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

**Name: Aishwarya Musuku**

**ID: 16344653**

1. In the question, I First Loaded the train dataset after that, determined all the missing values from each column and began to replace them. As a basic imputation technique, I used the mode to replace the missing values in the categorical columns. This reduces bias and preserves relationships by maintaining sample size and data distribution. Then I used mean of column values to fill in the missing values in the numerical columns. Mean helps in preserving sample size, data integrity, and lessens the possibility of bias introduction.

A screenshot of a computer

Description automatically generated

1. In this question, I removed the units from the attributes of Mileage, Engine, Power, New\_price columns.

Removed units ‘kmpl’ and ‘km/kg’ from Mileage.

Removed units ‘CC’ from Engine.

Removed units ‘bhp’ from Power.

Removed units ‘Lakh’ and ‘Cr’ from New\_Price.

A screenshot of a computer

Description automatically generated

1. In this question, I used pd.get\_dummies() function for encoding the categorical variables “Fuel\_Type” and “Transmission”.

A screenshot of a computer

Description automatically generated

1. In this question, I assigned the current year 2024 to a variable current\_year and then subtracted each car’s year from the current\_year variable to get the car’s Current\_Age. Then new column named Current\_Age is added to the datasetA screenshot of a computer

   Description automatically generated
2. In this question, all the operations select, filter, rename, mutate, arrange, and summarize with group by or implemented on the provided dataset.
3. **Select Operation**

A screenshot of a computer

Description automatically generated

1. **Filter Operation**

A screenshot of a computer

Description automatically generated

1. **Rename Operation**

A screenshot of a computer

Description automatically generated

1. **Mutate Operation**

A screenshot of a computer

Description automatically generated

1. **Arrange Operation**

A screenshot of a computer

Description automatically generated

1. **Group By**

A screenshot of a computer

Description automatically generated