

SECTION A: DATA ANALYSIS & SCIENCE

1. Imagine you're given a CSV file with 10,000 customer records, but 15% of the "Age" column values are missing. How would you handle this missing data before analysing customer demographics?

a) Fill missing ages with the average (mean) age:

If some values in the *Age* column are missing, one simple and commonly used method is to replace the missing values with the average age of all customers. First, I would calculate the mean age using the available data. Then, wherever the age value is missing, I would insert this average value.

b) Remove rows if age is not very important for the analysis:

If age is not a key factor in my analysis, I may choose to remove the rows where age is missing. For example, if I am focusing more on customer location, spending behavior, or product preferences, then missing age values may not affect my results much.

2. You need to identify which products are frequently purchased together. What kind of analysis would you perform, and what specific techniques or tools would you use?

a) Look into order data to see which products appear in the same order:

To find which products are frequently purchased together, I would first look at the order or transaction data. Each order usually contains a list of products bought by a customer at the same time. By checking which products appear together again and again in the same order, I can identify common combinations.

For example, if I notice that *Product A* and *Product B* often appear in the same order, it means customers usually buy them together. This basic observation helps in understanding customer buying behavior.

b) Use techniques or algorithms like association

After understanding the order data, I would use association techniques, which are commonly used in market basket analysis.

These techniques help in finding relationships between products, such as: If a customer buys one product, what other product they are likely to buy.

c) Use tools like pandas (Python)

To perform this analysis, I can use Python, especially the pandas library.

Pandas helps in:

- Reading order data files (like CSV files)
- Cleaning and organizing data
- Grouping orders and products easily

Since pandas is beginner-friendly and widely used, it is a good tool for learning and practicing this type of analysis.

3. Explain the difference between a bar chart and a histogram. When would you use each, and what does this tell you about the nature of the data they represent?

a) Bar Chart

A bar chart is mainly used to represent categorical data, where values are divided into different groups or categories.

- It is used for categories like products, cities, or departments.
- Each bar is separate from the others, showing that the categories are different.
- The height of each bar shows the value or count for that category.
- It is easy to compare different categories using a bar chart.

Example:

- Sales by product
- Number of customers in each city

b) Histogram

A histogram is used to show the frequency distribution of numerical data.

- It is used for continuous data, like age, salary, or marks.
- The bars are joined together, showing that the data flows continuously.
- It shows how data is distributed across different ranges or intervals.
- It helps us understand patterns like where most values lie.

Example:

- Age distribution of customers
- Salary distribution of employees

4. Write a simple SQL query to find all customers from "New York" who spent more than \$500 in January, ordered by their total spending (highest to lowest).

Select customers from city="New York" where spend>=500 and month="January" order by total_spend desc;

5. You analyse website traffic data and find that page load times increase dramatically when more than 100 users are active. What are at least three possible reasons for this correlation, and how would you investigate further?

Possible reasons why page load time increases after 100 users

1. Server overload

When many users are active at the same time, the server may not have enough power to handle all requests. This can cause the website to respond slowly.

2. Limited bandwidth

Internet bandwidth is shared among users. When more users access the website, the available bandwidth gets divided, which increases page load time.

3. Slow database performance

If many users request data at the same time, the database may take longer to respond, especially if queries are not optimized.

How I would investigate this issue

- I would check server performance metrics like CPU usage, memory usage, and response time.
- I would test website performance using load testing tools to see how it behaves under heavy user traffic.
- I would compare performance when users are below 100 and above 100 to clearly understand where the slowdown starts.

SECTION G: TECHNICAL MINDSET & PROBLEM SOLVING

1. Describe your process for learning a completely new technical skill or technology. How do you approach it, what resources do you use, and how do you know when you've understood it well enough?

1. First, I try to understand where the technology or skill is used in real life and which domain it belongs to.
2. Once I know its use, I search on Google or YouTube to understand the basic definition in simple words.
3. After that, I look for free beginner-level tutorials or YouTube courses to build a basic foundation.
4. While learning, I try to solve the questions shown in the tutorials and also watch the solutions to understand the logic.
5. When I feel comfortable, I search online for practice sets and try to solve more examples on my own.
6. If I don't understand something, I re-read the topic or watch another tutorial until it becomes clear.
7. Finally, I try to explain the concept to myself or to someone else. If I can explain it clearly and use it somewhere, I know I have understood it well.

2. You're stuck on a technical problem for several hours. What do you do?
Outline at least three specific strategies you would use to make progress.

- 1) First, I take a short break so my mind becomes fresh and calm.
- 2) After the break, I re-read the problem carefully and try to solve it again.
- 3) If I still face issues, I focus on small details mentioned in the question.
- 4) Then, I break the problem into smaller parts and try to solve each part one by one.
- 5) I take help from YouTube, Google, and learning websites that provide sample code and simple explanations.
- 6) Once I understand the concept, I try solving the problem again and check whether my logic matches the solution.
- 7) If I'm still stuck, I ask for help from friends or my mentor

3. How would you explain a technical concept from your domain (like a database, API, or machine learning model) to a non-technical family member?

- 1) Since I am still learning the concept myself, I avoid complex technical words and jargon.
- 2) I try to use simple language so the explanation is easy to understand.
- 3) I compare the concept with real-life examples or things they already know.
- 4) I focus on how it works or how the flow goes, not how it is built internally.
- 5) I ask very simple questions in between to check if they are understanding my explanation.
- 6) If they still look confused, I explain the same concept in a different and easier way.

4. What does "debugging" mean beyond just fixing code errors? Describe the mindset and systematic approach you would take to debug any complex system problem.

- 1) Debugging means finding the root cause of a problem.
- 2) It's about understanding why something is not working, not just fixing it.
- 3) I stay patient and calm, not frustrated.
- 4) I check the problem step by step, not all at once.
- 5) I test one change at a time to see what works.
- 6) I learn from mistakes so I don't repeat the same error again.

5. Why is documentation important in technical work, even if you're the only one who will ever see the code or project?

- 1) It helps me remember what I did earlier in the project.
- 2) It saves time when I return to the project after many days.
- 3) It makes my work clear and organized
- 4) It helps me understand my own logic better
- 5) If I make changes, I know why those changes were made.
- 6) It builds a good habit and improves professionalism.