**PDS \_ Assignment 2**

N. Devi Harshitha 16353896

**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**

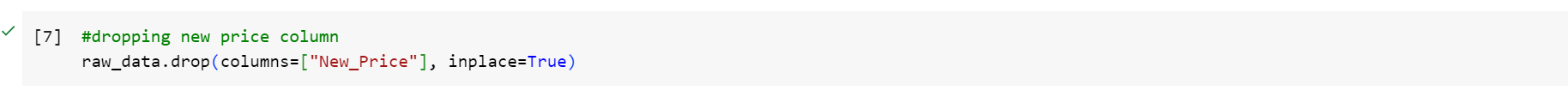
At first we are checking for the null values in entire dataset

A screenshot of a computer program

Description automatically generated

* Before imputing the missing values we need to perform a proper cleaning process.
* Median operations can be performed only on numeric datatype.
* Dropping the first column and the New\_Price column as they wont add any weightage to our analysis.
* As all the missing values are of numeric type, replacing with Median values

Dropping the column



Imputing the missing values with median

A screen shot of a computer code

Description automatically generated

**b. Removing the units from the attributes like Mileage, Power, Engine and New\_Price and keeping the numeric part for analysis**.

A screenshot of a computer

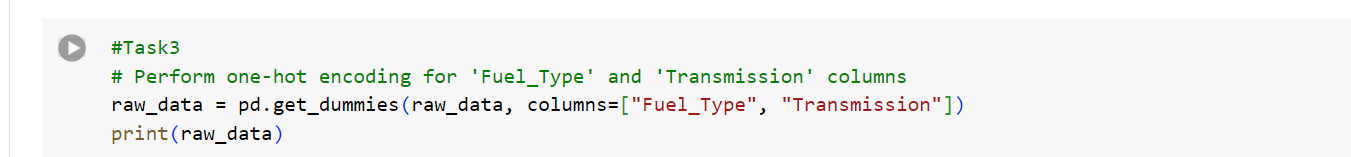
Description automatically generated

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

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

Rename:

A screenshot of a computer

Description automatically generated

Filter:

A screenshot of a computer

Description automatically generated

Arranging:

A screenshot of a computer

Description automatically generated

Grouping and Summarization:

A screenshot of a computer

Description automatically generated