VIRTAVIS
LEARNING

FREE GUIDE · 2026 EDITION

FOR ASPIRING AND EARLY-CAREER DATA ANALYSTS

# The Data Analyst Interview Playbook

50 questions, model answers and a 14-day plan, from first screen to final round.

- **Every interview stage** — from recruiter screen to final panel
- **50 questions** — with model answers and top-candidate tips
- **Case-study frameworks** — for vague, live business problems
- **SQL, Excel and stats** — checklists for live exercises
- **A 14-day plan** — to prepare without cramming
- **Worksheets** — project story and mock-interview scorecard




WRITTEN BY

**Sophie Lumpa**

Data and Business Intelligence Specialist
Founder, Virtavis Learning


**Check the numbers. Then say what they mean.**

virtavis.com

INSIDE THIS PLAYBOOK

# Contents


**01 How to use this playbook** 3
What hiring teams are really testing

**02 The interview journey** 4
Stages, formats and what to ask the recruiter

**03 What makes a top candidate** 5
Eight habits, and adequate vs top-tier answers

**04 A repeatable answer pattern** 6
Clarify, structure, execute, validate, interpret, recommend

**05 Case studies and live problem-solving** 7
Metric drops, new features, executive summaries and STAR

**06 50 questions and model answers** 8
Grouped by topic so you can go straight to what you need

- Foundations Q1–4 8
- Statistics and experiments Q9–16 9
- SQL Q21–30 12
- Visualisation and BI Q35–37 14
- Motivation and closing Q47–50 17
- Data quality and cleaning Q5–8 9
- Metrics and business cases Q17–20 11
- Excel Q31–34 13
- Behavioural and stakeholders Q38–46 15

**07 Skills demonstrations** 18
SQL, Excel and statistics checklists; portfolio minimum standard

**08 Your 14-day preparation plan** 19
Day by day, plus the night before

**09 Worksheets** 20
Project story worksheet, interview scorecard, questions to ask

**10 Keep going** 22
Courses, live cohorts and free resources


## A NOTE FROM SOPHIE

I wrote this playbook for anyone preparing for a data analyst interview, whether it is your first role, a career change or your next step up. It is not a script to memorise. It shows you what interviewers are listening for and gives you a way to practise it.

The habit that matters most is the one I teach in every Virtavis course: **check the numbers, then say what they mean.** Candidates who do that out loud stand out at every stage.




Sophie Lumpa

Data and Business Intelligence Specialist · Founder, Virtavis Learning

A qualified adult-education teacher with a BSc (Hons) in Business Management and Information Technology. I founded Virtavis to teach the data skills I use with clients.

SECTION 01

# How to use this playbook

Use it as a preparation system rather than a list to memorise. First understand the interview process, then practise technical explanations aloud, then build evidence from one or two projects. Revisit the 50 questions in timed practice.

### A note on the examples

The answers in this playbook are model responses to adapt, not scripts to memorise. Any project results or metrics used in your own interview must be truthful and defensible. If you are new to the field, use coursework, volunteering, personal projects, or simulated business cases honestly. Never claim a percentage improvement or result unless it is true of your own work.

## WHAT HIRING TEAMS ARE REALLY TESTING

### Problem framing

Do you clarify the business decision before jumping into SQL or charts?

### Technical judgement

Can you choose a sensible method, explain trade-offs, and spot edge cases?

### Data trust

Do you validate joins, definitions, missingness, duplicates, and unexpected results?

### Commercial thinking

Can you connect a metric to revenue, cost, retention, risk, or customer experience?

### Communication

Can you explain the insight, uncertainty, and recommended next step without hiding behind jargon?

### Ownership

Can you explain what you personally did, what changed, and what you would improve?

## WHO THIS PLAYBOOK IS FOR

### Career starters

Graduates and school leavers going for a first analyst role.

### Career changers

People moving into data from another job, with skills to translate.

### Working analysts

Analysts preparing for a step up or a move to a new team.



SECTION 02

# The interview journey

Most analyst hiring processes move through some version of these stages. Each one tests something different, so prepare for each on purpose.

## A REALISTIC INTERVIEW JOURNEY

| STAGE | WHAT IT TESTS | PREPARATION MOVE |
|---|---|---|
| 1 **Recruiter / HR screen** | Motivation, communication, role fit, logistics | Prepare a 60–90 second introduction and why this role/company. |
| 2 **Hiring-manager conversation** | Business judgement, priorities, project ownership | Prepare two project stories and questions about success in the role. |
| 3 **SQL / Excel exercise** | Query logic, accuracy, data manipulation | Practise joins, aggregation, nulls, dates, duplicates, lookups and pivots. |
| 4 **Statistics / analytics case** | Hypotheses, metrics, uncertainty, experiment design | Practise metric trees, segmentation, A/B tests and caveats. |
| 5 **Dashboard / take-home** | End-to-end work, clarity, reproducibility | State assumptions; validate outputs; submit concise findings and limitations. |
| 6 **Panel / final round** | Depth, collaboration, influence, consistency | Defend decisions calmly and explain alternatives you considered. |

## INTERVIEW FORMATS: WHAT TO EXPECT

### Structured behavioural interview
Same core competencies for candidates; answer with a specific example and your own contribution.

