Module 02

# Pandas: Series and DataFrame

Data Science Developer

Purwadhika
Startup and Coding School

# Outline

- What is Pandas ?
- Series
- DataFrame
- Creating Series and DataFrame

# What is Pandas ?

# What is Pandas

- Pandas is a high-level data manipulation tool developed by Wes McKinney
- Built on the Numpy package and its key data structure is called the Series and DataFrame.

# Series in Pandas

Purwadhika
Startup and Coding School

# Series

- A Series is very similar to a 1D NumPy array (in fact it is built on top of the NumPy array object).
- The differences:
  - a Series can have axis labels, meaning it can be indexed by a label, instead of just a number location.
  - can hold any arbitrary Python Object, not just numeric.

Index label

```
0    10
1    20
2    30
dtype: int32
```

Series

**VS**

```
array([10, 20, 30])
```

Arrays

Purwadhika
Startup and Coding School

# Data in a Series

A pandas Series can hold a variety of object types:

```
In [10]: pd.Series(data=labels)

Out[10]: 0    a
         1    b
         2    c
         dtype: object

In [11]: # Even functions (although unlikely that you will use this)
         pd.Series([sum,print,len])

Out[11]: 0      <built-in function sum>
         1    <built-in function print>
         2      <built-in function len>
         dtype: object
```

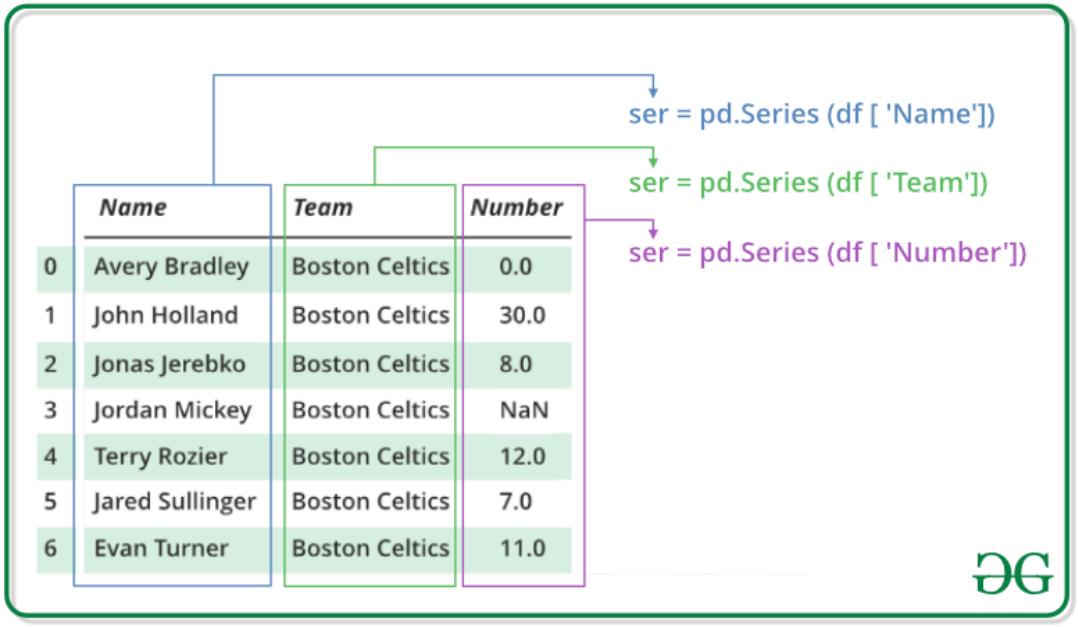
Purwadhika
Startup and Coding School

# DataFrame in Pandas

Purwadhika
Startup and Coding School

# DataFrame

- DataFrames are the workhorse of pandas and are directly inspired by the R programming language.
- A bunch of Series objects put together to share the same index.

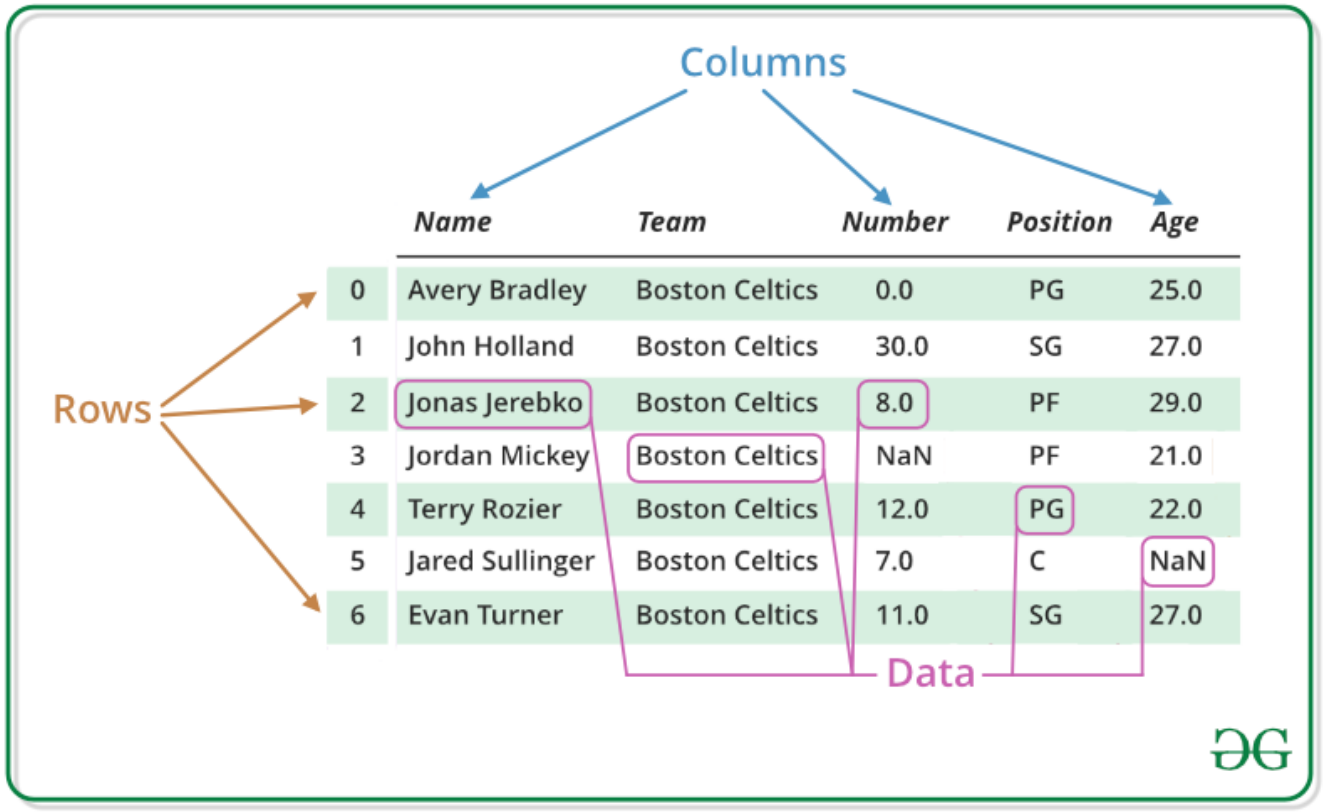



| | Name | Team | Number |
|---|---|---|---|
| 0 | Avery Bradley | Boston Celtics | 0.0 |
| 1 | John Holland | Boston Celtics | 30.0 |
| 2 | Jonas Jerebko | Boston Celtics | 8.0 |
| 3 | Jordan Mickey | Boston Celtics | NaN |
| 4 | Terry Rozier | Boston Celtics | 12.0 |
| 5 | Jared Sullinger | Boston Celtics | 7.0 |
| 6 | Evan Turner | Boston Celtics | 11.0 |

Purwadhika
Startup and Coding School

# DataFrame

- Is a structured Data: Tabular with columns and rows




Purwadhika
Startup and Coding School

# Creating a Series

# Using Pandas in Python

```
In [2]: import numpy as np
        import pandas as pd
```

# Creating a Series

## From a Python List

```
In [3]: labels = ['a','b','c']
        my_list = [10,20,30]
        arr = np.array([10,20,30])
        d = {'a':10,'b':20,'c':30}
```

**Using Lists**

```
In [4]: pd.Series(data=my_list)
Out[4]: 0    10
        1    20
        2    30
        dtype: int64
```

```
In [5]: pd.Series(data=my_list,index=labels)
Out[5]: a    10
        b    20
        c    30
        dtype: int64
```

```
In [6]: pd.Series(my_list,labels)
Out[6]: a    10
        b    20
        c    30
        dtype: int64
```

