Simple Data Manipulation with Pandas

Pandas is a powerful data manipulation library in Python, providing data structures and tools for cleaning and analyzing structured data. This document walks through basic operations for manipulating data using Pandas.

1. Reading CSV Files

To read a CSV file, use the `read_csv()` function provided by Pandas. This function reads the file into a DataFrame, which is a table-like structure.

```
import pandas as pd


# Reading a CSV file

df = pd.read_csv('data.csv')

print(df.head())  # Display the first 5 rows
```

2. Handling Missing Values

Missing values can be handled by using methods such as `dropna()` to remove them or `fillna()` to fill them with a specified value.

```
# Dropping rows with missing values

df_cleaned = df.dropna()


# Filling missing values with a specific value

df_filled = df.fillna(0)
```

3. Removing Duplicates

Duplicates can be removed using the `drop_duplicates()` function, which returns a DataFrame with duplicate rows removed.

```
# Removing duplicate rows
```

```
df_no_duplicates = df.drop_duplicates()
```

4. Filtering Data

Filtering allows you to select rows that meet certain criteria. This is done using boolean indexing.

```
# Filtering rows where column 'A' is greater than 5

filtered_df = df[df['A'] > 5]
```

5. Sorting Data

Sorting can be done using the `sort_values()` method, which allows you to sort the data by one or more columns.

```
# Sorting by column 'B'

sorted_df = df.sort_values(by='B', ascending=True)
```

6. Grouping Data

Grouping is used to group data by specific columns and apply aggregate functions to each group.

```
# Grouping by column 'C' and calculating the mean for each group

grouped_df = df.groupby('C').mean()
```