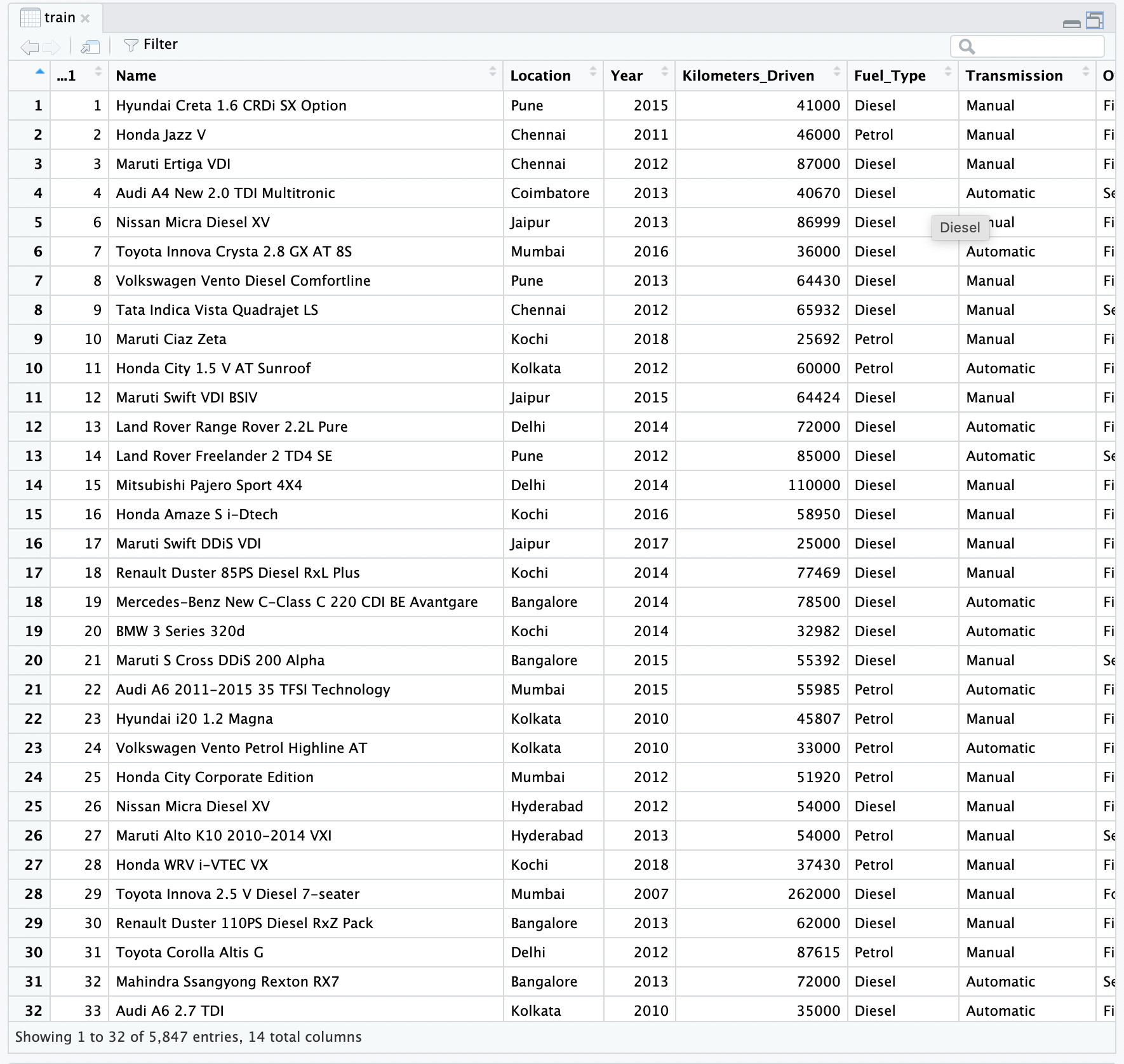
**Assignment – 2**

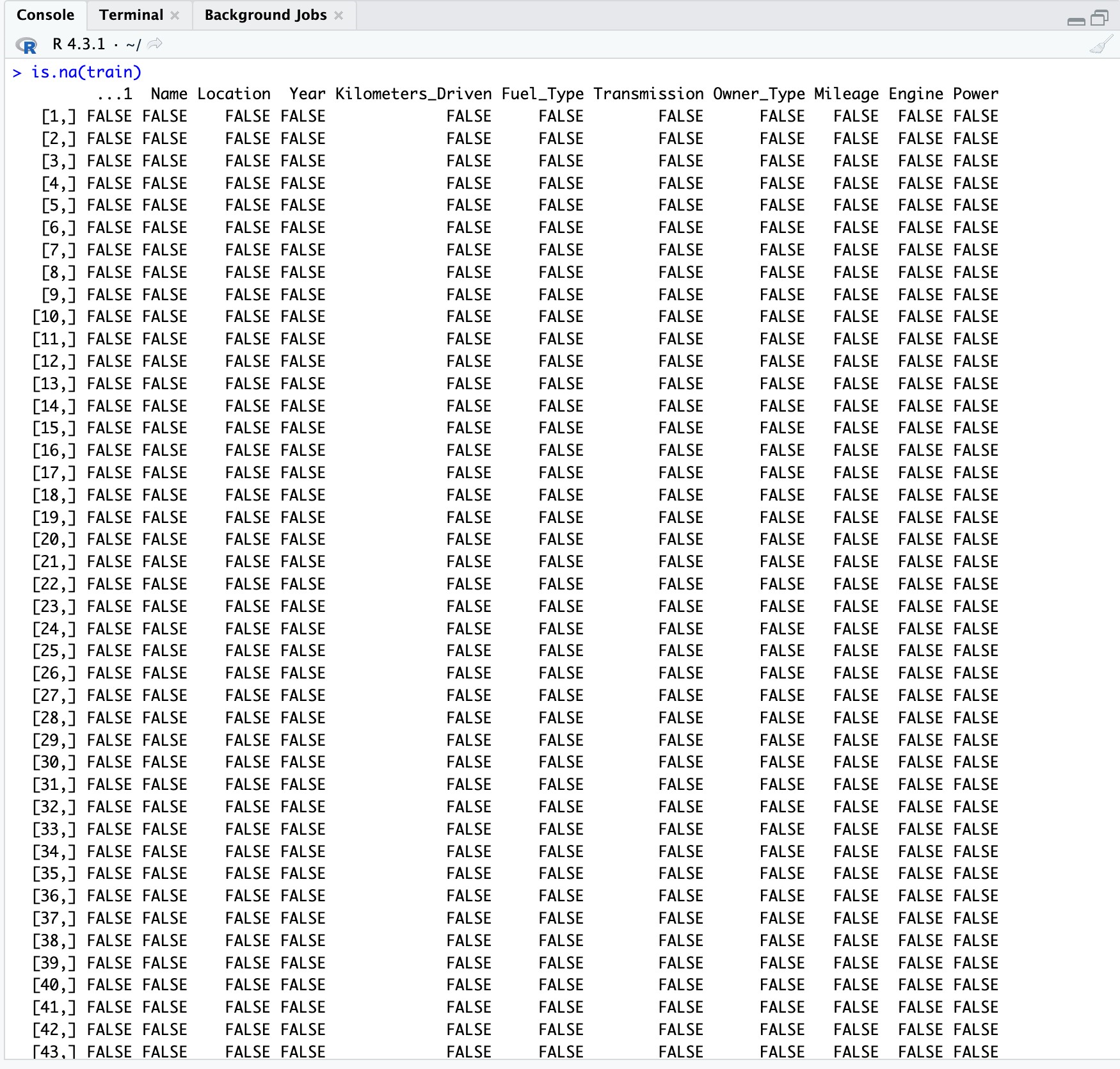
**a) Look for the missing values in all the columns and either impute them (replace with mean,**

