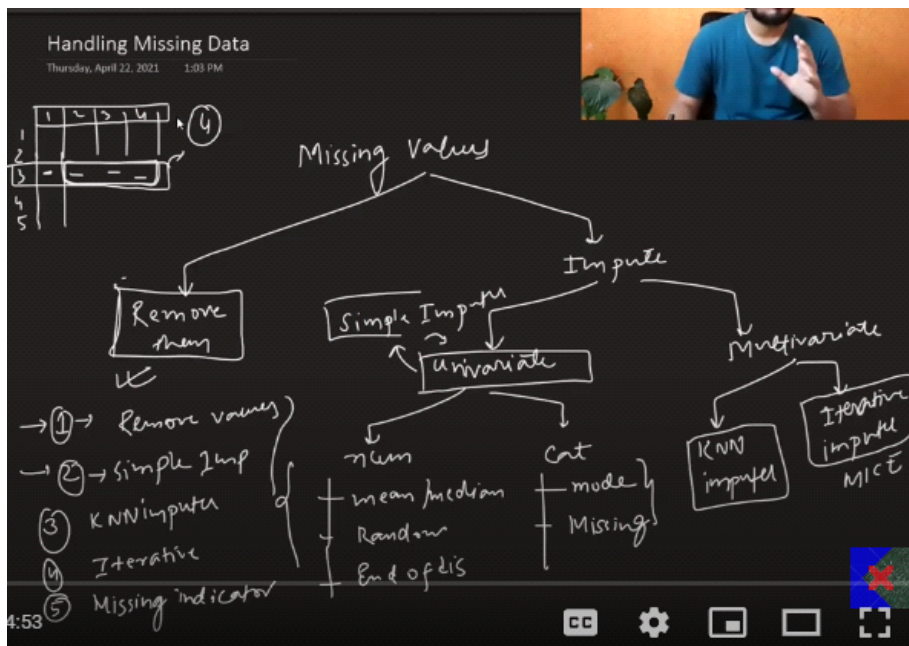


# Handling Missing Data

01 December 2023 11:04



Remove is not recommendable

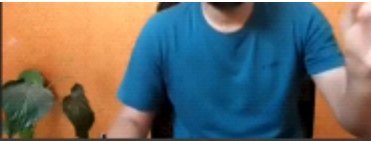
# Complete Case Analysis

01 December 2023

11:17

## Complete Case Analysis

Thursday, April 22, 2021 12:55 PM



Complete-case analysis (CCA), also called "list-wise deletion" of cases, consists in **discarding** observations where values in **any** of the variables are missing.

۲۵۳                      ۱۳۱

- Complete Case Analysis means literally analyzing only those observations for which there is information in **all** of the variables in the dataset.

# Assumptions for CCA

01 December 2023 11:18

Missing completely at random

The screenshot shows a presentation slide titled "Assumption For CCA" with a timestamp of "Thursday, April 22, 2021 12:58 PM". The slide features handwritten notes in white on a dark background. The notes include:

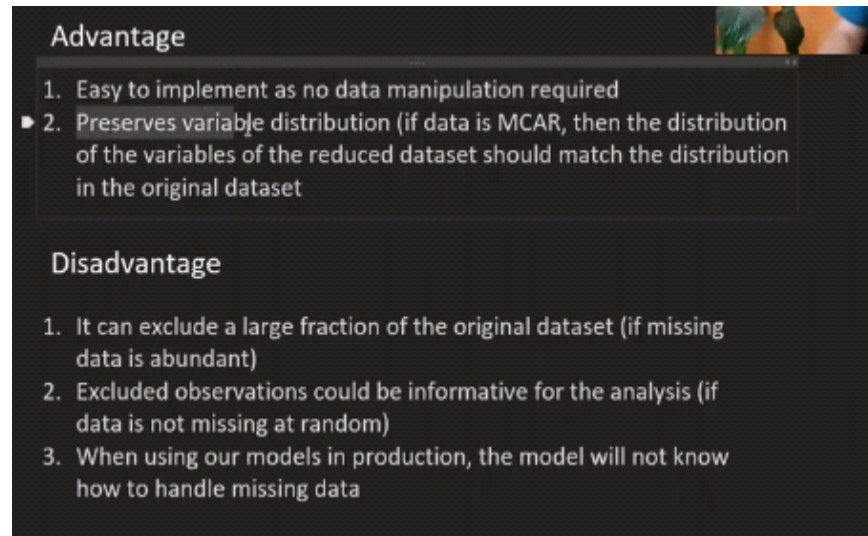
- 1) Missing completely at random
- A diagram showing a flow from "Age" (with "1000, 4" above it) to a circle containing "50", which then points to a circle containing "950, 4".
- A box labeled "MCAR" with an arrow pointing to a circle containing "CCA".
- A separate diagram on the right showing a circle with "1000" pointing to a circle with "950".

A video inset in the top right corner shows a person in a blue shirt gesturing with their hand. The left sidebar of the presentation software is visible, showing a menu with options like "Handling Missing Data", "Complete Case Analysis", "Assumption For CCA", "Advantage/Disadvantage", "When to use CCA?", and "Example".

# Advantages and Disadvantages

01 December 2023

12:06



The image is a screenshot of a presentation slide with a dark background. At the top right, there is a small, partially visible image of a person's face. The slide is divided into two main sections: 'Advantage' and 'Disadvantage'. The 'Advantage' section contains a list of two points, with the second point preceded by a small white arrow icon. The 'Disadvantage' section contains a list of three points.

## Advantage

1. Easy to implement as no data manipulation required
- ▶ 2. Preserves variable distribution (if data is MCAR, then the distribution of the variables of the reduced dataset should match the distribution in the original dataset)

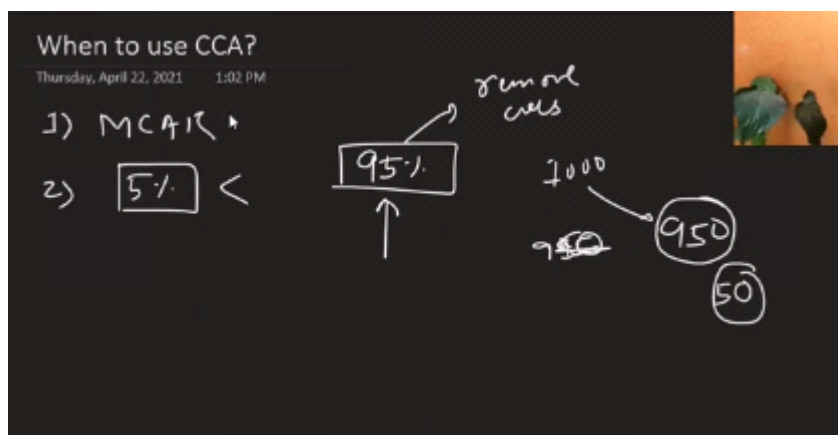
## Disadvantage

1. It can exclude a large fraction of the original dataset (if missing data is abundant)
2. Excluded observations could be informative for the analysis (if data is not missing at random)
3. When using our models in production, the model will not know how to handle missing data

# When to use CCA

01 December 2023

12:08



# Example

01 December 2023

12:09



readme



data\_scienc  
e\_job



Complete  
Case Anal...