1. To handle the missing values I have plot the graphs graphs to check the distribution.

A screenshot of a computer

Description automatically generated

A graph with blue lines

Description automatically generated

I removed the "New\_Price" column because it contained around 85% missing values, making it unreliable for analysis. For the "Engine," "Mileage," and "Power" columns, I used the median to fill missing values, as all three distributions are positively skewed and asymmetrical. This positive skewness indicates a higher concentration of lower values on the left side, with a gradual decline toward higher values, likely due to most cars experiencing a reduction in engine performance, mileage, and power ratings after a certain amount of usage. For the "Seats" column, I used the mode to fill in missing values, as the dataset predominantly includes cars with 5 seats, making the mode a suitable choice to represent typical seating.

A screenshot of a computer

Description automatically generated

B)A screen shot of a computer program

Description automatically generated

C)

A screenshot of a computer program

Description automatically generated

D)

A screenshot of a computer

Description automatically generated

E)

A screen shot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated