Assignment – 2

Deekshith Venkat Sai M 16342068

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

median, or mode) or drop them.

* Load the dataset using pandas. The cars dataset contains 5847 observations and 14 variables

A screenshot of a computer

Description automatically generated

* Now we need to find the missing values in the dataset.

A screenshot of a computer

Description automatically generated

* Before imputing the missing values we need to perform a proper cleaning process.Mean 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 Mean values
* The cleaned dataset is stored in a file in results folder a s “clean\_cars\_data”.

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

A screenshot of a computer

Description automatically generated

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

encoded value (0,1)

A screenshot of a graph

Description automatically generated

d. Create one more feature and add this column to the dataset. 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

The asked steps for the given data set are performed successfully.