**Assignment 2 – PDS**

**Chandra haas Paladugula  
16355595**

1. The columns with missing values are ‘Mileage’, ‘Engine’ , ‘Power’ , ‘Seats’, ‘New\_Price’.

I have replaced the missing values in ‘Mileage’, ‘Engine’, ‘Power’ with median. I have not chosen to mean, since all these values can be a decimal and vary for each car/model. Also that way it is less prone to the outliers.

For the ‘Seats’ column, I have chosen mode because the number of seats will be a whole number. Basically for the seats it should be an integer from (2-8) so most repeated value should be fine.

I have dropped the column ‘New\_Price’ since lot of values are missing in it. It had 5032 missing values.

A screenshot of a computer

AI-generated content may be incorrect.

1. The new features calculated are ‘Age\_of\_Car’ which is the difference of the current year and the ‘Year’ in the particular row. Others are ‘Power\_per\_Liter’ and ‘Mileage\_per\_Seat’.