

SQL & Python | Data Analysis, Reporting & Automation

Eight SQL queries and one Python script that turn raw invoice data into reporting tables and charts.
Fictional sample data.

What it does

- Checks the data first: duplicates, missing customers or products, impossible dates.
- Answers the monthly questions in SQL: sales and margin, category mix, top customers, budget variance, growth.
- Shows who owes money and for how long (receivables ageing and overdue ranking).
- One Python script runs every query and saves CSV tables and charts, so the monthly report is a single command.

Example query: receivables ageing

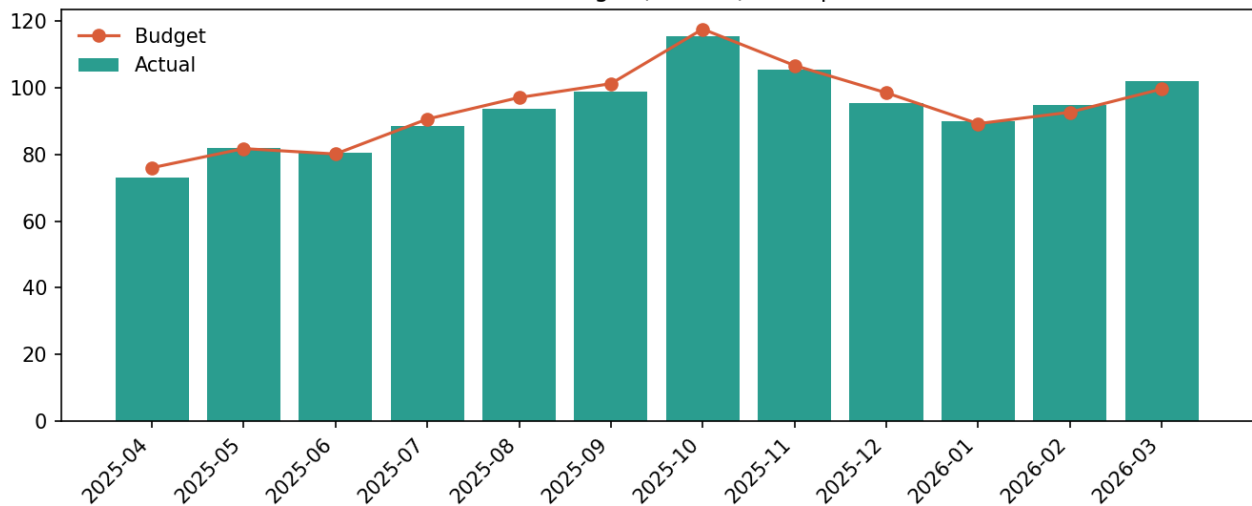
```
-- Open receivables by days past due, as on 31 Mar 2026
SELECT CASE WHEN julianday('2026-03-31')-julianday(due_date) <= 0 THEN '1 Not due'
            WHEN julianday('2026-03-31')-julianday(due_date) <= 30 THEN '2 1-30 days'
            WHEN julianday('2026-03-31')-julianday(due_date) <= 60 THEN '3 31-60 days'
            WHEN julianday('2026-03-31')-julianday(due_date) <= 90 THEN '4 61-90 days'
            ELSE '5 90+ days' END AS bucket,
        COUNT(*) AS invoices, ROUND(SUM(net_sales)/1e5,2) AS open_lakh
FROM sales WHERE paid_date IS NULL GROUP BY 1 ORDER BY 1;
```

Result from the sample data

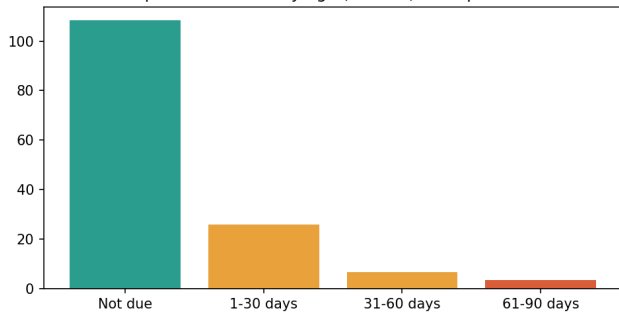
Bucket	Invoices	Open (Rs lakh)
Not due	414	108.37
1-30 days	109	25.85
31-60 days	29	6.71
61-90 days	14	3.55

Charts produced by analysis.py

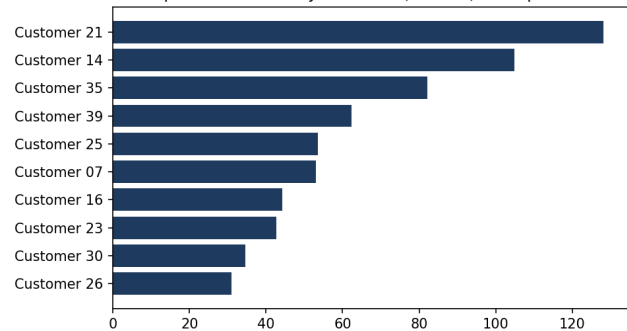
Net sales vs budget (Rs lakh) - sample data



Open receivables by age (Rs lakh) - sample data



Top 10 customers by net sales (Rs lakh) - sample data



Example query: month-on-month growth (window functions)

```
-- Month-on-month growth using LAG, plus running total
WITH m AS (SELECT substr(invoice_date,1,7) AS month, SUM(net_sales) AS sales FROM sales GROUP BY 1)
SELECT month, ROUND(sales/1e5,2) AS sales_lakh,
       ROUND(100.0*(sales-LAG(sales) OVER (ORDER BY month))/LAG(sales) OVER (ORDER BY month),1) AS mom_pct
,
       ROUND(SUM(sales) OVER (ORDER BY month)/1e5,2) AS ytd_lakh
FROM m ORDER BY month;
```

Data quality checks (all run before reporting)

Check	Issues found
duplicate invoice ids	0
sales without a customer	0
sales without a product	0
negative or zero net sales	0
paid before invoice date	0

Top 10 customers by net sales

Customer	Region	Net sales (Rs lakh)	Share %
Customer 21	North	128.22	11.5
Customer 14	West	104.84	9.4
Customer 35	South	82.17	7.3
Customer 39	West	62.26	5.6
Customer 25	South	53.61	4.8
Customer 07	West	53.00	4.7
Customer 16	South	44.24	4.0
Customer 23	North	42.63	3.8
Customer 30	North	34.56	3.1
Customer 26	Central	30.97	2.8

Skills shown

- SQL: joins, CTEs, window functions (RANK, LAG, running total), CASE, date arithmetic.
- Python: pandas, sqlite3, matplotlib, file handling, a repeatable monthly run.
- Reporting habits: validate first, one definition per number, outputs saved as files.

All data is fictional and created only to demonstrate the work. No real company information is shown.