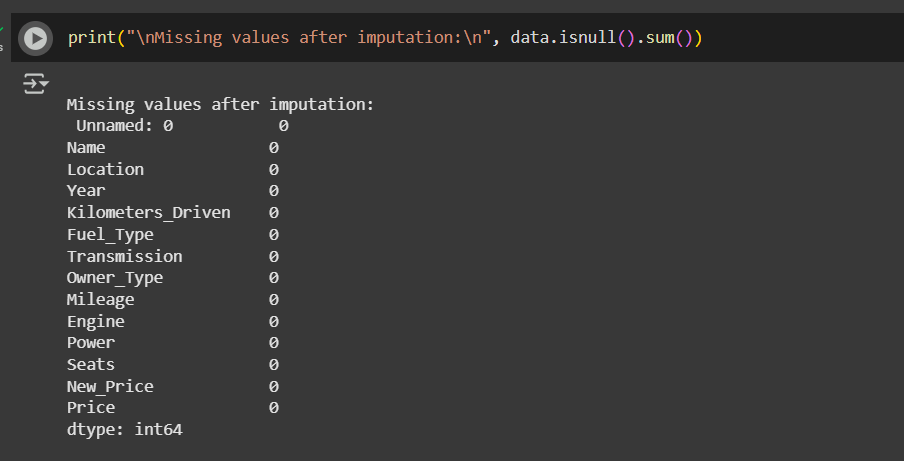
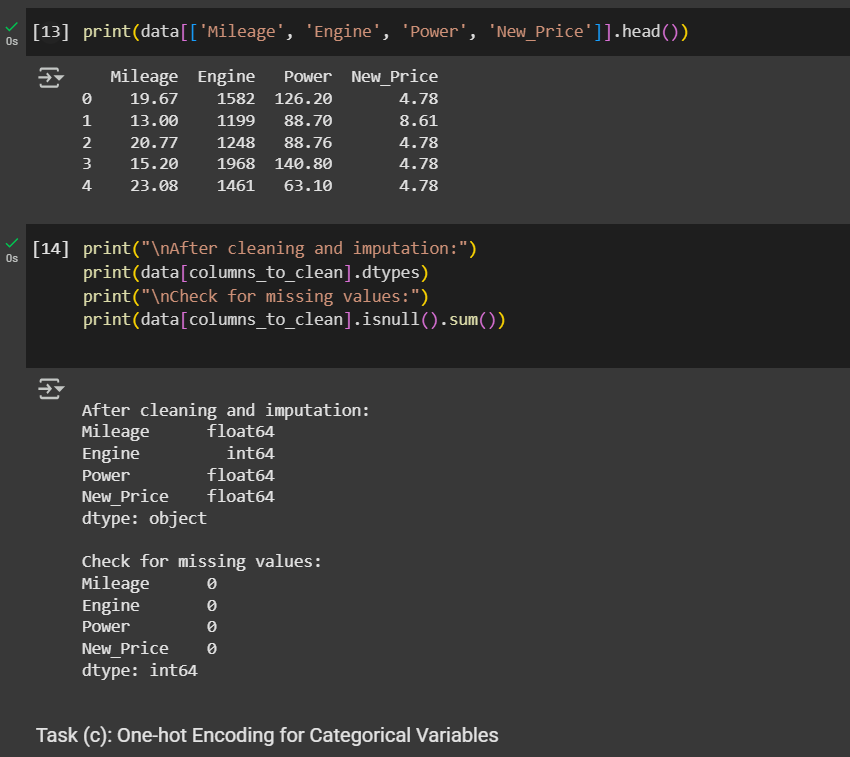
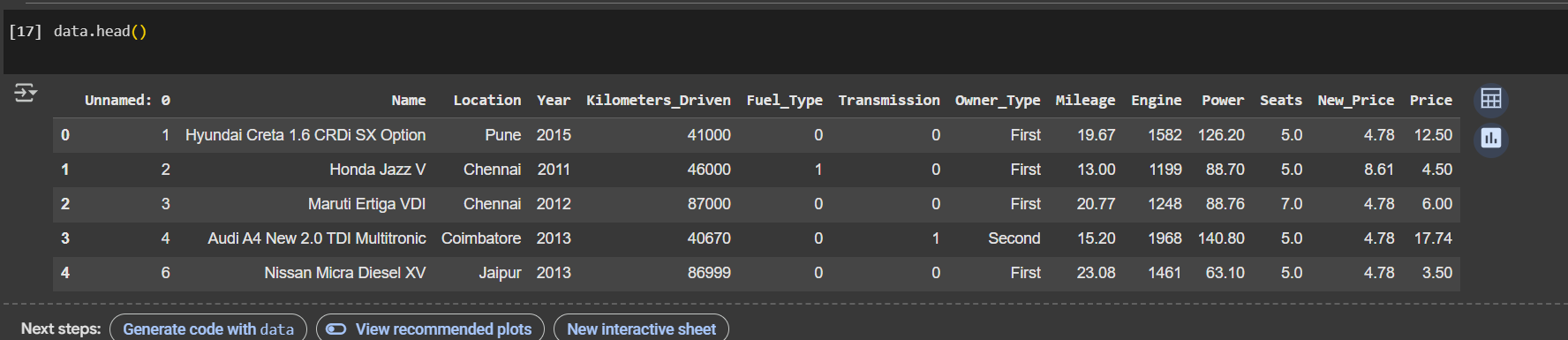
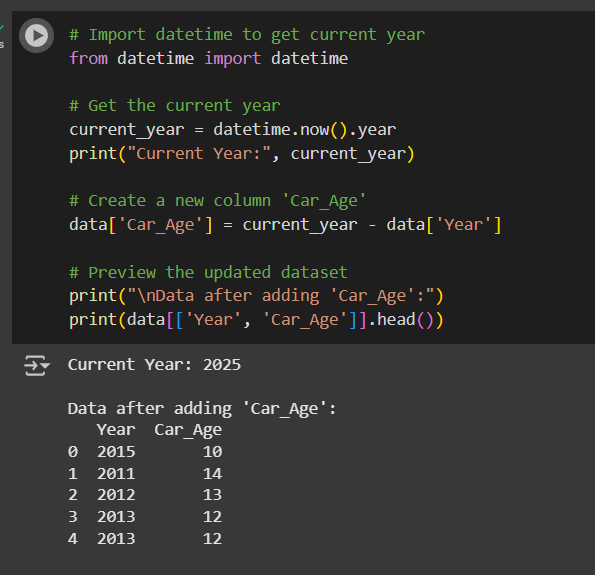
a)  Look for the missing values in all the columns and either impute them (replace with mean, median, or mode) or drop them. Justify your action for this task.     (4 points)    
  
