

# Data Entry, Data Validation & Data Cleaning

Stop bad data at the door with data validation and drop-down lists, then clean what already exists with Flash Fill, Text to Columns, Find & Replace, Go To Special, Remove Duplicates, Paste Special and Excel Tables.

● Live class 2

✦ Excel for Microsoft 365, 2026

Beginner

⌚ 90 min

## By the end of this lesson you can

- Enter data quickly and consistently with AutoFill, Flash Fill and custom lists
- Build drop-down lists, number and date limits and custom-formula rules with Data Validation
- Create dependent (cascading) drop-downs
- Clean messy text, split and combine columns, and remove duplicates safely
- Use Paste Special and Excel Tables as everyday cleaning tools

## Datasets used

`employees_small.csv`

## Garbage in, garbage out

Most time in real analytics is not spent on fancy charts; it is spent fixing data: extra spaces, inconsistent spellings ( `Bangalore` , `Bengaluru` , `bengaluru` ), numbers stored as text, duplicates and blanks. Surveys of data teams regularly report that cleaning and preparing data takes well over half of an analyst's time.

There are two ways to win this fight:

## Two ways to get clean data

### Prevent (at entry)

- Data validation rules
- Drop-down lists
- Input messages and error alerts
- Excel Tables with fixed columns
- Protected sheets (Lesson 11)

### Cure (after entry)

- TRIM, CLEAN, PROPER, SUBSTITUTE
- Flash Fill and Text to Columns
- Find & Replace, Go To Special
- Remove Duplicates
- Power Query for repeatable cleaning (Lesson 12)

Prevent problems at entry time, and clean whatever slips through. Both are needed.

## Fast, consistent data entry

### AutoFill and the fill handle

Select a cell and drag the small square at its bottom-right corner (the **fill handle**). Excel continues the pattern it recognises.

| You type            | Drag gives you                                              |
|---------------------|-------------------------------------------------------------|
| Jan                 | Feb, Mar, Apr ...                                           |
| Monday              | Tuesday, Wednesday ...                                      |
| 1 , 2 (select both) | 3, 4, 5 ...                                                 |
| Q1                  | Q2, Q3, Q4, Q1 ...                                          |
| 01-04-2026          | 02-04-2026, 03-04-2026 ... (right-drag for months or years) |
| Item 1              | Item 2, Item 3 ...                                          |
| =B2*12              | =B3*12 , =B4*12 ...                                         |

Double-click the fill handle to fill a formula down as far as the column next to it has data. This is the fastest way to copy a formula to 10,000 rows.

### TIP Custom lists

File → Options → Advanced → **Edit Custom Lists** lets you add your own sequences, for example North, South, East, West or your company's department order. AutoFill and custom sort then follow that order.

## Flash Fill (Ctrl + E)

Flash Fill learns a pattern from one or two examples you type. It is perfect for splitting names, extracting codes or reformatting text.

**X** Flash Fill: type the first answer in each column, then press Ctrl + E

|   | A                 | B                  | C                 | D                | E               |
|---|-------------------|--------------------|-------------------|------------------|-----------------|
| 1 | <b>Full name</b>  | <b>Email</b>       | <b>First name</b> | <b>Last name</b> | <b>Username</b> |
| 2 | Kavya Iyer        | kavya.iyer@abc.in  | Kavya             | Iyer             | kiyer           |
| 3 | Imran Reddy       | imran.reddy@abc.in | Imran             | Reddy            | ireddy          |
| 4 | Harpreet Kulkarni | harpreet.k@abc.in  | Harpreet          | Kulkarni         | hkulkarni       |
| 5 | Diya Iyer         | diya.iyer@abc.in   | Diya              | Iyer             | diyer           |

