Data Analyst Interview Question Bank

(Europe/Ireland • 3–4 Years Experience)

Curated from 15+ real interviews (mid-2025).

---

1 SQL & Databases

1. Explain the difference between INNER JOIN and LEFT JOIN and give a situation where each is essential.

2. What is a *window function*? Show how you'd use LAG() to calculate churn.

3. You discover duplicate customer records. Write SQL to surface them and outline fixes.

4. How do you calculate a rolling 7-day average of daily active users in PostgreSQL?

5. WHERE vs HAVING - why can't you always swap them?

6. Contrast subquery and CTE for readability *and* performance; when is one superior?

7. A query that worked yesterday is now slow. Name three diagnostic steps.

8. Define a *surrogate key*. When would a natural key be risky?

9. Given an orders table, return the top 5 products by revenue last month in Ireland.

10. Explain normalisation. Would you always store an analytics mart in 3NF?

11. Pivot a tall table of yearly sales into a wide table (years as columns).

12. UNION vs UNION ALL - demonstrate the performance impact with an example.

13. Describe the ACID properties. Which one is least critical in a read-only BI replica?

14. How do NULLs affect AVG() and how can COALESCE() help?

15. Write SQL for month-over-month growth of active subscriptions.

16. Clustered vs non-clustered indexes - how do they alter the execution plan?

17. What is an execution plan? Walk through reading one in SQL Server.

18. Rank customers by total spend and return the top 10 % for each country.

19. DELETE vs TRUNCATE ; differences in logging and rollback.

20. Explain window frames: ROWS BETWEEN vs RANGE BETWEEN on time-series.

21. Compute the median order value in a DB without a built-in MEDIAN().

22. When is a CROSS JOIN acceptable? Provide a real analytics example.

23. VARCHAR(255) vs TEXT in Postgres - does it matter?

24. Implement Slowly Changing Dimensions (SCD) Type 2 in pure SQL.

25. OLTP vs OLAP schemas - how does the difference shape query strategy?

Disclaimer: This question bank is curated from my personal Data Analyst interview experiences and serves as a guide, not an exhaustive list. Interview content can vary widely.

26. Describe how you would detect and fix deadlocks in SQL Server.

27. Write a query to build a cohort retention table (week 0, week 1, ...).

28. Explain table partitioning vs sharding. How does partition pruning improve speed?

---

2 Excel

1. Compare VLOOKUP, INDEX-MATCH, and XLOOKUP - which is most flexible and why?

2. Steps to build a dynamic KPI dashboard using PivotTables and slicers.

3. Use Power Query to clean and combine multiple CSVs - outline the process.

4. Create a rolling 90-day average with the AVERAGE + OFFSET combo.

5. Detect and remove duplicates - two different methods.

6. Explain absolute vs relative references in a complex formula.

7. Scenario: Finance wants a what-if model for pricing. Build it with Data Tables.

8. Connect Excel to an SQL database; discuss authentication and refresh traps.

9. Use conditional formatting to flag top 5 % outliers dynamically.

10. Why and how would you adopt structured references within Excel Tables?

11. Describe using Power Pivot and the Data Model for multi-table analysis.

12. Write a formula to extract the domain from thousands of email addresses.

13. Auditing: outline a checklist to trace errors in a 20-sheet workbook.

14. Automate a daily Excel-to-CSV export with VBA; high-level steps.

15. Explain the difference between Goal Seek and Solver with a margin-target example.

---

3 Python & Data Wrangling

1. You've got two big CSV files of customers and need to stitch them together in pandas without crashing your computer. How do you do it?

2. About one-tenth of the prices column is empty. What's an easy, sensible way to fill those gaps?

3. Marketing wants a line chart showing conversion rate for the past week ;by lunchtime. Walk me through the quickest path from raw data to final picture.

4. A few monster orders are making average order value look crazy high. How would you spot them and dial them back?

5. Yesterday's ETL script suddenly uses five times more RAM. What's your first detective move?

6. Product asks, "What's our monthly churn?" Sketch the pandas steps you'd take.

7. You need one Jupyter notebook that works for any country we sell in. How would you pass country in as a variable?

Disclaimer: This question bank is curated from my personal Data Analyst interview experiences and serves as a guide, not an exhaustive list. Interview content can vary widely.

8. For forecasting, you need sales from one day ago and seven days ago. Show how you'd add those columns.

9. We're switching from CSV to Parquet on S3. In plain language, why is that faster and cheaper?

10. A nightly job fails but doesn't tell anyone. How would you add logging or alerts so it shouts when it breaks?

11. Everyone has different Python versions installed. How do virtual environments save the day?

12. You're about to run a new data-cleaning script on production. What tests do you run first?

13. You need a quick interactive dashboard for a demo tomorrow. Would you pick Plotly or Streamlit, and why?

14. A long-running SQL query from Python keeps timing out. How could a connection pool help?

15. The KPI dashboard needs an hourly refresh. What would the skeleton of an Airflow DAG look like?

16. Parsing messy user-agent strings is too slow. What simple tricks in pandas or regex would speed it up?

17. Two people editing the same DataFrame keep overwriting each other. Explain deep vs shallow copy in plain words.

18. You need to save a giant machine-learning model (more than 2 GB). Which tool would you use and why?

19. Write a tiny decorator that times a function and prints how long it took.

---

4 Business & Product Cases

1. Deep-dive case: Our monthly active users in Ireland just dropped 12 %. What questions would you ask, what slices of data would you pull, and how would you turn your findings into a one-pager for leadership?

2. An A/B test shows a 1.5 % lift but a p-value of 0.07. Explain, in everyday language, whether we should launch or keep testing.

3. Finance is thinking about adding a middle-tier price plan. What numbers or analyses would you check first?

4. Churn has been stuck at 8 % for months. What early warning signs would you track and what quick wins would you try?

5. Users say the app feels slow. How would you measure the business impact and prioritise fixes?

6. The CEO wants a 12-month sales forecast. Which model would you build and what assumptions would you spell out?

7. Our main delivery partner just asked for higher rates. How would you use data to push back in negotiations?