

Pandas

1. What is data cleaning in Pandas?

Data cleaning in Pandas is the process of identifying and fixing errors in a dataset. It includes handling missing values, removing duplicate records, correcting data types, and fixing inconsistent or incorrect data to improve data quality.

2. What is the difference between loc and iloc?

loc is used for label-based data selection, meaning it selects rows and columns using names or labels.

iloc is used for position-based data selection, meaning it selects data using integer index positions.

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

Missing values are data entries that are not available or undefined and are represented as NaN in Pandas. Pandas provides functions like `isnull()`, `dropna()`, and `fillna()` to detect, remove, or replace missing values.

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

`groupby()` is a Pandas function used to group data based on one or more columns and perform aggregation operations such as sum, mean, and count. It is mainly used for data analysis and summarization.

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

Pandas helps in data preprocessing by cleaning data, handling missing values, removing duplicates, converting categorical data into numerical form, and selecting relevant features. This makes the data suitable for training AI and machine learning models.