### Business case
The interviewer may intentionally leave the prompt ambiguous. Clarify the goal, define a metric, segment the problem, test explanations, and recommend an action.

### Portfolio walkthrough
Be ready to explain why the question mattered, where data came from, how you cleaned it, what you found, and what decision it supports.

### Live technical screen
Talk through your thinking. Ask clarifying questions, state assumptions, test edge cases, and narrate trade-offs.

### Take-home assignment
Expect a dataset and a short brief. A clear README, reproducible steps, checks, concise charts, and limitations often matter as much as sophistication.

### Presentation / panel
Lead with the answer, support it with evidence, then handle challenges without becoming defensive.

### Before each interview, ask the recruiter

- Which stages are included, and how long is each one?
- Will the technical task use SQL, Excel, Python, a BI tool, or a mix?
- Is the case live, take-home, or presentation-based?
- Which SQL dialect or tools should I expect?
- May I use documentation, a calculator, or my own laptop?

SECTION 03

# What makes a top candidate

A strong candidate is not simply the person who knows the most functions. They make the work trustworthy, useful, and easy for others to act on.

**1 Starts with the decision**
Asks who will use the analysis, what decision is pending, and how success will be measured.

**2 Defines metrics precisely**
States numerator, denominator, time window, population, exclusions, and whether the metric is a rate, count, or value.

**3 Checks the data before trusting it**
Profiles data, checks grain and keys, reconciles totals, tests joins, and investigates anomalies.

**4 Explains trade-offs**
Can justify median vs mean, LEFT JOIN vs INNER JOIN, a chart choice, or an imputation decision.

**5 Separates evidence from inference**
Says what the data demonstrates, what it suggests, and what would need a controlled test or more data.

**6 Communicates the 'so what'**
Summarises finding › business impact › recommendation › risk/next check.

**7 Owns their work**
Clearly distinguishes personal contribution from team contribution and admits mistakes constructively.

**8 Shows learning agility**
Can describe how they learned a tool, handled feedback, and improved a process.

## ADEQUATE ANSWER VS TOP-TIER ANSWER

| AREA | ADEQUATE ANSWER | TOP-TIER ANSWER |
|---|---|---|
| **SQL** | Uses a join and GROUP BY to calculate revenue. | Confirms table grain and join keys, protects against duplicate inflation, handles nulls/refunds, and validates totals. |
| **Business case** | Lists possible causes for a metric drop. | Checks instrumentation and data freshness first, decomposes the metric, segments impact, ranks hypotheses, and proposes tests. |
| **Statistics** | Says $p < 0.05$ means significant. | Explains effect size, uncertainty, assumptions, multiple testing, practical significance, and the decision threshold. |
| **Behavioural** | Says "we analysed data and improved performance." | Gives context, personal actions, checks, measurable result where known, limitations, and learning. |
| **Dashboard** | Presents several attractive charts. | Organises around a user decision, defines metrics, highlights exceptions, supports drill-down, and avoids misleading encodings. |

SECTION 04

# A repeatable answer pattern

Whatever the question, the same six moves keep your answer clear and show your judgement.

**1**

### CLARIFY

Restate the question and ask about the decision, audience, timeframe, and constraints.

**2**

### STRUCTURE

Break the problem into a few logical steps before calculating.

**3**

### EXECUTE

Explain the method and why it fits; write readable code or show clear calculations.

**4**

### VALIDATE

Check grain, missing values, duplicates, edge cases, totals, and assumptions.

**5**

### INTERPRET

State the finding in plain language and explain the business implication.

**6**

### RECOMMEND

Suggest an action, follow-up test, or monitoring metric.

## A PRACTICAL WAY TO DEMONSTRATE SKILLS

| | |
|---|---|
| **SQL** | Practise SELECT, WHERE, CASE, joins, GROUP BY/HAVING, CTEs, window functions, date logic and NULL handling. |
| **Excel** | Show XLOOKUP or INDEX/MATCH, SUMIFS/COUNTIFS, IF, PivotTables, date cleanup, conditional formatting and data validation. |
| **Python** | Demonstrate pandas filtering, merge, groupby, missing-value handling, date conversion, basic plots and reproducible notebooks. |
| **BI** | Explain the audience, KPI definitions, layout hierarchy, filters, drill-downs, refresh logic and how users should act on the dashboard. |
| **Statistics** | Explain distributions, sampling, confidence intervals, hypothesis tests, correlation vs causation and A/B test design. |



SECTION 05

# Case studies and live problem-solving

When the prompt is vague, do not rush to a chart or query. Make the problem tractable.

## FRAMEWORK: A KEY METRIC FALLS 15% OVERNIGHT

1. **Verify the signal.** Check refresh status, tracking changes, data latency, schema changes, and whether the definition changed. Compare against source totals.
2. **Decompose the metric.** For example, revenue = orders × average order value; conversion rate = conversions ÷ eligible sessions. Confirm numerator and denominator.
3. **Segment systematically.** Break down by platform, geography, channel, customer cohort, product, and funnel step. Look for where the decline is concentrated.
4. **Rank explanations.** Distinguish instrumentation, operational, customer-behaviour, and external causes. Use evidence to prioritise rather than guessing.
5. **Recommend and monitor.** Explain what to investigate or roll back, who needs to act, what data would confirm the cause, and how recovery will be tracked.

## FRAMEWORK: EVALUATE A NEW FEATURE

Define the target user and intended behaviour; select a primary metric and guardrails; verify exposure and event logging; choose a suitable experiment or comparison; estimate effect and uncertainty; inspect segments without cherry-picking; recommend launch, iterate, or stop. Guardrails may include latency, complaints, refunds, or retention depending on the feature.

## FRAMEWORK: PRESENT A FINDING TO EXECUTIVES

**Answer first** · **Evidence** · **Business impact** · **Recommendation** · **Risks / next check**

Keep technical detail available for questions, not as the opening.

## STAR FOR BEHAVIOURAL ANSWERS

| S | T | A | R |
|---|---|---|---|
| **Situation** | **Task** | **Action** | **Result** |
| Give enough context to understand the problem. | State your responsibility and what success meant. | Spend most of the answer here: what you did, why, what you checked, and how you worked with others. | Share the outcome, evidence, learning, and what happened next. |

Aim for about 60–120 seconds for a standard behavioural answer. Do not invent impact metrics. If the result was qualitative, say what changed and how you know. If the project was academic or personal, label it honestly.

### SAMPLE STAR ANSWER · DATA QUALITY CHALLENGE

**Situation:** In a portfolio project, regional sales files used inconsistent date formats and category labels.

**Task:** I needed to create a comparable monthly sales view.

**Action:** I profiled each file, standardised dates and categories using a documented mapping, checked duplicate transaction IDs, and reconciled totals before and after cleaning. I kept an exceptions log rather than silently deleting ambiguous records.

**Result:** The monthly comparison became reproducible, and the exceptions were visible for review. I learned to confirm definitions early and preserve an audit trail.

*This example is intentionally framed as a portfolio project. Replace it with your own truthful details.*

SECTION 06

# 50 questions and model answers

These are concise answer frameworks. For technical questions, practise explaining your reasoning and edge cases. For experience questions, replace the illustrative wording with your own real evidence.

## FOUNDATIONS

What the job is and how you approach it.

Q1–4

### 01 What does a data analyst do?

**MODEL ANSWER**

A data analyst turns data into evidence that supports a decision. The work typically includes clarifying the question, collecting and preparing data, analysing patterns, validating results, communicating findings, and helping stakeholders decide what to do.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Connect the analysis to a real decision and explain how you would measure whether the recommendation worked.

### 02 Walk me through the data analysis process.

**MODEL ANSWER**

Define the business problem, identify the required data, assess quality, clean and transform it, explore the data, apply appropriate analysis, validate findings, communicate recommendations, and monitor outcomes.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Emphasise iteration: a surprising result may send you back to the data definition or collection process.

### 03 What is the difference between data analysis and data analytics?

**MODEL ANSWER**

Data analysis often refers to examining a dataset to answer a question; data analytics can refer more broadly to the methods and workflow used to turn data into decisions, including descriptive, diagnostic, predictive, and prescriptive work. Usage varies by organisation.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Acknowledge that organisations may use the terms interchangeably and focus on the work required.

### 04 What makes a good data analyst?

**MODEL ANSWER**

A good analyst combines technical competence, statistical reasoning, curiosity, attention to detail, business understanding, and clear communication. They question assumptions, validate outputs, and explain what the evidence can and cannot support.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Give a brief example showing these skills together, rather than listing traits only.

## DATA QUALITY AND CLEANING

Can the numbers be trusted?

Q5–8

### 05 How do you ensure data accuracy?

**MODEL ANSWER**

I check schema and types, missingness, duplicates, key uniqueness, valid ranges, and consistency between related fields. I reconcile aggregates to trusted source totals, investigate anomalies, document transformations, and ask a data owner when business meaning is unclear.

**LIFT IT TO TOP-CANDIDATE LEVEL** Describe specific checks and what you would do when a check fails.

### 06 What is data cleaning, and why does it matter?

**MODEL ANSWER**

Data cleaning identifies and addresses errors such as inconsistent formats, duplicate records, invalid values, and missing fields. Poor quality can bias analysis or distort business decisions, so cleaning should be documented and appropriate to the question.

**LIFT IT TO TOP-CANDIDATE LEVEL** Do not automatically delete outliers or impute values; explain the reason and impact of each treatment.

### 07 How do you handle missing data?

**MODEL ANSWER**

First quantify how much is missing and inspect whether missingness is concentrated in particular groups or time periods. Depending on why it is missing and the analytical goal, I may leave it explicit, exclude affected records, impute values, or obtain better source data. I document assumptions and test sensitivity.

**LIFT IT TO TOP-CANDIDATE LEVEL** Explain why a simple mean fill can distort variance or relationships.

### 08 How do you identify outliers?

**MODEL ANSWER**

I combine visual checks such as boxplots and scatterplots with statistical rules such as the IQR method or z-scores, then use domain context to investigate. An outlier may be an error, a rare valid event, or an important business signal.

**LIFT IT TO TOP-CANDIDATE LEVEL** State that detection is not the same as justification for removal.

## STATISTICS AND EXPERIMENTS

Reasoning with uncertainty.

Q9–16

### 09 Mean, median, or mode: which would you use?

**MODEL ANSWER**

Mean uses every numeric value but is sensitive to extremes; median is robust for skewed distributions; mode identifies the most common value and can be useful for categorical data. I choose based on the distribution and the question.

**LIFT IT TO TOP-CANDIDATE LEVEL** Use a concrete example, such as median income when a few very high values skew the mean.

## 10 What is standard deviation?

MODEL ANSWER

Standard deviation describes how spread out numeric observations are around the mean, in the same units as the data. A larger value indicates more variability, though interpretation depends on the distribution and context.

> LIFT IT TO TOP-CANDIDATE LEVEL Distinguish spread from uncertainty about an estimated mean.

## 11 Correlation vs causation?

MODEL ANSWER

Correlation describes association between variables; causation means a change in one variable produces a change in another. Confounding, reverse causality, selection effects, or coincidence can explain an association, so causal claims need stronger design or assumptions.

> LIFT IT TO TOP-CANDIDATE LEVEL Name a plausible confounder and how you might investigate it.

## 12 What is a p-value?

MODEL ANSWER

A p-value is the probability, assuming the null hypothesis and test assumptions are true, of observing results at least as extreme as those obtained. It is not the probability that the null is true and does not measure business importance.

> LIFT IT TO TOP-CANDIDATE LEVEL Pair it with effect size, confidence interval, study design, and practical significance.

## 13 What is a confidence interval?

MODEL ANSWER

A confidence interval gives a range of values produced by a method that, over repeated comparable samples, would contain the true parameter at the stated rate. It communicates uncertainty around an estimate.

> LIFT IT TO TOP-CANDIDATE LEVEL Explain both the point estimate and interval width; do not interpret it as a probability about a fixed parameter after observing the data.

## 14 What are Type I and Type II errors?

MODEL ANSWER

Type I error is a false positive: rejecting a true null hypothesis. Type II error is a false negative: failing to reject a false null. The business cost of each should inform test design and decision thresholds.

> LIFT IT TO TOP-CANDIDATE LEVEL Give a relevant example, such as launching a feature that does not help versus missing a genuinely helpful feature.

## 15 How would you design an A/B test?

MODEL ANSWER

Define a hypothesis, primary metric, guardrail metrics, eligible population, randomisation unit, sample size and duration. Check instrumentation and allocation, avoid peeking-driven decisions, analyse effect size and uncertainty, and assess harms or segment effects carefully.

> LIFT IT TO TOP-CANDIDATE LEVEL Mention interference, novelty effects, sample-ratio mismatch, and multiple comparisons where relevant.

### 16 How do you distinguish descriptive, diagnostic, predictive, and prescriptive analytics?

**MODEL ANSWER**

Descriptive summarises what happened; diagnostic investigates why; predictive estimates what may happen; prescriptive evaluates actions that may improve outcomes. The methods and confidence needed differ by question.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Explain that predictive accuracy alone does not prove an intervention will work.

## METRICS AND BUSINESS CASES

Connecting numbers to decisions.

Q17–20

### 17 What is a KPI? How would you choose one?

**MODEL ANSWER**

A KPI is a measure tied to an important objective. I start from the business goal, define the target behaviour or outcome, specify the formula and population, then add supporting and guardrail metrics so improvements do not hide trade-offs.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Avoid choosing a metric only because it is easy to measure; define how it could be gamed.

### 18 Which KPIs would you track for an e-commerce site?

**MODEL ANSWER**

Depending on the goal, I might track conversion rate, average order value, revenue per visitor, cart abandonment, customer acquisition cost, repeat purchase, customer lifetime value, refunds, and return on ad spend. I would define the time window and segment by channel and customer type.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Explain the trade-offs: discounting can lift conversion while reducing margin.

### 19 How would you analyse customer churn?

**MODEL ANSWER**

Agree on the churn definition and observation window, build a reliable customer-level dataset, examine churn by cohort and segment, compare usage and service signals, and distinguish correlation from causes. Recommend targeted retention tests and monitor both churn and guardrails.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Define the denominator and ensure customers have had enough time to churn before comparing cohorts.

### 20 A key metric fell 15% overnight. What do you do?

**MODEL ANSWER**

First verify freshness, instrumentation, definitions, and source totals. Then decompose the metric into its components, segment by platform, channel, region, cohort, and funnel step, rank hypotheses, and recommend the fastest reliable checks before proposing a fix.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Do not start by assuming customer behaviour changed; data pipeline issues can create false alarms.

## SQL

The most tested technical skill.

Q21–30

### 21 What is SQL and why do analysts use it?

MODEL ANSWER

SQL is a language for querying and manipulating relational data. Analysts use it to filter, join, aggregate, transform, and summarise data close to where it is stored.

LIFT IT TO TOP-CANDIDATE LEVEL Mention readable queries, validation, and awareness of table grain.

### 22 What is the difference between WHERE and HAVING?

MODEL ANSWER

WHERE filters individual rows before grouping; HAVING filters groups after aggregation. For example, WHERE can select completed orders before aggregation, while HAVING can retain customers whose total revenue exceeds a threshold.

LIFT IT TO TOP-CANDIDATE LEVEL Explain the order of operations rather than merely reciting definitions.

### 23 Explain INNER JOIN vs LEFT JOIN.

MODEL ANSWER

INNER JOIN keeps rows with matching keys in both tables. LEFT JOIN keeps every row from the left table and adds matching right-side values, leaving NULLs when no match exists.

LIFT IT TO TOP-CANDIDATE LEVEL Check key uniqueness and row counts: one-to-many joins can multiply rows and inflate sums.

### 24 UNION vs UNION ALL?

MODEL ANSWER

UNION stacks compatible result sets and removes duplicate rows; UNION ALL stacks them while preserving duplicates and is generally less work. Both require corresponding columns to have compatible types and positions.

LIFT IT TO TOP-CANDIDATE LEVEL Choose based on whether duplicates are meaningful, not simply because one sounds safer.

### 25 How would you find duplicate customer emails?

MODEL ANSWER

Group by a normalised email value and filter groups with COUNT(*) greater than one. Before merging or deleting records, confirm the business definition of duplicate and inspect source IDs and timestamps.

LIFT IT TO TOP-CANDIDATE LEVEL Show awareness that case, whitespace, shared family emails, and nulls affect interpretation.

### 26 What is a CTE?

MODEL ANSWER

A common table expression, introduced with WITH, names a query result for use within a larger statement. It can make multi-step logic easier to read, test, and maintain.

LIFT IT TO TOP-CANDIDATE LEVEL Break a complex problem into named steps and verify each intermediate result.

### 27 What are window functions?

MODEL ANSWER

Window functions calculate values across related rows without collapsing each group to one row. Examples include ROW_NUMBER, RANK, running totals, and LAG for comparing with a previous period.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Explain the PARTITION BY and ORDER BY logic and handle ties deliberately.

### 28 How do you calculate month-over-month growth?

MODEL ANSWER

Aggregate the metric by month, ensure the series includes months with no activity if appropriate, use LAG to retrieve the prior month, then calculate (current – prior) / prior. Define how to handle a zero or missing prior value.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Avoid silently treating a zero baseline as ordinary percentage growth; report it separately.

### 29 What is the difference between a primary key and a foreign key?

MODEL ANSWER

A primary key uniquely identifies a row in a table. A foreign key references a key in another table to represent a relationship and, when enforced, help maintain referential integrity.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Explain how key design affects join correctness and duplicate inflation.

### 30 How do you use CASE in SQL?

MODEL ANSWER

CASE applies conditional logic and returns a value for the first matching WHEN branch, with ELSE as a fallback. It is useful for categorisation, conditional aggregation, and business rules.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Order overlapping conditions carefully and decide how NULLs should be classified.

PRACTISE THIS

**SQL for Data Analytics** builds these skills step by step, with challenges and model answers in every section.

Start the course ›

## EXCEL

Still in almost every analyst job ad.

Q31–34

### 31 What is a PivotTable?

MODEL ANSWER

A PivotTable quickly summarises data by dimensions such as region or month and measures such as revenue or order count. It supports slicing and comparing groups without building every formula manually.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Check source range, data types, duplicate grain, and whether the aggregation should be sum, count, or distinct count.

## 32 VLOOKUP vs INDEX/MATCH or XLOOKUP?

**MODEL ANSWER**

VLOOKUP searches the first column of a table and returns a value from a specified column to its right. INDEX/MATCH and XLOOKUP are more flexible for lookup direction and exact-match patterns; availability depends on the spreadsheet version.

> **LIFT IT TO TOP-CANDIDATE LEVEL** State whether you use exact matching and how you handle missing lookup keys.

## 33 How would you clean dates stored in different formats?

**MODEL ANSWER**

Identify the source locale and intended format, parse values explicitly, convert them to a consistent date type, and flag invalid or ambiguous records. Validate ranges and compare counts before and after conversion.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Do not assume 03/04 means the same date in every locale.

## 34 What is data validation in Excel?

**MODEL ANSWER**

Data validation restricts or checks entries using rules such as allowed lists, numeric ranges, or date limits. It can reduce input errors, though it does not replace quality checks on imported data.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Describe how you would handle existing invalid values and maintain a clear exception list.

**PRACTISE THIS**

**Excel for Data Analytics** practises lookups, PivotTables and cleaning on one realistic business dataset.

Start the course ›

# VISUALISATION AND BI

Making the answer easy to see.

Q35–37

## 35 How do you choose the right chart?

**MODEL ANSWER**

Use a line chart for trends over time, a bar chart for category comparisons, a scatter plot for relationships, and a histogram for distributions. Keep labels clear and avoid unnecessary decoration or misleading scales.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Start with the question and audience; check whether the visual makes the key comparison easy.

## 36 What makes a useful dashboard?

**MODEL ANSWER**

A useful dashboard is built around a defined audience and decision. It has clear KPI definitions, a sensible visual hierarchy, appropriate filters, reliable refreshes, accessible design, and enough context to interpret changes.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Describe a specific action the dashboard enables and how usage or value would be evaluated.

### 37 Power BI vs Tableau?

MODEL ANSWER

Both are business intelligence tools for modelling, exploring, and communicating data. The choice depends on the existing ecosystem, licensing, data connections, governance, user skills, and deployment needs, not a universal winner.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Discuss the task you can demonstrate rather than claiming expertise in tools you have not used.

PRACTISE THIS

**Power BI for Data Analytics** takes you from raw data to a finished dashboard, with challenges and a final project.

Start the course ›

## BEHAVIOURAL AND STAKEHOLDERS

How you work with people and pressure.

Q38–46

### 38 How would you explain a technical result to a non-technical stakeholder?

MODEL ANSWER

Lead with the business question and main finding, use a simple visual or analogy, explain the practical implication and uncertainty, then invite questions. Keep implementation details available without making them the headline.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Check understanding and tailor the recommendation to what the stakeholder can control.

### 39 Tell me about a challenging project.

MODEL ANSWER

Use STAR: describe the context and goal, clarify your own responsibility, explain the steps you took and why, then share the outcome and learning. Use a project you can discuss in detail, including limitations.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Prepare follow-ups on alternatives, validation, your personal contribution, and what you would do differently.

### 40 Tell me about a time your analysis was wrong.

MODEL ANSWER

I would explain how the issue surfaced, take responsibility, assess who or what was affected, correct the analysis, communicate the change, and add a check to reduce recurrence. The focus is on learning and restoring trust.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Choose a genuine example and show accountability rather than blaming a colleague or the data.

### 41 How do you handle a stakeholder who disagrees with your findings?

MODEL ANSWER

I would listen to the concern, clarify which assumption or conclusion is disputed, revisit definitions and calculations, and share the evidence and limitations. If disagreement remains, I would propose a targeted check or test.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Stay curious and evidence-led; do not treat a challenge as a personal contest.

## 42 How do you prioritise competing requests?

**MODEL ANSWER**

I compare business impact, urgency, risk, effort, dependencies, and who is blocked. I confirm priorities with stakeholders, agree on scope and timelines, and communicate what will be delayed if a new request takes precedence.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Make the trade-off explicit and avoid promising every request can be done immediately.

## 43 How do you work with incomplete data and a tight deadline?

**MODEL ANSWER**

I identify which missing fields affect the decision, quantify the gap, seek a source fix or cross-check, and choose a transparent treatment. I share a provisional answer only with its assumptions and limitations clearly stated.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Separate what can be answered now from what needs further evidence.

## 44 How do you learn a new analytical tool?

**MODEL ANSWER**

I start with the task the tool should solve, learn core concepts from documentation or a structured tutorial, reproduce a small example, apply it to a real dataset, and document what I learned. I validate results against a known baseline.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Give a concrete example of how you moved from tutorial to a working deliverable.

## 45 How do you stay organised and reproducible?

**MODEL ANSWER**

I use clear file and query naming, version control where available, documented assumptions, repeatable transformations, data-quality checks, and a short readme or handover. I keep raw data separate from transformed outputs when practical.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Explain how another analyst could reproduce your result without relying on your memory.

## 46 What would you do if the data contradicted a senior stakeholder's expectation?

**MODEL ANSWER**

I would verify the data and methodology, check whether definitions or context explain the difference, then present the evidence respectfully with uncertainty and alternative explanations. I would work with the stakeholder on the next check or decision.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Show independence without being combative; seniority does not replace evidence.

## MOTIVATION AND CLOSING

Why you, why here, and the final questions.

Q47–50

### 47 Why should we hire you?

**MODEL ANSWER**

I bring a structured approach to questions, careful data handling, willingness to explain my reasoning, and focus on turning findings into useful decisions. I can demonstrate those strengths through a project and I am clear about the areas I am still developing.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Make the answer specific to the role's stated needs and support each claim with evidence.

### 48 Why do you want this data analyst role?

**MODEL ANSWER**

Connect the organisation's product, customers, or mission to the kinds of questions the role solves. Then link two or three relevant skills or projects to the job description and explain what you hope to contribute and learn.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Avoid generic praise; mention a real business problem or data challenge relevant to the employer.

### 49 What questions would you ask the interviewer?

**MODEL ANSWER**

I would ask how the team defines success in the first six months, which decisions analysts influence, how data quality and metric definitions are governed, how work is prioritised, and what distinguishes excellent performance on this team.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Ask questions that reveal the team's real operating environment, not only perks.

### 50 What is your greatest weakness?

**MODEL ANSWER**

Choose a genuine, manageable development area that is relevant but not disqualifying. Explain the steps you are taking to improve and give evidence of progress. For example: "Earlier in my projects, I sometimes spent too long polishing charts before confirming which decision the audience needed to make. I now agree the key question and success criteria first, share an early draft, and use feedback to focus the analysis. That has helped me prioritise clarity and usefulness over decoration." Adapt this to a real weakness and a real improvement you can discuss.

> **LIFT IT TO TOP-CANDIDATE LEVEL** Be specific, show self-awareness without rehearsed clichés, and explain a repeatable behaviour you have changed. Do not disguise a strength as a weakness or claim a progress story that is not true.

SECTION 07

# Skills demonstrations

In a live exercise, correctness matters, but so does the path you take. Use a consistent checklist.

## SQL EXERCISE CHECKLIST

- Restate the output: one row per what — customer, order, month, or region?
- Filter at the right stage; use WHERE for rows and HAVING for groups.
- Handle NULLs, duplicate keys, ties, time zones, and incomplete periods intentionally.
- Confirm the grain and keys of each table before joining.
- Use explicit column names and readable aliases; avoid SELECT * in a final analytical query.
- Sanity-check row counts and totals against a simpler query or source report.

### EXAMPLE SQL · COMPLETED REVENUE BY CUSTOMER, INCLUDING CUSTOMERS WITH NO COMPLETED ORDERS

```sql
SELECT
    c.customer_id,
    COALESCE(SUM(o.order_total), 0) AS completed_revenue
FROM customers AS c
LEFT JOIN orders AS o
    ON  o.customer_id = c.customer_id
    AND o.status = 'completed'
GROUP BY c.customer_id
ORDER BY completed_revenue DESC;
```

**Why this is a strong pattern:** The LEFT JOIN preserves customers without completed orders; placing the order-status condition in the JOIN avoids filtering those customers out; COALESCE makes their total zero. In a real schema, confirm whether order_total already accounts for refunds, discounts, and currency.

## EXCEL EXERCISE CHECKLIST

- Inspect the headers, types, blanks, duplicates, and date formats before building a summary.
- Use a PivotTable for quick aggregation and SUMIFS/COUNTIFS for criteria-based checks.
- Use exact-match lookups unless approximate matching is explicitly required.
- Validate formulas on a few manually checked rows and reconcile totals.
- Label units and time periods; make assumptions visible rather than burying them in formulas.

## STATISTICS CHECKLIST

- Clarify the hypothesis, population, outcome, and unit of randomisation.
- Choose a primary metric and guardrails before looking at results.
- Explain effect size and uncertainty, not only a p-value.
- Check data quality, sample balance, exposure, test duration, and multiple comparisons.
- Make the decision rule explicit and explain what evidence would change your recommendation.

## PORTFOLIO PROJECT: MINIMUM STANDARD

- A clear question and why it matters.
- Reproducible cleaning and transformation steps.
- A small set of charts chosen to answer specific questions.
- A recommendation, limitations, and sensible next steps.
- A source note and a description of the dataset and its limitations.
- At least three quality checks, such as duplicate checks, missingness, and total reconciliation.
- Three to five findings, each with evidence and a practical implication.
- A README that lets someone reproduce the work.

FREE AT VIRTAVIS.COM

Get the **portfolio kit**, the **practice dataset** and the **SQL, Excel and Python cheat sheets** to build a project that meets this standard.

Get the free cheat sheet ›

SECTION 08

# Your 14-day preparation plan

Two weeks, one focus a day. Pick the tools named in the job ad and spend the technical days on those.

| KNOW THE ROLE | |
|---|---|
| **Day 1** | Read the job description; list the required tools, domain, and competencies. |
| **Day 2** | Write a 90-second introduction and a role-specific "why this job?" answer. |

| TECHNICAL PRACTICE | |
|---|---|
| **Day 3** | Practise SQL filtering, aggregation, CASE, NULLs, and joins. |
| **Day 4** | Practise CTEs, window functions, dates, duplicates, and query validation. |
| **Day 5** | Practise Excel formulas, lookups, PivotTables, and cleaning inconsistent data. |
| **Day 6** | Review descriptive statistics, sampling, confidence intervals, and hypothesis tests. |
| **Day 7** | Design an A/B test and practise a metric-drop investigation aloud. |

| EVIDENCE AND STORIES | |
|---|---|
| **Day 8** | Build or polish a portfolio project with checks and a clear business question. |
| **Day 9** | Prepare two STAR stories: quality/problem-solving and stakeholder communication. |
| **Day 10** | Prepare two more STAR stories: prioritisation/teamwork and failure/learning. |
| **Day 11** | Create a 3–5 minute project walkthrough with one recommendation. |

| REHEARSE | |
|---|---|
| **Day 12** | Complete a timed technical practice and explain your reasoning as you work. |
| **Day 13** | Run a mock interview; record yourself and identify filler, gaps, and unclear claims. |
| **Day 14** | Review company context, prepare interviewer questions, and rest well. |

### The night before

- Review the job description and your own CV; be able to discuss every claim.
- Check the interview link, time zone, location, tools, and any permitted materials.
- Prepare a blank page for assumptions, metrics, and structure.
- Prepare three thoughtful questions for the team.
- Do not memorise every answer. Practise clear reasoning and sleep.

SECTION 09

# Worksheets

Print these pages or copy them into your notes. Fill one worksheet per project you plan to talk about.