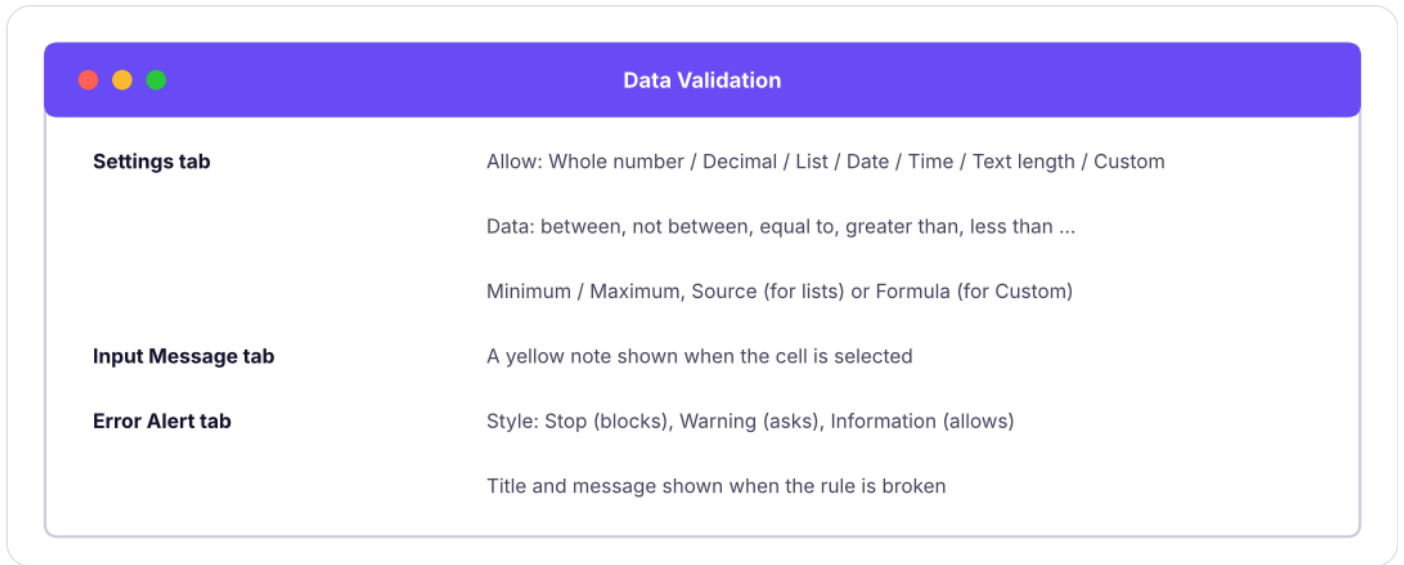
Purple cells are typed by you; green cells are filled by Flash Fill.

### WATCH OUT Flash Fill is not a formula

Flash Fill writes **static values**. If the source changes, the results do not update. Use formulas (Lesson 5) or Power Query (Lesson 12) when the data will be refreshed.

## Data Validation: rules for what can be typed

**Data → Data Validation** (shortcut Alt, A, V, V) attaches a rule to cells. Anyone who types something that breaks the rule is stopped or warned.



The three tabs of the Data Validation dialog.

## The validation types you will use most

| Allow        | Example rule                       | Business use                      |
|--------------|------------------------------------|-----------------------------------|
| Whole number | between 18 and 65                  | Age of an employee                |
| Decimal      | between 0 and 0.5                  | Discount rate                     |
| List         | Sales, Finance, IT, HR, Operations | Department drop-down              |
| Date         | between 01-04-2026 and 31-03-2027  | Entries inside the financial year |
| Time         | between 09:00 and 18:00            | Office hours                      |
| Text length  | equal to 6                         | PIN code, employee code           |
| Custom       | =COUNTIF(\$A\$2:\$A\$500, A2)=1    | No duplicate IDs                  |

## Building a drop-down list, step by step

**STEP BY STEP****Department drop-down**

1. On a sheet called **Lists**, type the allowed values in a column (Sales, Finance, IT, HR, Operations) and convert it to a Table (Ctrl + T) named **tblDept**. A Table grows automatically when you add a department.
2. Select the cells where people will enter a department (for example C2:C500).
3. Data → Data Validation → Allow: **List**.
4. Source: `=INDIRECT("tblDept[Department]")` or simply select the range `=Lists!$A$2:$A$6`.
5. Keep **In-cell dropdown** ticked. Add an Input Message ("Choose a department") and an Error Alert (Stop, "Please pick a department from the list").

**X Data entry sheet with validation rules (rules shown in row 2)**

|   | A                      | B               | C                    | D                  | E                        | F                    | G                    |
|---|------------------------|-----------------|----------------------|--------------------|--------------------------|----------------------|----------------------|
| 1 | Emp ID                 | Name            | Department           | Age                | Join date                | Discount             | PIN code             |
| 2 | <i>Text length = 4</i> | <i>Any text</i> | <i>List: tblDept</i> | <i>Whole 18-65</i> | <i>Date ≥ 01-01-2000</i> | <i>Decimal 0-0.5</i> | <i>Text length =</i> |
| 3 | E101                   | Kavya Iyer      | HR                   | 29                 | 28-02-2016               | 0.1                  | 40000                |
| 4 | E102                   | Imran Reddy     | Fin                  | 17                 | 14-12-2019               | 0.75                 | 1100                 |

Row 3 passes every rule (green). In row 4, **Fin** is not in the list, 17 is under the minimum age, 0.75 is above the maximum discount and 11002 has only 5 characters, so Excel would block each of these entries (red).

