



Semester :VI\_\_\_\_\_

Subject :\_\_\_\_\_ DAV \_\_\_\_\_

Academic Year: 2023 - 2024

### **PANDAS:**

Pandas is a powerful and open-source Python library. The Pandas library is used for data manipulation and analysis. Pandas consist of data structures and functions to perform efficient operations on data.

Pandas is a powerful and versatile library that simplifies the tasks of data manipulation in Python. Pandas is well-suited for working with tabular data, such as spreadsheets or SQL tables.

The Pandas library is an essential tool for data analysts, scientists, and engineers working with structured data in Python.

### **What is Python Pandas used for?**

It is built on top of the NumPy library which means that a lot of the structures of NumPy are used or replicated in Pandas.

The data produced by Pandas is often used as input for plotting functions in Matplotlib, statistical analysis in SciPy, and machine learning algorithms in Scikit-learn.

Here is a list of things that we can do using Pandas.

- Data set cleaning, merging, and joining.
- Easy handling of missing data (represented as NaN) in floating point as well as non-floating point data.
- Columns can be inserted and deleted from DataFrame and higher-dimensional objects.
- Powerful group by functionality for performing split-apply-combine operations on data sets.
- Data Visualization.

### **Installing Pandas**

The first step in working with Pandas is to ensure whether it is installed in the system or not. If not, then we need to install it on our system using the pip command.

*pip install pandas*

### **Importing Pandas**

After the Pandas have been installed in the system, you need to import the library. This module is generally imported as follows:

*import pandas as pd*

### **Data Structures in Pandas Library**

Pandas generally provide two data structures for manipulating data. They are:

- Series
- DataFrame

### **Pandas Series**

A Pandas Series is a one-dimensional labeled array capable of holding data of any type (integer, string, float, Python objects, etc.). The axis labels are collectively called indexes.

### **Example:**

```
import pandas as pd
import numpy as np
```



Semester :VI\_\_\_\_\_

Subject :\_\_\_\_\_ DAV \_\_\_\_\_

Academic Year: 2023 - 2024

```
# Creating empty series
ser = pd.Series()
print("Pandas Series: ", ser)

# simple array
data = np.array(['h', 'e', 'l', 'l', 'o'])

ser = pd.Series(data)
print("Pandas Series:\n", ser)
```

**Output:**

Pandas Series:

```
0    h
1    e
2    l
3    l
4    o
```

**Pandas DataFrame**

Pandas DataFrame is a two-dimensional data structure with labeled axes (rows and columns). Pandas DataFrame is created by loading the datasets from existing storage (which can be a SQL database, a CSV file, or an Excel file).

Pandas DataFrame can be created from lists, dictionaries, a list of dictionaries, etc.

**Example:**

```
import pandas as pd
```

```
# Calling DataFrame constructor
df = pd.DataFrame()
print(df)
```

```
# list of strings
lst = ['APSIT', 'is', 'the', 'best', 'place', 'for', 'Everyone']
```

```
# Calling DataFrame constructor on list
df = pd.DataFrame(lst)
print(df)
```

**Output:**

```
0
0  APSIT
1    is
2   the
3  best
4 place
5   for
6 Everyone
```