

# Data Analyst Interview Preparation: Q&A Format

To make this easier to follow, I've restructured the entire guide into a clear, straightforward question-and-answer format. Each section starts with a brief intro, followed by individual Q&A pairs. Questions are numbered for quick reference, and answers include theory, practical examples, and mock datasets where relevant. This covers beginner-to-intermediate levels, with a focus on common interview topics. Practice explaining your answers verbally for interviews.

## SQL for Data Analytics

SQL questions often test querying skills for insights. Here are 35 common ones, from basics to case studies, with queries and explanations.

### Q1: What is SQL?

**A:** SQL stands for Structured Query Language, used for managing relational databases. It allows querying, inserting, updating, and deleting data. For analysts, it's key for pulling data for reports.

### Q2: What are the basic elements of an SQL query?

**A:** A typical query includes SELECT (columns to retrieve), FROM (table source), WHERE (row filters), GROUP BY (for aggregations), HAVING (group filters), and ORDER BY (sorting). Example: `SELECT name FROM employees ORDER BY salary DESC;` retrieves names sorted by salary.

### Q3: How do you select all data from a table?

**A:** Use `SELECT * FROM employees;`. This fetches everything, but avoid it in large datasets for performance reasons.

### Q4: How do you select specific columns?

**A:** `SELECT name, salary FROM employees;`. You can alias columns like `SELECT name AS employee_name;`.

### Q5: What is the syntax for the WHERE clause?

**A:** It filters rows before grouping. Example: `SELECT * FROM employees WHERE salary > 50000;`. Supports operators like `=, >, and LIKE` for patterns (e.g., `name LIKE 'A%'` ).

### Q6: How do you handle multiple conditions in WHERE?

**A:** Use AND/OR: `SELECT * FROM employees WHERE dept_id = 10 AND salary > 50000;` .  
OR allows alternatives.

**Q7: How does LIKE work for text filtering?**

**A:** Matches patterns: `SELECT * FROM employees WHERE name LIKE 'A%';` (names starting with A). In mock data (Alice, Bob), it returns Alice.

**Q8: How do you UPDATE rows?**

**A:** `UPDATE employees SET salary = 60000 WHERE id = 1;` . Always include WHERE to prevent updating all rows.

**Q9: How do you DELETE rows?**

**A:** `DELETE FROM employees WHERE dept_id = 30;` . Targets inactive departments, for example.

**Q10: How does ORDER BY work for sorting?**

**A:** `SELECT * FROM employees ORDER BY salary DESC;` . DESC for descending; ASC (default) for ascending.

**Q11: How do you combine tables using Joins?**

**A:** INNER JOIN matches both tables: `SELECT e.name, d.name FROM employees e INNER JOIN departments d ON e.dept_id = d.id;` . Using mock data: Returns Alice-HR, Carol-HR, Bob-Sales.

**Q12: What is a LEFT JOIN?**

**A:** Includes all from left table plus matches from right: `SELECT * FROM departments d LEFT JOIN employees e ON d.id = e.dept_id;` . Useful for finding unmatched records.

**Q13: How do you use aggregates like SUM and COUNT?**

**A:** `SELECT dept_id, SUM(salary) AS total FROM employees GROUP BY dept_id;` . Mock result: HR total 105000, Sales 60000.

**Q14: What is the difference between HAVING and WHERE?**

**A:** WHERE filters individual rows pre-grouping; HAVING filters aggregated results post-grouping. Example: `SELECT dept_id, COUNT(*) FROM employees GROUP BY dept_id HAVING COUNT(*) > 1;` .

**Q15: What are subqueries?**

**A:** Nested queries: `SELECT name FROM employees WHERE salary > (SELECT AVG(salary) FROM employees);`. Finds employees above average salary.

**Q16: What is a self-join?**

**A:** Joins a table to itself for hierarchies: `SELECT e1.name AS manager, e2.name AS employee FROM employees e1 JOIN employees e2 ON e1.id = e2.manager_id;`.

**Q17: What is the ROW\_NUMBER() window function?**

**A:** Assigns row numbers: `SELECT name, salary, ROW_NUMBER() OVER (ORDER BY salary DESC) AS rank FROM employees;`. Mock: Bob rank 1, Carol 2, Alice 3.