## YOUR PROJECT STORY WORKSHEET

| | |
|---|---|
| **Business question / decision** | |
| **Data sources, grain, and limitations** | |
| **My responsibility** | |
| **Methods and why I chose them** | |
| **Quality checks and assumptions** | |
| **Finding and supporting evidence** | |
| **Recommendation / result / learning** | |

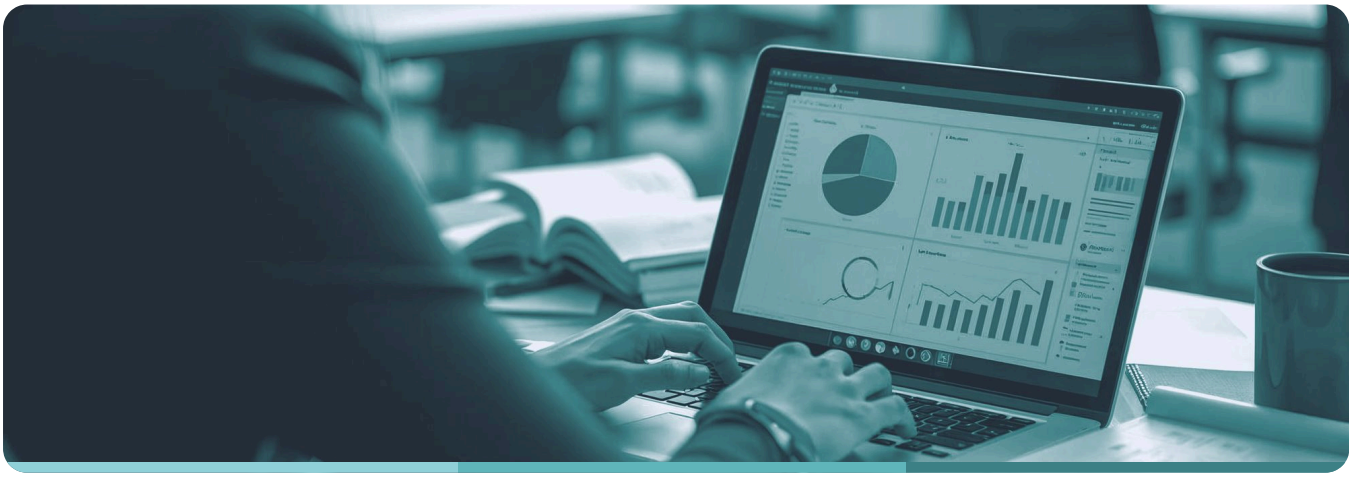
## FINAL INTERVIEW SCORECARD

After a mock interview, score each area from 1 (needs work) to 5 (ready). Write one specific improvement for the lowest-scoring area.

| SKILL | 1 | 2 | 3 | 4 | 5 | EVIDENCE / NEXT IMPROVEMENT |
|---|---|---|---|---|---|---|
| Clarifies the business question | ☐ | ☐ | ☐ | ☐ | ☐ | |
| SQL / spreadsheet accuracy | ☐ | ☐ | ☐ | ☐ | ☐ | |
| Statistical reasoning | ☐ | ☐ | ☐ | ☐ | ☐ | |
| Data quality and validation | ☐ | ☐ | ☐ | ☐ | ☐ | |
| Business interpretation | ☐ | ☐ | ☐ | ☐ | ☐ | |
| Communication and storytelling | ☐ | ☐ | ☐ | ☐ | ☐ | |
| STAR examples and ownership | ☐ | ☐ | ☐ | ☐ | ☐ | |
| Handles challenge and uncertainty | ☐ | ☐ | ☐ | ☐ | ☐ | |

## QUESTIONS TO ASK THE HIRING TEAM

1. What decisions would you expect this role to influence in the first six months?
2. How are key metrics defined and governed across teams?
3. What are the biggest data-quality or reporting challenges today?
4. How does the team balance quick-turn requests with deeper analysis?
5. What does an excellent analysis deliver here beyond a dashboard or report?
6. How are findings adopted and their impact measured?



SECTION 10 · KEEP GOING

# Turn preparation into proof.

The best interview answer is not the longest or most technical. It is the answer that makes the problem clear, demonstrates sound judgement, earns trust through validation, and helps the business decide what to do next. The fastest way to give that answer is to have done the work.

## BUILD THE SKILLS BEHIND THE ANSWERS

Every Virtavis course is built around real business problems, with videos, a handbook, section quizzes, challenges with model answers and a final project. Find them on Udemy, single or as a bundle.

**SQL for Data Analytics**
Questions 21–30 and the live SQL exercise.

**Excel for Data Analytics**
Questions 31–34, lookups, PivotTables and cleaning tasks.

**Power BI for Data Analytics**
Questions 35–37, dashboard take-homes and walkthroughs.

**Python for Data Analytics**
pandas tasks and reproducible notebooks.

**Data Analytics capstone projects**
End-to-end projects you can talk through with confidence.

Start the course ›



LEARN LIVE WITH SOPHIE

### Small-group cohorts

Weekly live sessions, feedback on your work, and a community of people working towards the same goal. Ideal if you want structure and accountability while you prepare.

Register interest › virtavis.com/cohorts

FREE AT VIRTAVIS.COM

- Practice dataset
- Portfolio kits
- SQL, Excel, DAX and Python cheat sheets
- KPI cheat sheet

Get the free cheat sheet › virtavis.com/free

VIRTAVIS LEARNING

**Check the numbers. Then say what they mean.**

virtavis.com

© 2026 Sophie Lumpa, Virtavis Learning. Free to share with credit. Icons: Lucide (ISC licence). Photos: Unsplash (Apex Virtual Education, Christina @ wocintechchat.com, Vitaly Gariev).