SELECTED WORK

Data Audit and Validation

A practical spreadsheet data-quality sample

PROJECT TYPE Reusable data-quality workflow for spreadsheet-based research and operational datasets.

Audit sequence

Stage	Checks	Output
1. Structure	Column names, data types, row count, unique identifiers	Data inventory
2. Completeness	Missing values and blank patterns	Missingness summary
3. Validity	Ranges, dates, categories and logic rules	Exception list
4. Consistency	Spelling, units, labels and duplicated categories	Standardisation rules
5. Uniqueness	Duplicate rows and repeated identifiers	Duplicate review
6. Verification	Re-run checks after cleaning	Validation record

Example cleaning log

Issue	Rule	Action	Verification
Mixed Yes/No labels	Allowed: Yes, No	Standardise case and spelling	Frequency table
Age outside range	18 to 75	Flag for source review	Min/max check
Duplicate ID	One row per ID	Compare records before removal	Unique-count check
Blank outcome	Required for analysis	Report; do not invent value	Missingness check

Validation checklist

- Record the original row count and confirm the final row count.
- Preserve an untouched source file.
- Document every recoding or removal rule.
- Do not silently replace missing values.
- Check that category totals match the valid record count.
- Review extreme values in context before deciding they are errors.
- Recalculate derived fields after source columns are corrected.
- Use a separate sheet or file for the cleaning log.

Analysis-ready handoff

A clean dataset should be accompanied by a data dictionary, an issue log and a short note explaining exclusions. This makes later analysis reproducible and prevents unexplained changes.

RESPONSIBLE PRACTICE	Cleaning improves data quality; it must not be used to manufacture favourable results or hide valid observations.