**Assignment – 2**

**Name:** Sai Pavan Pratapagiri

**ID:** 16343743

**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.**

* 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 won’t 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**. **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”).**

A screen shot of a computer program

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

**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

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated