

Session4-Pandas (DataFrames)

DAwithPython S4

Training Clarusway

Pear Deck - April 28, 2022 at 8:53PM

Part 1 - Summary

Use this space to summarize your thoughts on the lesson

Part 2 - Responses

Slide 1



**Data Analysis with
Python**

Session-4

CLARUSWAY®
HELPING YOU REACH YOUR GOALS



Use this space to take notes:

Slide 2



CLARUSWAY®
WAY TO REINVENT YOURSELF



Use this space to take notes:

Slide 3

► Table of Contents

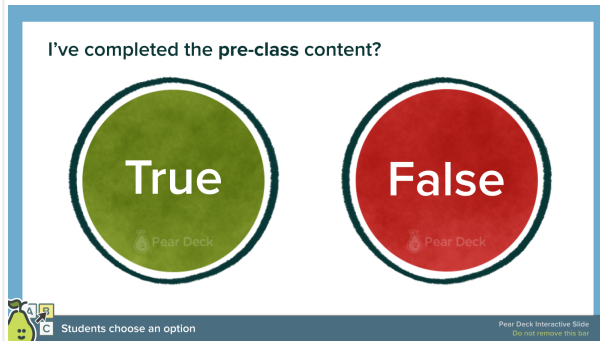
- What is Pandas DataFrames?
- Creating Pandas DataFrames
- Basic Methods & Attributes
- Indexing & Selection Pandas DataFrames
- Multi-Index & Index Hierarchy

Link(s) on this slide:

- [http://localhost:8888/notebooks/DS/Instructor_Files/DAwPython/Cohort011_new/DAwPy_S4_\(Pandas_DataFrames%2C_Selection_and_Indexing_w_loc%2Ciloc%2C%20multi\).ipynb#10](http://localhost:8888/notebooks/DS/Instructor_Files/DAwPython/Cohort011_new/DAwPy_S4_(Pandas_DataFrames%2C_Selection_and_Indexing_w_loc%2Ciloc%2C%20multi).ipynb#10)

Use this space to take notes:

Slide 4



Use this space to take notes:

Your Response

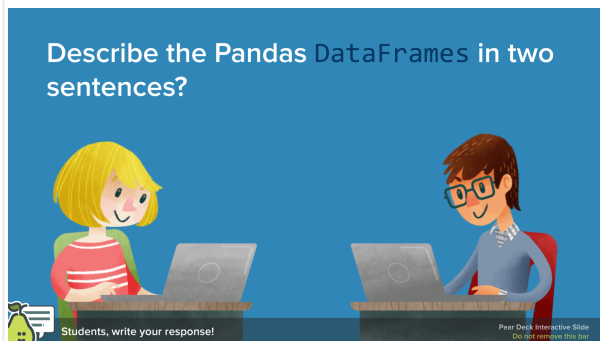
You Chose

- **True**

Other Choices

- False

Slide 5



Use this space to take notes:

Your Response

Answer 1:
two dimensional, contains heterogenous
data label index

Slide 6

► Pandas DataFrames



What is Pandas DataFrames?

- DataFrames are the **workhorse** of Pandas.
- We can think of a DataFrame as a **bunch of Series** objects put together to **share the same index**.
- A DataFrame is a **two-dimensional** data structure where data is aligned **in rows and columns**.
- Three principal components; the **data, rows, and columns** form the Pandas DataFrame.



Use this space to take notes:

Slide 7

► Pandas DataFrames



What is Pandas DataFrames?

	Column names								
	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0
1	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	NaN
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0	6-8	235.0	LSU	1170960.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0	6-2	190.0	Louisville	1824360.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN	6-9	260.0	Ohio State	2569260.0
6	Evan Turner	Boston Celtics	11.0	SG	27.0	6-7	220.0	Ohio State	3425510.0



Use this space to take notes:

Slide 8

Pandas DataFrames

Creating Pandas DataFrames

`pandas.DataFrame(data=None, index=None, columns=None, dtype=None, copy=None)`

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

Use this space to take notes:

Slide 9

Pandas DataFrames

Creating Pandas DataFrames

Row Oriented

Dictionary

```
sales = [{'account': 'Jones LLC', 'Jan': 150, 'Feb': 200, 'Mar': 140},
         {'account': 'Alpha Co', 'Jan': 200, 'Feb': 210, 'Mar': 215},
         {'account': 'Blue Inc', 'Jan': 50, 'Feb': 90, 'Mar': 95}]
df = pd.DataFrame(sales)
```

default

	account	Jan	Feb	Mar
0	Jones LLC	150	200	140
1	Alpha Co	200	210	215
2	Blue Inc	50	90	95

from_records

List

```
sales = [['Jones LLC', 150, 200, 50],
         ['Alpha Co', 200, 210, 90],
         ['Blue Inc', 140, 215, 95]]
labels = ['account', 'Jan', 'Feb', 'Mar']
df = pd.DataFrame.from_records(sales, columns=labels)
```

from_records

Column Oriented

from_dict

```
sales = {'account': ['Jones LLC', 'Alpha Co', 'Blue Inc'],
         'Jan': [150, 200, 50],
         'Feb': [200, 210, 90],
         'Mar': [140, 215, 95]}
df = pd.DataFrame.from_dict(sales)
```

from_items

```
sales = [{'account': 'Jones LLC', 'Alpha Co', 'Blue Inc'},
         {'Jan': [150, 200, 50]},
         {'Feb': [200, 210, 90]},
         {'Mar': [140, 215, 95]}]
df = pd.DataFrame.from_items(sales)
```

Use this space to take notes:

Slide 10

▶ Pandas DataFrames

Basic Methods & Attributes

- .dtype
- .size
- .ndim
- .head
- .tail
- .shape
- .sample

- .sort_index()
- .sort_values()
- .isin
- .index
- .keys()
- .values
- .items()

»

10

Use this space to take notes:

Slide 11

▶ Pandas DataFrames

Basic Methods & Attributes

- .columns
- .reset_index
- .set_index()
- .iloc[]
- .loc[]
- .rename()

- .info()
- .describe()
- .value_counts()
- .unique()
- .nunique()
- .drop()

»

11

Use this space to take notes:

Slide 12	Your Response

Draw lines to match the attributes/methods to their definitions:

df.values

df.shape

df.drop

Return a Numpy representation of the DataFrame.

Return an int representing the number of axes / array dimensions.

Return an int representing the number of elements in this object.

Return a tuple representing the dimensionality of the DataFrame.

Return the first n rows.

Drop specified labels from rows or columns.

df.head

df.ndim

df.size

Draw lines to match the attributes/methods to their definitions:

df.values

df.shape

df.drop

Return a Numpy representation of the DataFrame.

Return an int representing the number of axes / array dimensions.

Return an int representing the number of elements in this object.

Return a tuple representing the dimensionality of the DataFrame.

Return the first n rows.

Drop specified labels from rows or columns.

df.head

df.ndim

df.size

Use this space to take notes:

Slide 13


▶

Data Analysis with Python

jupyter

python

CO



let's start the hands-on phase

13

Use this space to take notes:

Slide 14	Your Response

Did you find this lesson interesting and challenging?

Too hard Just right Too easy

Students, drag the icon!

Pear Deck Interactive Slide
Do not remove this bar

Did you find this lesson interesting and challenging?

Too hard Just right Too easy

Students, drag the icon!

Pear Deck Interactive Slide
Do not remove this bar

Use this space to take notes:

Slide 15

THANKS!

Any questions?

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

CLARUSWAY®
HELPING LEADERSHIP FLOURISH

Use this space to take notes: