**1. What are missing values and how do you handle them?**  
Missing values are the empty or null entries in a dataset where data is not available.  
**Handling methods:**

* Remove rows/columns with missing data.
* Fill missing values with mean, median, mode, or a constant value.

**2. How do you treat duplicate records?**  
Duplicate records are repeated rows in a dataset.  
**Treatment:**

* Detect them using df.duplicated().
* Remove them using df.drop\_duplicates().

**3. Difference between dropna() and fillna() in Pandas?**

* dropna() → Removes rows or columns with missing values.
* fillna() → Fills missing values with a specific value (like mean, median, or 0).

**4. What is outlier treatment and why is it important?**  
Outlier treatment means handling extreme or unusual data points that are very different from others.  
**Importance:** Outliers can affect analysis and model accuracy.  
**Methods:** Remove them or replace them using techniques like IQR or capping.

**5. Explain the process of standardizing data.**  
Standardizing means converting data into a common scale (usually mean = 0 and standard deviation = 1).  
It helps models perform better by treating all features equally.

**6. How do you handle inconsistent data formats (e.g., date/time)?**  
Convert all data into a **consistent format** using tools like:

* pd.to\_datetime() for dates.
* Changing units or formats (like kg → grams, USD → INR).

**7. What are common data cleaning challenges?**

* Missing or duplicate values.
* Inconsistent formats.
* Outliers or wrong data.
* Human errors in data entry.

**8. How can you check data quality?**

* Use df.info() and df.describe() to check data type and summary.
* Check for missing, duplicate, and invalid values.
* Verify data accuracy and consistency.