

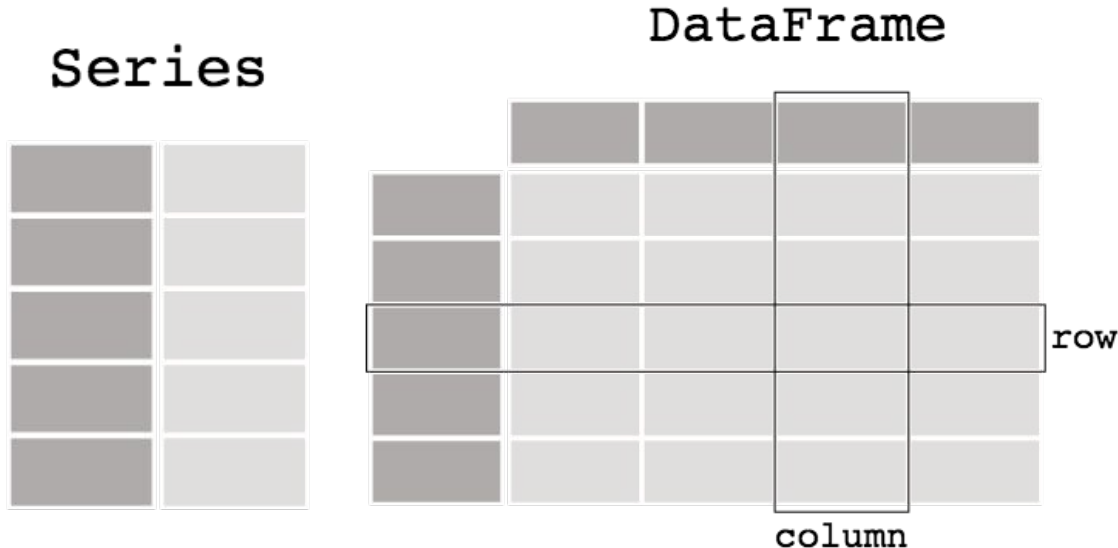
Pandas

What is Pandas?

The **Pandas library** is built on NumPy and provides easy-to-use data structures and data analysis tools for the Python programming language. It is used to handle tabular data, such as data stored in spreadsheets or databases. pandas will help you to explore, clean, and process your data.

Pandas data structures:

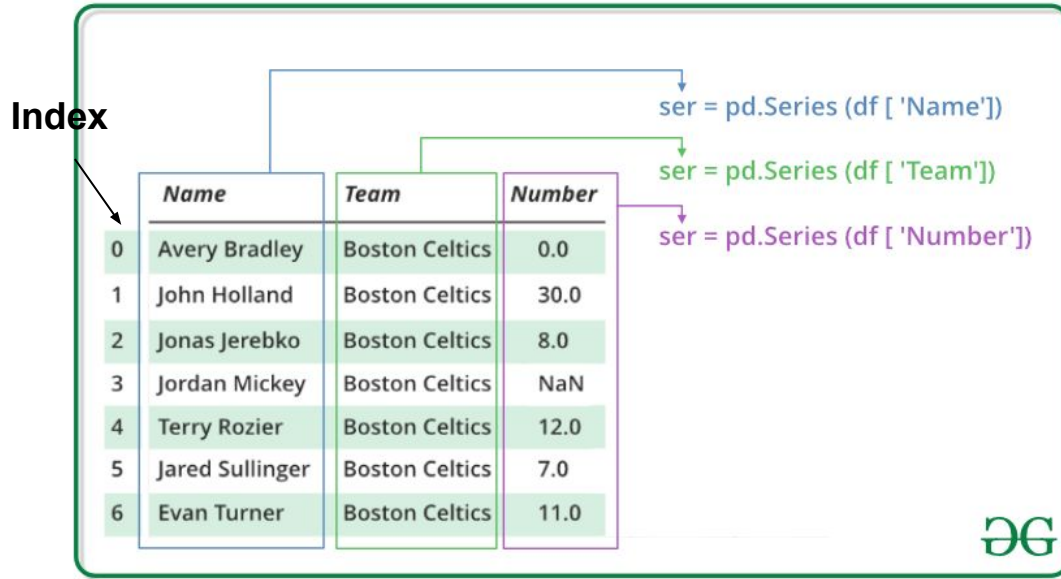
- Series
- DataFrame



Pandas data structures - Series

Pandas Series is a **one-dimensional labeled array** capable of holding data of any type (integer, string, float, python objects, etc.). The axis labels are collectively called *index*.

Pandas Series is nothing but a column in an excel sheet.



Pandas data structures - DataFrame

Pandas DataFrame is a two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns).

Pandas DataFrame consists of three principal components, the data, rows, and columns.

The diagram illustrates the components of a Pandas DataFrame. It shows a table with 7 rows and 5 columns. The columns are labeled 'Name', 'Team', 'Number', 'Position', and 'Age'. The rows are indexed from 0 to 6. The 'Index' label points to the row numbers. The 'Columns' label points to the column headers. The 'Rows' label points to the row indices. The 'Data' label points to the data cells. The table content is as follows:

	<i>Name</i>	<i>Team</i>	<i>Number</i>	<i>Position</i>	<i>Age</i>
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

The diagram also includes a logo in the bottom right corner.



NumPy

What is NumPy?

NumPy stands for Numerical Python. NumPy is a Python library used for working with **arrays**.

It also has functions for working in domain of linear algebra, fourier transform, and matrices.

Working on NumPy arrays is 50x faster than looping through lists.

The NumPy array

NumPy's main object is the homogeneous multidimensional **array**. It is a table of elements (usually numbers), all of the same type, indexed by a tuple of non-negative integers. In NumPy dimensions are called *axes*.

This is a 2x3 array.

```
[[1., 0., 0.],  
 [0., 1., 2.]]
```