 I have done the imputation as dropping rows/columns might lead to significant data loss and reduce dataset representativeness.

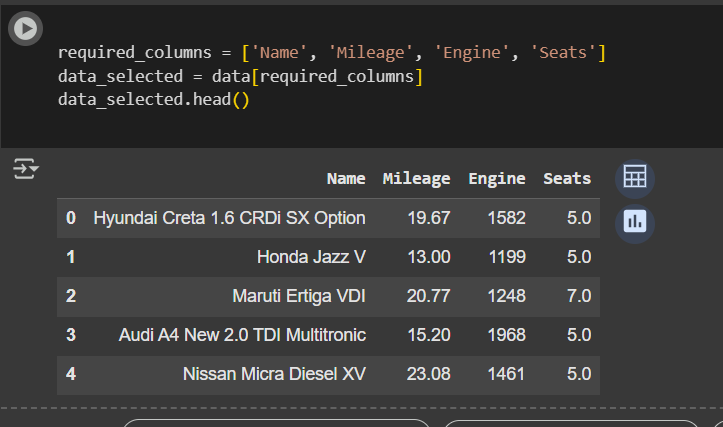
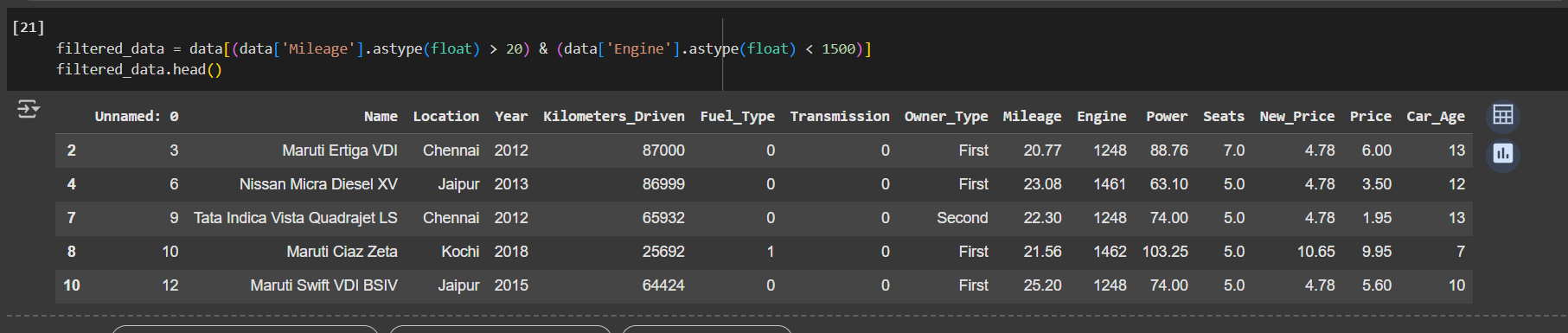
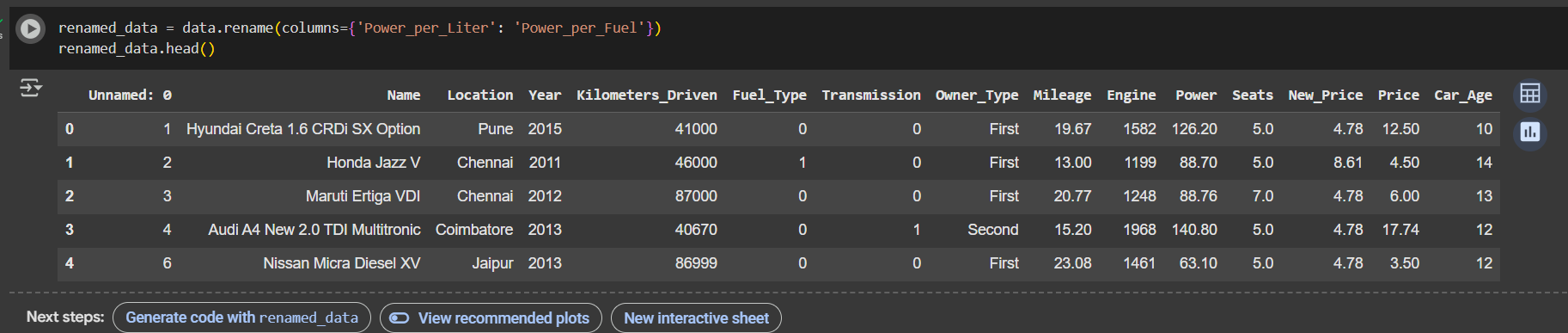
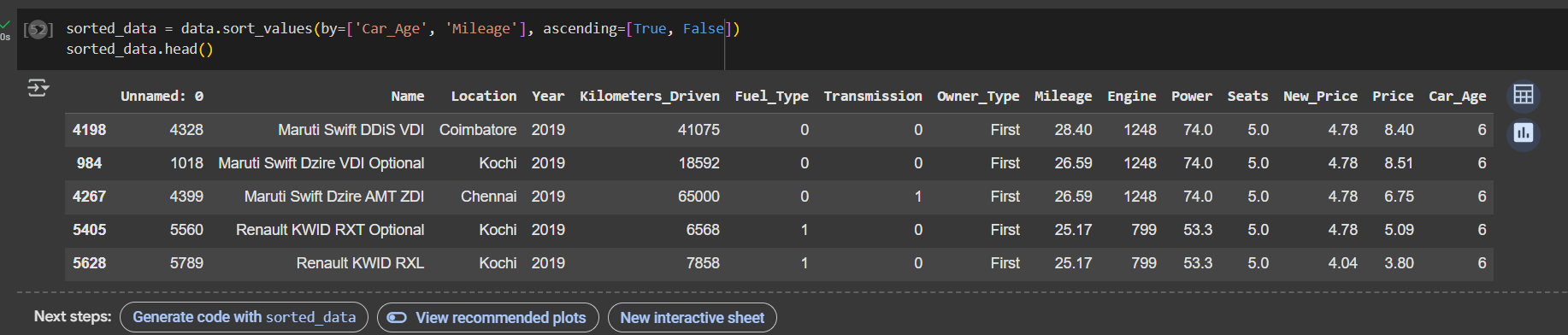
I have used Median and Mode imputation as they maintain the original distribution and structure of the dataset, ensuring that subsequent analyses are minimally impacted.  
  


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”). (4 points)   


C) Change the categorical variables (“Fuel\_Type” and “Transmission”) into numerical one hot encoded value. (4 points).



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.   (4 points)     
 I have added a another feature by calculating age of car.  


e) Perform select, filter, rename, mutate, arrange and summarize with group by operations (or their equivalent operations in python) on this dataset. (4 points)    
  
selecting column  
  
filter  
  
Rename  
  
arrange  
   
summarize  