**Custom formula rules**

The **Custom** option accepts any formula that returns TRUE (allow) or FALSE (block). Write the formula for the **first** selected cell; Excel adjusts it for the others.

| Rule                                            | Custom formula (cells A2:A500 selected)             |
|-------------------------------------------------|-----------------------------------------------------|
| No duplicates                                   | =COUNTIF(\$A\$2:\$A\$500,A2)=1                      |
| Must start with "E" and be 4 characters         | =AND(LEFT(A2,1)="E",LEN(A2)=4)                      |
| Numbers only (for a text-formatted column)      | =ISNUMBER(--A2)                                     |
| No spaces                                       | =ISERROR(FIND(" ",A2))                              |
| Weekdays only (dates)                           | =WEEKDAY(A2,2)<6                                    |
| Total of a budget column cannot exceed 1,00,000 | =SUM(\$B\$2:\$B\$50) ≤ 100000                       |
| Email must contain "@" and "."                  | =AND(ISNUMBER(FIND("@",A2)),ISNUMBER(FIND(".",A2))) |

## Dependent (cascading) drop-downs

A second list that changes with the first: choose a **Category**, and the **Sub-category** list shows only matching items.

**X** Lists sheet: each column is a named range (Formulas → Create from Selection → Top row)

|   | A                | B                      | C                 |
|---|------------------|------------------------|-------------------|
| 1 | <b>Furniture</b> | <b>Office_Supplies</b> | <b>Technology</b> |
| 2 | Chairs           | Binders                | Phones            |
| 3 | Tables           | Paper                  | Accessories       |
| 4 | Bookcases        | Storage                | Machines          |
| 5 | Furnishings      | Art                    | Copiers           |
| 6 |                  | Appliances             |                   |

### STEP BY STEP Make it work

1. Select the lists (with headers) → Formulas → **Create from Selection** → Top row. Excel creates the names Furniture, Office\_Supplies and Technology.
2. Category cell A2: Data Validation → List → Source: Furniture,Office\_Supplies,Technology.
3. Sub-category cell B2: Data Validation → List → Source: =INDIRECT(A2).
4. INDIRECT turns the text in A2 into a reference to the named range with that name.

**2026 UPDATE****Dynamic-array version (Microsoft 365)**

With a two-column mapping table (Category, SubCategory), put `=FILTER(Map[SubCategory],Map[Category]=A2)` in a helper cell and point the validation source to it with the spill operator, for example `=$H$2#`. No named ranges are needed, and new items appear automatically.

## Finding data that already breaks the rules

Validation only checks **new** typing. Data pasted in, or entered before the rule existed, is not checked. Use **Data → Data Validation → Circle Invalid Data** to draw red circles around existing violations.

**COMMON MISTAKES****Validation is not security**

Pasting over a validated cell replaces the rule. Use Paste Special → Values when pasting into validated areas, and protect the sheet (Lesson 11) if people must not change rules.

## Cleaning text with formulas

These functions fix most text problems. Each formula below is written for row 2 and copied down.

**X Common text-cleaning formulas**

|   | A           | B                      | C                              | D                                   | E                       |
|---|-------------|------------------------|--------------------------------|-------------------------------------|-------------------------|
| 1 | Raw value   | <code>=TRIM(A2)</code> | <code>=PROPER(TRIM(A2))</code> | <code>=SUBSTITUTE(A2,"-","")</code> | <code>=UPPER(A2)</code> |
| 2 | bengaluru   | bengaluru              | Bengaluru                      | bengaluru                           | BENGALURU               |
| 3 | NEW DELHI   | NEW DELHI              | New Delhi                      | NEW DELHI                           | NEW DELHI               |
| 4 | pune-411001 | pune-411001            | Pune-411001                    | pune411001                          | PUNE-411001             |
| 5 | mUMBAI      | mUMBAI                 | Mumbai                         | mUMBAI                              | MUMBAI                  |

**B2** `=TRIM(A2)`

**C2** `=PROPER(TRIM(A2))`

**D2** `=SUBSTITUTE(A2,"-","")`

**E2** `=UPPER(A2)`

*Each pattern is copied down/across to the other cells in its column.*

- TRIM removes spaces at the start and end and squeezes repeated inner spaces to one.
- PROPER capitalises the first letter of each word; UPPER and LOWER change the case of everything.
- SUBSTITUTE(text, old, new) replaces every occurrence of old.

