# What should be the allowed percentage of Missing Values?

**missing_values, dimensionality**

---

**shuvayan** **Jul '15**

Deleting columns containing more than a particular number of missing values is one of the techniques of dimensionality reduction.
What percent of values should be missing in the column to drop it completely??

---

**karthe1** **Jul '15**

**Awesome Hackers!**

**@shuvayan** - Theoretically, 25 to 30% is the maximum missing values are allowed, beyond which we might want to drop the variable from analysis. Practically this varies.At times we get variables with ~50% of missing values but still the customer insist to have it for analyzing. In those cases we might want to treat them accordingly.

---

**kunal** **Jul '15**

**@shuvayan**

As **@karthe1** suggested, this varied from case to case and the amount of information you think the variable has. For example, if you are working on some dataset which contains a column for date of marriage. It may be blank for 50% (or even more) of the population, but might have very high information about the lifestyle of the person. In such cases, you would still use the variable.

If the information contained in the variable is not that high, you can drop the variable if it has more than 50% missing values. I have seen projects / models where imputation of even 20 - 30% missing values provided better results - the famous Titanic dataset on Kaggle being one such case. Age is missing in ~20% of cases, but you benefit by imputing them rather than ignoring the variable.

Hope this helps.

Kunal