CareerSwami · Data Analytics & PowerBI

FREE TWO-WEEK INTERNSHIP

Project 1 — Explain why a campus store is busier and earning less

Four weeks of data. Visits up every week, revenue drifting down. Books are the worst of it — more people looking, fewer buying, more coming back. Something changed in week three and nobody knows what.

2 weeks	1 project	6–8 hrs
START TO SUBMIT	ONE SUBMISSION	PER WEEK

WHAT IS IN THIS BRIEF

1. The situation — the problem you are being handed
2. The four things hiding in this dataset — the cases your work must survive
3. What you are given — your data and starting files
4. What finished looks like — the two things you hand in
5. How the two weeks run — week by week
6. Before you start — rules, tools and your starter kit
7. How the work is judged — the marking scheme
8. How to submit — one link, pasted into your dashboard

Read 1 to 4 before you do anything. They take five minutes and they are what the marking scheme is built on. Sections 5 to 8 are the how — come back to them once you know what you are solving.

1 The situation

Treat this as a real client brief. The store is fictional, but the pattern — a headline number going up while the business gets worse — is the most common thing an analyst is asked to explain.

A campus store sells books, stationery and accessories. You have four weeks of data: visits, orders, revenue and returns, split by category. The manager's summary is that things are going well, because visits have gone from six hundred and fifty a week to eight hundred and eighty.

Look at books and the story falls apart. Week one: two hundred visits, forty orders, eight thousand rupees. Week four: two hundred and eighty visits, twenty-eight orders, five thousand six hundred. Forty percent more people looking and thirty percent fewer buying. Returns in books have gone from two a week to five.

Accessories are doing the opposite and stationery is flat, so the total hides all of it. Your job is to take the raw file, clean it properly, and build something that shows a manager where the problem actually is — without claiming to know things this data cannot tell you.

WHAT YOU ARE BEING ASKED TO DO

Clean the data, build a model you can explain, and produce a dashboard that shows what is really happening by category — with a written note on the questions this dataset cannot answer.

2 The four things hiding in this dataset

Each of these is something that a total or an average conceals. Your quality log and your annotations should show you found all four.

Books — conversion collapsing

Visits 200 → 280. Orders 40 → 28.

Why the obvious reading is wrong — In week one, one visit in five became an order. By week four it is one in ten. Reporting visits and revenue side by side hides this entirely; only the ratio shows it, and the ratio is not in the file — you have to create it.

Returns rising on falling orders

Books: 2 returns on 40 orders, then 5 on 28.

Why the obvious reading is wrong — Five looks small next to two. As a share of orders it has gone from 5% to 18%, which is a different kind of problem altogether. Rates, not counts — and say so when you present it.

Gross revenue only, with no refund amounts

Revenue in the file is before refunds.

Why the obvious reading is wrong — You know how many items came back and nothing about what they were worth. Any claim about net revenue is unsupported. Name this limitation in the memo rather than quietly estimating it.

The dirty copy

One duplicate row and one blank orders cell.

Why the obvious reading is wrong — Load it without looking and the duplicate inflates one week and the blank cell either becomes a zero or silently drops a row from an average. The two are different mistakes with different consequences, and your quality log must record what you found, what you did, and why.

Week three is where books turn: visits keep climbing but orders fall from forty-four to thirty-five. Something changed then. You cannot tell what from this data — and saying so clearly is worth more than a confident guess.

3 What you are given

Twelve rows, four weeks, three categories, and two deliberate defects in the dirty copy. The ratios that make it legible are not in the file — you create them.

The columns you are given

`week` `category` `visits` `orders` `revenue` `returns`

Books across four weeks, so the pattern is visible

`Week 1 · Books`	200 visits · 40 orders · ₹8,000 · 2 returns
`Week 2 · Books`	220 visits · 44 orders · ₹8,800 · 2 returns
`Week 3 · Books`	250 visits · 35 orders · ₹7,000 · 4 returns
`Week 4 · Books`	280 visits · 28 orders · ₹5,600 · 5 returns
`Conversion`	20% → 20% → 14% → 10%
`Return rate`	5% → 4.5% → 11% → 18%
`What you cannot say`	Why — this data contains no cause

Four steps the work must show

`Define` `Clean` `Model` `Explain`

All figures are fictional. Revenue is gross; refund amounts are not supplied, so net revenue cannot be calculated and must not be presented. Keep one untouched copy of the raw data. Where you must assume something, record the assumption — a recorded assumption is good practice; an assumption presented as a finding is not.

4 What finished looks like

At the end of two weeks someone should be able to open your work and know what is wrong with the store and what you had to assume.

- **A brief, a data dictionary and a quality log** — what question you are answering and for whom; every column defined, with its type, its grain and what it does not include; and a record of every data problem you found, what you did about it, and what that decision could distort.

- **Cleaned data, a transformation log and a model diagram** — the clean dataset with the raw file preserved untouched alongside it, a step-by-step record of every transformation so the cleaning can be re-run, and a diagram of how your tables relate and where your calculated measures sit.

