

The Excel Data Survival Guide

Eight ways Excel quietly
destroys your data — and
the settings that stop it.

What Excel does to your data before you type anything

Every item below happens on open or on paste, with no warning and no undo prompt. None of them are your mistake. They are defaults — and once the file is saved, most are permanent.

WHAT BREAKS	YOU TYPED → YOU GET	WHO THIS HURTS
Leading zeros vanish	01234 → 1234	Postcodes, SKUs, member and employee IDs, sort codes, French and Italian phone numbers.
Long numbers become scientific notation	009876543210... → 9.87E+11	Barcodes, EAN and ISBN numbers, IMEI, tracking references, bank account numbers.
Digits past the 15th are erased	...3456789012345678 → ... 3456789012340000	16-digit card numbers and long barcodes. This one cannot be recovered by any formula.
Text is converted to a date	SEP1 → 01-Sep · 1-2 → 02-Jan	Gene names, part numbers, dimensions, ratios, dosages, well-plate references.
Non-breaking spaces arrive invisibly	"Smith·" ≠ "Smith"	Anything pasted from a web page or a PDF. Every lookup against it silently returns #N/A.
Accented characters turn to mojibake	Niñal → NiÃ±al	Any CSV that was written as UTF-8 and opened without saying so.
The whole file lands in column A	a;b;c → one cell	Files written in a region that uses the semicolon as its list separator.
Numbers stored as text are ignored	SUM(...) → 0	Anything exported from an accounting or ERP system. The total is wrong, and it looks right.

THE ONE THAT CANNOT BE UNDONE

Excel stores at most 15 significant digits. A 16-digit card number or long barcode opened as a number loses everything past the 15th permanently — the file on disk is still correct, but what you save from Excel is not. No formula recovers it. Re-import the original as Text.

Eight rules that prevent all of it

Prevention is the only real fix. Cleaning afterwards recovers some damage; these rules mean there is none to recover. Learn rules 1 to 3 and you have covered most of it.

01 Never double-click a CSV you care about.

Double-clicking hands the file to Excel with every autocorrect switched on. Open Excel first with a blank workbook, then import the file deliberately.

02 Import with Data → From Text/CSV, and choose Transform Data.

Transform Data opens Power Query, where you set the type of every column before a single value is converted. Load applies the guesses immediately.

03 Set every ID column to Text, explicitly.

Postcode, SKU, part number, phone, ISBN, batch code, member ID. This one step prevents the first four failures on the previous page.

04 No Power Query? Rename the file to .txt first.

That forces the old Text Import Wizard, which lets you mark each column as Text at step 3. A .csv opens silently with defaults; a .txt asks you first.

05 For accented or non-Latin text, set File Origin to 65001 (UTF-8).

This is the entire fix for mojibake. The file was never broken; Excel guessed the wrong encoding on the way in.

06 Export with Save As → CSV UTF-8, not plain CSV.

Plain "CSV (Comma delimited)" writes your system codepage and corrupts every non-ASCII character on the way out.

07 A semicolon file is a regional setting, not a broken file.

Import it and pick the delimiter, rather than changing your Windows regional settings and breaking every other file you own.

08 Paste into a pre-formatted Text column.

Format the destination as Text before pasting, and use Paste Special → Values. Formatting a column afterwards does not undo a conversion.

Cleaning what has already arrived

These are the formulas the workbook uses. They work identically in Excel, Google Sheets, Numbers and LibreOffice — no macros, no add-ins, nothing version-locked.

THE CLEAN CHAIN — USE THIS, NOT TRIM ALONE

```
=TRIM(CLEAN(SUBSTITUTE(SUBSTITUTE(A2," NBSP "," ")," ZWSP ","")))
```

NBSP and **ZWSP** stand for the actual invisible characters — type them into the formula by copying one out of your own data, or use the workbook, which already has them. The two are replaced *first*. TRIM cannot see them, and CLEAN does not remove a non-breaking space because it is a printing character. Only once they are ordinary spaces can CLEAN strip control codes and TRIM collapse the runs and cut the ends.

WHY NOT CHAR(160)

```
=SUBSTITUTE(A2,CHAR(160)," ") NOT PORTABLE
```

Almost every article online recommends this. It is wrong outside Excel: in LibreOffice `CODE(CHAR(160))` returns 239, so the formula matches nothing and reports your data clean when it is not — and UNICHAR does not exist there at all. Putting the literal character inside the quotes needs no charset function and behaves the same everywhere.

VERIFIED

IS THIS TEXT PRETENDING TO BE A NUMBER?

```
=AND(ISTEXT(A2),ISNUMBER(VALUE(A2)))
```

TRUE means SUM is silently skipping it. This is why an exported ledger totals to something smaller than the column obviously contains.

WOULD EXCEL READ THIS AS A DATE?

```
=OR(AND(OR(ISNUMBER(SEARCH("/",A2)),ISNUMBER(SEARCH("-",A2))),
  ISNUMBER(VALUE(SUBSTITUTE(SUBSTITUTE(A2,"/",""),"-","")))),
  ISNUMBER(MATCH(UPPER(LEFT(A2,3)),{"JAN";"FEB";...;"DEC"},0)))
```

Shape-matching, not DATEVALUE. DATEVALUE is locale-dependent — when this was tested it matched 3/4 but not SEP1 or 1-2, both of which Excel does convert. This version does not change with your regional settings.

KEEP THE LEADING ZEROS ON PURPOSE

```
=IF(ISNUMBER(VALUE(A2)),TEXT(VALUE(A2),"@"),A2)
```

Where a value is genuinely a number that must keep its zeros, store it as text deliberately rather than leaving the decision to Excel.

Using the workbook that came with this

Everything in this guide is already built into the spreadsheet in your download. Open it, paste a column, and it tells you which of these problems your data has.

01 Paste Here

Drop your messy column into cell A2. Up to 300 rows. Both columns are already formatted as Text, so what you paste is what gets stored.

02 Diagnosis

Nine checks per row, one column each, plus a running count of how many rows have a problem at all. YES means that specific failure is present in that value.

03 Cleaned

The repaired values: cleaned, proper case, digits only, and a text-safe version that keeps leading zeros. Copy the column you need, then *Paste Special* → *Values*.

04 Compare Two Lists

Paste a second list into column B and it reports what is in A but not B, and the other way round — the reconciliation people usually do by eye.

05 How It Works

Every formula in the file, written out and explained, so you can lift them into your own workbook instead of depending on this one.

ONE HABIT WORTH KEEPING

Clean before you compare. COUNTIF and VLOOKUP match the value exactly as stored, so a single trailing space reports a row as missing when it is sitting right there. Almost every "the lookup is broken" hour ends here.