- CLEAN removes non-printing characters such as line breaks from copied web data.
- The non-breaking space (character 160) from websites is not removed by TRIM. Use `=TRIM(SUBSTITUTE(A2,CHAR(160)," "))`.

## Text to Columns

**Data → Text to Columns** splits one column into several, either by a **delimiter** (comma, space, -, |) or by **fixed width**.

**X Before: one column with a delimiter**

|   | A                 | B     | C          |
|---|-------------------|-------|------------|
| 1 | <b>Order info</b> |       |            |
| 2 | IN-2025-00042     | Pune  | Technology |
| 3 | IN-2025-00043     | Delhi | Furniture  |

### STEP BY STEP Split by a delimiter

1. Select the column. Data → Text to Columns.
2. Choose **Delimited** → Next → tick **Other** and type | (or comma, tab, space).
3. In step 3, set each column's type: keep IDs and codes as **Text** to protect leading zeros; set dates to **Date: DMY**.
4. Choose the destination cell (so the original is not overwritten) → Finish.

Text to Columns is also a quick fix for numbers stored as text: select the column, Text to Columns, and click **Finish** immediately. Excel re-reads every value as a number.

### 2026 UPDATE TEXTSPLIT, TEXTBEFORE, TEXTAFTER (Microsoft 365)

Formula versions that update automatically: `=TEXTSPLIT(A2,"|")` spills the parts across columns; `=TEXTBEFORE(A2,"-")` and `=TEXTAFTER(A2,"-", -1)` return the text before the first or after the last hyphen. See Lesson 5.

## Find & Replace and wildcards

Ctrl + H replaces text in the selection (or the whole sheet). Click **Options** to match case, match the entire cell, or search in formulas versus values.

| Wildcard | Meaning                            | Example                          |
|----------|------------------------------------|----------------------------------|
| *        | Any number of characters           | Bang* finds Bangalore, Bangaluru |
| ?        | Exactly one character              | E10? finds E101 to E109          |
| ~        | Treat the next character literally | ~* finds a real asterisk         |

Example: replace Bang\* with Bengaluru (with **Match entire cell contents** ticked) to standardise every spelling in one step.

## Go To Special: select blanks, constants, formulas, errors

F5 → Special (or Ctrl + G → Special) selects cells by type. The most useful trick in data cleaning is **filling blanks with the value above**:

**X** Before: category written only once per block

|   | A               | B              | C            |
|---|-----------------|----------------|--------------|
| 1 | <b>Category</b> | <b>Product</b> | <b>Sales</b> |
| 2 | Furniture       | Chair A        | 12000        |
| 3 |                 | Table B        | 18500        |
| 4 |                 | Bookcase C     | 9400         |
| 5 | Technology      | Phone X        | 25200        |
| 6 |                 | Router Y       | 4300         |

### STEP BY STEP Fill blanks with the value above

1. Select A2:A6. Press F5 → Special → Blanks → OK. Only the empty cells are selected (yellow).
2. Type = and press the **Up arrow** (the formula becomes =A2).
3. Press Ctrl + Enter to put the formula in every selected blank at once.
4. Select the column, copy, and Paste Special → **Values** to freeze the result.

Other Go To Special options: **Constants** (typed values only, great for clearing inputs but keeping formulas), **Formulas** (to check or protect them), **Errors**, **Visible cells only** (copy a filtered list without hidden rows; shortcut Alt + ;), **Row differences**.

## Removing duplicates

**Data → Remove Duplicates** deletes rows that repeat in the columns you choose and keeps the **first** occurrence.

### X Duplicate check before removing

|   | A      | B                 | C      | D             |
|---|--------|-------------------|--------|---------------|
| 1 | Emp ID | Name              | City   | Is duplicate? |
| 2 | E101   | Kavya Iyer        | Mumbai | First         |
| 3 | E102   | Imran Reddy       | Delhi  | First         |
| 4 | E101   | Kavya Iyer        | Mumbai | Duplicate     |
| 5 | E103   | Harpreet Kulkarni | Mumbai | First         |
| 6 | E102   | Imran Reddy       | Delhi  | Duplicate     |