**Q18: What is the difference between RANK() and DENSE\_RANK()?**

**A:** `RANK()` skips ties (e.g., two 1st become next 3rd); `DENSE_RANK()` doesn't skip. For ties at 60000, `RANK` next=3, `DENSE` next=2.

**Q19: How do LAG and LEAD work?**

**A:** Compare to previous/next row: `SELECT salary, LAG(salary) OVER (ORDER BY id) AS prev_salary FROM employees;`.

**Q20: What is NTILE for bucketing?**

**A:** Divides into groups: `SELECT name, NTILE(2) OVER (ORDER BY salary) AS bucket FROM employees;`. Splits into high/low salary buckets.

**Q21: Case study: Find the largest salary per department.**

**A:** `SELECT d.name, MAX(e.salary) FROM departments d JOIN employees e ON d.id = e.dept_id GROUP BY d.name;`. Mock: HR 55000, Sales 60000.

**Q22: How to find the 2nd highest salary in a department?**

**A:** Use window: `WITH RankedSalaries AS (SELECT salary, DENSE_RANK() OVER (ORDER BY salary DESC) AS rnk FROM employees WHERE dept_name = 'engineering') SELECT salary FROM RankedSalaries WHERE rnk = 2;`. Handles ties.

**Q23: Case study: Find neighborhoods with zero users.**

**A:** Anti-join: `SELECT n.neighborhood_id FROM neighborhoods n LEFT JOIN users u ON n.neighborhood_id = u.neighborhood_id WHERE u.user_id IS NULL;`.

**Q24: How to select a random manufacturer?**

**A:** `SELECT manufacturer FROM cars ORDER BY RANDOM() LIMIT 1;`. (PostgreSQL syntax.)

**Q25: Case study: 2022 sign-ups spend per product.**

~~ANSWER: `SELECT p.product_id, SUM(u.spend) AS total_spend FROM products p JOIN users u ON p.product_id = u.product_id WHERE u.signup_year = 2022 GROUP BY p.product_id;`~~

**A:** JOIN and aggregate: `SELECT p.name, SUM(price * quantity) FROM users u JOIN purchases p ON u.id = p.user_id WHERE YEAR(u.signup_date) = 2022 GROUP BY p.name; .`

**Q26: Case study: Daily average downloads by plan type.**

**A:** `SELECT download_date, plan_type, ROUND(COUNT(*)::DECIMAL / COUNT(DISTINCT account_id), 2) AS avg_downloads FROM accounts a JOIN downloads d ON a.account_id = d.account_id GROUP BY download_date, plan_type; .`

**Q27: How to find duplicates in a table?**

**A:** `SELECT name, COUNT(*) FROM employees GROUP BY name HAVING COUNT(*) > 1; .`

**Q28: How to create a pivot-like view for sales by month?**

**A:** Conditional aggregation: `SELECT YEAR(date), SUM(CASE WHEN MONTH(date)=1 THEN sales ELSE 0 END) AS Jan FROM sales GROUP BY YEAR; .`

**Q29: How to calculate a running total?**

**A:** `SELECT id, salary, SUM(salary) OVER (ORDER BY id) AS running_total FROM employees; .`

**Q30: How to get the top 3 salaries overall?**

**A:** `SELECT salary FROM (SELECT salary, ROW_NUMBER() OVER (ORDER BY salary DESC) AS rn FROM employees) WHERE rn <= 3; .`

**Q31: What is the difference between UNION and JOIN?**

**A:** UNION stacks rows from similar tables; JOIN combines columns based on keys.

**Q32: How does EXISTS subquery work?**

**A:** `SELECT * FROM employees e WHERE EXISTS (SELECT 1 FROM bonuses b WHERE b.emp_id = e.id); .` Finds employees with bonuses.

**Q33: What is a CTE (Common Table Expression)?**

**A:** Temporary result set: `WITH high_earners AS (SELECT * FROM employees WHERE salary > 50000) SELECT * FROM high_earners JOIN departments ON ...; .`

**Q34: How to extract dates?**

**A:** `SELECT EXTRACT(YEAR FROM created_at) FROM transactions; .`

**Q35: How to create a histogram in SQL?**

**A:** `SELECT FLOOR((date - now)/10000) * 10000 AS bin_start, COUNT(*) AS bin_end FROM transactions GROUP BY`

**A:** `SELECT FLOOR(Salary/10000)*10000 AS Bucket, COUNT(*) FROM employees GROUP BY bucket;` . Buckets salaries in 10k ranges.

