

UNIT 2: DATA

2.3: DATA MODELLING

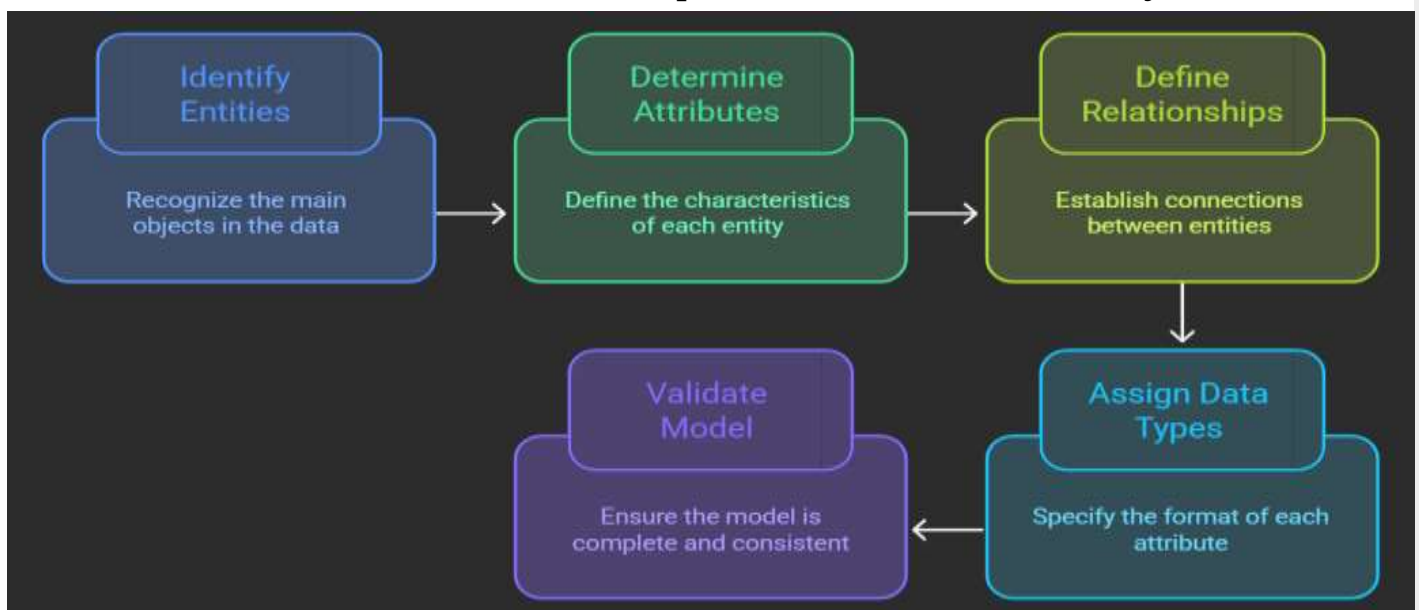


2.3: DATA MODELLING

Data modelling is the process of creating a structured representation of real-world information to organize, analyze, and communicate data effectively. It is important because it helps in predicting trends, making decisions, and ensuring consistency in datasets. Core concepts include identifying entities, attributes, relationships, constraints, and selecting appropriate data types. Data models can be conceptual, logical, or physical depending on abstraction level.

Algorithm / Process:

1. Identify the main entities in the data (e.g., students, courses).
2. Determine the attributes for each entity (e.g., student name, age, course ID).
3. Define relationships between entities (e.g., students enroll in courses).
4. Assign appropriate data types to each attribute (e.g., text, number, date).
5. Validate the model for completeness and consistency.



UNIT 2: DATA

2.3: DATA MODELLING



2.3.1: Data Vs Information

Data consists of raw facts, numbers, or text without context, whereas information is data that has been processed and organized to provide meaning. This distinction is crucial because only information can guide decisions and problem-solving. Core concepts include collection, processing, summarization, and visualization of data. Converting data into information may involve sorting, filtering, calculating, or creating charts. Understanding this concept ensures that students can distinguish between unprocessed data and actionable insights.



2.3.2 Data Analysis using MS Excel

Data analysis is the process of examining, cleaning, and transforming data to extract meaningful insights. It is important because it helps identify trends, patterns, and anomalies to support decision-making. Core concepts include sorting, filtering, using formulas, applying conditional formatting, and visualizing data using charts. Advanced analysis can include pivot tables, trendlines, and statistical functions. Effective analysis converts large, complex datasets into actionable information.



For Example, retail store collects sales data for each product. The manager sorts and filters the data to see which items sell the most. They use formulas to calculate total revenue and average sales per day. Charts help visualize monthly sales trends and seasonal patterns. Based on the analysis, the store decides which products to restock and which to promote.

UNIT 2: DATA



2.3: DATA MODELLING

2.3.3 Data Collection: MS Excel

Data collection is the process of gathering raw facts and figures from various sources for analysis. It is important because accurate and complete data is the foundation for reliable conclusions. Core concepts include identifying sources, selecting relevant data, recording systematically, and ensuring consistency. Proper collection in Excel involves creating structured tables and using data validation to prevent errors.



