## Handling Missing data:

only began to the on Missing Values Impute (Fill) -ccA Multivariate Univariate (Focus on multi column) (Focus on one column) Iterative Impulere Categorical Numeri cal Imputer -mode - Mean - use 'missing' word -Median - Random value

### Complete Cane Analysis: (CCA)

It means analyzing only those rows for which there is information in all of the variables in the dataset.

(columns)

Herre you discard the rows where values in any of the columns are missing.

- You can only do this, when the values that are missing are in random rows.
- Not like the first 50 rows, on middle 50 rows on last 50 rows.
- The values should be kandomly missing.

- Because only if you delete the rows of the dataset at random, only then the distribution of the data remains same.
  - Advantage ) Easy to delete : -
- DisAdvantage ->
  DisAdvantage -
  - 2) Much Into will be lost (Because the nows you are deleting many numbers of columns might have valuable information)
  - 3) If you can't train your model with missing data, then the model will not know how to handle it.

## mone ( bata missing at remain news ? baru ad ti blooks natur

- D MCAR -> (Column values should be missing at random rows)
- 2) Remove the rows on columns where maximum values are missing

How to check missing pencentage in Pandas?

- df. is null (). mean() \* 100
- 3) Apply CCA on those columns where missing values at random nows

## Handling Missing Numerical Data: non off style us of the support

Techniques - Mean, Median Imputations

- Antitary value (Randons ral) (A ponticular value)
- End of distribution part (
- Random is all savious of (c

# Mean-Median Imputation: Alexandry discussed

Disadvage -> ) Change the shape of the distribution

- 2) some extra outliers tormed
  - 3) Conrelation chages with other columns

#### When to use?

- D) MCAR (Data missing at random rows) (6000 and 41 bloods model
- (2000) mobined to prission and bloods coulor minutes) 94911 (1.

  2) When missing data < 5%.

  2) When missing data < 5%.

  3. 2000 and on order municipal samples on 2000 and over all comparts.

## Arbitary Value Imputation:

Mostly we update eategorical missing values with this technique

- Suppose we change missing values to "missing" word.
- To use in numerical column, we can put any value that is not available in that column.

- Lary to apply to the me man pridom and an area

Disadvantage:

- D Greath (Alf) change shape
- 2) Variance changes the state of mitotigmi interviews
- do opportunition changes. Low IM lost of another will

When to use?

- In this case, you can use this technique when data are not reardomly missing " (not reardomly missing ease) Distavantage - PAT vorionie, conrelation energe

Because using Anothery Impulation we are sung such

End of Distribution Imputations

- If your data is normally distributed, there use (mean + 30, mean - 30 to replace missing values and a pt) said and mi

. math tunde mechanisser

- If the data is skewed, then use (IAR Proximity)

on, Q3 + 1'5 IQR

Using Antitary Impulation one End of distribution we are presenting one we are making aware our model about the missing values.

Because using Antitary Impulation, we are using such a value to replace missing value which is not available in the other nows. In End of distribution imputation we are replacing missing bit values by making them like outliers. So that ML Model can have a separate knowledge ab observation about them.

Advantage: - Basy to use

Disadvantage - Pdf, variance, conrelation changes

When your data of a painticular column is not missing at nardom nows when your data of a painticular column is not missing at nardom nows wather missing in nows like (Top 50 nows, middle 50, Lower 50) something like that.

81-1151GR on 634 151GR

## Handling Categorical Missing data:

Categorical Value

Random

Value

Value

Specific value (Missing)

### Observation needed to use Mode:

- Data should be missing at rundom nows (MCAR)
- Mode value should be much higher than the other values.counts.

## When your missing data >20% ? (>10%)

- Can you specific values ("missing") to replace.
- So that your column will have now new category "Missing"
- Yours model then can observe this.