

Session3-Pandas (Series)

Pear Deck Session
Training Clarusway
Pear Deck - November 30, 2022 at 7:37PM

Part 1 - Summary

Use this space to summarize your thoughts on the lesson

Part 2 - Responses

Slide 1

Data Analysis with Python

Session-3

CLARUSWAY®

WAY TO REINVENT YOURSELF

Use this space to take notes:

Slide 2



Use this space to take notes:

Slide 3

▶ Table of Contents

- ▶ Introduction to Pandas
- ▶ Pandas Series
 - What is Pandas Series
 - Creating Pandas Series
 - Basic Methods & Attributes
 - Indexing & Selection Pandas Series



Use this space to take notes:

Slide 4

Your Response

Slide 4

Your Response

I've completed the pre-class content?

True **False**

 Pear Deck

 Students choose an option

Pear Deck Interactive Slide
Do not remove this bar

Use this space to take notes:

Slide 5

► Introduction

What is Pandas?



- ▶ Free software library written for the Python programming language for **data manipulation and analysis**.
- ▶ Offers data structures & operations for **manipulating numerical tables & time series**.
- ▶ Used for **machine learning** in form of DataFrames.
- ▶ Allows **importing data** of various file formats such as csv, excel etc.

5

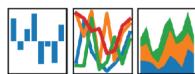
Use this space to take notes:

Slide 6

► Introduction

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



	BandName	WavelengthMax	WavelengthMin
0	CoastalAerosol	450	430
1	Blue	510	450
2	Green	590	530
3	Red	670	640
4	Nearinfrared	880	850
5	ShortWavelength_1	1650	1570
6	ShortWavelength_2	2290	2110
7	Orus	1360	1360

- ▶ You can think of Pandas as an extremely powerful version of **Excel**, with a lot more features.

Use this space to take notes:

Slide 7

► Introduction

- ▶ In this section of the course, we'll cover:

- Series
- DataFrames
- GroupBy
- Useful Operations
- Handling with Missing Data
- Handling with Outliers
- Combining DataFrames
- Working with Text/Time Data
- Data IO and RegEx



Use this space to take notes:

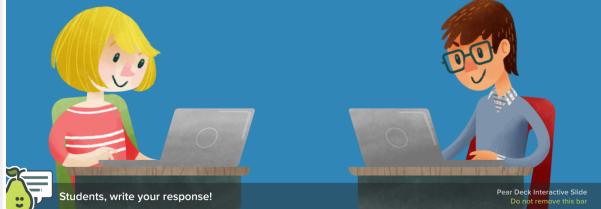
Slide 8

Your Response

Slide 8

Your Response

What are the differences and similarities between NumPy arrays and Pandas Series?



Use this space to take notes:

Slide 9

► Pandas Series

What is Pandas Series?



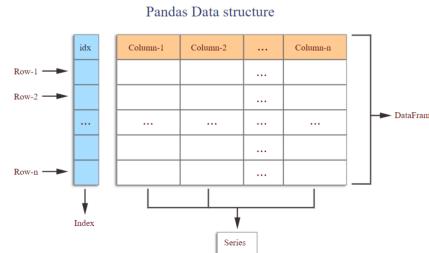
- ▶ Pandas series is a **one-dimensional labeled array** capable of **holding data of any type** (integer, string, float, python objects, etc.).
- ▶ You can think of the pandas series as a **column with labels** in an excel sheet.
- ▶ A Series is very **similar to a NumPy array** (in fact it is built on top of the NumPy array object).
- ▶ What differentiates the NumPy array from a Series, is that a **Series can have axis labels**, meaning it can be indexed by a label, instead of just a number location.



Use this space to take notes:

Slide 10

► Pandas Series



10

Use this space to take notes:

Slide 11

► Pandas Series

The diagram compares Jupyter format and Standard Python format for a Pandas DataFrame and Series.

Jupyter format: Shows a DataFrame with columns: YEAR, MODA, TEMP, MAX, MIN. The first few rows are: 0 20160601 65.5 73.6 54.7, 1 20160602 65.8 80.8 55.0, 2 20160603 68.4 77.0 55.6, 3 20160604 57.5 70.9 43.2, 4 20160605 51.4 58.3 43.2, 5 20160606 52.2 59.7 43.0, 6 20160607 56.9 65.1 45.9, 7 20160608 54.2 60.4 47.5, 8 20160609 49.4 54.1 45.7, 9 20160610 49.5 55.9 43.0.

Standard Python format: Shows a DataFrame with columns: YEAR, MODA, TEMP, MAX, MIN. The first few rows are: 0 20160601 65.5 73.6 54.7, 1 20160602 65.8 80.8 55.0, 2 20160603 68.4 77.0 55.6, 3 20160604 57.5 70.9 47.3, 4 20160605 51.4 58.3 43.2, 5 20160606 52.2 59.7 42.8, 6 20160607 56.9 65.1 45.9, 7 20160608 54.2 60.4 47.5, 8 20160609 49.4 54.1 45.7, 9 20160610 49.5 55.9 43.0.

Pandas DataFrame: pandas.core.frame.DataFrame

Pandas Series: pandas.core.series.Series

Annotations in the diagram point to specific elements:
- "Column labels" points to the column names in both formats.
- "Data" points to the numerical values in the DataFrame.
- "Index" points to the index numbers in the DataFrame.
- "Pandas Series" points to the Series object.
- "Index" points to the index numbers in the Series.
- "Data" points to the numerical values in the Series.
- "Column label" points to the column name in the Series.
- "Data type" points to the data type information in the Series.



11

Use this space to take notes:

Slide 12

► Pandas Series ➤

Series

	apples
0	3
1	2
2	0
3	1

Series

	oranges
0	0
1	3
2	7
3	2

DataFrame

	apples	oranges
0	3	0
1	2	3
2	0	7
3	1	2

Use this space to take notes:

Slide 13

► Pandas Series ➤

Creating Pandas Series

```
pandas.Series(data=None, index=None, dtype=None,  
              name=None, copy=False, fastpath=False)
```

- ▶ “data” parameter can be;
 - NumPy Array
 - List
 - Dictionary
 - Scalar value

Use this space to take notes:

Slide 14

► Pandas Series

Basic Methods & Attributes

- .dtype
- .size
- .ndim
- .head
- .tail
- .shape
- .sample
- .sort_index()
- .sort_values()
- .isin
- .index
- .keys()
- .values
- .items()



14

Use this space to take notes:

Slide 15

► Data Analysis with Python



let's start the
hands-on phase



15

Use this space to take notes:

Slide 16

Your Response

Slide 16

Did you find this lesson interesting and challenging?

Too hard Just right Too easy

Students, drag the icon!

Peer Deck Interactive Slide
Do not remove this bar

Your Response

Did you find this lesson interesting and challenging?

Too hard Just right Too easy

Students, drag the icon!

Peer Deck Interactive Slide
Do not remove this bar

Use this space to take notes:

Slide 17

THANKS!
Any questions?

You can find us at:
steve_w@clarusway.com
michael_g@clarusway.com

CLARUSWAY®
Leading to Success Worldwide



Use this space to take notes: