VIRTAVIS
LEARNING

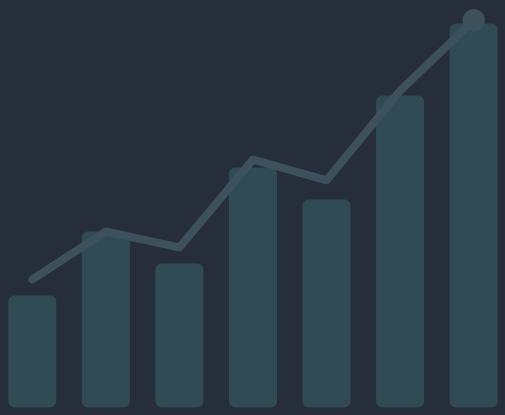
FREE PRACTICE DATASET

FOR EXCEL, SQL, POWER BI AND PYTHON PRACTICE

Kettle & Bean Co.
Practice Dataset

A full year of realistic sales data from a UK coffee and tea retailer, January to December 2023. Messy in all the right places.

- **10 linked tables**
  8,667 orders and 19,747 order lines
- **5,027 customers**
  12 shops plus a website
- **Four ready-to-use formats**
  Excel, CSV, SQLite and PostgreSQL
- **A full data dictionary**
  every column explained in plain English
- **Check figures**
  so you know your answer is right
- **20 practice questions**
  from warm-up to a board-ready brief



CREATED BY

Sophie Lumpa

Data and Business Intelligence Specialist
Founder, Virtavis Learning

Check the numbers. Then say what they mean.

virtavis.com

SECTION 01

Start here

Kettle & Bean Co. is a fictional company, but its data behaves like the real thing: walk-in customers with no account, cancelled orders, discount codes, late deliveries and a few honest mistakes.

MEET KETTLE & BEAN

Kettle & Bean began in 2016 as a market stall in Bristol. By the end of 2023 it sells coffee, tea, brewing equipment, kitchenware, gifts and hampers through **12 shops in five regions** and a **website** that delivers across the UK.

This sample covers every order placed from **1 January to 31 December 2023**: 8,407 completed and 260 cancelled. Shop sales are retail only, as café drinks go through a separate till.

WHAT A YEAR OF DATA CAN TELL YOU

- Which shops, regions and products earn the most, and which are slipping
- How the website compares with the high street
- Whether discount codes and the loyalty scheme pay off
- What gets returned, and why
- How sales build towards Christmas

WHAT IS IN THE DOWNLOAD

FILE	BEST FOR	WHAT IT IS
`Kettle_and_Bean_2023.xlsx`	Excel	One sheet per table, each an Excel Table, plus a Start here sheet. Open it in Excel or upload it to Google Sheets.
`csv/ folder`	Power BI, Python, any tool	One CSV file per table (UTF-8). In Power BI use Get data › Folder, or load the files one by one.
`kettle_and_bean_2023.db`	SQL with no install	A SQLite database. Open it in the free DB Browser for SQLite, DBeaver, or Python's sqlite3.
`kettle_and_bean_2023_postgresql.sql`	SQL in PostgreSQL	Creates every table and loads the data. Run it in pgAdmin's Query Tool on an empty database.

All four hold exactly the same data. Prices are in pounds sterling (£) and dates are written year-month-day (2023-12-24).

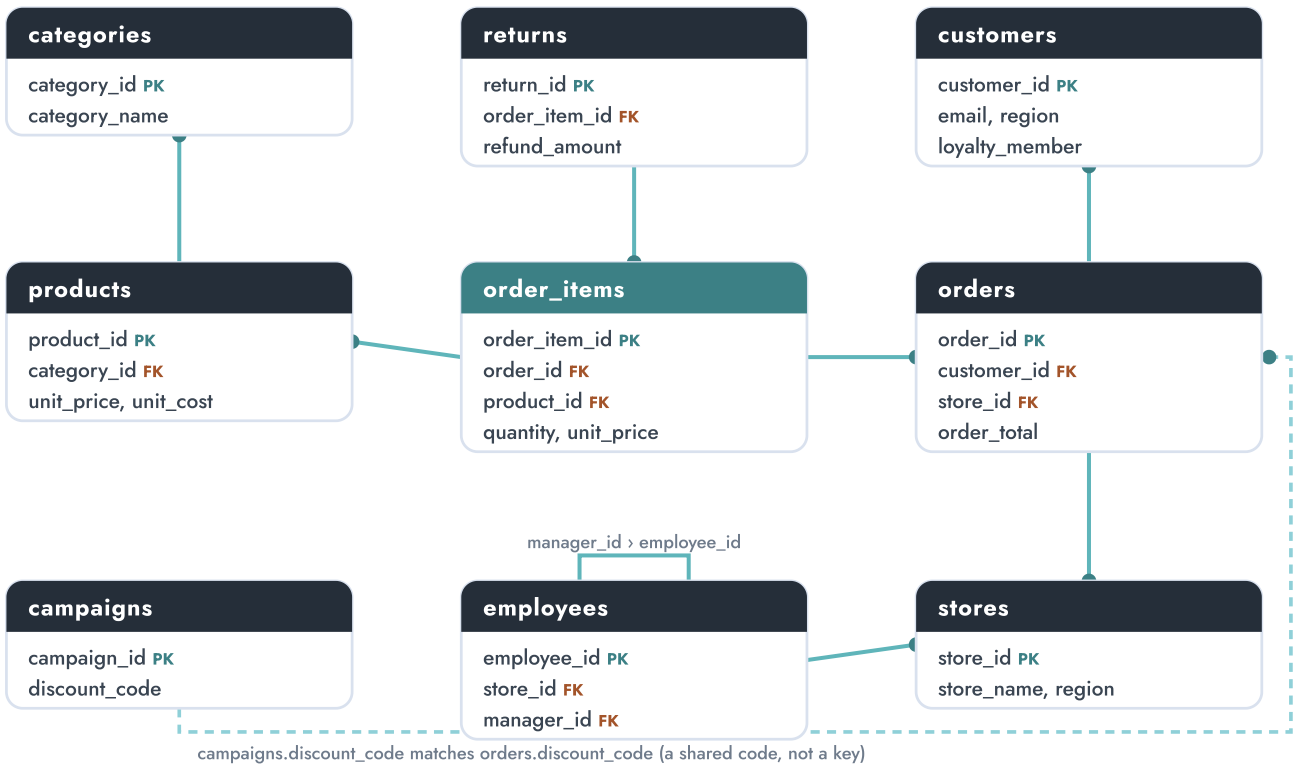
FIRST STEPS IN YOUR TOOL

Excel	Start with the **sales_lines** sheet for PivotTables. Then practise XLOOKUP between **orders**, **customers** and **products**, and Power Query to combine them.
SQL	Start with `SELECT * FROM orders LIMIT 10;` Then join **order_items** to **products** and **orders**. The **sales_lines** view shows a finished join you can check yours against.
Power BI	Load the nine core tables (skip sales_lines) and build a star schema: **order_items** as the fact table, the rest as dimensions. Add a date table for time intelligence.
Python	`pd.read_csv("csv/orders.csv", parse_dates=["order_date", "delivery_date"])` Then merge, group and plot. In Colab, upload the csv folder first.

SECTION 02

How the tables fit together

Nine core tables linked by keys, plus one ready-made flat table. Learn this picture and every question becomes a matter of which tables to join.



PK primary key: one unique value per row. **FK** foreign key: points to the primary key of another table. **sales_lines** is not shown: it is order_items with the names already joined on.

SIX RULES TO KNOW BEFORE YOU START

One row per...

orders holds one row per order; **order_items** one row per product on that order. Joining them repeats each order on every line, so never add up **order_total** after the join.

Revenue

Line revenue = quantity × unit_price × (1 − discount_pct). The **sales_lines** table has it ready as **line_revenue**.

Completed only

Count sales from **status = Completed**. Cancelled orders stay in the data, just as they would in a real system.

Blank customer

963 shop orders have no customer_id: walk-in shoppers who did not give an account. They are real sales, so keep them in revenue.

Prices

order_items.unit_price is the price charged on the day. **products.unit_price** is the 2023 list price.

Returns

Returns belong to 2023 orders. 80 arrived in early January 2024, after Christmas.

SECTION 03

Data dictionary

Every column in every table. Keep this page open while you work: most mistakes start with a column that means something slightly different from what you assumed.

orders — 8,667 rows

One row per order placed in 2023, in a shop or online.

order_id	Whole number	Unique id.
customer_id	Whole number	Links to customers. Blank for walk-in shoppers.
store_id	Whole number	Links to stores. Online orders use the website, store 3.
order_date	Date	Date the order was placed.
channel	Text	In store or Online.
status	Text	Completed or Cancelled.
discount_code	Text	Code used, if any.
delivery_date	Date	Date delivered. Completed online orders only.
order_total	£ amount	Amount charged for the order, after discounts.

products — 118 rows

Every product on sale in 2023, with its 2023 price and unit cost.

product_id	Whole number	Unique id.
sku	Text	Stock code, e.g. COF-HBB-250.
product_name	Text	Name as shown on the till and website.
category_id	Whole number	Links to categories.
unit_price	£ amount	2023 list price, including VAT.
unit_cost	£ amount	What Kettle & Bean pays per unit. Some are missing.
launch_date	Date	First day on sale.
is_discontinued	True / False	True if withdrawn by 31 December 2023.
discontinued_date	Date	Last day on sale, if withdrawn.

order_items — 19,747 rows

One row per product line on an order.

order_item_id	Whole number	Unique id for each line.
order_id	Whole number	Links to orders.
product_id	Whole number	Links to products.
quantity	Whole number	Units bought.
unit_price	£ amount	Price per unit on the day, before discount.
discount_pct	Decimal	Discount as a decimal: 0.10 is 10% off.

customers — 5,027 rows

Everyone who had signed up by 31 December 2023.

customer_id	Whole number	Unique id.
first_name, last_name	Text	Customer name (fictional).
email	Text	Contact email (all @example.com). Some are blank.
city, region	Text	Where the customer lives.
signup_date	Date	Date the account was created.
signup_channel	Text	In store or Online.
loyalty_member	True / False	True if in the loyalty scheme.

stores — 13 rows

The 12 shops and the website, with region, type and opening date.

store_id	Whole number	Unique id for each store. The website is store 3.
store_name	Text	Shop name, or Website.
city	Text	Town or city. Blank for the website.
region	Text	South West, London, South East, Midlands, North, or Online.
store_type	Text	Flagship, High Street or Online.
opened_date	Date	Date the shop opened.

categories — 6 rows

The six product categories.

category_id	Whole number	Unique id.
category_name	Text	Coffee, Tea, Brewing Equipment, Kitchenware, Gifts & Hampers, Accessories.

returns — 820 rows

Returned items from 2023 orders, with reason and refund.

return_id	Whole number	Unique id.
order_item_id	Whole number	The order line that came back.
return_date	Date	Date the return was processed.
reason	Text	Changed mind, Faulty, Not as described, Damaged in transit or Wrong item sent.
refund_amount	£ amount	Amount refunded.

campaigns — 3 rows

Marketing campaigns running in 2023 and their discount codes.

campaign_id	Whole number	Unique id.
campaign_name	Text	Name of the campaign.
channel	Text	Email or Social.
start_date, end_date	Date	When the campaign runs.
budget	£ amount	Total budget for the whole campaign.
discount_code	Text	Code customers enter. Matches orders.discount_code.

employees — 81 rows

Head office, regional and shop staff, with who reports to whom.

employee_id	Whole number	Unique id.
first_name, last_name	Text	Employee name (fictional).
job_title	Text	Role, from Chief Executive to Barista.
store_id	Whole number	Their shop. Blank for head office and regional staff.
manager_id	Whole number	employee_id of their manager. Blank for the Chief Executive.
hire_date	Date	Start date.
leave_date	Date	Last day, if they left during or before 2023.

sales_lines — 19,747 rows

Ready-made flat table: one row per order line with names and revenue already added. Built for PivotTables.

order_item_id … discount_pct		One row per order line, with order, store, customer, product and category columns already joined on.
line_revenue	£ amount	quantity × unit_price × (1 − discount_pct), rounded to the penny.

SECTION 04

Check your numbers

Analysts earn trust by checking a result against a figure they already trust. Use these to test your formulas, queries and measures before you build on them.

CHECK FIGURES FOR 2023

8,407
completed orders
status = Completed

£334,089.83
revenue from order lines
each line rounded to the penny, then added

£39.74
average order value
mean order_total, completed orders

29.3%
of completed orders online
channel = Online

3,846
customers placed an order
distinct customer_id in orders

£16,034.62
refunded on 820 returns
sum of refund_amount

Two revenue figures, both "right"
Add up **order_total** for completed orders and you get £334,083.08. Add up the order lines and you get £334,089.83 (£334,089.61 if you skip rounding each line to the penny). Real company data often disagrees with itself like this. Finding out why, and saying which figure you used, is part of the job.

DATA HEALTH CHECK: SIX PROBLEMS TO FIND

The data has been left as a real system would produce it. Before you analyse anything, find these problems, decide what to do about each one, and write down what you decided.

Missing emails
93 customers have no email address.

Possible duplicate customers
27 email addresses belong to more than one customer_id. Duplicates, or a shared family inbox?

Impossible quantities
4 order lines have a quantity of zero.

Time travel
2 online orders were delivered before they were placed.

Totals that do not add up
7 orders have an order_total that differs from the sum of their lines. Overall the two revenue figures are £6.75 apart.

Missing costs
6 products have no unit_cost, so their margin cannot be worked out yet.

HABIT TO BUILD

For every problem, say three things: **how many rows**, **how much money** it affects, and **what you did** (fixed, excluded or flagged).

SECTION 05

20 practice questions

Work through them in order: each level builds on the one before. Warm-up answers are at the bottom, so you can check you are set up correctly.

WARM-UP

Single table, one answer

Excel · SQL · Python

1. How many orders were completed in 2023, and how many were cancelled?
2. Which shop took the most in-store revenue?
3. Which product category earned the most revenue?
4. What was the average value of a completed order?
5. Which month had the highest revenue?

CORE

Join, group and compare

Any tool

6. Show revenue by region and by channel. What share of revenue came from the website?
7. List the top ten products by revenue and by units sold. Are they the same products?
8. Work out the return rate (items returned ÷ items sold) for each category. Which is highest, and what is the main reason?
9. What share of orders used a discount code? Are discounted orders bigger or smaller on average?
10. Compare loyalty members with other customers: average order value and orders per customer.
11. How many days does an online order take to arrive, on average, month by month? When does it slip?

STRETCH

Windows, ranks and self-joins

SQL · Power BI · Python

12. Calculate month-on-month revenue growth for the whole company.
13. Rank the shops within each region by revenue.
14. What percentage of customers who ordered in 2023 ordered more than once?
15. List every employee next to their manager's name. How many staff does each shop have?
16. Did the Black Friday 2023 weekend beat an average weekend? By how much?
17. Four shops opened in 2022. How did their monthly revenue build up through 2023?

BUSINESS BRIEF

Result › so what › now what

Your choice

18. The Sales Director asks: "How did 2023 go?" Write five sentences: three findings with numbers, one risk and one recommendation.
19. Build a one-page dashboard for shop managers: revenue, orders, average order value and return rate, by month and by shop.
20. Pick one data problem from Section 04. Write a short note to the person who owns the system: what is wrong, how big it is, and what you need from them.

WARM-UP ANSWERS

1. 8,407 completed, 260 cancelled
2. London Covent Garden, £29,106.75
3. Coffee, £94,243.00
4. £39.74
5. December 2023, £57,377.32

Model answers for questions 6 to 20 are not included. The Virtavis courses work through questions like these step by step, with a model answer for every challenge.

SECTION 06 · KEEP GOING

This is one year. The courses have three.

Every course in the Virtavis Data Analyst Path is built on Kettle & Bean Co. In the full dataset you get 2023 to 2025, the York shop opening, targets, supplier price lists and the stories hidden inside them. You join the company as a junior analyst and each lesson starts with a request from a colleague who needs an answer.

ONLINE COURSES

LEARN WITH THE SAME COMPANY

Videos, a handbook, section quizzes, challenges with model answers and a final project in every course. Find them on Udemy, single or as a bundle.

Excel for Data Analytics
XLOOKUP, PivotTables, Power Query and a dashboard.

SQL for Data Analytics
Queries, joins, window functions and KPIs in PostgreSQL.

Power BI for Data Analytics
Power Query, a star schema, DAX and a board dashboard.

Python for Data Analytics
pandas, Plotly, cohorts and customer segments.

Statistics for Data Analytics
A/B testing, customer churn and machine learning.

The Virtavis Data Analyst Path
Take them in order: Excel › SQL › Power BI › Python.

PREFER TO LEARN LIVE?

Small-group cohorts

Weekly live sessions with Sophie, feedback on your work and a small group learning alongside you. Details at virtavis.com.

MORE FREE AT VIRTAVIS.COM

- SQL, Excel, DAX and Python cheat sheets
- KPI cheat sheet
- Portfolio kits
- Interview playbook
- Monthly data challenge

Pick up where this year ends
All three years of Kettle & Bean, guided projects and model answers.

Start the course ›

VIRTAVIS LEARNING

Check the numbers. Then say what they mean.

virtavis.com

© 2026 Sophie Lumpa, Virtavis Learning. Kettle & Bean Co., its people and its customers are fictional; any likeness is coincidence. Use the data freely for learning and in your portfolio, with credit to Virtavis Learning. To share it, please link to virtavis.com rather than passing on the files. Icons: Lucide (ISC licence).