

Data Analyst interview Questions (2025)



SQL Questions (1–7)

1. Find customers who purchased more than 3 times in the last month.

SELECT customer_id, COUNT(*) AS purchase_count

FROM orders

WHERE order_date >= DATEADD(month, -1, GETDATE())

GROUP BY customer id

HAVING COUNT(*) > 3;

2. Write a query to find the second highest salary.

SELECT MAX(salary)

FROM employees

WHERE salary < (SELECT MAX(salary) FROM employees);

3. What is the difference between RANK(), DENSE_RANK(), and ROW_NUMBER()?

- RANK(): Skips numbers after ties.
- DENSE RANK(): No gaps in ranking.
- ROW_NUMBER(): Unique sequential number regardless of ties.

4. Find duplicate records in a table.

SELECT customer_id, COUNT(*)

FROM customers

GROUP BY customer_id

HAVING COUNT(*) > 1;

5. What's the difference between WHERE and HAVING?

- WHERE: Filters before aggregation.
- HAVING: Filters after aggregation.

6. Get the average order value for each customer.

SELECT customer_id, AVG(order_amount) AS avg_value

FROM orders

GROUP BY customer_id;

7. How do you optimize a slow SQL query?

- Create indexes
- Avoid SELECT *
- Use EXPLAIN PLAN
- Limit subqueries
- Partition large tables

Python & Pandas Questions (8–12)

8. Drop missing values from a DataFrame.

df.dropna(inplace=True)

9. Group by and calculate total sales by region.

df.groupby("Region")["Sales"].sum()

10. Find outliers using the IQR method.

Q1 = df['amount'].quantile(0.25)

Q3 = df['amount'].quantile(0.75)

IQR = Q3 - Q1

outliers = df[(df['amount'] < Q1 - 1.5*IQR) | (df['amount'] > Q3 + 1.5*IQR)]

11. Merge two DataFrames.

pd.merge(df1, df2, on='customer_id', how='inner')

12. How do you handle large datasets in Python?

- Use dask or modin for parallel processing
- Load data in chunks with read_csv(chunksize=10000)
- Optimize data types (e.g., convert object to category)

ii Excel/Power BI Questions (13–15)

13. What Excel functions do you use in analysis?

- VLOOKUP, INDEX-MATCH
- IF, IFS, SUMIFS, COUNTIFS
- Pivot Tables, Charts, Slicers

14. Difference between Calculated Column and Measure in Power BI?

- Column: Calculated row-by-row and stored.
- Measure: Calculated at query time (more efficient for aggregations).

15. What are slicers and filters in Power BI?

- Slicers: Visual tools for filtering.
- Filters: Apply filtering at report, page, or visual level.

Business Case & Product Questions (16–19)

16. What metrics would you track for Amazon delivery performance?

- On-Time Delivery Rate
- · Average Delivery Time
- Return Rate
- Customer Satisfaction Score

17. Design a dashboard to monitor sales performance.

Metrics:

- Total Sales, Profit
- Orders by Region/Category
- Top-Selling Products
- · Filters: Time, Region, Category

18. How would you reduce cart abandonment on Amazon?

- Analyze drop-off steps in checkout funnel
- A/B test different UX changes
- Use ML model to predict high-risk customers

19. How would you evaluate if a new feature increased sales?

Use A/B Testing

- Pre/post analysis of KPIs
- Control for seasonality and external factors

A/B Testing Questions (20–21)

20. Explain p-value in A/B testing.

- Probability of seeing the observed difference (or more extreme) under the null hypothesis.
- A low p-value (e.g. < 0.05) suggests the difference is statistically significant.

21. How would you calculate statistical significance in Python?

from scipy.stats import ttest ind

t_stat, p_val = ttest_ind(group_A, group_B)



Behavioral (Leadership Principles) (22–25)

22. Tell me about a time you used data to solve a business problem.

In my previous project, I used Power BI to identify why return rates were high in one region. After root-cause analysis, we changed the vendor, reducing returns by 30%.

23. Describe a time when you had to dive deep.

I noticed a discrepancy in weekly revenue numbers. I traced it to a duplicate data load and wrote a validation script to catch it before dashboard refresh.

24. Tell me about a time you took ownership.

When a data pipeline broke, even though I wasn't the owner, I debugged it and restored the process to avoid dashboard downtime.

25. Have you ever disagreed with your manager about a decision?

Yes, we disagreed on the metric to use for campaign success. I backed my point with data and showed how conversion rate was a more accurate metric than just clicks.

Dipankar Pal

(Power BI Developer)

Email - dippal351@gmail.com

