tell me different ways of handling missing values and when every technique has to be used

Here are some different ways of handling missing values and when to use each technique:

Delete the missing values. This is the simplest way to handle missing values, but it can lead to a loss of information, especially if there are a lot of missing values. This technique is best used when the missing values are few and far between, and when deleting them will not have a significant impact on the accuracy of the analysis.

Impute the missing values. This involves filling in the missing values with estimated values. There are many different imputation methods, each with its own advantages and disadvantages. Some common imputation methods include:

* Mean or median imputation: This involves replacing the missing values with the mean or median of the non-missing values for the same variable. This is a simple and easy-to-implement method, but it can underestimate the variability of the data.
* Mode imputation: This involves replacing the missing values with the most common value for the same variable. This method is best used for categorical variables.
* Random sample imputation: This involves replacing the missing values with random samples from the non-missing values for the same variable. This method is better at preserving the variability of the data than mean or median imputation, but it can be more computationally expensive.
* Multiple imputation: This involves creating multiple imputed datasets, each with different estimated values for the missing values. The results of the analysis are then averaged across the imputed datasets. This method is more complex to implement, but it is the most robust to missing values.

Use "missingness" as a feature. In some cases, it may be useful to use the fact that a value is missing as a feature in the analysis. This can be done by creating a new variable that indicates whether or not the value is missing. This technique is best used when there is a reason to believe that the missing values are not random. For example, if you are analyzing a dataset of medical records, and age is missing for some patients, you may want to create a new variable that indicates whether or not the patient's age is missing. This would allow you to investigate whether the missing values are related to other variables in the dataset, such as the patient's health condition.

Which technique to use?

The best way to handle missing values will depend on the specific dataset and the analysis being performed. However, some general guidelines can be given:

* If there are a few missing values and deleting them will not have a significant impact on the accuracy of the analysis, then deleting the missing values is a simple and effective solution.
* If there are a lot of missing values, or if deleting them would have a significant impact on the accuracy of the analysis, then imputation is a better option.
* If the missing values are random, then simple imputation methods such as mean, median, or mode imputation can be used.
* If the missing values are not random, then more complex imputation methods such as multiple imputation should be used.
* If there is a reason to believe that the missing values are related to other variables in the dataset, then "missingness" can be used as a feature in the analysis.

It is also important to note that there is no one-size-fits-all solution to handling missing values. The best approach will vary depending on the specific dataset and the analysis being performed.