**D2** =IF(COUNTIF(\$A\$2:A2,A2)>1,"Duplicate","First")

**D3** =IF(COUNTIF(\$A\$2:A3,A3)>1,"Duplicate","First")

**D4** =IF(COUNTIF(\$A\$2:A4,A4)>1,"Duplicate","First")

**D5** =IF(COUNTIF(\$A\$2:A5,A5)>1,"Duplicate","First")

**D6** =IF(COUNTIF(\$A\$2:A6,A6)>1,"Duplicate","First")

The expanding range \$A\$2:A2 counts how many times the ID has appeared **so far**, so the first appearance says "First" and later ones say "Duplicate". This lets you review duplicates before deleting anything.

#### WATCH OUT Always work on a copy

Remove Duplicates permanently deletes rows. Copy the sheet first, or use the formula above, conditional formatting (Home → Conditional Formatting → Highlight Duplicate Values) or =UNIQUE() (Lesson 6) to inspect duplicates without losing data.

## Paste Special: the cleaning Swiss-army knife

Ctrl + Alt + V opens Paste Special.

| Option                                    | What it does                                     | Typical use                                                                      |
|-------------------------------------------|--------------------------------------------------|----------------------------------------------------------------------------------|
| <b>Values (V)</b>                         | Pastes results, not formulas                     | Freeze formula results; break links to other files                               |
| <b>Formats (T)</b>                        | Copies only formatting                           | Apply a header style to another table                                            |
| <b>Transpose (E)</b>                      | Rows become columns                              | Turn a wide header row into a list                                               |
| <b>Multiply / Divide / Add / Subtract</b> | Combines the copied number with each target cell | Convert rupees to lakhs: copy 100000, select the amounts, Paste Special → Divide |
| <b>Skip blanks</b>                        | Empty copied cells do not overwrite targets      | Merge two partially filled columns                                               |
| <b>Paste Link</b>                         | Creates =Sheet1!A1 references                    | Live link to a source cell                                                       |

## Excel Tables: structure that maintains itself

Convert any clean range with **Ctrl + T**. Tables are the best habit you can build in Excel:

- **Auto-expand:** new rows and columns join the table, so formulas, charts, pivot tables and validation lists include them automatically.
- **Structured references:** formulas read like English, for example `=SUM(tblSales[Amount])` or `=[@Price]*[@Qty]`.
- **Calculated columns:** type a formula once and it fills the whole column.
- **Total Row:** Table Design → Total Row adds a row with drop-down summaries that respect filters (they use `SUBTOTAL`).
- **Banded rows, filter buttons and a frozen header** come for free.

**X** An Excel Table named `tblOrders` with a calculated column and Total Row

|   | A            | B       | C     | D   | E      |
|---|--------------|---------|-------|-----|--------|
| 1 | OrderID      | Product | Price | Qty | Amount |
| 2 | 1001         | Chair   | 3500  | 4   | 14,000 |
| 3 | 1002         | Desk    | 12500 | 1   | 12,500 |
| 4 | 1003         | Lamp    | 1200  | 6   | 7,200  |
| 5 | 1004         | Shelf   | 5400  | 2   | 10,800 |
| 6 | <b>Total</b> |         |       | 13  | 44,500 |

**E2** =C2\*D2

**E3** =C3\*D3

**E4** =C4\*D4

**E5** =C5\*D5

**D6** =`SUBTOTAL(109, D2:D5)`

E6    =SUBTOTAL(109,E2:E5)

In a real Table, the Amount formula appears as =[@Price]\*[@Qty]. The Total Row uses SUBTOTAL(109, ...), which sums only visible rows, so totals change when you filter.

#### DO IT WITH AI    Do it with AI

Copilot in Excel or any chat assistant can draft cleaning steps. A good prompt: *"I have a column of Indian city names with inconsistent spelling, extra spaces and mixed case (examples: ' bengaluru', 'BANGALORE', 'Bengaluru '). Give me an Excel formula to standardise them to proper case, and a list of Find & Replace steps to map old names to new ones."*

Always test the answer on a copy of your data.

## Summary

#### KEY TAKEAWAYS

- **Prevent** bad data with Data Validation (lists, limits, custom formulas) and **cure** the rest with cleaning tools.
- Validation lists should point to a Table so they grow automatically; use INDIRECT or FILTER for dependent drop-downs.
- Validation does not check pasted or old data: use **Circle Invalid Data**.
- TRIM, CLEAN, PROPER, SUBSTITUTE fix most text; Text to Columns and Flash Fill split data.
- **Go To Special → Blanks** then =↑ and Ctrl + Enter fills blank cells with the value above.
- Inspect duplicates with COUNTIF before using Remove Duplicates; always keep a copy.
- Turn every dataset into an **Excel Table** (Ctrl + T).