## Excel: Common Formulas and Features

Excel questions assess data handling. Here are 20 Q&As with practical examples using mock sales data: | Product | Region | Amount | | A | North | 100 | | B | South | 200 |.

### Q1: What is a cell address?

**A:** It's the column-row intersection, e.g., D5.

### Q2: What is the difference between relative and absolute referencing?

**A:** Relative (=A1) adjusts when copied; absolute (=A\$1) stays fixed. Example: = \$B\$1 \* A1 for fixed multiplier.

### Q3: How do you freeze panes?

**A:** View > Freeze Panes to lock headers while scrolling.

### Q4: How do you protect cells from copying?

**A:** Format Cells > Protection > Locked, then Review > Protect Sheet.

### Q5: What is the difference between a formula and a function?

**A:** Formula: =A1+B1 (custom calculation); Function: =SUM(A1:B1) (built-in).

### Q6: What is the order of operations in formulas (PEDMAS)?

**A:** Parentheses, Exponents, Division/Multiplication, Addition/Subtraction. Example: = (A1\*10 +5)/2.

### Q7: What are COUNT, COUNTA, and COUNTBLANK?

**A:** COUNT counts numbers; COUNTA non-empty cells; COUNTBLANK empty ones. In mixed data [1, text, blank], COUNT=1, COUNTA=2.

### Q8: What is the shortcut for filters?

**A:** Ctrl+Shift+L to add dropdown filters to headers.

### Q9: How do you insert a hyperlink?

**A:** Ctrl+K to link text to a URL.

### Q10: How do you merge text with CONCATENATE?

**A:** =CONCATENATE(A1, " ", B1) or simpler =A1 & " " & B1 (e.g., "Product A North").

### Q11: How do you split a column using Text to Columns?

**A:** Data > Text to Columns > Delimited (e.g., by space) to separate full names

~~A. Data > Text to Columns > Delimited (e.g., by space) to separate full names.~~

**Q12: Explain VLOOKUP with an example.**

**A:** =VLOOKUP(lookup\_value, table\_array, col\_index\_num, FALSE) for exact match.

Example: =VLOOKUP("A", B2:C3, 2, FALSE) returns North for Product A. Limitation: Only looks rightward.

**Q13: What is the difference between VLOOKUP and LOOKUP?**

**A:** VLOOKUP uses tables and column index; LOOKUP is simpler for sorted data but less versatile.

**Q14: How does INDEX-MATCH work?**

**A:** =INDEX(return\_range, MATCH(lookup\_value, lookup\_range, 0)). More flexible than

VLOOKUP: =INDEX(B:B, MATCH("A", A:A, 0)) gets Region for A.

**Q15: How do you create a PivotTable?**

**A:** Select data > Insert > PivotTable. Drag Region to Rows, Amount to Values (Sum). Result: North 100, South 200.

**Q16: How does Conditional Formatting work?**

**A:** Home > Conditional Formatting > Highlight Cells > Greater Than 150 (colors high values). For duplicates: Highlight Duplicates.

**Q17: What is SUMIF vs. SUMIFS?**

**A:** SUMIF one criterion: =SUMIF(Region:Region, "North", Amount:Amount) =100. SUMIFS multiple: =SUMIFS(Amount:Amount, Region:Region, "North", Product:Product, "A").

**Q18: How do you use an IF statement?**

**A:** =IF(Amount>150, "High", "Low"). For South (200): "High".

**Q19: How do you create data validation for a drop-down?**

**A:** Data > Data Validation > List (source e.g., A1:A3 for options).

**Q20: When should you use different charts?**

**A:** Bar for categories (regions); Line for trends (monthly sales). Base on PivotTable data.

**Power BI / Tableau: Beginner Dashboard and Calculations**

These tools focus on visualization. 12 Q&As for entry-level scenarios, like building a sales dashboard

questions.

