

Data Analysis for Non-Coders — Workbook

This workbook turns the course into reps. Work through it with one real dataset of your own (or the practice file provided in the templates) open beside you. Each section maps to one course module: you will frame a question, clean a messy table, summarize it with pivots and the five core functions, and package the finding into a brief or dashboard. Do the exercises in the spreadsheet, not in your head, that is where the skill actually forms.

Thinking Like an Analyst Before You Touch a Spreadsheet

Practice turning fuzzy requests into sharp, answerable questions and stress-testing what your data can actually support.

Exercise: Sharpen Three Fuzzy Requests

Take three real or realistic requests and rewrite each as a single answerable question containing a metric, a population, a timeframe, and a comparison. Underline the fuzzy word in the original before you rewrite it.

- Rewrite 'How are sales doing?' into a one-sentence question with metric, population, timeframe, and comparison.

- Rewrite 'Is our marketing working?' the same way, naming the exact metric you would count.

- Rewrite a fuzzy request from your own work or studies, then state which columns you would need to answer it.

- For one of the three, write the single number or small table that would count as a complete answer.

Worksheet: Question Definition Sheet

Fill this in for the one analysis you will carry through the whole workbook. Keep it at the top of your spreadsheet and check every later step against it.

Original request (in quotes, as you heard it)

Fuzzy word(s) identified

Metric (exactly what you will count or measure)

Population (which rows qualify)

Timeframe (explicit start and end dates)

Comparison / benchmark (vs last year, target, etc.)

Final one-sentence question

What a complete answer looks like (a number, a ranked list, a chart)

Exercise: Find the Blind Spot

For your chosen dataset, interrogate what is missing before you trust it. The goal is to know the edges so you can say 'this suggests' rather than 'this proves'.

- List at least two things this dataset cannot tell you (who is not in it, what was not recorded).
 - Name one event in the period (a promotion, holiday, outage) that could distort the numbers.
 - Could a single large customer or one unusual day be skewing the totals? How would you check?
 - If AI offered a 'because' explanation, write one plausible alternative explanation yourself.
-

Checklist: Pre-Analysis Readiness Check

- I have written my final question as one sentence at the top of the sheet
- My question names a metric, population, timeframe, and comparison
- I have identified at least two things the data cannot tell me
- I have chosen my tool for this task (Excel, Sheets, or AI) on purpose
- I have confirmed no confidential columns will be pasted into a public AI chat
- I can state in one sentence what a complete answer will look like

Cleaning and Structuring Real Data

Reshape data into the tidy structure analysis needs, then run the five-problem cleaning drill on a real messy table.

Exercise: Tidy-Data Diagnosis

Open a real spreadsheet (or the messy-data-cleaning-drill template) and judge it against the three tidy-data rules: one variable per column, one record per row, one value per cell.

- Is row 1 a single clean header row with no title banner or blank rows above it? If not, what must you remove?
 - Find any column that is really many values (a month-per-column layout). How would you unpivot it?
 - Find any cell holding multiple facts (like '12 / red / large'). Which columns would you split it into?
 - List every subtotal row or blank spacer row sitting inside the data that must be deleted.
-

Exercise: The Five-Problem Cleaning Drill

Work the messy dataset through all five common problems IN ORDER. Always duplicate the sheet first and name the original 'RAW - do not edit'. Use built-in menus, no formula writing required to begin.

- Duplicates: how many duplicate rows did Remove Duplicates find, and what was the row count before and after?
 - Inconsistent text: pick one column, sort it, and list every variant of a single value (e.g. USA / U.S.A. / United States) plus the clean version you chose.
 - Stray spaces and blanks: which columns had leading/trailing spaces, and for each column does a blank mean zero, unknown, or delete?
 - Dates: are any dates stored as text (left-aligned)? Describe how you converted them to real dates.
-

Worksheet: Cleaning Log

Record every cleaning action so your work is reversible and auditable. One row per change.
Column name

Problem found (duplicate / inconsistent text / spaces / blank / bad date / other)

Action taken (menu used or rule applied)

Rows affected

Row count before

Row count after

Verified? (spot-checked a few known rows: yes/no)

Checklist: Clean-Data Sign-Off

- I made a copy and kept an untouched 'RAW - do not edit' sheet
- Row 1 is a single clean header row with no gaps or merged cells
- Each column is one variable and each cell holds a single value
- Duplicates removed and row count change recorded
- One column's text variants unified via Find and replace
- Leading/trailing spaces removed and blanks handled per a stated rule
- All dates are real dates that sort chronologically (not left-aligned text)
- I used AI only to catalogue/suggest fixes and applied them myself in the sheet

Summarizing and Finding Patterns

Build pivot tables, apply the five core functions, and use AI to interpret the summary without getting fooled.

Exercise: Five Pivot Tables, Five Answers

On your clean table, build five pivot tables that each answer one version of your question. Drag a grouping field to Rows, a measure to Values, and change the summary (Sum, Average, Count) to match the question.

- Total by category: which category had the highest total, and what was it? (Rows = category, Values = Sum)

- Typical value: switch Values to Average. Did the ranking change? What does that tell you?