### Quick quiz

6 questions · tap an answer to check it

**1. Which validation setting stops users entering a duplicate employee ID in A2:A500?**

- A** List with source A2:A500
- B** Custom: =COUNTIF(\$A\$2:\$A\$500,A2)=1
- C** Text length equal to 4
- D** Whole number between 1 and 500

**Answer: B.** COUNTIF counts how many times the value appears; allowing only a count of 1 blocks duplicates.

**2. You pasted 2,000 rows into a validated column. How do you find values that break the rule?**

- A** Re-create the validation rule
- B** Data Validation → Circle Invalid Data
- C** Press F9
- D** Use Remove Duplicates

**Answer: B.** Validation only checks typing. Circle Invalid Data marks existing violations.

**3. What does =TRIM(" New Delhi ") return?**

- A** "New Delhi"
- B** "New Delhi"
- C** "NewDelhi"
- D** " New Delhi "

**Answer: B.** TRIM removes leading and trailing spaces and reduces repeated inner spaces to one.

**4. Which feature fills blank cells in a column with the value from the cell above in a few keystrokes?**

- A** Flash Fill
- B** Remove Duplicates
- C** Go To Special → Blanks, then = + Up arrow + Ctrl + Enter
- D** Text to Columns

**Answer: C.** Go To Special selects only the blanks; Ctrl + Enter writes the relative formula into all of them.

**5. What is the main downside of Flash Fill?**

- A** It only works on numbers
- B** Results are static values that do not update when the source changes
- C** It needs a macro
- D** It cannot split text

**Answer: B.** Use formulas or Power Query for refreshable cleaning.

**6. Why should a drop-down list's source be an Excel Table?**

- A** The list grows automatically when new items are added
- B** Tables are required for validation
- C** It makes the drop-down faster
- D** It prevents duplicates

**Answer: A.** Tables expand automatically, so new departments appear in the drop-down without editing the rule.

### EXERCISE Validation set-up

Using `employees_small.csv`, add validation so that: department must be one of Sales, Finance, IT, HR, Operations; rating must be a whole number from 1 to 5; join date cannot be in the future; and employee IDs are unique.

#### Show solution (worked answer)

- Department (C2:C21): List, source Sales, Finance, IT, HR, Operations.
- Rating (G2:G21): Whole number, between 1 and 5.
- Join date (E2:E21): Date, less than or equal to `=TODAY()`.
- Emp ID (A2:A21): Custom, `=COUNTIF($A$2:$A$21,A2)=1`. Add an Error Alert (Stop) for each with a helpful message, then run Circle Invalid Data to check the existing rows.

### EXERCISE Clean the city column

A column contains: `mumbai`, `MUMBAI`, `Bombay`, `Mumbai`, `mumbai city`. Produce one clean value, `Mumbai`, for all five.

#### Show solution (worked answer)

1. Helper formula `=PROPER(TRIM(A2))` gives `Mumbai`, `Mumbai`, `Bombay`, `Mumbai`, `Mumbai City`.
2. Paste Special → Values over the original.
3. Find & Replace with Match entire cell contents: `Bombay` → `Mumbai`, `Mumbai City` → `Mumbai`. For data that refreshes every week, build the same steps in Power Query so they repeat automatically.

## INTERVIEW CORNER

### Questions you may be asked

**Q: How would you stop users from entering invalid data in a shared tracker?** Use Data Validation (drop-down lists from a Table, number and date limits, custom formulas for uniqueness), add input messages, set error alerts to Stop, and protect the sheet so the rules cannot be removed. Check older data with Circle Invalid Data.

**Q: How do you remove duplicates without losing information?** First flag them with COUNTIF or conditional formatting and review which record to keep; decide the key columns that define a duplicate; work on a copy; then use Remove Duplicates (or UNIQUE) on those columns.

**Q: What is the difference between TRIM and CLEAN?** TRIM removes extra spaces; CLEAN removes non-printing characters such as line breaks. Web data often needs both, plus SUBSTITUTE(..., CHAR(160), " ") for non-breaking spaces.

## Practise more online

Free, trusted places to test yourself on this topic. Links open the original site in a new tab.