Preserve the raw data. An analysis you cannot re-run from the original file is not reproducible, and the first question a reviewer asks is what the numbers looked like before you touched them.

5 How the two weeks run

Work through it in two parts — but you submit **once**, at the end, not once per week.

PART 1 · WEEK 1

Understand the data before you chart any of it

Define the question, document every column honestly, and go through the dirty file looking for what is wrong with it. Resist building a visual until you can describe the grain of the data in one sentence.

Produce

- A brief: the question, the audience, what a useful answer looks like
- A data dictionary: every column, its type, its grain, what it excludes
- A quality log: each issue found, the fix applied, the risk that fix carries

> Your dictionary must state that revenue is gross and that refund amounts are absent. The quality log must name both seeded defects — the duplicate row and the blank orders cell — and say what you did with each.

PART 2 · WEEK 2

Clean it reproducibly and model it

Transform the data in recorded steps rather than by hand, keep the raw file intact, and build the model with the ratio measures the raw columns do not contain.

Produce

- A clean dataset, with the raw file preserved separately and untouched
- A transformation log: every step, in order, repeatable
- A model diagram showing tables, relationships and calculated measures

> Your measures must include conversion rate and return rate, not just counts. A transformation you did by typing over a cell is not reproducible — redo it as a step.

6 Before you start

- Work in rates as well as counts. Conversion and return rate are the two measures that make this dataset legible, and neither is in the file.

- Never edit the raw file. Clean into a copy and record the steps so the whole thing can be re-run.
- A blank cell and a zero are different. Decide which yours is, say why, and note what the alternative would have changed.
- Totals across all three categories hide the entire story. Any dashboard whose headline is a single revenue number has missed the point.
- This data contains no causes. You may note that week three is where books turn; you may not say what caused it.

Your tools

Power BI Desktop is free on Windows. Excel, Google Sheets, Tableau Public or Python with pandas are all acceptable alternatives — the thinking is what is marked, not the logo. No paid tool is required, and no ID document is ever required.

CS_data-analytics-powerbi_Starter_Kit.zip — campus_store_raw.csv, campus_store_dirty.csv, a data dictionary template, a quality log, a transformation log, an evidence log, and a file index.
https://careerswami.com/wp-content/uploads/careerswami-projects/CS_data-analytics-powerbi_Starter_Kit.zip

7 How the work is judged

WHAT IS ASSESSED	WEIGHT
Problem and requirements — did you understand what you were solving	20%
The core artefact — the thing you actually made	40%
Validation and evidence — how you know it works	25%
Clarity and reproducibility — can someone else open and follow it	15%

Strong	Complete, correct, evidence-backed and easy to inspect.
Developing	A useful partial result, with the gaps explained.
Needs revision	Core deliverable missing, claims unverified, or the work cannot be opened.

Credit reflects demonstrated work, not visual polish. A small, honest result beats a large, unsupported one.

What counts as evidence

expected result → actual result → the table, measure or screenshot it came from → conclusion. Then say what the evidence does not establish.

Why this one is worth doing. Dashboard tools are easy to learn and every applicant has one on their CV. What a hiring manager is checking is whether you noticed the number that was going up while the business got worse, and whether you were willing to write down what the data could not tell you. Those two habits are the whole of analytical credibility, and they are visible in a quality log within a minute.

8 How to submit

One link, pasted into your dashboard

1. Put your work somewhere with a link — Google Drive, Figma, GitHub, Notion, Loom. Whatever fits the work.
2. **Set it so anyone with the link can view.** A locked link is the single most common reason a submission comes back.
3. Include a short README or cover page saying what to open first, and where each deliverable lives.
4. Open your dashboard, paste the link into the submit box, and add a line or two on anything you want the reviewer to know.

Name files `CS_data-analytics-powerbi_YourName_v1`. Keep your own copy. A screenshot on its own is not enough for anything interactive. A video is optional — annotated screenshots or narrated slides are fine.

Submitted and heard nothing, or the link came back? Email **support@careerswami.com** with the email address you enrolled with.

Work honestly. Use fictional data and your own work. Credit anything you reused, and disclose AI assistance — including what you checked yourself. Never invent interviews, users, tests or results. Nothing here requires you to spend money, contact real organisations, or deploy anything publicly. Never upload identity documents, real customer records or credentials.

Stuck, confused, or something is broken? Write to us.

`support@careerswami.com`

Or WhatsApp **+91 7383138418**. We answer on working days. Use it for anything at all — a part of this brief that does not make sense, a starter file that will not open, a dashboard link that has stopped working, a deadline you need moved, or a submission you are not sure about.

To get a useful answer fast, spend twenty minutes narrowing the problem down first, then tell us four things: what you expected, what actually happened, what you already tried, and your one specific question. Include your track name and the email you enrolled with.

Asking is not penalised. Submitting partial work with a clearly stated blocker always scores better than a finished answer you copied, and better than a silent no-show.

Optional stretch

Optional, and only once the main work is done — cap it at an hour. Add a small-multiples view comparing all three categories on the same scale, or model what net revenue would look like under two refund assumptions, clearly labelled. Label it clearly as a stretch.