Macro Programming

Vignesh V
Department of Computer Applications
vigneshv@pes.edu

PES UNIVERSITY

Macro Programming

Handling Missing Values and Removing Duplicates

Vignesh V
Department of Computer Applications

Introduction

Real-life datasets often have:

- **Missing Values** – Empty cells, incomplete entries, or errors.
- **Duplicates** – Same records repeated multiple times.

Why handle them?

- To improve **data quality**
- To ensure **accurate calculations & reports**

Excel provides **built-in tools** to manage both problems.

Identifying Missing Values

Common forms of missing data:

- **Blank Cells** → No data entered.
- **Errors** → Example: `#N/A`, `#VALUE!`.

How to detect?

1. **Filter** → Filter column → Check “Blanks.”
2. **Conditional Formatting** → Highlight empty cells.
3. **Go To Special** → Home → Find & Select → Go To Special → Blanks.

Example:

E1 fx

	A	B
1	**Student**	**Marks**
2	A	90
3	B	
4	C	78
5	D	#N/A

Here, B has a blank mark, and D has an error.

Handling Missing Values – Techniques

1. Delete rows/columns with blanks

- Best when large portions of data are missing.
- *Example:* If 80% of values in a column are blank → remove that column.

2. Replace with Zero (0)

- Useful for numeric fields where missing = no value.
- *Example:* Sales data – Missing entry means no sales → replace with `0`.

Handling Missing Values – Techniques

3. Fill using Mean/Median/Mode

- Replace blanks with **average, middle value, or most common value**.
- *Example:*
  - Student marks: [90, **blank**, 78, 85]
  - Average = (90+78+85)/3 = 84.3
  - Replace blank with **84**.

4. Forward Fill / Backward Fill

- Fill blanks using the value before or after.

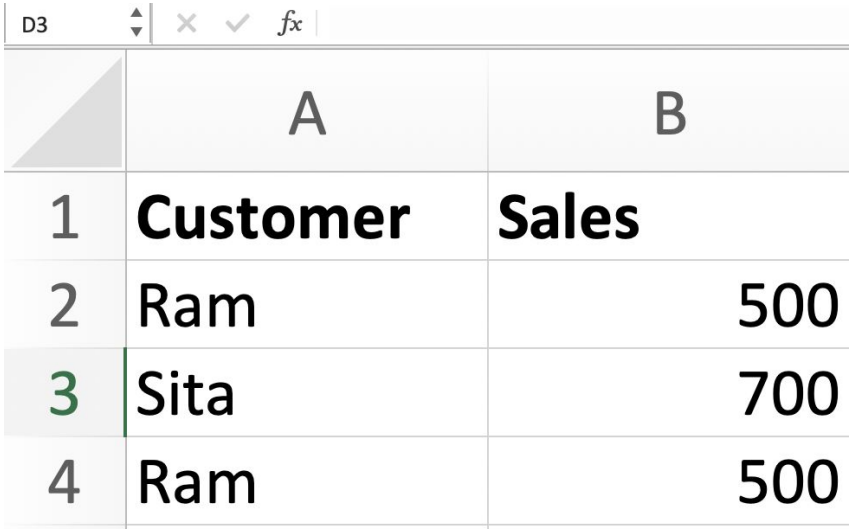
Handling Missing Values – Techniques

- ***Example:***
  - Dates: [1-Jan, blank, blank, 4-Jan]
  - Forward Fill → [1-Jan, 1-Jan, 1-Jan, 4-Jan].

Removing Duplicates – Why?

Duplicate data → incorrect analysis.

Example:



If not removed → Total Sales = 1700 (instead of 1200).

This inflates reports and leads to wrong business decisions.

Removing Duplicates – Steps

Select dataset → **Data Tab** → **Remove Duplicates**.

Choose column(s) to check.

- *Example:* If Customer ID repeats → check only “Customer ID.”

Excel removes duplicates and shows how many were deleted.

Best Practices

- Always create a **backup** before cleaning.
- Don’t delete missing values blindly → Analyze pattern.
- Use **Replace with Mean/Median** for numeric data.
- After removing duplicates → Cross-check with record count.

Macro Programming

Practical Activity

Student ID	Name	Marks
101	A	90
102	B	
103	C	85
104	B	
105	D	78
102	B	

Practical Activity

Task:

1. Identify missing marks → Fill with average = (90+85+78)/3 = 84.3 → **84**.
2. Remove duplicate Student ID (102).
3. Final dataset will have **clean values + unique records**.

Summary

Missing Values → Handle using delete, replace, or fill methods.

Duplicates → Remove using Excel's built-in tool.

Clean data ensures:

- Accuracy
- Reliability
- Better analysis & decision making.

THANK YOU

Vignesh V
Department of Computer Applications
vigneshv@pes.edu