**Q1: What is a dashboard in Power BI?**

A: A single-page summary of visuals. Create by building reports and publishing to Power BI Service. Scenario: Sales dashboard with key metrics.

**Q2: How do you share a Power BI dashboard?**

A: Publish to Power BI Service and share a link (requires same domain access).

**Q3: What are DAX basics?**

A: Data Analysis Expressions for calculations. Syntax like Excel; functions include CALCULATE. Context: Row or filter level.

**Q4: How do you create a DAX measure for total sales?**

A: Modeling > New Measure: Total Sales = SUM(Sales[Amount]). Use in visuals for dynamic aggregation. Scenario: Display on sales dashboard.

**Q5: How do filters work in Power BI?**

A: Apply at slicer, visual, page, or report level for interactivity. Scenario: Add year slicer to filter sales chart.

**Q6: What are the key differences between Power BI and Tableau?**

A: Power BI excels in Microsoft integration and DAX; Tableau in advanced visualizations and big data handling.

**Q7: What is Power Query for data prep?**

A: Get Data > Transform to clean (e.g., remove duplicates, filter rows). Scenario: Prep Excel sales data for dashboard.

**Q8: How do you create a dashboard in Tableau?**

A: Connect data > New Worksheet (drag fields to Rows/Columns) > New Dashboard > Add sheets. Scenario: Multi-sheet sales overview.

**Q9: What is a Tableau parameter?**

A: Dynamic value: Right-click > Create Parameter (e.g., sales threshold for filtering).

**Q10: How do you create a calculated field in Tableau?**

A: Analysis > Create Calculated Field: Profit = [Revenue] - [Cost]. Scenario: Add to viz for

~~A. Analysis - Create calculated field. Profit = [Revenue] - [Cost]. Scenario. Add to viz for profit trends.~~

### **Q11: What is data blending in Tableau?**

**A:** Combine unrelated data sources (e.g., SQL + Excel) for a unified dashboard.

### **Q12: What are context filters in Tableau?**

**A:** Right-click filter > Add to Context for performance in complex dashboards (applies first).

## **Statistics for Data Analysis**

Stats questions evaluate analytical thinking. 22 Q&As covering probability, testing, and more, with examples like coin flips or sales data.

### **Q1: What is standard deviation and variance?**

**A:** Variance is average squared difference from mean; SD is its square root. Example: Data [1,3,5], mean=3, variance=2.67, SD≈1.63.

### **Q2: What is skewness?**

**A:** Measures asymmetry: Positive (right tail, mean > median); negative (left); zero (symmetric).

### **Q3: What is a histogram?**

**A:** Bar chart of frequency bins showing distribution (e.g., sales amounts by range).

### **Q4: What is the difference between descriptive and inferential statistics?**

**A:** Descriptive summarizes sample (e.g., mean sales=150); inferential predicts population (e.g., 95% CI: 140–160).

### **Q5: What are sampling types?**

**A:** Random (equal chance); stratified (by subgroups like age); cluster (group sampling).

### **Q6: What is the Central Limit Theorem?**

**A:** Sample means approximate normal distribution for  $n>30$ , regardless of population shape.

### **Q7: What are joint, marginal, and conditional probability?**

~~$\Delta \cdot \text{Marginal } P(\Delta)$ ; joint  $P(\Delta \text{ and } R)$ ; conditional  $P(\Delta|R) = P(\Delta \text{ and } R)/P(R)$ . Example:~~

A: Marginal ( $P(A)$ ), joint ( $P(A \text{ and } B)$ ), conditional ( $P(A|B) = P(A \text{ and } B)/P(B)$ ). Example:  $P(\text{Rain}|\text{Cloudy})$ .

**Q8: What is a probability distribution?**

A: Assigns probabilities to outcomes. Normal: Bell curve, 68% within 1 SD of mean.

**Q9: What is the binomial distribution?**

A: For fixed trials with success probability (e.g., 5 coin flips,  $P(3 \text{ heads}) \approx 0.31$ ).

**Q10: What is the Poisson distribution?**

A: For rare events (e.g.,  $\lambda=2$  calls/hour,  $P(0 \text{ calls}) = e^{-\lambda} \approx 0.14$ ).

