

Training Report Day-20

28 June 2024

Pandas is a powerful and widely-used open-source library in Python for data manipulation and analysis. It provides data structures and functions needed to work with structured data seamlessly and efficiently. Here are some key features and components of the Pandas library:

keyboard_arrow_down

Key Features of Pandas

1. Data Structures:

- **Series:** A one-dimensional array-like object containing a sequence of values. It is similar to a list or a one-dimensional NumPy array but with labeled axes (index).
- **DataFrame:** A two-dimensional table of data with labeled axes (rows and columns). It can be thought of as a collection of Series objects sharing the same index.

2. **Data Alignment:** Automatic alignment of data for arithmetic operations on Series and DataFrame objects, making it easy to manage missing data.
3. **Data Cleaning:** Functions for handling missing data, such as filling, replacing, and dropping missing values.
4. **Data Transformation:** Tools for reshaping, pivoting, and transforming data, including functions for merging, joining, and concatenating DataFrame objects.
5. **Data Aggregation and Grouping:** Powerful group-by functionality to split data into groups, apply functions to each group, and combine the results.
6. **Time Series Handling:** Specialized tools for working with time series data, including date range generation and frequency conversion.
7. **Input and Output:** Functions for reading and writing data to and from various file formats, including CSV, Excel, SQL databases, and more.
8. **Visualization:** Integration with plotting libraries like Matplotlib to provide basic data visualization capabilities directly from DataFrame and Series objects.

```
import numpy as np
import pandas as pd
l1 = [1, 2, 3, 4, 5, 6]
labels = ['a', 'b', 'c', 'd', 'e', 'f']
d1 = {"A":10, "B":20, "C":30, "D":40, "E":50}
print(type(s1))
s2 = pd.Series(labels)
s2
s2[4]
```