

NEVER MISS A STEP

Data Cleaning Checklist

A step-by-step checklist to make sure your data is genuinely ready for analysis — work through each section in order, and don't skip the backup step.

BEFORE YOU START

- ☐ **Make a backup copy** of the raw file before changing anything
Never clean data you don't have an unedited copy of.
- ☐ **Understand what each column represents** before deciding how to treat it
A column you don't understand is a column you shouldn't "fix" yet.

STRUCTURAL CHECKS

- ☐ **Check for duplicate rows**
Verify using a unique identifier (ID), never by name alone — two different people can share a name.
- ☐ **Check for missing values** in every column
Decide, per column: fill with a justified value, flag clearly, or exclude — and note which you chose and why.
- ☐ **Confirm each column's data type is correct**
Dates stored as text, or numbers stored as text, will silently break later calculations.

CONSISTENCY CHECKS

- ☐ **Standardize date formats** across the entire column
Pick one format and apply it everywhere — mixed formats are a common hidden error.
- ☐ **Standardize text casing** (e.g., "Lagos" vs. "LAGOS" vs. "lagos")
Use PROPER, UPPER, or LOWER consistently across a column.
- ☐ **Trim extra spaces** from text fields
Invisible trailing spaces can make two "identical" values fail to match.
- ☐ **Check categorical fields use a consistent, limited set of values**
E.g., a "Sex" column should only ever contain the values you expect — nothing else.

VALIDATION

- ☐ **Look for outliers** that may indicate an entry error
An age of 200, or a negative amount, is worth a second look before you trust it.
- ☐ **Check logical consistency across related fields**
E.g., a discharge date should never be earlier than an admission date.
- ☐ **Spot-check a sample of rows manually**
Automated checks catch a lot, but a few rows read by eye often catch what rules miss.

DOCUMENTATION

- ☐ **Keep a short log of every change you made**
What you changed, and why — so anyone (including future you) can retrace your steps.
- ☐ **Note anything you chose not to fix, and why**
Some issues are fine to leave for a specific analysis — just make that a decision, not an oversight.

FINAL CHECK BEFORE ANALYSIS

- ☐ **Re-run your duplicate and missing-value checks one more time**
Cleaning steps can sometimes introduce new gaps — confirm the final file is actually clean.
- ☐ **Confirm row and column counts make sense**
A sudden, unexplained drop in row count usually means something was deleted by accident.

Want structured practice cleaning real, messy datasets with a mentor guiding you? That's exactly where our programs start. See quantifiedinsights.net/programs or book a free consultation.