**Q11: What is a p-value?**

A: Probability of data under null hypothesis;  $<0.05$  suggests rejecting null (e.g.,  $p=0.03$  for  $\text{mean} \neq 50$ ).

**Q12: What are Type I and Type II errors?**

A: Type I: False positive (reject true null,  $\alpha=0.05$ ); Type II: False negative (miss effect,  $\beta$ ).

**Q13: What is parametric vs. non-parametric testing?**

A: Parametric assumes distribution (e.g., t-test for normal); non-parametric doesn't (e.g., Chi-square).

**Q14: What is regression?**

A: Models relationship:  $Y = \beta_0 + \beta_1 X + \text{error}$ .  $\beta_1$ : Change in Y per X unit.

**Q15: What are residuals?**

A: Observed minus predicted values; random scatter indicates good fit.

**Q16: What is the difference between correlation and causation?**

A: Correlation ( $r=-1$  to  $1$ ) shows association (e.g.,  $r=0.8$  sales/temp); causation needs experiments (e.g., A/B test), as confounders like heat could explain ice cream/crime link.

**Q17: What are the steps in hypothesis testing?**

A: State null/alternative, choose test, calculate statistic/p-value, decide (reject if  $p < \alpha$ ).

**Q18: What is the difference between t-test and z-test?**

A: T-test for small samples/unknown variance; z-test for large/known (e.g., compare group means).

**Q19: What are the basics of A/B testing?**

A: Randomize users to groups; measure metric (e.g., conversion); run 2+ weeks; check

A: Randomize users to groups, measure metric (e.g., conversion), run 2+ weeks, check significance with t-test (power=0.8).

#### **Q20: How do you calculate sample size for A/B testing?**

A:  $n = (Z\alpha + Z\beta)^2 * 2\sigma^2 / \delta^2$ , where  $\delta$  is minimum detectable effect,  $\alpha=0.05$ ,  $\beta=0.2$ .

#### **Q21: What is a confidence interval?**

A: Range for parameter: Mean  $\pm$  1.96\*SE for 95% CI.

#### **Q22: What is ANOVA?**

A: Compares means across >2 groups using F-statistic (variance between vs. within).

### **Business Case Questions: Real-World Scenarios**

These test end-to-end problem-solving. 16 Q&As with steps, using frameworks like hypothesize-analyze-recommend.

#### **Q1: Sales growth dropped 20% QoQ—what do you do?**

A: Step 1: Clarify metric/scope. Step 2: Hypothesize (seasonal, competition). Step 3: SQL segment: `SELECT region, SUM(sales) FROM sales GROUP BY region;` (e.g., South down 30%). Step 4: Recommend A/B promo test in affected areas.

#### **Q2: Customer churn is 15% monthly—analyze and fix.**

A: Step 1: Retention=85%. Step 2: Cohort analysis in Excel Pivot (new users higher churn?). Step 3: T-test cohorts. Step 4: Recommend onboarding A/B with emails.

#### **Q3: A/B test shows +10% clicks for variant B ( $p=0.03$ )—next steps?**

A: Step 1: Validate sample/power. Step 2: Check secondary metrics (e.g., conversion). Step 3: If significant, implement. Step 4: Monitor long-term effects.

#### **Q4: E-commerce conversion is low at 2%—optimize.**

A: Step 1: Funnel metrics. Step 2: SQL: `SELECT step, COUNT(*) FROM funnel GROUP BY step;` (drop at cart). Step 3: Hypothesize high prices. Step 4: A/B discount; recommend simplified checkout.

#### **Q5: New feature adoption is only 10%—measure success.**

A: Step 1: Track usage. Step 2: Cohort: Adopters +15% retention (t-test). Step 3: Survey barriers. Step 4: Recommend pop-up tutorials.

#### **Q6: Analyze seasonal sales spikes.**

A: Step 1: Time-series: `SELECT MONTH(date), AVG(sales) GROUP BY MONTH;` Step 2:

Regression forecast. Step 3: Recommend inventory ramp-up for peaks.

**Q7: User engagement dropped after update.**

A: Step 1: Metrics like DAU/MAU. Step 2: Segment mobile/web. Step 3: A/B revert. Step 4: Recommend rollback if causal.

