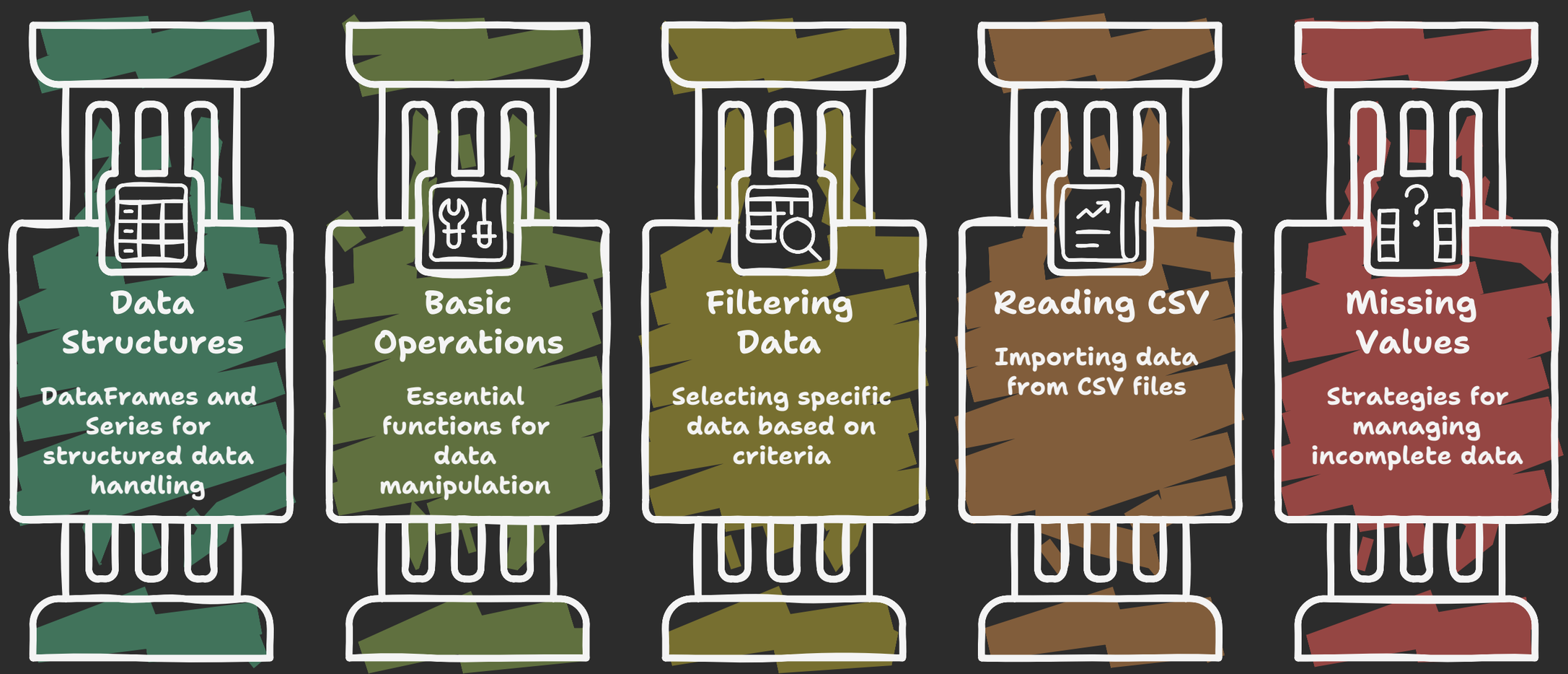
Introduction to Pandas



1. Import Pandas

To start using Pandas, you need to import it into your Python environment. The most common convention is to use the alias pd.

```
import pandas as pd
```

Explanation:

- pd is an alias for the Pandas library, allowing you to reference it easily throughout your code. This is a widely accepted convention in the Python community, making your code more readable and understandable to others familiar with Pandas.

2. Create DataFrame

You can create a DataFrame in Pandas using a dictionary. Here's an example:

```
import pandas as pd

data = {
    "Name": ["Aman", "Sita", "Rahul"],
    "Age": [22, 21, 23],
    "Marks": [85, 90, 78]
}

df = pd.DataFrame(data)

print(df)
```

Explanation:

- The dictionary data contains keys that represent column names and values that represent the data for those columns.
- The pd.DataFrame(data) function converts the dictionary into a DataFrame, which is a two-dimensional labeled data structure with columns and rows.

3. Basic Operations

Pandas provides several basic operations to interact with your DataFrame.

View First Rows

To view the first few rows of the DataFrame, you can use:

```
print(df.head())
```

To get a concise summary of the DataFrame, use:

```
print(df.info())
```

Statistics

To generate descriptive statistics, you can use:

```
print(df.describe())
```

Explanation:

- head(): Displays the first five rows of the DataFrame by default.
- info(): Provides information about the DataFrame, including the number of entries, column names, data types, and memory usage.

- describe(): Returns a statistical summary of the numerical columns, including:
    - count: Number of non-null entries.
    - mean: Average of the values.
    - std: Standard deviation of the values.
    - min: Minimum value.
    - max: Maximum value.

4. Select Data

You can select specific columns from the DataFrame.

Select Column

To select a single column, use:

```
print(df["Name"])
```

Select Multiple Columns

To select multiple columns, use:

```
print(df[["Name", "Marks"]])
```

5. Filtering Data 🔥

Filtering allows you to extract rows that meet certain conditions. For example, to filter students with marks greater than 80:

```
high_marks = df[df["Marks"] > 80]
print(high_marks)
```

Explanation:

- The condition inside the brackets (df["Marks"] > 80) creates a boolean mask that returns True for rows where the condition is met.
- The result is a filtered DataFrame containing only the rows that satisfy the condition.

6. Reading CSV File

Pandas makes it easy to read data from CSV files. You can load a dataset using:

```
df = pd.read_csv("data.csv")
print(df.head())
```

Explanation:

- Reading CSV files is a common task in real-world projects, as many datasets are stored in this format.
- The pd.read_csv() function loads the dataset into a DataFrame, allowing you to manipulate and analyze the data easily.

7. Handling Missing Values

Missing values can be a common issue in datasets. You can check for missing values using:

```
print(df.isnull().sum())
```

Remove Missing Values

To remove rows with missing values, use:

```
df = df.dropna()
```

Fill Missing Values

Alternatively, you can fill missing values with a specific value or the mean of the column:

```
df["Marks"].fillna(df["Marks"].mean(), inplace=True)
```

Explanation:

- isnull().sum(): Returns the count of missing values in each column.
- dropna(): Removes any rows that contain missing values.
- fillna(): Replaces missing values with the specified value, in this case, the mean of the "Marks" column.

Conclusion

Pandas is an essential tool for data analysis in Python, offering a wide range of functionalities for data manipulation and analysis. By understanding how to import Pandas, create DataFrames, perform basic operations, filter data, read CSV files, and handle missing values, you can effectively work with structured data and derive meaningful insights. As you continue to explore Pandas, you'll discover even more powerful features that can enhance your data analysis capabilities.