

1. What is data cleaning in Pandas?

Data cleaning means fixing or removing **incorrect**, **missing**, or **duplicate data** to make the dataset ready for analysis.

Common data cleaning tasks:

- Removing duplicates
- Handling missing values
- Correcting data types

Example:

```
import pandas as pd

df = pd.DataFrame({
    "Name": ["Asha", "Binu", "Asha"],
    "Age": [20, None, 20]
})

df = df.drop_duplicates()
df = df.fillna(0)
print(df)
```

2. Difference between loc and iloc

loc	iloc
Label-based indexing	Integer-based indexing
Uses row/column names	Uses index positions

Example:

```
import pandas as pd

df = pd.DataFrame({
    "Name": ["Asha", "Binu", "Chitra"],
    "Age": [20, 22, 21]
})

print(df.loc[0, "Name"])
print(df.iloc[1, 1])
```

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

Missing values are empty or undefined

values in a dataset, represented as NaN.

Pandas handles missing values by:

- Removing them
- Replacing them
- Detecting them

Example:

```
import pandas as pd
import numpy as np

df = pd.DataFrame({
    "Marks": [80, np.nan, 90]
})

print(df.isnull())
df["Marks"] = df["Marks"].fillna(df["Marks"].mean())
print(df)
```

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

groupby() is used to **group data based on a column** and apply **aggregate functions** like sum, mean, or count.

Example:

```
import pandas as pd

df = pd.DataFrame({
    "Department": ["IT", "IT", "HR"],
    "Salary": [30000, 35000, 25000]
})

result = df.groupby("Department")["Salary"].mean()
print(result)
```

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

Pandas helps AI models by:

- Cleaning and organizing data
- Handling missing values
- Filtering useful features
- Converting raw data into structured format

Example:

```
import pandas as pd

df = pd.read_csv("data.csv")
df = df.dropna()
df["Age"] = df["Age"].astype(int)

print(df.head())
```

Pandas prepares **clean and structured data** for AI algorithms.