**Q8: Pricing A/B: +10% price but -5% volume—impact?**

A: Step 1: New revenue = old \* 0.95 \* 1.10 (neutral). Step 2: Elasticity analysis. Step 3: Recommend segmented pricing.

**Q9: How to predict churn?**

A: Step 1: Features (usage, tenure). Step 2: Logistic regression or Excel Pivot by usage. Step 3: Recommend alerts for low-usage users.

**Q10: Calculate marketing ROI.**

A: Step 1: `SELECT campaign, SUM(revenue)/SUM(cost) FROM marketing GROUP BY campaign;`. Step 2: Attribution. Step 3: Recommend scaling high-ROI campaigns.

**Q11: Handle inventory overstock.**

A: Step 1: ABC/Pareto (80% value in 20% items). Step 2: Demand forecast. Step 3: Recommend JIT for low-value items.

**Q12: A/B test email subject lines.**

A: Step 1: Metric open rates. Step 2: 50/50 randomization. Step 3: T-test  $p < 0.05$ . Step 4: Recommend winner rollout.

**Q13: Perform customer segmentation.**

A: Step 1: K-means clustering. Step 2: RFM Pivot (recency, frequency, monetary). Step 3: Recommend targeted VIP emails.

**Q14: Detect fraud in transactions.**

A: Step 1: Z-score  $> 3$  for anomalies. Step 2: SQL flag high-value buys. Step 3: Recommend alert system.

**Q15: Measure product recommendation impact.**

A: Step 1: A/B with/without. Step 2: Lift in basket size. Step 3: Recommend personalization by history.

**Q16: High churn in global expansion.**

A: Step 1: Compare regions. Step 2: Stats + survey for cultural issues. Step 3: Recommend

UI localization.

---

## Comprehensive Survey of Data Analyst Interview Questions and Answers

This detailed survey expands on the Q&A above, providing deeper context, additional examples, and professional insights for thorough preparation. It draws from industry best practices, emphasizing how these skills apply in real roles at companies like Google or Amazon. For each section, I've included tips on common pitfalls, variations in interview styles (e.g., live coding vs. verbal explanation), and extensions for intermediate candidates. Mock datasets are reused for consistency, and where relevant, I've added tables for quick reference (e.g., function comparisons). This superset ensures you can handle follow-ups like "Why this query?" or "What if the data is dirty?"

### In-Depth SQL for Data Analytics

SQL interviews often involve whiteboarding queries or using tools like SQLPad. Beginners should master basics; intermediates add optimization (e.g., indexes for large tables). Common pitfalls: Forgetting aliases in joins or mishandling NULLs in aggregates. Practice with 100+ LeetCode problems.

The mock employees table (| id | name | dept\_id | salary |) and departments (| id | name |) from earlier illustrate most examples. For case studies, assume larger datasets like a sales DB with 1M rows—queries must be efficient (use WHERE early).

Expanding on Q1-Q10 (basics): These test foundational syntax. Interviewers may ask to trace execution order (FROM > WHERE > GROUP BY > HAVING > SELECT > ORDER BY). Variation: "Write a query to find employees earning more than their manager" (self-join + comparison).

For Q11-Q20 (joins/windows): Joins are 40% of questions. LEFT JOINs detect gaps (e.g., depts without employees). Windows avoid subqueries for rankings—key for analytics (e.g., YoY growth: LAG over PARTITION BY year). Pitfall: Over-partitioning slows queries; use only when needed.

Q21-Q35 (case studies/advanced): These mimic Uber/Facebook problems. For random selection (Q24). MvSQL uses RAND(). Duplicates (Q27) often lead to "How to remove?"

$(=ROW\_NUMBER() \text{ OVER (PARTITION BY name) } =1)$ . CTEs (Q33) make code readable—always use for multi-step logic. Histogram (Q35) is great for EDA; extend to percentiles with PERCENT\_RANK().

SQL Concept	Use Case	Common Mistake	🔗
JOINS	Combine customer/orders	Forgetting ON clause	
Aggregates	Sales totals by region	GROUP BY missing columns	
Windows	Top N per group	Wrong ORDER BY in OVER	
Subqueries	Correlated vs. non	Performance on large data	

## Detailed Excel Skills

Excel is ubiquitous for quick analysis; interviews may share a sample file via screen-share. Focus on efficiency—e.g., array formulas for intermediates. Pitfalls: Circular references or #N/A in VLOOKUP (use IFERROR).

