Introduction to Pandas Library in Python

Pandas is an open-source library in Python used for data analysis and manipulation. It provides powerful tools to work with structured data, such as tables. The core data structures of Pandas are DataFrame and Series, which facilitate flexible and efficient data handling.

Core Data Structures in Pandas

DataFrame:

A two-dimensional data structure similar to tables in databases or spreadsheets in Excel. It consists of rows and columns and can contain data of different types (numbers, text, dates, etc.).

Series:

A one-dimensional data structure resembling a single column in a DataFrame. It can hold data of one type.

Basic Operations Using Pandas

Importing Data:

From CSV and Excel Files: Use the pd.read_csv() function to read data from CSV files, and pd.read_excel() to read data from Excel files. These functions quickly import data into a DataFrame.

```
import pandas as pd
df = pd.read_csv('data.csv')
```

Exploring Data:

Initial View: Use head() to display the first few rows of the DataFrame and tail() to display the last few rows.
Data Information: The info() method provides a summary of the DataFrame, such as data types and size.

Cleaning Data:

Handling Missing Values: Use dropna() to remove rows or columns with missing values and fillna() to fill missing values with specified values.
Removing or Correcting Invalid Values: Use functions like replace() to correct invalid values.

Applying Functions:

Using apply(): Apply functions to DataFrame or Series data. It helps in modifying values or calculating new values based on a specific function.

Merging Tables:

Joining and Concatenating: Use merge() and concat() to combine DataFrames. These operations are useful for merging data from multiple sources.

Data Analysis:

Basic Statistics: Functions like describe() provide a statistical summary of the data, including mean, standard deviation, min, and max.
Aggregation: Use groupby() to group data by specific values and calculate statistics like sum or mean.