1. What is the difference between structured and unstructured data?

Answer:

- **Structured data** is organized into rows and columns, like in relational databases (e.g., SQL tables).
- **Unstructured data** doesn't follow a clear format — examples include emails, images, videos, and social media posts.

2. What is data cleaning, and why is it important?

Answer:
Data cleaning is the process of fixing or removing incorrect, corrupted, or incomplete data. It's important because poor data quality can lead to inaccurate analysis and bad business decisions.

3. What are the most commonly used tools for data analysis?

Answer:
Popular tools include:

- **Excel** for quick analysis and pivot tables
- **SQL** for querying databases
- **Python (pandas, NumPy)** or **R** for advanced analysis
- **Power BI** or **Tableau** for data visualization

4. What is the difference between INNER JOIN and LEFT JOIN in SQL?

Answer:

- **INNER JOIN** returns rows that have matching values in both tables.
- **LEFT JOIN** returns all rows from the left table, and the matching rows from the right table. If there is no match, it returns NULL for the right side.

5. How do you handle missing data?

Answer:
Methods include:

- Removing rows or columns with missing values
- Filling missing values with the mean, median, or mode
- Using more advanced imputation methods (e.g., KNN or regression)

6. What is a pivot table, and how is it useful?

Answer:
A pivot table is a tool (mainly in Excel) used to summarize, sort, reorganize, and analyze data. It helps to view data patterns by aggregating data based on different categories.

7. What is the purpose of data visualization?

Answer:
Data visualization helps to present data clearly using charts, graphs, and dashboards. It makes complex data easier to understand and supports data-driven decision-making.

8. Explain a time when you used data to solve a problem.

Answer (sample):
In a university project, I analyzed survey data to understand student satisfaction. After cleaning the data in Excel and analyzing trends using Python, I found that course workload was the biggest concern. The insights helped the academic team adjust course schedules.