Introduction to Pandas in Python

DATA ANALYSIS AND MANIPULATION MADE EASY PRESENTED BY Mistry Meet S.

What is Pandas?

* - Pandas is a Python library for data manipulation and analysis.
* - Provides data structures like Series and DataFrame.
* - Built on NumPy, integrates well with other Python libraries.

Why Use Pandas?

* - Simple and powerful data manipulation.
* - Handles missing data efficiently.
* - Supports multiple file formats (CSV, Excel, JSON, SQL, etc.).
* - Optimized for performance and large datasets.

# Pandas Installation

* Run the following command in your terminal:
* `pip install pandas`
* Requires Python and works well with Jupyter Notebook.

# Pandas Data Structures

* \*\*Series (1D labeled array)\*\*
* ```python
* import pandas as pd
* s = pd.Series([10, 20, 30, 40])
* print(s) ```
* \*\*DataFrame (2D table)\*\*
* ```python
* data = {'Name': ['Alice', 'Bob'], 'Age': [25, 30]}
* df = pd.DataFrame(data)
* print(df)
* ```

# Reading & Writing Data

* \*\*Reading CSV:\*\*
* ```python
* df = pd.read\_csv('data.csv')
* ```
* \*\*Writing to CSV:\*\*
* ```python
* df.to\_csv('output.csv', index=False)
* ```

# Data Exploration

* - `df.head()` – View first few rows.
* - `df.info()` – Get dataset summary.
* - `df.describe()` – Statistics of numeric columns.
* ```python
* print(df.head())
* print(df.info())
* print(df.describe())
* ```

# Data Cleaning

* - Handling missing values:
* ```python
* df.dropna() # Remove missing values
* df.fillna(0) # Replace missing values
* ```
* - Removing duplicates:
* ```python
* df.drop\_duplicates()
* ```
* - Changing data types:
* ```python
* df['column'] = df['column'].astype(int)
* ```

# Data Manipulation

* - Filtering data:
* ```python
* df[df['Age'] > 25]
* ```
* - Sorting:
* ```python
* df.sort\_values('Age', ascending=False)
* ```
* - Adding new columns:
* ```python
* df['New\_Column'] = df['Age'] \* 2
* ```

# Grouping and Aggregation

* - `groupby()` for summarizing data:
* ```python
* df.groupby('Category').sum() ```
* - Aggregating functions:
* ```python
* df['Column'].mean()
* df['Column'].max()
* ```

# Merging & Joining DataFrames

* \*\*Merging (like SQL JOIN)\*\*
* ```python
* df\_merged = pd.merge(df1, df2, on='ID', how='inner')
* ```
* \*\*Concatenation\*\*
* ```python
* df\_combined = pd.concat([df1, df2])
* ```

# Visualization with Pandas

* Pandas integrates with Matplotlib for quick plots.
* ```python
* import matplotlib.pyplot as plt
* df['Age'].plot(kind='hist')
* plt.show()
* ```

# Summary

* - Pandas is a powerful library for data handling.
* - Provides easy-to-use functions for data cleaning, manipulation, and analysis.
* - Supports integration with visualization tools.

MINI PROJECT LINK

TITLE :- Introduction to Pandas in Python

LINK :-

https://github.com/meet1402/mistry-meet.git

# THANK YOU