**How to install Pandas**

Open VS Code🡪Create new file give any name to file(pandas.py)

Go to Terminal🡪Go to New Terminal

PS C:\Users\amitk> pip install pandas (press Enter)

It will automatically install numpy.

Note:-Once you install pandas than numpy also install automatically.

**🐼 Introduction to Pandas in Python**

Pandas is an open-source Python library used for data manipulation and data analysis. It is built on top of NumPy and provides powerful and easy-to-use data structures.

🔑 Key Features of Pandas:

* Data structures: Series (1D), DataFrame (2D)
* Handle missing data
* Perform data filtering, grouping, merging
* Support for CSV, Excel, SQL, JSON, etc.
* Efficient handling of large datasets

**Data structures in Pandas**

1. **Series (1D labeled array)**

import pandas as pd

data = pd.Series([10, 20, 30, 40])

print(data)

**Output:**

0 10

1 20

2 30

3 40

dtype: int64

**2. DataFrame (2D labeled table)**

import pandas as pd

data = {'Name': ['Amit', 'Ashutosh'], 'Age': [25, 26]}

df = pd.DataFrame(data)

print(df)

**Output:**

Name Age

0 Amit 25

1 Ashutosh 26

**Basic Operations in Pandas**

print(df.head()) # First 5 rows

print(df.info()) # Summary of the DataFrame

print(df.describe()) # Statistical summary

print(df['Name']) # Access a column

**Reading and Writing Data**

df = pd.read\_csv('data.csv') # Read from CSV

df=to\_excel('data.xlsx') # Write to Excel

**Importing and exporting files**

* 1. Importing (Reading) Files

Read a CSV file

import pandas as pd

df = pd.read\_csv(' store\_dat.csv') # Reads the CSV file into a DataFrame

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

Read an Excel file

df = pd.read\_excel('Detail\_data.xlsx')

print(df.head())

Read a JSON file

df = pd.read\_json('data.json')

print(df.head())

**Basic functionalities of a data object**

In Pandas, the fundamental data objects are:

1. Series – a one-dimensional labeled array.
2. DataFrame – a two-dimensional labeled data structure (like a table or spreadsheet).
3. Index – an immutable array used to label rows and columns.

1. Creation

You can create a DataFrame or Series using:

import pandas as pd

# Series

s = pd.Series([10, 20, 30], index=['a', 'b', 'c'])

# DataFrame

data = {'Name': ['Amit', 'Ravi', 'sumit', 'ram', 'shyam', 'mohan' ,'Amit'], 'Age': [28, 30,32,25,16,12,35]}

df = pd.DataFrame(data)

print(df)

* 1. Viewing Data

df.head() # First 5 rows

df.tail(3) # Last 3 rows

df.info() # Summary (data types, non-null values)

df.describe() # Statistical summary of numeric columns

* 1. Accessing Data

df['Name'] # Access column

df.Name # Shortcut to access column

df.iloc[0] # Access by index position

df.loc[0] # Access by index label

* 1. Filtering / Selection

df[df['Age'] > 25] # Filter rows by condition

df[(df['Age'] > 25) & (df['Name'] == 'Amit')] # Multiple conditions

* 1. Modifying Data

df['Age'] = df['Age'] + 1 # Update column values

df['Country'] = 'India' # Add new column

* 1. Handling Missing Data

df.dropna() # Drop rows with NaN

df.fillna(0) # Replace NaN with a value

df.isna() # Check for NaN values

**Example:-**

import pandas as pd

# Series

s = pd.Series([10, 20, 30], index=['a', 'b', 'c'])

# DataFrame

data = {'Name': ['Amit', 'Ravi', 'Sumit', 'ram', 'shyam', 'mohan' ,'Amit'], 'Age': [28, 30,None,25,None,12,35]}

df = pd.DataFrame(data)

print(df.dropna())

print(df.fillna(0) )

print(df.isna()  )

* 1. Aggregation and Statistics

df.mean() # Column-wise mean

df['Age'].sum() # Sum of a column

df.groupby('Name').mean() # Group by and aggregate

**Example:-**

import pandas as pd

# Sample DataFrame with numeric columns

data = {

'Math': [85, 90, 78, 92],

'Science': [88, 85, 82, 95],

'English': [75, 80, 78, 85]

}

df = pd.DataFrame(data)

print("Original DataFrame:")

print(df)

# Calculate column-wise mean (default behavior)

mean\_values = df.mean()

print("\nColumn-wise Mean:")

print(mean\_values)

**Example:-**

import pandas as pd

# Sample DataFrame with numeric columns

data = {

    'Math': [85, 90, 78, 92],

    'Science': [88, 85, 82, 95],

    'English': [75, 80, 78, 85]

}

df = pd.DataFrame(data)

print("Original DataFrame:")

print(df)

total\_Math = df['Math'].sum()

total\_Science = df['Science'].sum()

total\_English = df['English'].sum()

print("\nSum of Math column:")

print(total\_Math)

print("\nSum of Science column:")

print(total\_Science)

print("\nSum of English column:")

print(total\_English)

**Example:-**

import pandas as pd

# Sample data

data = {

    'Name': ['Alice', 'Bob', 'Alice', 'Bob', 'Charlie', 'Alice'],

    'Subject': ['Math', 'Math', 'Science', 'Science', 'Math', 'English'],

    'Score': [85, 78, 92, 81, 89, 75]

}

df = pd.DataFrame(data)

print(df)

avg\_scores = df.groupby('Name')['Score'].mean()

print(avg\_scores)