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



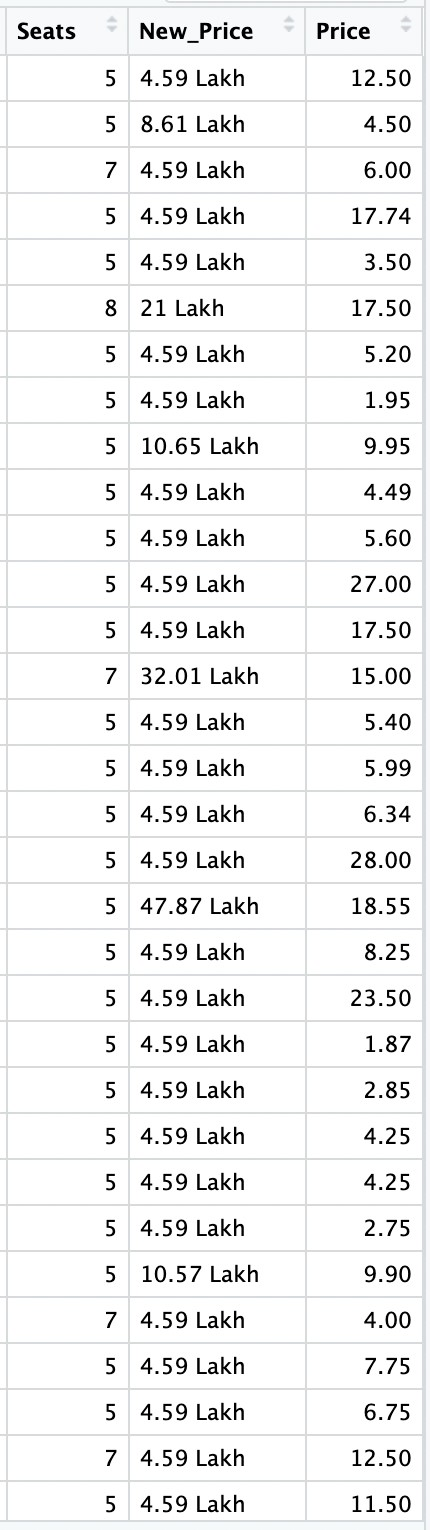


+`







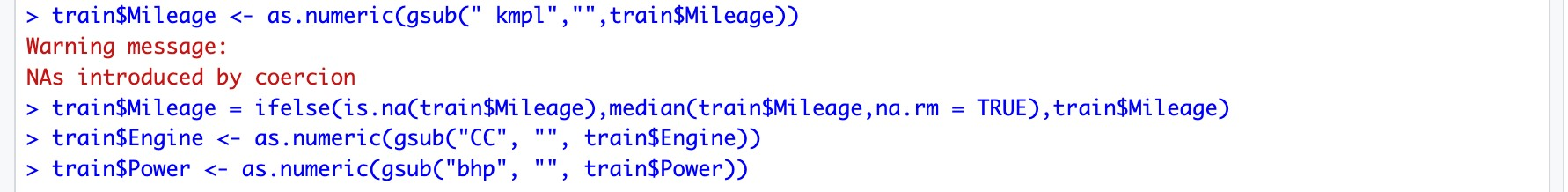


Since the values have higher variance, I have substituted the median for the NA values. Because there are a lot of NA values in the new\_price dataset and replacing with the mean is not optimal, eliminating the rows eliminates lost data.

**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**

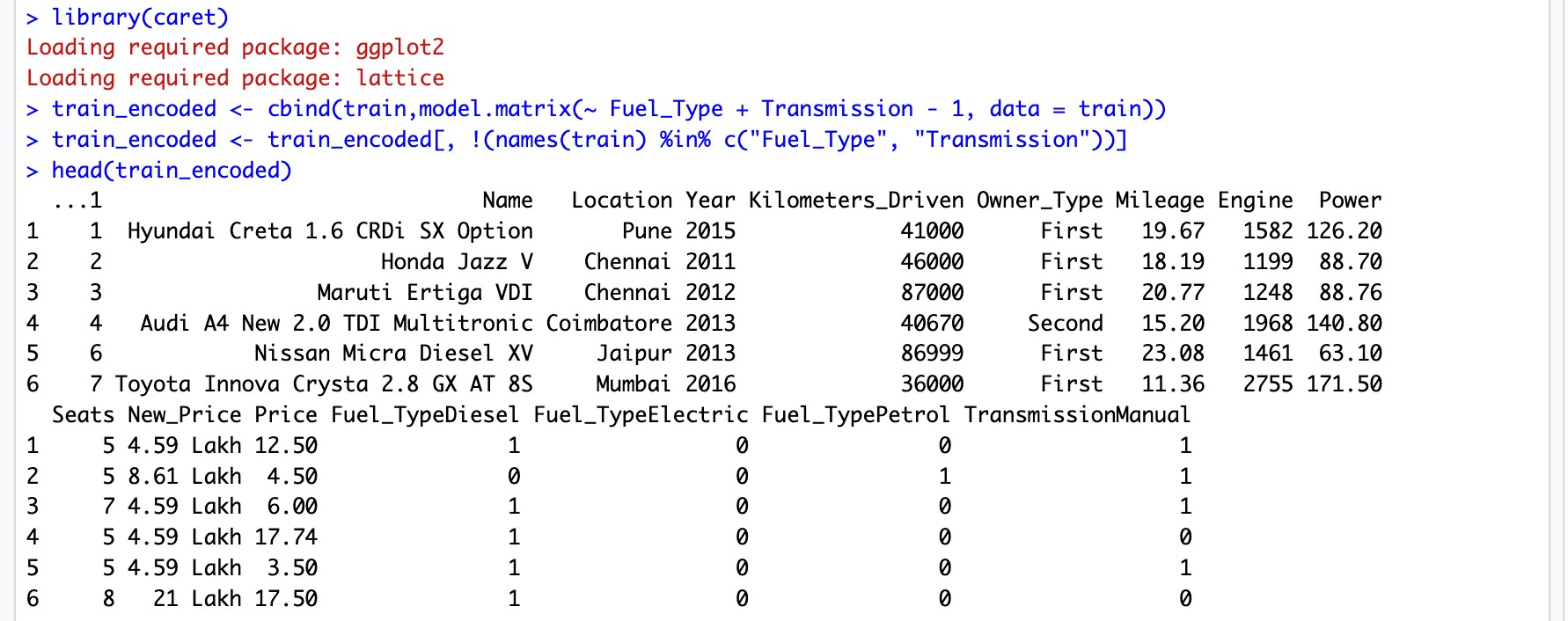
**“New\_price”).**

+



**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.**



