Tidy Data in Plain English

Principles for Clean Data Structure

Tidy data means each value has one clear place. Each column means one thing. Each row represents one case. Each table describes one kind of object.

A tidy dataset has one column for each variable. A variable means one attribute such as price, date, or score. A tidy dataset has one row for each observation. An observation means one record such as one student, one customer, or one event.

Tidy data separates different concepts into different tables. A table for people stays separate from a table for transactions. A table for classes stays separate from a table for grades. Links between tables use clear keys such as an ID field.

Tidy data avoids mixed meanings inside one cell. One cell holds one value. Not two values in one place. Not lists inside a cell. Not stacked meanings inside text.

Tidy data avoids spreading one variable across many columns. A single "measure" column works better than many repeated metric columns when the value represents the same concept. Long form supports analysis and reshaping with less friction.

Tidy data makes analysis easier. Cleaning becomes simpler. Visualization becomes clearer. Models train with fewer surprises. Collaboration becomes smoother because structure remains consistent.

A dataset becomes tidy through deliberate design, not through convenience. You name columns with purpose. You separate variables with care. You treat structure as part of the analysis, not an afterthought.

Strong projects begin with tidy data.

Weak structure creates confusion.

Clear structure creates insight.