

Hey there, let's get this test started! I'll be asking you a series of questions to see if you have the skills of a data analyst. We'll start with the Junior phase, move on to Mid-Level, and then tackle the Senior-level questions. Each phase will have at least 35 questions.

Phase 1: Junior Data Analyst (0-2 years of experience)

1. What's the difference between a row and a column in a spreadsheet?
2. How would you count the number of empty cells in a column in Excel?
3. What's a primary key in a database?
4. Can you explain what a pivot table is and when you'd use it?
5. How do you calculate the average of a column in SQL?
6. What does the **GROUP BY** clause do in SQL?
7. How would you import a CSV file into a Pandas DataFrame in Python?
8. What's the difference between a list and a tuple in Python?
9. Can you explain what a data type is? Give a few examples.
10. What's the purpose of data cleaning?
11. What are the three measures of central tendency?
12. How do you find the maximum value in a column in SQL?
13. What is a "null" value in a database?
14. How would you handle a "divide by zero" error in a calculation?
15. What's the difference between **WHERE** and **HAVING** in SQL?
16. What's the **COUNT()** function in SQL used for?
17. How would you check for duplicate records in a table?
18. What's the purpose of a visualization like a bar chart or a line graph?
19. How would you create a new column in a Pandas DataFrame based on a condition?
20. What is a SQL **JOIN**? Name the most common types.
21. What's the difference between an inner join and a left join?
22. How do you filter data in SQL using the **WHERE** clause?
23. What's a common way to deal with missing data?
24. How do you sort data in a SQL query?
25. What's the difference between **SUM** and **COUNT**?
26. What does **SELECT *** do in SQL?
27. How do you find the number of unique values in a column in SQL?
28. What is the **ORDER BY** clause used for?
29. How would you create a simple bar chart in Python using a library like Matplotlib?
30. What's the difference between **SELECT** and **FROM** in a SQL query?
31. How do you find the top 5 rows with the highest sales in a table?
32. What is a data dictionary?
33. How would you merge two dataframes in Pandas?
34. What's the purpose of a **LIMIT** clause in SQL?
35. How would you create a calculated field in a pivot table?

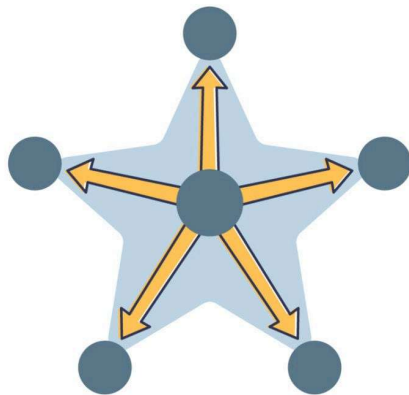
Phase 2: Mid-Level Data Analyst (2-5 years of experience)

1. What's the difference between a **UNION** and **UNION ALL** in SQL?
2. Can you explain the concept of a **subquery** and provide an example of when you'd use one?
3. How would you handle **outliers** in a dataset? Name a few methods.
4. What's the difference between **correlation** and **causation**?
5. Explain the **normalization** process in a database. Why is it important?
6. What is a **window function** in SQL? Give an example of a common use case.
7. How would you use a **case statement** in a SQL query?
8. Describe a time when you had to deal with a messy dataset. What steps did you take?
9. What's the purpose of creating a **dashboard**? What makes a good one?
10. How would you deal with a **time-series** dataset where you have missing values on some dates?
11. What's the difference between a **stored procedure** and a simple query?
12. Can you explain the difference between **left join** and **full outer join**?
13. How would you calculate the **running total** of a column in SQL?
14. What are some of the common pitfalls of a **self-join**?
15. How would you calculate the **percentile** of a column in SQL?
16. What's the difference between **ETL** and **ELT**?
17. How do you handle a situation where a report you built is showing incorrect data?
18. What's the purpose of a **cross join**? When is it useful?
19. How would you identify the **top N** records for each group in a dataset?
20. Can you explain the concept of **data integrity**?
21. How do you deal with **skewed data** in a visualization?
22. What is the **ACID** property in a database?
23. Describe a situation where you used a **temporary table** or a **CTE (Common Table Expression)**.
24. What are the pros and cons of using a **NoSQL database** compared to a SQL database?
25. How would you **automate** a weekly report you create?
26. Explain the concept of a **foreign key**.
27. What are some of the key things to look for during an **exploratory data analysis (EDA)** phase?
28. How would you use a **lag or lead function** in SQL?
29. How do you perform a **sentiment analysis** on text data?
30. How would you calculate the **month-over-month** growth of a metric?
31. What is the difference between a **data warehouse** and a **data lake**?
32. How would you measure the effectiveness of a new marketing campaign?
33. Describe a time when you had to present your findings to a non-technical audience. How did you do it?
34. What's the difference between a **view** and a **table** in a database?
35. What are some common data validation techniques you use?

