CLASS XII - IP (2020-2021)
BBPSPP

Chapter 3 - Python Pandas

By Beena Nair

LEARNING OBJECTIVES

This presentation will help you to analyse and comprehend about the following topics:

1. **Introduction to Pandas**
2. **Data Structures in Panda**

- Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

To be able to use Pandas, we need to **import pandas** module in the current shell environment.

- Using Pandas we can perform 5 major steps:
  - Load the data
  - Prepare the data
  - Manipulate the data
  - Design a model
  - Analysis of data

- Pandas deals with the following three data structures —
  - Series
  - DataFrame
  - Panel

These data structures are built on top of Numpy array, which means they are fast.

DATA STRUCTURES	DIMENSION	DESCRIPTION
SERIES	One dimensional	1D labeled homogeneous array, size-immutable. Homogenous means data of same data type.

For example- A series having the following collection of integers

34	44	66	101	520	610	703	90	2	720

DATA STRUCTURES	DIMENSION	DESCRIPTION
DATA FRAMES	Two dimensional	General 2D labeled, size-mutable tabular structure with potentially heterogeneously typed columns. Hetrogenous means data of different data type.

For example : A table of data being represented in rows and columns as shown below:

NAME	AGE	GENDER	MARKS	STREAM
Sutapa	16	Female	87.5	HUMANITIES
Suraj	17	Male	89.3	COMMERCE

DATA STRUCTURES	DIMENSION	DESCRIPTION
PANEL	Three dimensional	General 3D labeled, size-mutable array.

Panel is a three-dimensional data structure with heterogeneous data. It is hard to represent the panel in graphical representation. But a panel can be illustrated as a container of DataFrame.

Watch this video to understand the basic concept of PYTHON PANDA-

https://www.youtube.com/watch?v=B42n3Pc-N2A

BBPSPP

To create an empty series

```
import pandas as pd
d=pd.Series()
d
```

It will print the output as

```
Series([], dtype:float64)
```

By default the series will be of float data type.

To create a series using List

```
import pandas as pd

d=pd.Series(1,2,3,4)

print(d)
```

It will print the output as

```
0       1

1       2

2       3

3       4

dtype: int64
```

The first series starting from 0 to 3 is the index number for the values of the series

To create a series using range() method

```
import pandas as pd

d=pd.Series(range(5))

print(d)
```

It will print the output as

```
0    1

1    2

2    3

3    4

dtype: int64
```

The range method is used to display the series starting from 0 to n-1 i.e 5-1=4

To create a series with user specific index value

```
import pandas as pd

d=pd.Series([10,20,30,40], index=[1,3,5,7])

print(d)
```

It will print the output as

```
1       10

3       20

5       30

7       40

dtype: int64
```

The index value has changed.

To access the series with user specific index value

```
import pandas as pd

d=pd.Series([10,20,30,40], index=[1,3,5,7])

d[3]
```

It will print the output as

```
20

dtype: int64
```

The series value 20 is at index position 2. Hence 20 will be displayed.

Try yourself :

https://www.tutorialspoint.com/python_pandas/python_pandas_series.htm

ASSIGNMENT

Given the data :
Weight measurements for 14 values of muffins
78, 72, 69, 81, 63, 67, 65, 79, 74, 71, 83, 71, 79, 80

1. Write python program to print a blank panda series.
2. Write python program to print the weight measures of muffins using panda series.
3. Write python program to print the weight measures of muffins using panda series with index values starting from A to M.
4. What is the default data type of python panda series.
5. Find the error in the statement
   D=pd.Series(2,4,6,8, index=range(4))

ALL THE ABOVE QUESTIONS TO BE PART OF PRACTICAL FILE ALSO.YOU WRITE THE CODE USING W3SCHOOL.COM OR JUPYTER.ORG SITES

BBPSPP