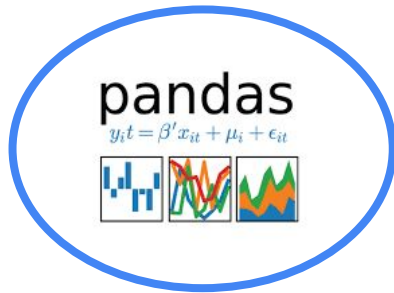


Pandas

04.12.19 / Us (again)



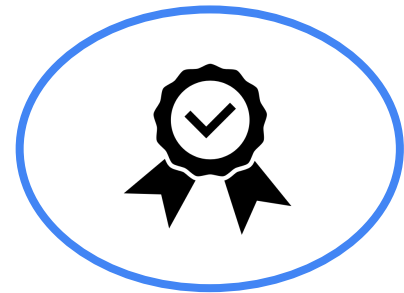
Today's Focus



Pandas



Matplotlib II



Challenges

Pandas



- Python library designed for data manipulation and analysis
- High-performance, easy-to-use data structures: DataFrames and Series
- Objective: Carry out entire data analysis workflow in Python (no need for domain specific languages like R!)
- Really easy to import from and export to Excel spreadsheets

What is a DataFrame?

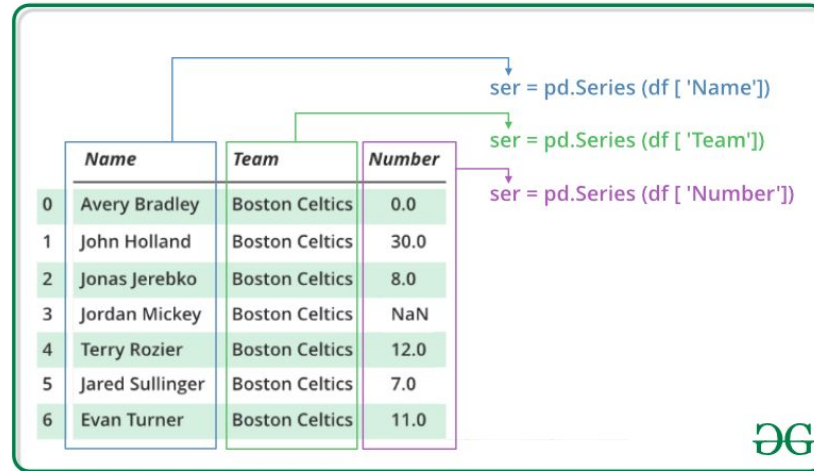
- Python object: 2d table (like a spreadsheet with column names and row labels)

The diagram illustrates a DataFrame as a 2D table. It features a table with 6 rows and 5 columns. The columns are labeled 'Name', 'Team', 'Number', 'Position', and 'Age'. The rows are indexed from 0 to 6. Annotations include: 'Columns' with arrows pointing to the column headers; 'Rows' with arrows pointing to the row indices; and 'Data' with a box highlighting a subset of cells (Jonas Jerebko, Boston Celtics, 8.0, PF, 29.0) and a line pointing to the 'Data' label.

	Name	Team	Number	Position	Age
0	Avery Bradley	Boston Celtics	0.0	PG	25.0
1	John Holland	Boston Celtics	30.0	SG	27.0
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN
6	Evan Turner	Boston Celtics	11.0	SG	27.0

What is a Series?

- Python object: 1d array (like a single spreadsheet column)



Pandas vs NumPy

- Pandas provides us with some powerful objects like DataFrames and Series
- NumPy provides multi-dimensional arrays, DataFrames are 2d tables
- Pandas works when data is in tabular format
- NumPy works when data is numeric

There is no best alternative, which one you should use depends on the use case
(sometimes you even need to use both)

**Let's jump to the Jupyter
Notebook, see you there!!**

Google Drive link with files:
<https://tinyurl.com/sr3qbnx>

Thanks!

Python team

Wiki:

<https://wiki.tum.de/display/ldv/Info>

Mail:

pythonworkshop.tum@gmail.com

Web:

<https://www.ei.tum.de/startseite/>

Git:

<https://gitlab.ldv.ei.tum.de/daedalus/python>