LAB- WK-3-L1-Data Collection

Your school wants to record students' favorite sports. The sheet **Sports Collection** already has headers Name, Class, Favorite Sport and some blank rows. Complete the dataset by entering data for all students.

1. Open StudentData.xlsx and go to the Sports Collection sheet.
2. Check that headers are correct.
3. Enter data for 10 students under each column.
4. Use **Data → Data Validation → Allow: Text** for the Favorite Sport column.
5. Make headers bold for clarity. Save the workbook.

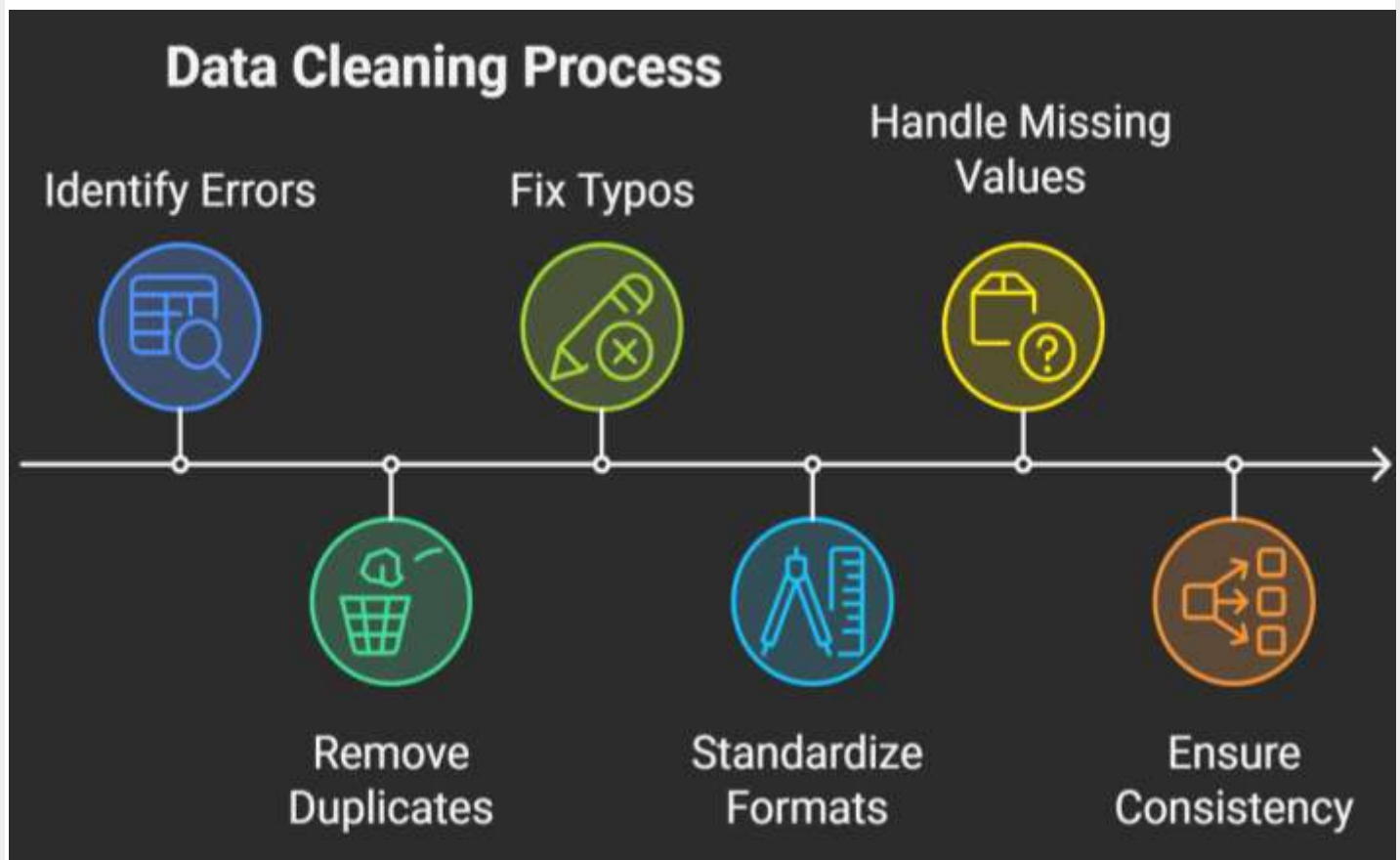
UNIT 2: DATA



2.3: DATA MODELLING

2.3.4 Data Cleaning

Data cleaning is the process of identifying and correcting errors, inconsistencies, or missing values in a dataset. It is important because dirty data can lead to misleading results and poor decisions. Core concepts include removing duplicates, fixing typos, standardizing formats, handling missing values, and ensuring consistency. Clean data ensures reliability for analysis and accurate reporting.



