



# Chapter.02 데이터 분석 라이브러리

## 리-05. Pandas DataFrame



Pandas DataFrame : pandas 라이브러리가 사용하는 기본 자료구조.

	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

Source : <https://www.geeksforgeeks.org/creating-a-pandas-dataframe/>

- DataFrame은 2차원 테이블 구조를 말합니다.
- 1차원 구조인 **Series** 도 있습니다. (1 row, 1 column)
- row, column으로 모든 원소를 구분합니다. (indexing)
- index, columns, values라는 객체 변수를 가지고 있습니다.
- Relational DB와 완전히 호환됩니다.
- 하나의 column을 기준으로 모든 원소의 data type이 동일합니다. (모두 numpy array가 가지는 data type과 동일)

- DataFrame은 numpy array를 상위 호환하는 개념으로 universal function이 사용 가능합니다.  
→ 내부 구현체로 numpy array를 사용하기 때문에!