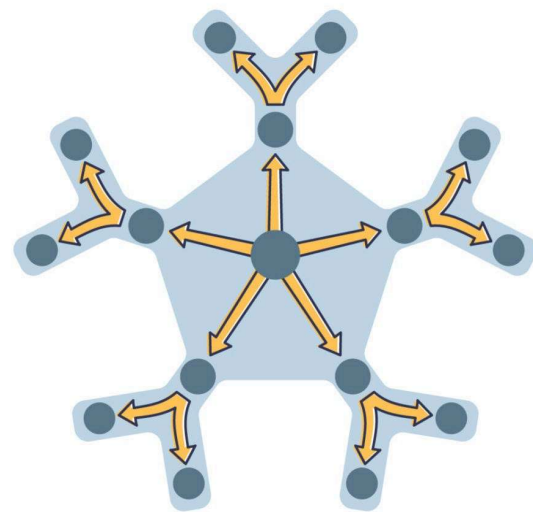
36. How would you write a query to find the second-highest salary in a table?

Phase 3: Senior Data Analyst (5+ years of experience)

1. What's the difference between a **star schema** and a **snowflake schema**? When would you use each?



STAR SCHEMA



SNOWFLAKE SCHEMA

Licensed by Google

2. How would you approach a project to **optimize a slow-running SQL query**?
3. Describe your process for building a **data model** from scratch.
4. How would you set up an **A/B test** to measure the impact of a new website feature?
5. Can you explain the concept of **data governance** and why it's crucial for a company?
6. How would you use **recursive CTEs** in SQL? Give a real-world example.
7. What's the difference between **online analytical processing (OLAP)** and **online transaction processing (OLTP)**?
8. How would you identify and communicate the **business impact** of your analysis?
9. Explain the concept of **surrogate keys** and **natural keys**.
10. Describe a time when you had to **resolve a data inconsistency issue** across multiple data sources.
11. What's the difference between a **data pipeline** and an **ETL process**?
12. How would you handle **schema evolution** in a data warehouse?

13. What is **data lineage**, and why is it important?
14. Can you describe the difference between **descriptive, diagnostic, predictive, and prescriptive analytics**?
15. How would you implement a **slowly changing dimension (SCD) Type 2** in a data warehouse?
16. What's your experience with **data visualization tools** like Tableau, Power BI, or Looker?
17. How would you use **change data capture (CDC)** in a data pipeline?
18. What are some of the key considerations for **data security and privacy** in your analysis?
19. How would you deal with a **large dataset** that doesn't fit into memory?
20. Describe your experience with **cloud platforms** like AWS, GCP, or Azure.
21. How do you **manage a backlog** of data requests and prioritize them?
22. What's the difference between a **view** and a **materialized view**?
23. How do you ensure the **quality and accuracy** of your data and analyses?
24. How would you design a **reporting system** for a new business unit?
25. What's the difference between **statistical significance** and **business significance**?
26. How would you use **indexes** to improve query performance?
27. Describe a time when you had to **mentor a junior analyst**. What did you teach them?
28. What are some of the **best practices for version control** in a data analysis project?
29. How would you approach a project that requires a **real-time dashboard**?
30. What's your experience with **data modeling techniques** beyond simple schemas?
31. How would you explain **data reliability and scalability** to a CEO?
32. What's the difference between **relational and non-relational databases** from a data analyst's perspective?
33. How would you conduct a **root cause analysis** for a sudden drop in a key business metric?
34. Describe a time you had to **advocate for a new tool or technology** for your team.
35. What is the role of **metadata** in a data analytics ecosystem?
36. What are the key components of a robust **data strategy**?
37. How do you stay up-to-date with the latest trends and tools in the data analytics field?

