🧩 Handling Missing Data – : Complete Case Analysis (CCA)

Missing data is **very common** in real-world datasets. If not handled properly, it can distort model performance and lead to wrong conclusions.

❓ What Is Complete Case Analysis (CCA)?

Also called **listwise deletion**, CCA simply:

- **Removes any row (data point) that contains missing values** in **any** of the selected columns.

Example:

If your data looks like:

Name	Age	Salary
A	25	50k
B		55k
C	30	

After applying CCA → Only row A remains.

✅ When Should You Use CCA?

Use CCA **only when**:

1. The **missing data is completely random** (no hidden patterns).
2. You can **afford to lose data** without hurting the model.
3. The dataset is **large enough** (dropping rows won't cause data scarcity).

✅ Advantages of CCA

Advantage	Why it's good
🔧 **Easy to implement**	No complex logic or math involved
✅ **No artificial values added**	No risk of wrong imputation
📉 **Keeps original data values**	Doesn't change existing information

---

⚠ Disadvantages of CCA

Disadvantage	Risk
📉 **Loss of data**	Can drastically reduce dataset size
❌ **Bias**	If data is not missing completely at random
📊 **Less statistical power**	Small sample → weak conclusions

CCA Works Best When:

- The **percentage of missing data is very low** (e.g., <5%)
- You're doing **exploratory analysis**
- The **missingness is random**

Key Assumption:

- ***Missing Completely at Random (MCAR)***
  Means: the chance of a value being missing has **nothing to do with the data itself**.
- If this assumption fails → CCA can introduce **bias** and **mislead** your model.
