Data Science-Introduction to Pandas

Here's a summary of the introduction to Pandas in data analysis for data science, along with examples:

What is Pandas?

Pandas is a Python library that provides high-performance, easy-to-use data structures and operations for working with structured data, including tabular data such as spreadsheets and SQL tables. It's particularly well-suited for data cleaning, filtering, grouping, merging, and analyzing data.

Key Features of Pandas:

1. **DataFrames**: The foundation of the library is the DataFrame data structure, which is a two-dimensional table of data with rows and columns.
2. **Series**: A one-dimensional labeled array of values, similar to a column in a spreadsheet.
3. **Handling Missing Data**: Pandas provides various methods for handling missing data, such as identifying, dropping, or imputing it.

Basic Operations:

1. **Creating a DataFrame**: You can create a DataFrame from a dictionary, another DataFrame, or using the `read_csv` function to load data from a CSV file.
2. **Inspecting Data**: Use methods like `head()`, `tail()`, and `info()` to quickly inspect your data.

Example: Creating a Simple DataFrame

```
 import pandas as pd

# Create a dictionary with data
data = {'Name': ['John', 'Anna', 'Peter'],
        'Age': [28, 24, 35],
        'City': ['New York', 'Paris', 'London']}

# Create a DataFrame from the dictionary
df = pd.DataFrame(data)

print(df)
```

Output:

```
     Name  Age       City
0   John   28   New York
1   Anna   24     Paris
2  Peter   35    London
```

Example: Handling Missing Data

```
 import pandas as pd

# Create a DataFrame with missing data
df = pd.DataFrame({'Name': ['John', 'Anna', None, 'Peter'],
                   'Age': [28, 24, np.nan, 35]})

print(df)

# Identify the missing values using `isnull()`
missing_values = df.isnull()

print(missing_values)
```

Output:

```
     Name     Age
0   John   28.0
1   Anna   24.0
2   NaN    NaN
3  Peter   35.0

[False False  True False]
```

This is just a brief introduction to Pandas, but it should give you an idea of how to get started with using the library for data analysis in Python.

Note: I assumed `numpy` is already installed and imported as `np`.

Curated by Brajesh Kumar