Using the sales mock: VLOOKUP/INDEX-MATCH (Q12-Q14) are staples; 80% of roles ask them. PivotTables (Q15) for ad-hoc reporting—refresh after data changes. Conditional tools (Q16-Q18) for dashboards; combine with charts.

Q19-Q20 (advanced): Data validation prevents errors; charts should tell a story (e.g., why bar over pie?). Extensions: Power Query for ETL (like Power BI's), or XLOOKUP (newer Excel) as VLOOKUP successor.

Function	Syntax Example	Scenario	🔗
VLOOKUP	=VLOOKUP(A2, B:C, 2, FALSE)	Lookup price by ID	
INDEX-MATCH	=INDEX(C:C, MATCH(A2, B:B, 0))	Two-way lookup	
SUMIFS	=SUMIFS(Amount, Region, "North")	Multi-filter sum	
IF	=IF(Amount>150, "High", Amount)	Categorize values	

## Power BI / Tableau Fundamentals

Visualization interviews test storytelling. Beginners: Build a simple dashboard in 10 mins.

Power BI is more queried (Microsoft ecosystem); Tableau for design pros. Pitfalls: Ignoring relationships (causes wrong sums).

Q1-Q7 (Power BI): DAX context is tricky—row vs. filter (CALCULATE modifies). Scenario extension: Use measures in cards for KPIs. Power Query (Q7) handles 90% of cleaning.

Q8-Q12 (Tableau): Parameters (Q9) for what-if analysis. Blending (Q11) vs. joining: Use blending for secondary sources. Context filters (Q12) optimize LOD expressions for intermediates.

Tool	Strength	Beginner Tip	?
Power BI	DAX, integration	Start with import > model relationships	?
Tableau	Viz intuitiveness	Use Show Me for auto-charts	?

## Statistics Essentials

Stats questions probe assumptions (e.g., "Is normality met?"). Use R/Python for demos if allowed, but explain manually. Pitfalls: Confusing p-value with effect size.

Q1-Q10 (descriptive/probability): Distributions underpin modeling. Binomial/Poisson for counts; CLT justifies sampling.

Q11-Q22 (inferential): Hypothesis steps (Q17) are framework gold. A/B (Q19) needs randomization to avoid bias. Sample size (Q20) formulas from stats texts—tools like Optimizely calculate it.

Test	When to Use	Output	?
T-test	Compare 2 means	t-stat, p-value	?
ANOVA	>2 groups	F-stat	?
Correlation	Association	r value	?

## Business Scenarios in Depth

Cases are behavioral + technical: 30 mins to structure response. Use STAR (Situation-Task-Action-Result) implicitly. Frameworks: MECE for hypotheses.

For Q1-Q8 (metrics/churn/A/B): Always segment (e.g., by cohort/demographic). Churn (Q2): RFM analysis extension. A/B (Q3): Discuss multiple testing correction.

Q9-Q16 (prediction/ROI/etc.): Churn prediction (Q9) leads to ML basics (e.g., logistic).

Segmentation (Q13): Elbow method for K-means. Fraud (Q14): Beyond stats, domain knowledge (e.g., velocity checks).

Scenario Type	Key Steps	Metrics	Open
Churn	Cohort SQL, t-test	Retention rate	
A/B	Randomize, power calc	Lift, p-value	
ROI	Attribution model	CAC/LTV	
Segmentation	Clustering	RFM scores	

This survey equips you for 90% of interviews—review with timers. For more, simulate with mock datasets in free tools.

## Key Citations

- InterviewQuery SQL Questions

- DataCamp SQL Guide

• DataCamp SQL Quizzes

- Medium 25 SQL Questions
- DataCamp Excel Questions
- Simplilearn Excel Questions
- Simplilearn Power BI
- Edureka Tableau
- DataCamp Statistics
- InterviewQuery Stats/A/B
- DataLemur A/B
- Medium Business Frameworks
- InterviewQuery Product Cases