Session4-Pandas (DataFrames)

DAwithPython S4
Training Clarusway
Pear Deck - April 28, 2022 at 8:53PM

Part 1 - Summary

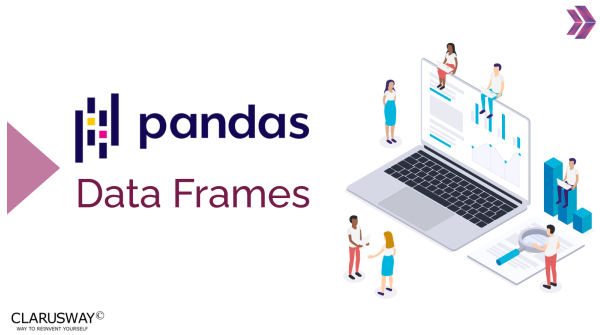
Use this space to summarize your thoughts on the lesson

Part 2 - Responses

Slide 1

Use this space to take notes:

Slide 2



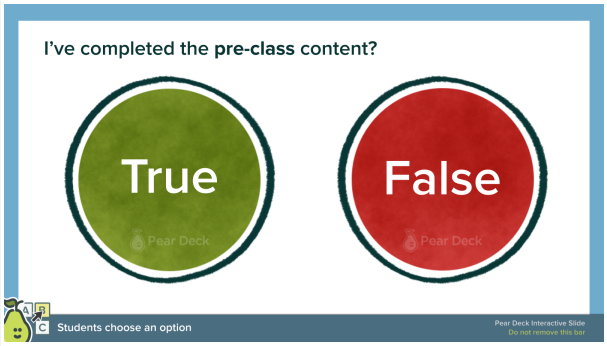
Use this space to take notes:

Slide 3

Table of Contents

- 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

Use this space to take notes:

Slide 4	Your Response
	You Chose • **True** Other Choices • False

Use this space to take notes:

Slide 5	Your Response
	**Answer 1:** two dimensional, contains heterogenous data label index

Use this space to take notes:

Slide 6

Pandas DataFrames

What is Pandas DataFrames?

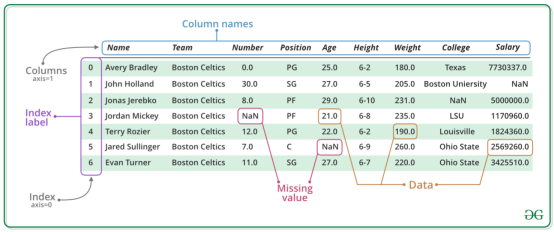
- DataFrames are the **workhorse** of Pandas.
- We can think of a DataFrame as a **bunch of Series** objects put together to **share the same index**.
- A DataFrame is a **two-dimensional** data structure where data is aligned **in rows and columns**.
- Three principal components; the **data, rows, and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

Use this space to take notes:

Slide 7

Pandas DataFrames

What is Pandas DataFrames?



	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 Uniersity	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

Dictionary | **List**

Row Oriented

```
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

```
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

```
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)
```

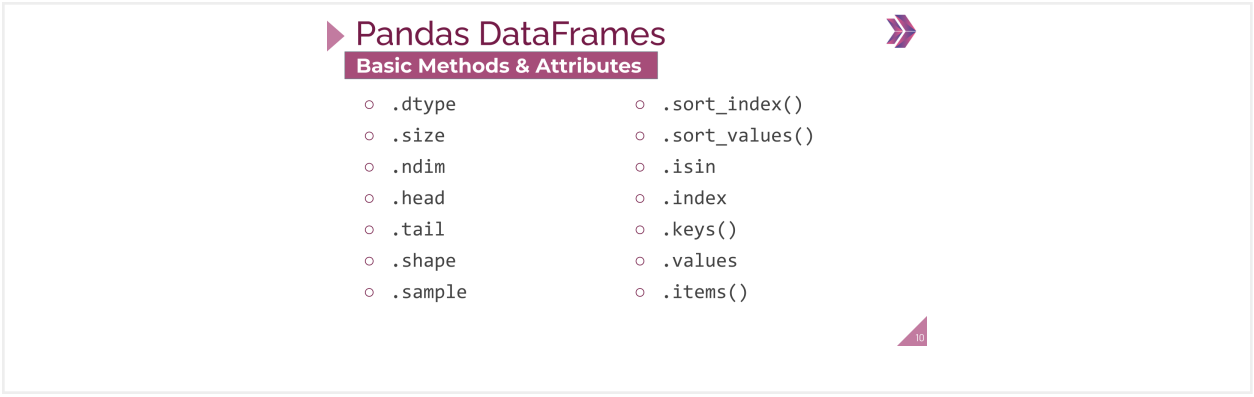
from_dict

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

from_items

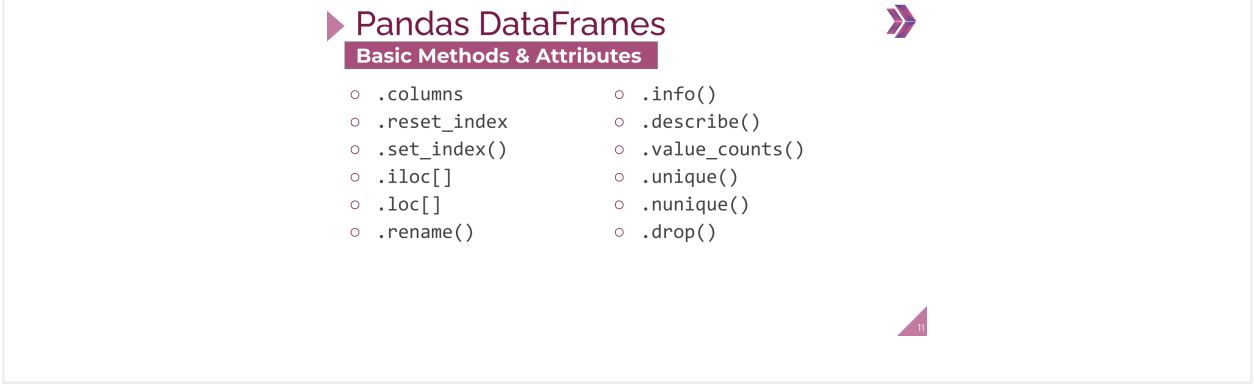
Use this space to take notes:

Slide 10



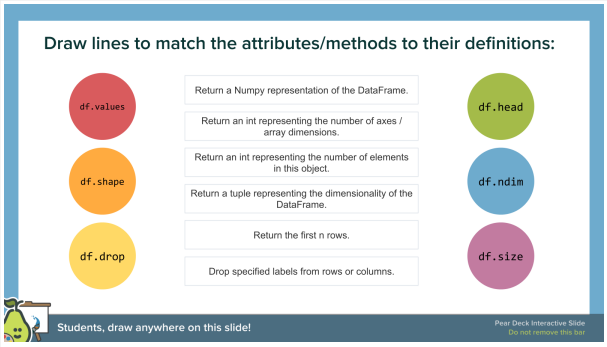
Use this space to take notes:

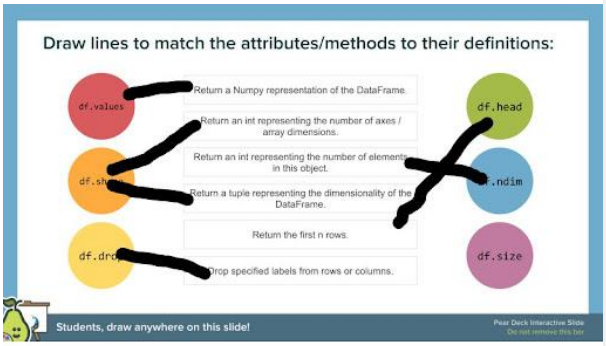
Slide 11



Use this space to take notes:

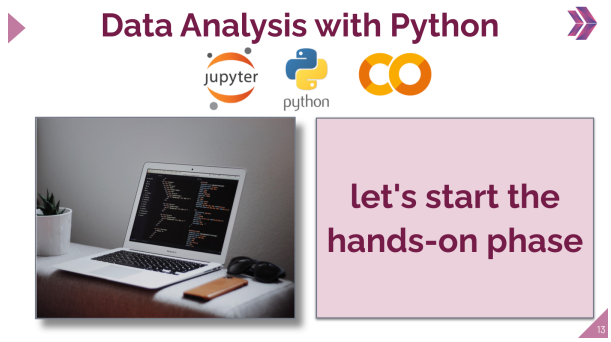
Slide 12	Your Response





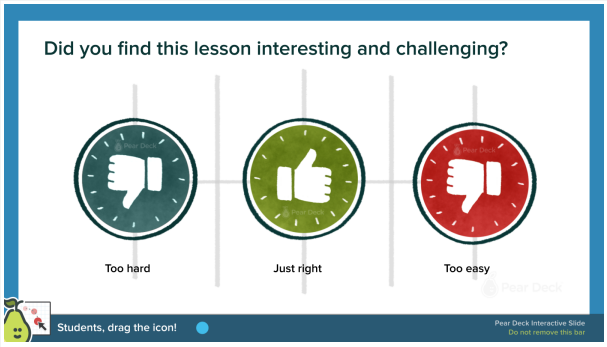
Use this space to take notes:

Slide 13



Use this space to take notes:

Slide 14	Your Response



Use this space to take notes:

Slide 15

Use this space to take notes: