DATA CLEANING MISTAKES & FIXES

Hi-res PDF: bit.ly/data-cleaning-mistakes

1 Ignoring blank cells: Disrupt formulas & analyses

Option 1:

Blank value is excluded from the AVERAGE formula.

=AVERAGE(C3:C7)

	A	B	C
2		Product	Sales $
3		Magic Wand	200
4		Fairy Dust	300
5		Bacon Scented Candle	
6		Web Shooter	100
7		Bat Signal	400
8		Average	250

Higher average sales ***per product*** could be misleading.

Option 2:

0 value is included in the AVERAGE formula.

=AVERAGE(C3:C7)

	A	B	C
2		Product	Sales $
3		Magic Wand	200
4		Fairy Dust	300
5		Bacon Scented Candle	0
6		Web Shooter	100
7		Bat Signal	400
8		Average	200

Note:

Either option can be correct. You need to decide which is your desired result.

Option 1 Fix:

1. Select the column with blanks: Home > Find & Select > Go To Special > Blanks (or **CTRL+G**)
2. Fill in an appropriate placeholder: e.g., 0 > Press **CTRL+ENTER**

2 Ignoring duplicates: Distorted calculations & reports

Data with duplicates:

ID	Name	Product
1	John Smith	Magic Wand
2	Anna Belle	Fairy Dust
3	Chris P. Bacon	Bacon Scented Candle
4	Peter Parker	Web Shooter
5	Bruce Wayne	Bat Signal
2	Anna Belle	Fairy Dust
3	Chris P. Bacon	Bacon Scented Candle
4	Peter Parker	Web Shooter

Product	Count
Bacon Scented Candle	2
Bat Signal	1
Fairy Dust	2
Magic Wand	1
Web Shooter	2
Grand Total	8

Incorrect Count

Duplicate–free data

ID	Name	Product
1	John Smith	Magic Wand
2	Anna Belle	Fairy Dust
3	Chris P. Bacon	Bacon Scented Candle
4	Peter Parker	Web Shooter
5	Bruce Wayne	Bat Signal

Product	Count
Bacon Scented Candle	1
Bat Signal	1
Fairy Dust	1
Magic Wand	1
Web Shooter	1
Grand Total	5

Correct Count

Fix:

1. Select the column with duplicate data > Home tab > Conditional Formatting > Highlight Cell Rules > Duplicate Values
2. Assess the highlighted duplicate values
3. Data tab > Remove Duplicates > Delete duplicate values

3 Inconsistent date formats: Cause unexpected sorting errors

	B	C
2	Long Date	Date
3	January 1, 2025	1/1/2025
4	January 2, 2025	2025-01-02
5	March 1, 2025	3/1/2025
6	January 3, 2025	Jan 03, 2025

March incorrectly appears before Jan when sorted from A to Z, as dates are formatted as text.

=DATEVALUE([@Date])

	A	B	C	D
2		Desired Date	Date	Fixed Date
3		January 1, 2025	1/1/2025	1/1/2025
4		January 2, 2025	2025-01-02	1/2/2025
5		January 3, 2025	Jan 03, 2025	1/3/2025
6		March 1, 2025	3/1/2025	3/1/2025

Correctly sorted dates

DATEVALUE Masterclass: https://bit.ly/fndatevalue

Step by Step

Problem: Dates in the Date column are stored as text, therefore sort incorrectly and will not work in math operations.

Solution:

1. Add a **new column** - "Fixed Date"
2. Enter the formula **=DATEVALUE([@Date])** in the first cell of the Fixed Date Column
3. **Copy the formula down** the column
4. **Format the new column** by going to the Home tab > Number Format > Short Date
5. **Sort on the "Fixed Date" column** from oldest to newest to get the correctly sorted data!

Pro Tip:

Use the Text to Columns tool to bulk-standardize dates with mixed delimiters (e.g., 01/01/2025 & 2025-01-02)

Steps: Select the column > Data tab >Text to Columns > Delimited >don't select any delimiter > Press Next > Choose the correct date format e.g., MDY in the final step.