Purwadhika
Startup and Coding School

# Creating a Series

## From a Numpy Array

```
In [7]: pd.Series(arr)

Out[7]: 0    10
        1    20
        2    30
        dtype: int64

In [8]: pd.Series(arr,labels)

Out[8]: a    10
        b    20
        c    30
        dtype: int64
```

Purwadhika
Startup and Coding School

# Creating a Series

## From a Dictionary

```
d = {'a':10,'b':20,'c':30}
```

```
In [9]: pd.Series(d)
```

```
Out[9]: a    10
        b    20
        c    30
        dtype: int64
```

# Creating a DataFrame

# Creating a DataFrame

## From lists

```
In [4]: a= [i for i in range (1,5)]
        b= [i for i in range (5,9)]
        c= [i for i in range (9,13)]
```

```
In [5]: pd.DataFrame(data=[a,b,c],
                     index='a b c'.split(),
                     columns= 'w x y z'.split())
```

Out[5]:

| | w | x | y | z |
|---|---|---|---|---|
| **a** | 1 | 2 | 3 | 4 |
| **b** | 5 | 6 | 7 | 8 |
| **c** | 9 | 10 | 11 | 12 |

```
In [8]: pd.DataFrame(data=list(zip(a,b,c)),
                     index='w x y z'.split(),
                     columns='a b c'.split())
```

Out[8]:

| | a | b | c |
|---|---|---|---|
| **w** | 1 | 5 | 9 |
| **x** | 2 | 6 | 10 |
| **y** | 3 | 7 | 11 |
| **z** | 4 | 8 | 12 |

Purwadhika
Startup and Coding School

# Creating a DataFrame

## From a list

```
In [7]: my_list= [[1,2,3,4],
                  [5,6,7,8],
                  [9,10,11,12]]
        pd.DataFrame(my_list,
                     index='a b c'.split(),
                      columns= 'w x y z'.split())
```

Out[7]:

| | w | x | y | z |
|---|---|---|---|---|
| **a** | 1 | 2 | 3 | 4 |
| **b** | 5 | 6 | 7 | 8 |
| **c** | 9 | 10 | 11 | 12 |

```
In [13]: pd.DataFrame(my_list,
                      index='a b c'.split(),
                       columns= 'w x y z'.split()).T
```

Out[13]:

| | a | b | c |
|---|---|---|---|
| **w** | 1 | 5 | 9 |
| **x** | 2 | 6 | 10 |
| **y** | 3 | 7 | 11 |
| **z** | 4 | 8 | 12 |

Purwadhika
Startup and Coding School

# Creating a DataFrame

## From an array

In [2]:
```
from numpy.random import randn
np.random.seed(101)
```

In [3]:
```
df = pd.DataFrame(randn(5,4),index='A B C D E'.split(),columns='W X Y Z'.split())
```

In [4]:
```
df
```

Out[4]:

| | W | X | Y | Z |
|---|---|---|---|---|
| **A** | 2.706850 | 0.628133 | 0.907969 | 0.503826 |
| **B** | 0.651118 | -0.319318 | -0.848077 | 0.605965 |
| **C** | -2.018168 | 0.740122 | 0.528813 | -0.589001 |
| **D** | 0.188695 | -0.758872 | -0.933237 | 0.955057 |
| **E** | 0.190794 | 1.978757 | 2.605967 | 0.683509 |

# Creating a DataFrame

## From a dictionary

In [35]:
```
my_d={'a':[1,2,3,4],
      'b':[5,6,7,8],
      'c':[9,10,11,12]}
pd.DataFrame(my_d,
             index= 'w x y z'.split())
```

Out[35]:

| | a | b | c |
|---|---|---|---|
| **w** | 1 | 5 | 9 |
| **x** | 2 | 6 | 10 |
| **y** | 3 | 7 | 11 |
| **z** | 4 | 8 | 12 |

Purwadhika
Startup and Coding School

# References


- Pandas Basics. https://www.learnpython.org/en/Pandas_Basics
- Python|Pandas DataFrame. https://www.geeksforgeeks.org/python-pandas-dataframe/
- Python|Pandas Series. https://www.geeksforgeeks.org/python-pandas-series/


Purwadhika
Startup and Coding School