



PROMPT WO-045

Clean and Validate Data

Purpose

Use this prompt to clean, validate, and prepare spreadsheet data for analysis. It identifies duplicates, missing values, formatting inconsistencies, invalid entries, and other data quality issues while recommending practical improvements to ensure accurate reporting and decision-making.

DIFFICULTY
 **Beginner**

TIME SAVED
 **30–60 min**

BEST FOR
Data Quality

Best AI Tools

- ChatGPT
- Claude
- Microsoft Copilot
- Google Gemini

Copy & Paste Prompt

You are an experienced Data Quality Analyst.

Review and clean the spreadsheet data using the information below.

Spreadsheet Purpose:
[Describe what the spreadsheet is used for]

Columns Included:
[List all column headings]

Data Sample:
[Paste the data or describe the dataset]

Known Issues:
[List any known problems or leave blank]

Desired Outcome:



[Explain how the cleaned data will be used]

Software:

[Excel, Google Sheets, CSV, etc.]

Please:

- Review the dataset for data quality issues.
- Identify duplicate records.
- Highlight missing or incomplete values.
- Detect inconsistent formatting (dates, numbers, text, currencies, etc.).
- Identify invalid entries or possible errors.
- Recommend validation rules to prevent future errors.
- Suggest the most efficient cleaning process.
- Summarise the improvements made and any remaining risks.

Return only the completed data quality report.

Example

INPUT

Spreadsheet Purpose: Customer database

Columns: Customer ID, Name, Email, Phone, Join Date

Known Issues: Duplicate records and inconsistent phone number formatting.

EXPECTED OUTPUT

Data Quality Report

Executive Summary

The dataset was reviewed for quality and consistency. Several issues were identified that could affect reporting accuracy and customer communications.

Issues Identified

Duplicate Records

- 12 duplicate customer records detected.
- Recommendation: Remove duplicates using Customer ID as the unique identifier.

Missing Values

- 8 records missing email addresses.
- 3 records missing join dates.

Formatting Issues



- Phone numbers stored in multiple formats.
- Dates recorded using both DD/MM/YYYY and MM/DD/YYYY formats.
- Mixed use of uppercase and lowercase text in customer names.

Invalid Entries

- Two email addresses appear to contain invalid formats.
- One future join date detected.

Recommended Validation Rules

- Require unique Customer IDs.
- Make email addresses mandatory.
- Restrict date fields to valid historical dates.
- Apply consistent phone number formatting.
- Use drop-down lists where appropriate.

Suggested Cleaning Process

1. Remove duplicate records.
2. Standardise date formats.
3. Correct phone number formatting.
4. Validate email addresses.
5. Complete or remove incomplete records.
6. Apply validation rules to prevent future errors.

Remaining Risks

The dataset will require periodic reviews to maintain quality as new records are added.

 **Builder's Tip:** *Poor-quality data leads to poor decisions. Spending a little time cleaning data before analysing it can dramatically improve the accuracy of reports, dashboards, and forecasts.*

Related Prompts

- WO-041 – Analyse Spreadsheet Data
- WO-042 – Create an Excel Formula
- WO-043 – Build a Pivot Table
- WO-050 – Audit an Excel Workbook

Tags

DATA
CLEANING

DATA
VALIDATION

EXCEL

SPREADSHEET

DATA QUALITY