

파이썬을 활용한 데이터 전처리

1 pandas 란?

Pandas 란?

1.

Pandas 란?



데이터 조작과 분석 위한
Python 소프트웨어 라이브러리

“표(행렬) 형태의 데이터를 조작하는데 매우 특화됨”

“대용량 데이터 셋을 이용한 다양한 통계 처리 기능 제공”

“엑셀이 제공하는 거의 모든 기능을 구현 가능!”

“데이터 베이스 (SQL) 문도 손쉽게 구현 가능!”

Pandas 란?

1.
Pandas 란?



?

Pandas 란?

1.
Pandas 란?



?

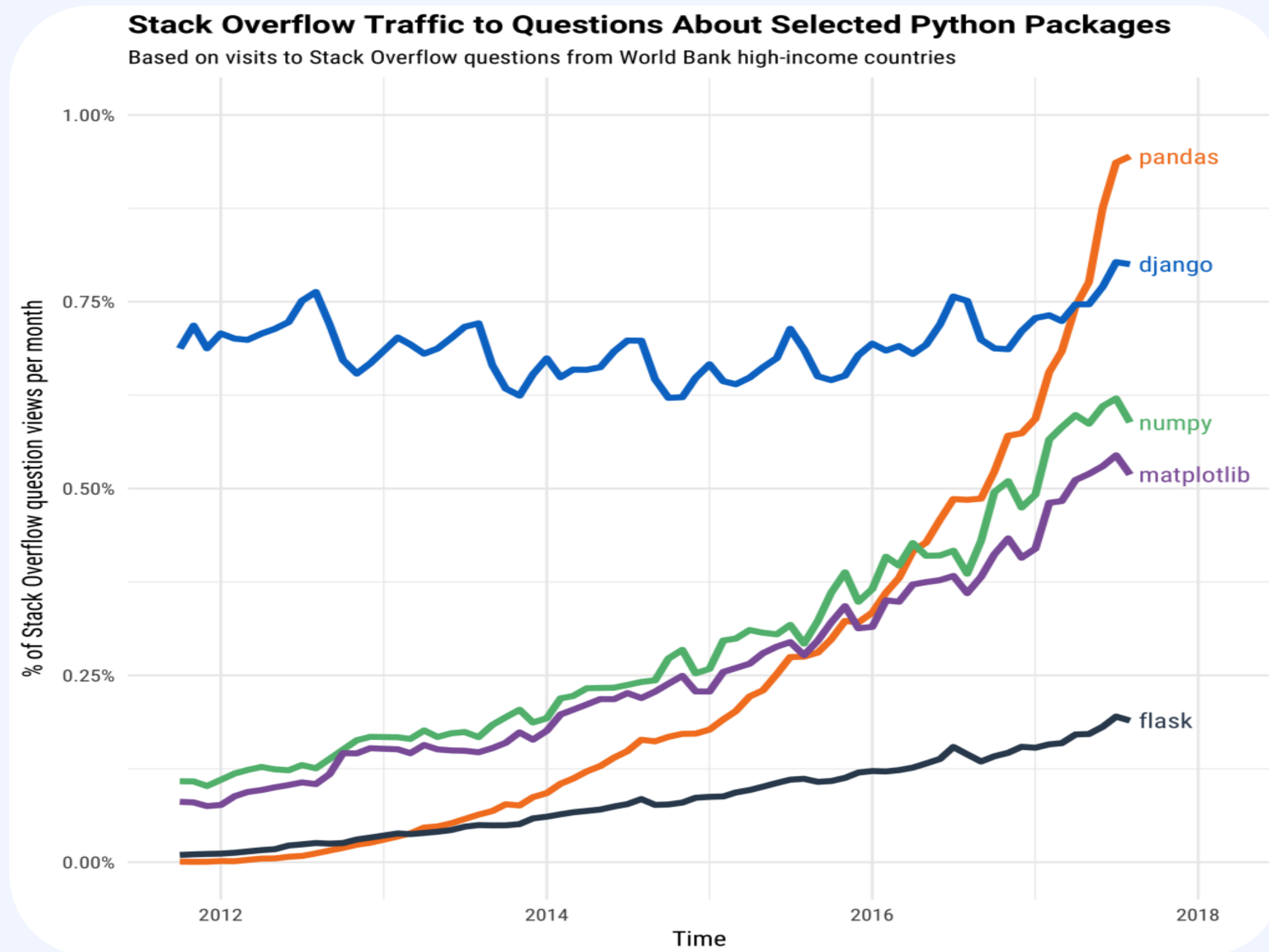
Panel Data System
(=Python Data Analysis)

