

DAY 4 – Excel Data Cleaning Notes

Topics:

- Removing Duplicates
- Handling Nulls / Blanks
- Introduction to Power Query
- Cleaning Dataset using Power Query

1. Removing Duplicates in Excel:

Steps:

1. Select dataset (Ctrl + A)
2. Go to Data → Remove Duplicates
3. Select columns to check
4. Click OK

2. Handling Nulls / Blanks:

Method 1: Replace manually

Home → Find & Select → Replace

Find: (empty)

Replace: Unknown → Replace All

Method 2: Formula

Replace blank text:

=IF(TRIM(A2)="", "Unknown", A2)

Replace blank numbers:

=IF(B2="", MEDIAN(\$B:\$B), B2)

3. Power Query – How to Open:

1. Select data
2. Data → From Table/Range
3. Power Query Editor opens

4. Cleaning in Power Query:

Step 1: Remove Duplicates

Home → Remove Rows → Remove Duplicates

Step 2: Replace Blank Values

Right-click column → Replace Values

Find: null

Replace: Unknown

Step 3: Fix Data Types

Click the type icon (ABC/123/calendar)

Choose correct type:

- Customer → Text
- Region → Text
- Quantity → Whole Number
- Price → Whole Number
- OrderDate → Date

Step 4: Sort Data

Home → Sort Ascending

Step 5: Close & Load

Home → Close & Load To...

Load as Table → New Worksheet

Name sheet: cleaned_powerquery

Summary:

You learned how to:

- Remove duplicates
- Handle blanks and nulls
- Use Power Query for automated cleaning
- Fix data types
- Load cleaned data back to Excel