1. What is data cleaning in Pandas?

Data cleaning is the process of **fixing or removing incorrect, incomplete, or messy data** so it can be used for analysis or AI models.

In **pandas**, data cleaning includes:

- Handling missing values
- Removing duplicates
- Fixing data types
- Correcting inconsistent values
- Handling outliers

2. Difference between `loc` and `iloc`

Feature	`loc`	`iloc`
Selection type	Label-based	Index-based
Uses	Row/column names	Integer positions
End index	Included	Excluded

3. What are missing values and how does Pandas handle them?

Missing values are empty or undefined values in a dataset, usually represented as **NaN**.

Pandas handles missing values by:

- Detecting → `isna()`
- Removing → `dropna()`
- Filling → `fillna()`

4. What is `groupby()` and why is it used?

`groupby()` is used to **group data based on a column and apply aggregation functions**.

Used for:

- Data summarization
- Feature engineering
- Statistical analysis
- Aggregation -Calculate:
  - `mean()`
  - `sum()`
  - `count()`
  - `min()`, `max()`

- Data Analysis & Insights

Used in:

- Sales analysis
- Salary analysis
- Medical data analysis

🔷 Syntax

```python
df.groupby('column_name')
```

5. How does Pandas help in data preprocessing for AI models?

Pandas plays a **crucial role in preparing raw data** before feeding it into AI/ML models. Most models require **clean, structured, numerical data**, and Pandas provides powerful tools to achieve this.

✅ Key Ways Pandas Helps

1️⃣ Cleaning & Handling Missing Data

- Detect missing values (`isnull()`)
- Remove or fill them (`dropna()`, `fillna()`)

2️⃣ Handling Duplicates & Inconsistent Data

- Remove duplicate rows (`drop_duplicates()`)
- Standardize values (e.g., lowercase text, fix typos)

3️⃣ Encoding Categorical Variables

- Convert text data into numbers (`get_dummies()`, label encoding)

4 Feature Selection & Transformation

- Select relevant columns
- Create new features from existing ones

5 Data Aggregation & Analysis

- Summarize data using `groupby()`, `mean()`, `sum()`

6 Preparing Data for ML Libraries

- Easy conversion to NumPy arrays
- Seamless integration with scikit-learn, TensorFlow, PyTorch