4 Incorrect data types: Cause calculation errors

=SUM([Sales $])

	A	B	C
3		Product	Sales $
4		Wand	100
5		Web Shooter	200
6		Unicorn Horn	300
7		Total	500

The total formula ignores $100 sales for Wand as it is formatted as text

Fix:

1. **Create another column** – Formatted Sales $
2. Use the formula **=VALUE([Sales $])** to arrive at proper number formatted values
3. **Calculate the total** sales.

=VALUE([@[Sales $]])

	B	C	D
3	Product	Sales $	Formatted Sales $
4	Wand	100	100
5	Web Shooter	200	200
6	Unicorn Horn	300	300
7	Total	500	600

Correct total is arrived at using the VALUE function, which converts text values to a number.

Pro Tip:

Automate data type corrections using Power Query:

Go to Data tab > Get & Transform Data > From Table/Range. In the Power Query editor change the column data type to "Whole Number" and clean error values by replacing them with 0.

5 Not trimming extra spaces: Cause lookup mismatches & errors

=XLOOKUP(E3,B3:B6,C3:C6)

	A	B	C	D	E	F
2		Sales Rep	Sales $		Sales Rep	Sales $
3		John Smith	100		Anna Belle	#N/A
4		Anna Belle	200			
5		Peter Parker	300			
6		Mary Jane	400			

Fix:

1. Insert a new column
2. Use **TRIM** function on the original data
3. **Paste as values** the trimmed data to the original column. (Microsoft 365 users can paste values with **CTRL + SHIFT + V**)

TRIM Masterclass: https://bit.ly/fntrim

CLEAN Masterclass: https://bit.ly/fcclean

Pro Tip:

If your data contains non-printable characters, such as tabs or line breaks, combine TRIM with CLEAN for a deep cleanse: **=TRIM(CLEAN(B2))**

Non-printing characters often sneak into data when copying from external sources like websites or PDFs.

=TRIM(B4)

	A	B	C	D
2		Sales Rep	Sales $	Trimmed Name
3		John Smith	100	John Smith
4		Anna Belle	200	=TRIM(B4)
5		Peter Parker	300	Peter Parker
6		Mary Jane	400	Mary Jane

=XLOOKUP(F3,B3:B6,C3:C6)

	F	G	H
2	Sales Rep	Sales $	
3	Anna Belle	200	

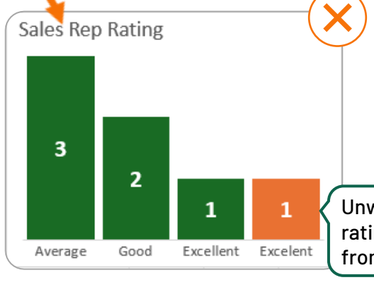
6 Not running a spell check: Causes grouping & filtering errors

Sales Rep	Rating
John Smith	Good
Jane Doe	Excellent
Chris P. Bacon	Average
Peter Parker	Average
Anna Belle	Average
Peter Parker	Good
Mary Jane	Excelent

Incorrect spelling results in incorrect grouping in charts

Pro Tip

Use Data Validation lists to standardize data entry

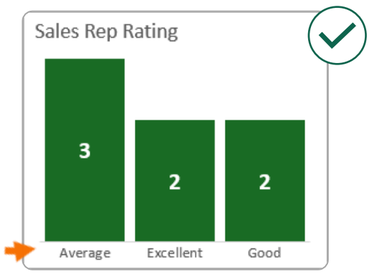


Unwanted additional rating category resulting from the spelling error.

Fix:

1. Select all data (CTRL + A)
2. Review tab > Spell Check
3. Review incorrect spellings **> select 'Change All'** (change all incorrect instances) or select '**Change** 'to change 1 instance

Sales Rep	Rating
John Smith	Good
Jane Doe	Excellent
Chris P. Bacon	Average
Peter Parker	Average
Anna Belle	Average
Peter Parker	Good
Mary Jane	Excellent



Best Practices:

1 Standardize:

Standardizing text case:

- Inconsistent case can cause issues.
- Correct one off errors with UPPER, LOWER, or PROPER functions.

2 Automate:

Avoid manual, error-prone data cleaning:

Automate repetitive data cleaning with Power Query

3 Document:

Document your cleaning steps:

- Log all data cleaning steps for transparency and reproducibility.
- Use Comments or a separate sheet to explain complex transformations