PANDAS LIBRARY IN PYTHON

BY ; AYUSH VISHWAKARMA
ROLL NO: 03
ENROLLMENT NO: 2301031800008
BATCH: C1

Introduction To Pandas

Pandas is an open-source Python library for data manipulation and analysis.

It provides data structures and functions needed to efficiently handle structured

Built on top of NumPy, it is widely used in data science and machine learning.

Installation

01. To install Pandas, use the following command:

02. pip install pandas

* Import Pandas in Python:

03. Import pandas as pd

Key Features Of Pandas

Fast and efficient for data manipulation.

Data alignment and handling of missing data.

Integration with libraries like NumPy, Matplotlib, and Scikit-

Label-based slicing and indexing similar to SQL and Excel.

Reading and writing data from various file formats (CSV, Excel,

APPLICATIONS OF PANDAS

01

Data Analysis: Cleaning and processing large datasets.

02

Data Analysis: Cleaning and processing large datasets.

03

Financial Analysis: Stock market and investment

04

Web Scraping: Storing and analyzing extracted

DATA STRUCTURE IN PANDAS

Series (1D labeled array): A Series is similar to a column
in Excel or a list in Python

```
import pandas as pd
s = pd.Series([10, 20, 30, 40])
print(s)
```

DataFrame (2D labeled table): A DataFrame is a table with rows and columns

```
data = {'Name': ['Alice', 'Bob'], 'Age': [25, 30]}
df = pd.DataFrame(data)
print(df)
```

READING AND WRITING DATA

1. Reading Data from Files
CSV: df = pd.read_csv('data.csv')
Excel: df = pd.read_excel('data.xlsx')
JSON: df = pd.read_json('data.json')

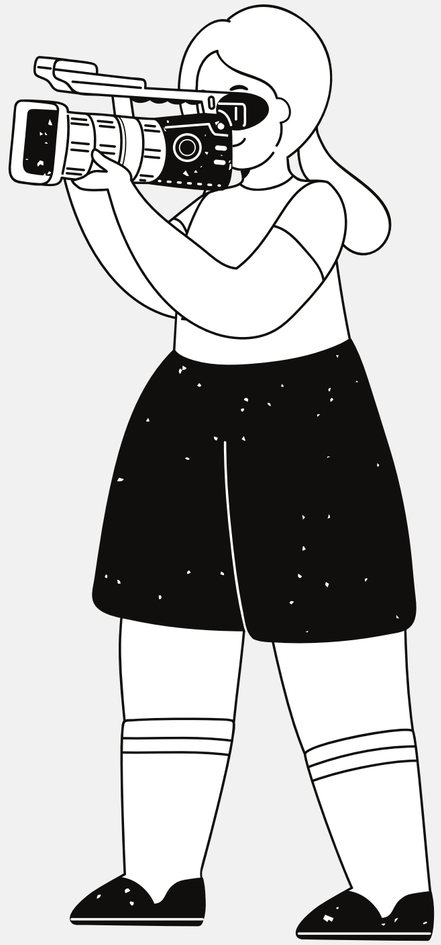
2. Writing Data to Files
CSV: df.to_csv('output.csv', index=False)
Excel: df.to_excel('output.xlsx', index=False)

HANDLING MISSING VALUES

Checking Missing Data:
df.isnull().sum() #
Count missing values
Filling Missing Data:
df.fillna(0) # Fill
missing values with 0
Dropping Missing Data:
df.dropna() # Remove
rows with missing values

DATA VISUALIZATION WITH PANDA

Line Plot:
df.plot(kind='line')
Bar Chart:
df.plot(kind='bar')
Histogram:
df.hist()



CONCLUSION

Pandas is a powerful library for data manipulation and analysis.

It provides efficient data structures like Series and DataFrame.

Used extensively in data science, finance, and machine learning.

Learning Pandas is essential for anyone working with large datasets