Pandas 사용량 추이

1.

Pandas 란?

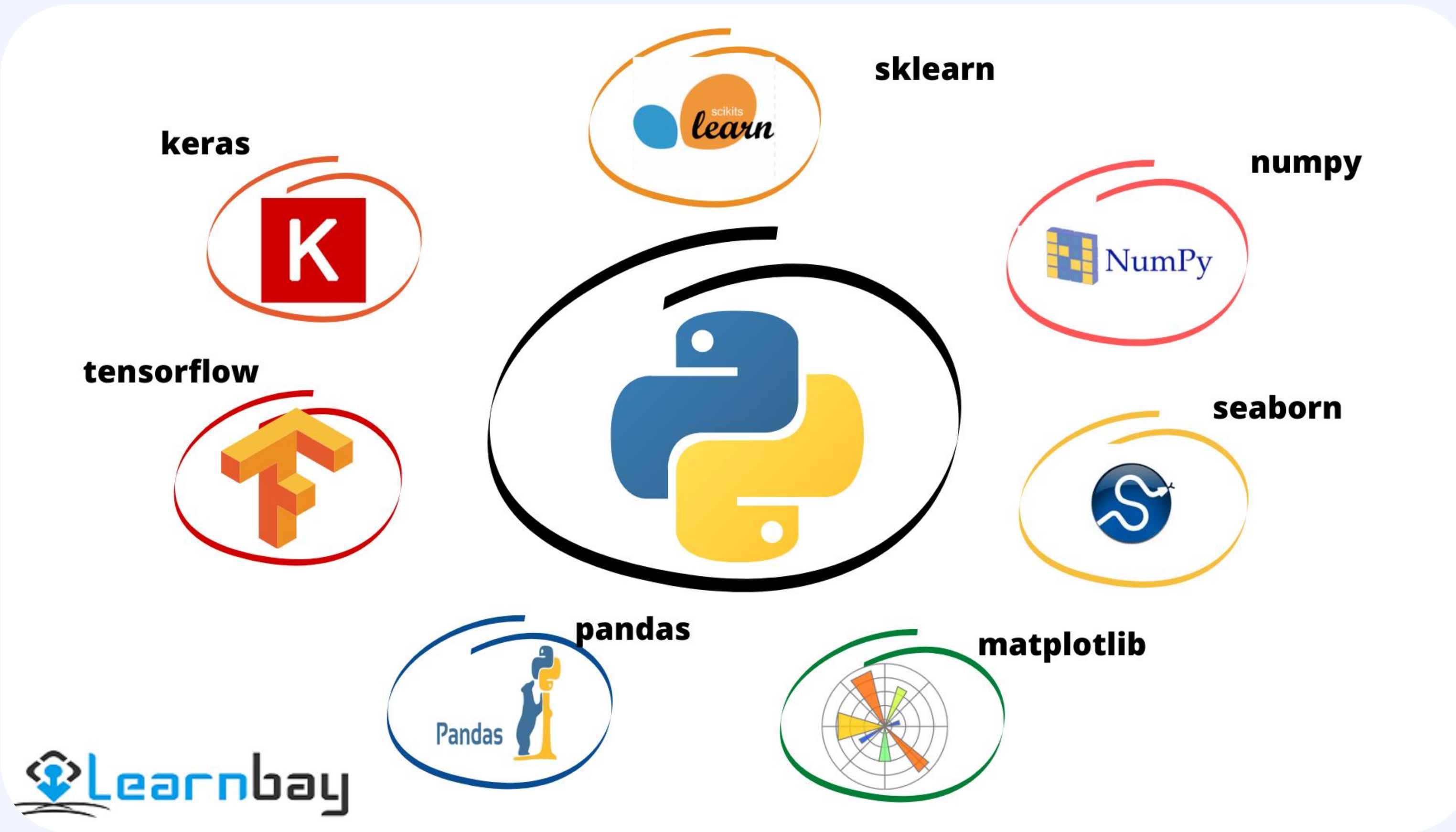
Python 으로 데이터 분석 하시는 분들 중,
Pandas 를 사용하지 않는 분은 없으실 겁니다.



Python 핵심 라이브러리

1.

Pandas 란?



Pandas 창시자

1.

Pandas 란?



Wes McKinney (1985~)

금융회사에서 일하다
간단한 작업도 Excel, SQL 로
하루 종일 씨름하는 것이
답답해서 만들었습니다.

필요한 것이 있으면 **직접**
만들어 버리세요!

*Pandas 개발 당시 Java 초급 코스만 수료

Pandas 데이터

1.
Pandas 란?

Pandas 에서 사용되는 데이터 유형

시리즈 (Series)

| | | |
|-------------|--|--|
| 인덱스 (Index) | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

1차원 배열

데이터 프레임 (DataFrame)

| | | | | |
|-------------|--|-------------|--|--|
| 인덱스 (Index) | | 컬럼 (Column) | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

2차원 배열

Pandas 데이터

1.

Pandas 란?

엑셀로 작성한 파일이 호환 됩니다

| | A | B | C | D | E |
|---|----|----|-----|----|---|
| 1 | ID | 국어 | 영어 | 수학 | |
| 2 | 1 | 80 | 85 | 75 | |
| 3 | 2 | 90 | 100 | 95 | |
| 4 | 3 | 75 | 70 | 65 | |
| 5 | | | | | |
| 6 | | | | | |
| 7 | | | | | |

Test.xlsx

```
In [5]: import pandas as pd
df = pd.read_excel('C:\Temp\Test.xlsx')
```

```
In [6]: df
```

```
Out[6]:
```

| | ID | 국어 | 영어 | 수학 |
|---|----|----|-----|----|
| 0 | 1 | 80 | 85 | 75 |
| 1 | 2 | 90 | 100 | 95 |
| 2 | 3 | 75 | 70 | 65 |

Pandas 로 할 수 있는 일

1.

Pandas 란?

엑셀로 하는 일 (전처리)



데이터 베이스 (SQL)

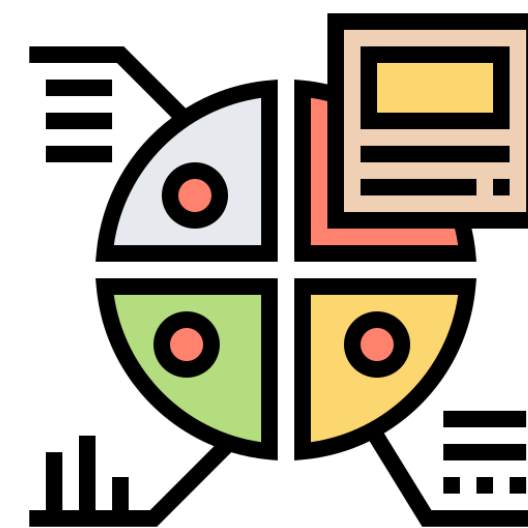


크롤링 (웹 수집)



Pandas 로
할 수 있는 일

시각화



감사합니다.