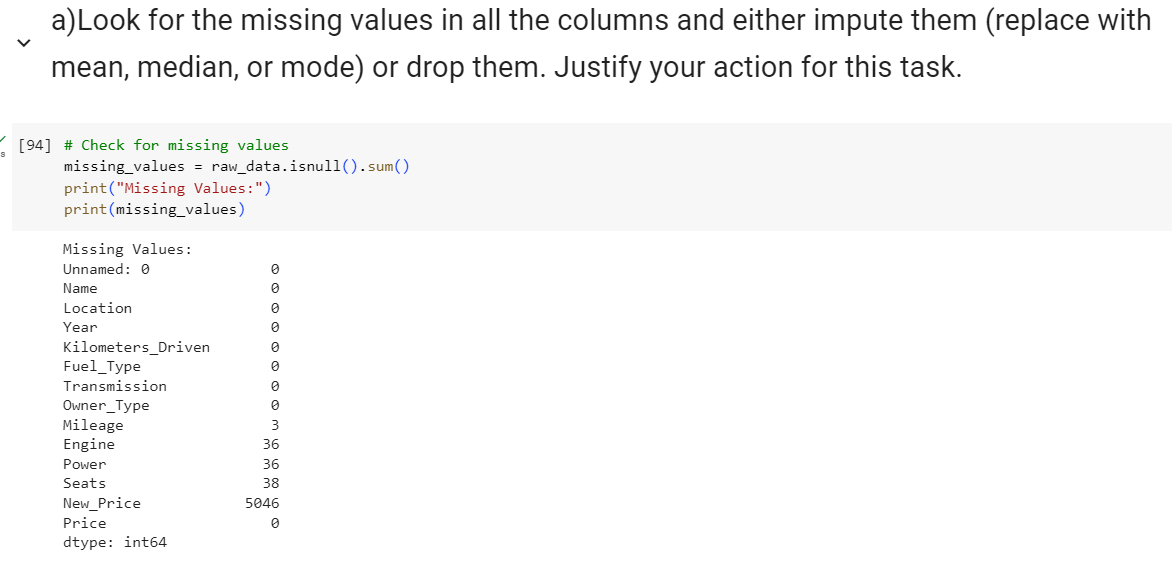
1. Look for the missing values in all the columns and either impute them (replace with mean, median, or mode) or drop them. Justify your action for this task.

Initially I have checked for missing values and got the below results.



Thus, Mileage, Engine, Power, Seats and New\_Price columns have missing values in them. Columns with missing values can be imputed or dropped based on the requirement.

Boxplots of Mileage, Engine and Power indicate that they have many outliers. In such case, considering mean to impute missing values can be misleading as mean just gives the central tendency of data. To deal with such skewed data, median is appropriate one to use. So, I have replaced all the missing values in Mileage, Engine, Power columns with median value of respective columns.

When column has small no. of numerical values/categorical values, mode can be used. So, it is best to replace missing values in seats column using mode value of that column.

New\_price column has 5046 missing values in it (which means almost 86% of the values are missing). So it is better to drop this column as it does help us to get proper analysis.