

SI No	Method	Description	Advantages / Disadvantages
1	Mean / Median / Mode Imputation	Mean: When data is normally distributed Median: When data is skewed or not Normally Distributed Mode: When data is Categorical	- Easy to Use - Median is robust to outliers - Quickest way to handle missing values - It affects the variance of the dataset
2	Random Sample Imputation	A Random value from the existing set of values is taken and used to fill the missing values	- Easy and variance is same as the original dataset - Implementation is a bit complicated
3	Capturing NaN Value with a new feature	Used when the data is missing due to some cause A new feature / column is created in the dataframe where 1 is stored for missing value and 0 otherwise <code>df['C_NAN'] = np.where(df['C'].isnull(), 1, 0)</code>	- Easy to Implement - Captures the missingness of the values - Increases the features / No of columns of the dataset
4	End of Distribution	Fill missing values with some extreme values of the feature Helpful when there are more extreme values in the dataset. Aim is to include them To find the extreme values use Mean +/- 3 * Standard Deviation	
5	Arbitrary Value Imputation	Each NaN value is replaced by an arbitrary value Decision about the arbitrary value is purely judgement based It must be a value that is not very frequent in the dataset Generally it can be something like the minimum or maximum value of the dataset Eg: Min sales for a month is 20,000. I can never be below that so fill missing months with 20,000	
6	Frequent Category Imputation	When the variable is categorical, the best way to fill the missing values is with the Mode	- Easy and fast way to handle missing categorical values - This value may end up over representing the dataset
7	Treat NaN as a New Category	This is done in order to fill the values by a new category that can indicate that the values are missing So if there are already 3 categories, then a 4th category is added for all the missing cells. It can be simply labelled as "missing"	
8	KNN Imputation	KNN is a machine learning model, that relates to the prediction of classes based on the K Nearest Neighbours The same model or logic can be employed to even fill the missing categorical values	
9	Drop all NaN Values	When we deal with a feature that has approx. 60% or more values missing, it is advisable to drop such features from the dataset	