- How many: switch Values to Count. Which category simply had the most records (a different question)?

- Cross-tab: add a second field to Columns (e.g. month or year). What pattern appears that a single breakdown hid?

Exercise: The Five-Function Workout

Write each of the five core functions once on your data and test it on a row where you already know the answer before filling it down. If a function errors, paste the formula and error into AI and ask why.

- SUM and AVERAGE: write both on one numeric column. Does the average look distorted by an outlier? Check the median too.

- COUNTIF (or SUMIF): write one that counts (or totals) rows meeting a condition. State the condition and

the result.

- XLOOKUP (or VLOOKUP): join two tables by a shared key (e.g. pull a price by product code). Confirm the result on one known row.
- IF: create a label column (e.g. =IF(amount>1000,"large","small")) so you can group by it in a pivot. What share came out 'large'?

Worksheet: Finding Capture Sheet

For each pivot or formula result worth keeping, capture the number AND the interpretation. Force yourself to apply the 'so what' test in the last field.
Question this answers

Pivot / formula used

The key number(s)

What stands out (one sentence)

Alternative explanation I should rule out

How I will verify it (second method)

So what? (the consequence or action this implies)

Checklist: Interpretation Guardrails

- I pasted only my summary (pivot/numbers) into AI, never raw confidential rows
- I asked AI for the three most notable things AND one alternative explanation
- I treated every AI 'cause' as a hypothesis to verify in the sheet
- I re-checked any specific number AI repeated back against my actual table
- I asked AI to 'argue the opposite' to surface weaknesses in the reasoning
- Any finding that drives a decision was confirmed two different ways
- Each kept finding survives the 'so what' test (it points to an action or consequence)

Presenting Insights People Act On

Choose honest charts, write an answer-first one-page brief with SCQA, and assemble a self-updating dashboard.

Exercise: Match the Chart to the Question

For your finding, choose the chart that fits the question and strip it to the essentials. Avoid the four misleading visuals (truncated y-axis, misleading dual-axis, 3D pie, cherry-picked time range).

- Write the one-sentence question your chart must answer.
- Name the chart type that fits (bar for comparison, line for time, scatter for relationship, big number for a single metric) and why.
- Did you sort the data meaningfully and does the bar-chart axis start at zero? Note what clutter you removed.
- Which of the four misleading visuals were you tempted by, and how did you avoid it?

Worksheet: SCQA One-Page Brief Builder

Draft your brief by filling each field, then assemble it answer-first: the Answer and your single best chart go near the top, with the title stating the finding itself.

Page title (the finding stated as a sentence, e.g. 'West region losses put the annual target at risk')

Situation (one sentence of shared context)

Complication (what changed or went wrong)

Question (the question that raises)

Answer (your finding, plainly stated)

Support point 1 (with number)

Support point 2 (with number)

The one chart that proves it (and its takeaway caption)

Recommendation / decision requested

Worksheet: Dashboard Layout Plan

Plan a one-screen dashboard before you build it, so it reads in five seconds and refreshes when new data is pasted in.

Top-row number 1 (most important metric, goes top-left)

Top-row number 2

Top-row number 3

Trend chart (which metric over time)

Comparison chart (which categories)

Filter / slicer control (region, month, channel)

Data source sheet name (raw data kept separate)

Refresh routine (how new data gets in and what you click)

Last-updated cell location

Checklist: Presentation Sign-Off

- Each chart type matches the question it answers
- No misleading visuals: bar axes start at zero, no 3D pie, no cherry-picked range
- My brief leads with the answer and the page title states the finding
- Every number on the brief was fact-checked against the spreadsheet
- The brief ends with a clear recommendation or decision requested
- Raw data sits on its own sheet; results never typed over it
- Dashboard has 3 to 5 key numbers up top, a trend chart, and a comparison chart
- Pasting new data and clicking Refresh updates every number and chart
- A 'last updated' date is visible on the dashboard

Your Action Plan

1. Pick one real dataset (or open the spreadsheet-practice-dataset template) and write your final one-sentence question at the top of the sheet.
2. List two things the data cannot tell you and one event that could distort it; decide whether your conclusion needs a caveat.
3. Duplicate the sheet, keep a 'RAW - do not edit' copy, and confirm row 1 is a single clean header row with no gaps.
4. Run the five-problem cleaning drill in order (duplicates, inconsistent text, spaces, blanks, dates) and fill in the Cleaning Log.
5. Build at least five pivot tables, switching Sum / Average / Count, to ask your question several ways and find the pattern.
6. Write the five core functions (SUM, AVERAGE, COUNTIF/SUMIF, XLOOKUP, IF), testing each on a known row before filling down.
7. Paste only your summary into ChatGPT or Claude; ask for what stands out, an alternative explanation, and 'argue the opposite'.
8. Capture each finding with its 'so what', then choose the one honest chart that proves your main point.
9. Write a one-page SCQA brief that leads with the answer, fact-check every number, and end with a recommendation.
10. If the question recurs, assemble a one-screen dashboard wired to the data sheet so a paste-and-Refresh keeps it current.