LAB- WK-3-L2-Data Cleaning

The sheet ScoresCleaning contains students' names and test scores. Some names are inconsistent, and a few scores are missing. Clean the dataset to make it consistent and complete.

UNIT 2: DATA

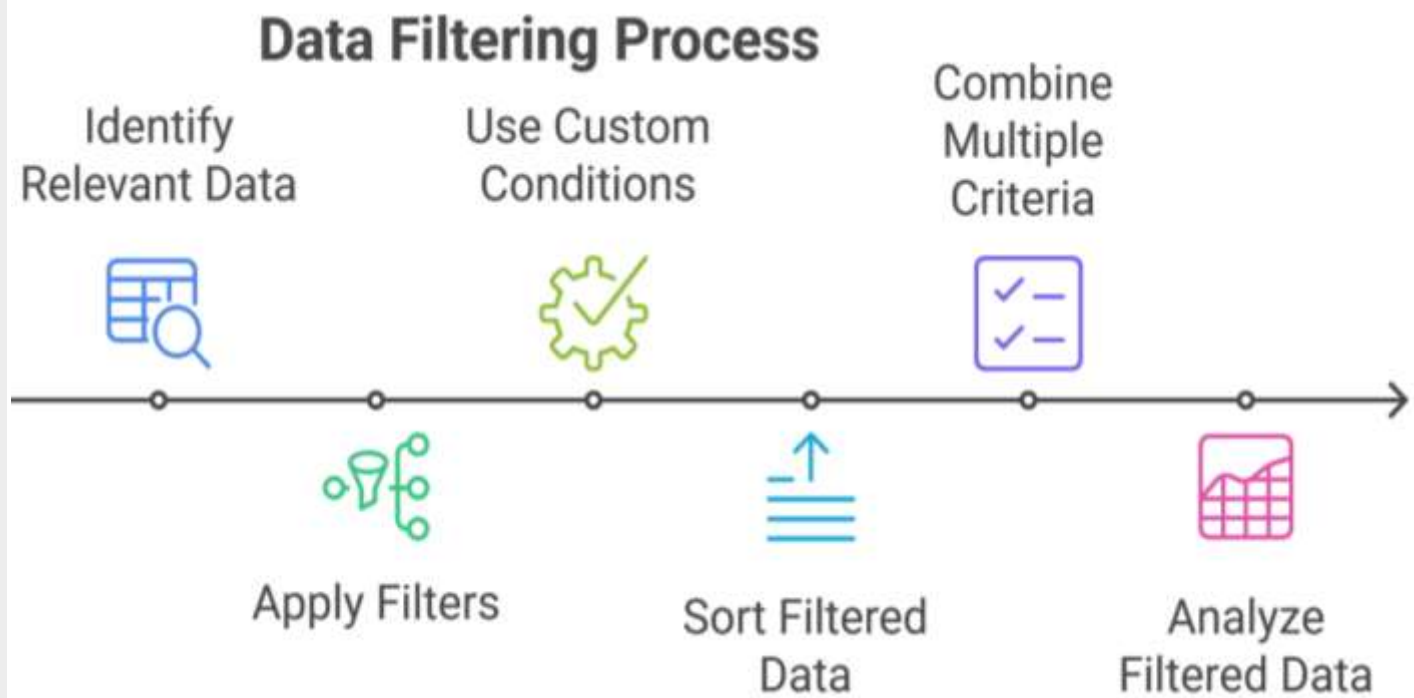


2.3: DATA MODELLING

1. Open StudentData.xlsx and go to the ScoresCleaning sheet.
2. Check headers: **Student Name, Class, Test 1, Test 2, Test 3.**
3. Use **TRIM()** to remove extra spaces in the **Student Name** column.
4. Standardize names using **UPPER()** or **LOWER()** functions.
5. Highlight missing scores using **Conditional Formatting → Highlight Cells Rules → Blank Cells.**
6. Fill in missing scores or replace them with 0.
7. Save the workbook.

2.3.5 Data Filtering: MS Excel

Data filtering is the process of displaying only the rows in a dataset that meet certain criteria. It is important because it allows users to focus on relevant information and quickly analyze subsets of data. Core concepts include applying filters, using custom conditions, sorting within filtered data, and combining multiple criteria. Filtering helps identify trends, outliers, and relationships without altering the underlying dataset.



UNIT 2: DATA

2.3: DATA MODELLING



LAB- WK-3-L2-Data Filtering

The sheet ScoresFiltering contains students' test scores and classes. Filter the dataset to show only students who scored more than 80 in Test 1 and are in Class 10.

1. Open StudentData.xlsx and go to the ScoresFiltering sheet.
2. Check headers: **Student Name, Class, Test 1, Test 2, Test 3.**
3. Click anywhere in the dataset and go to **Data → Filter.**
4. Click the dropdown in **Test 1** and select **Number Filters → Greater Than → 80.**
5. Click the dropdown in **Class** and select **Class 10.**
6. Review the filtered results; only students meeting both criteria should appear.
7. Clear filters to view all data again.
8. Save the workbook.

LESSON SUMMARY

