# Methods of Missing Value Imputation

In any dataset given, handling missing values are a crucial step of data pre-processing. As the existence of missing values may and will lead to inaccurate research, prediction and needless to say will produce errors while running to program.

Hence we must handle all missing values carefully. Not all of them should be treated the same way as I’ll explain down below.

Programming Language used – Python 3

Libraries used – Pandas, Numpy

* Generally for object type of data in missing values we use the (mode) method for missing value imputation where the missing cells are replaced by the value that has the most reoccurrence in the data set.
* For numerical data(integer and float) we use both (median) and (mean) methods in general terms. The median method gives us the centralized value of the column to replace the cell with while the mean method gives us the average value of the column.
* “ ffill “ and “ bfill “ method is used to fill the missing value of the rows with previous or the next row values respectively.
* “ limit function “ helps us to determine till which row our replacement will execute and stops thereafter.
* “ Interpolate function “ is used for predictive analysis where the missing values are replaced with the near most accurate value.
* Not all columns with missing values are Important for our analysis. Columns with more than 70% missing values are of no use to us and can drop them with the “dropna” method.
* With the help of the “ threshold function, “ we can determine that what number of valid values we will require in our columns.

**Special Note** – We can also use machine learning models to recover missing values like KNN, Linear regression, basically used for predictive modeling of our dataset.