

Handling Missing Data: Business Problems

Methods for Handling Missing Values

In a dataset, rows with missing values can often present data analysis and modelling challenges. Methods to handle missing values are as follows:

Imputation

You have already studied various methods to impute missing values with statistical functions like mean, median, or standard deviation.

- If the Standard deviation is similar to or near the Average, you can replace the missing value with the Median.
- If the Standard deviation is significantly less than the average value or has a small value. In that case, values are clustered near to the Average. In these scenarios, you can replace the missing value with the Average.

Delete the rows

You can easily delete the row with missing values if the number of missing values is less than the total number of rows in our dataset.

Discuss with your stakeholders/managers

It is not always feasible to replace missing values with imputation or remove rows containing missing values since doing so may lead to incorrect analysis. When this happens, it becomes necessary to seek input from stakeholders or managers to determine the best course of action. For example, imputing missing values in a column with increasing year-on-year values using the average or Median may not give a correct analysis. Discussing alternative methods with stakeholders is essential to preserve the trend in the data.