**Principles of Data Science (5530-0001)-Assignment 2**

**Name: Ankitha Srirama Reddy**

**ID: 16355578**

I am considering the given data set as the raw data. I am dropping the first column i.e., unnamed column with the serial number. I am dropping this column because the serial numbers are not in order.

**A screenshot of a computer

Description automatically generated**

1. In this question, I am identifying the missing values in all columns. For the categorical columns, I am imputing the missing values with the mode because if we use the mode to impute missing values, you can preserve the data that is currently in that category field. By guaranteeing that the imputed values are representative of the complete collection of data, this reduces the likelihood of bias.

For the numerical columns, I am imputing the missing values with the mean because by imputing missing values using the mean, the distribution and central tendency of the available data in the numerical column are maintained. This technique ensures that the imputed values correspond to the distribution of the entire data set and reduces the likelihood of introducing discernible bias.

A screenshot of a computer

Description automatically generated

1. In this question, I am removing the units from the attributes of Mileage, Engine, Power, New\_price columns and keeping only the numerical values and storing this modified data as the cleaned data (data\_clean.csv) under data\_clean folder.

A screenshot of a computer

Description automatically generated

1. In this question first I am loading the clean data and encoding the

categorical variables “Fuel\_Type” and “Transmission” into numerical values. And storing this modified data in data\_encoded.csv under the results folder.

A screenshot of a computer

Description automatically generated

1. In this question, I am first loading the data\_encoded.csv data obtained from the previous step. Then I am adding a new column Current\_Age and storing this final data as the data\_final\_results.csv under the results folder.

A screenshot of a computer

Description automatically generated