**RESULTS:**

**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**

**• The data is normally distributed or approximately symmetric.**

**• The missing values are missing completely at random (MCAR) or missing at random (MAR).**

**• Imputing with the mean will not introduce substantial bias in your analysis.**

**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”).**

**Mielage**

**A screenshot of a computer

Description automatically generated**



**Engine**

**A screenshot of a computer

Description automatically generated**



**Power**

**A screenshot of a computer

Description automatically generated**



**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**

