

CONTENTS /

Pandas 라이브러리 소개

#PROFESSIONAL IT EDU-PLATFORM

code.presso()

Python 기반 라이브러리

#PROFESSIONAL IT EDU-PLATFORM
code.presso()

데이터 수집 및 분석



데이터 시각화

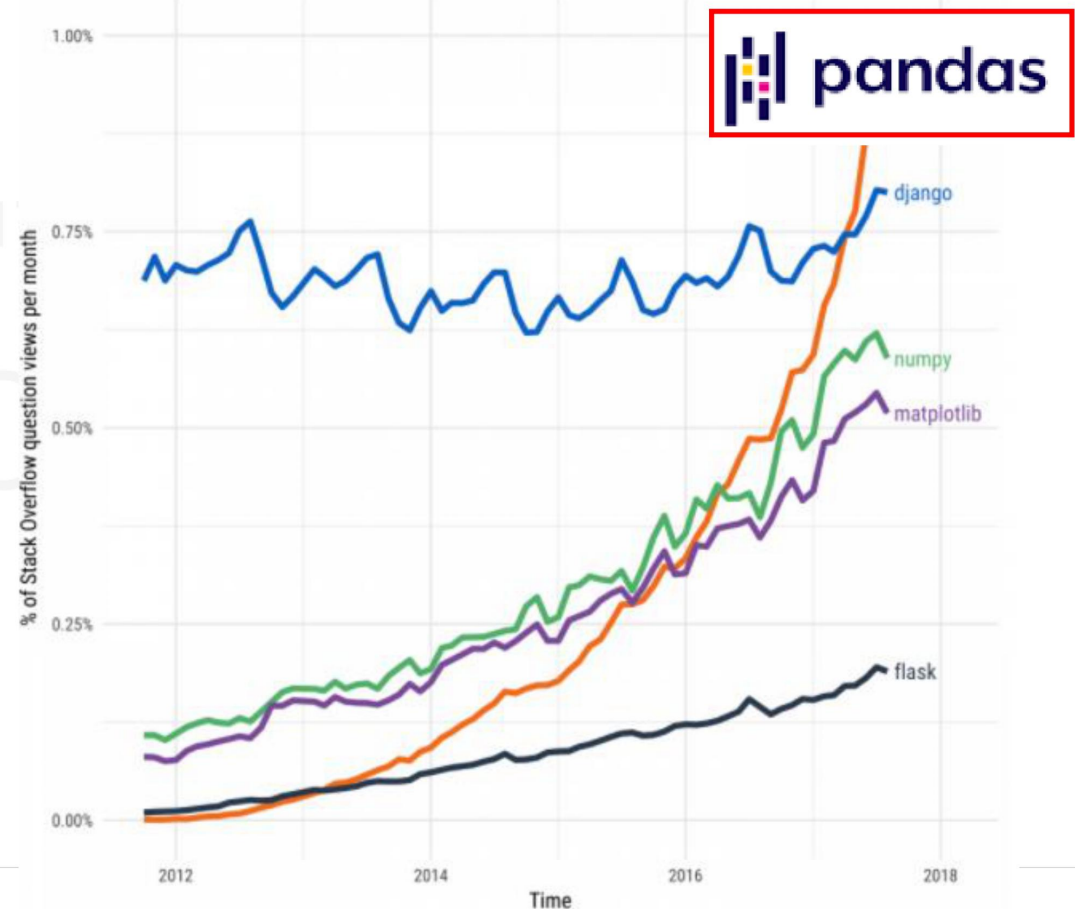


웹 어플리케이션 개발



Stack Overflow Traffic to Questions About Selected Python Packages

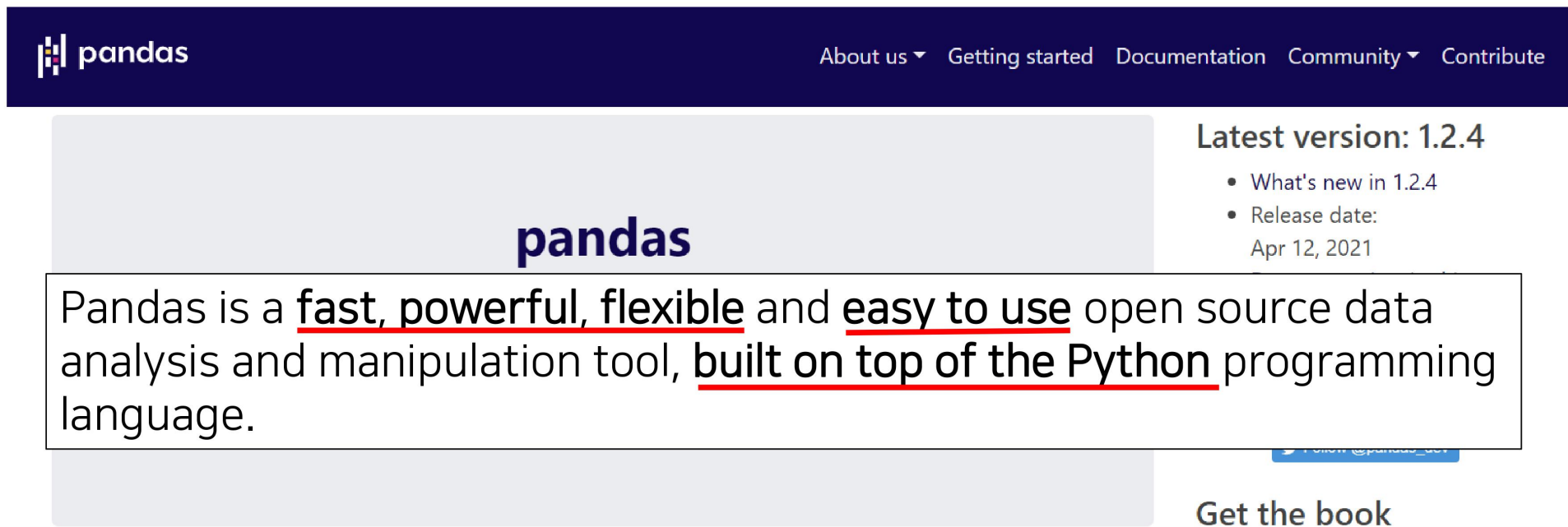
Based on visits to Stack Overflow questions from World Bank high-income countries



| Pandas 라이브러리 소개

- 데이터 분석을 위한 파이썬 기반의 라이브러리

<https://pandas.pydata.org> ✓



The screenshot shows the pandas website homepage. At the top is a dark blue navigation bar with the pandas logo on the left and links for 'About us', 'Getting started', 'Documentation', 'Community', and 'Contribute' on the right. Below the navigation bar, the word 'pandas' is displayed in a large, bold, dark blue font. To the right of the logo, the text 'Latest version: 1.2.4' is shown, followed by a bulleted list: 'What's new in 1.2.4' and 'Release date: Apr 12, 2021'. A white text box with a black border contains the description: 'Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.' Below this text box, a blue button with the text 'Get the book' is visible.

pandas

About us ▾ Getting started Documentation Community ▾ Contribute

Latest version: 1.2.4

- What's new in 1.2.4
- Release date: Apr 12, 2021

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Get the book

| Pandas 라이브러리 소개

- 데이터 분석을 위한 파이썬 기반의 라이브러리

<https://pandas.pydata.org>

Pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

- ✓ 2차원 데이터를 효율적으로 가공 및 처리할 수 있는 강력한 라이브러리

| 2차원 데이터?

'titanic' 데이터셋

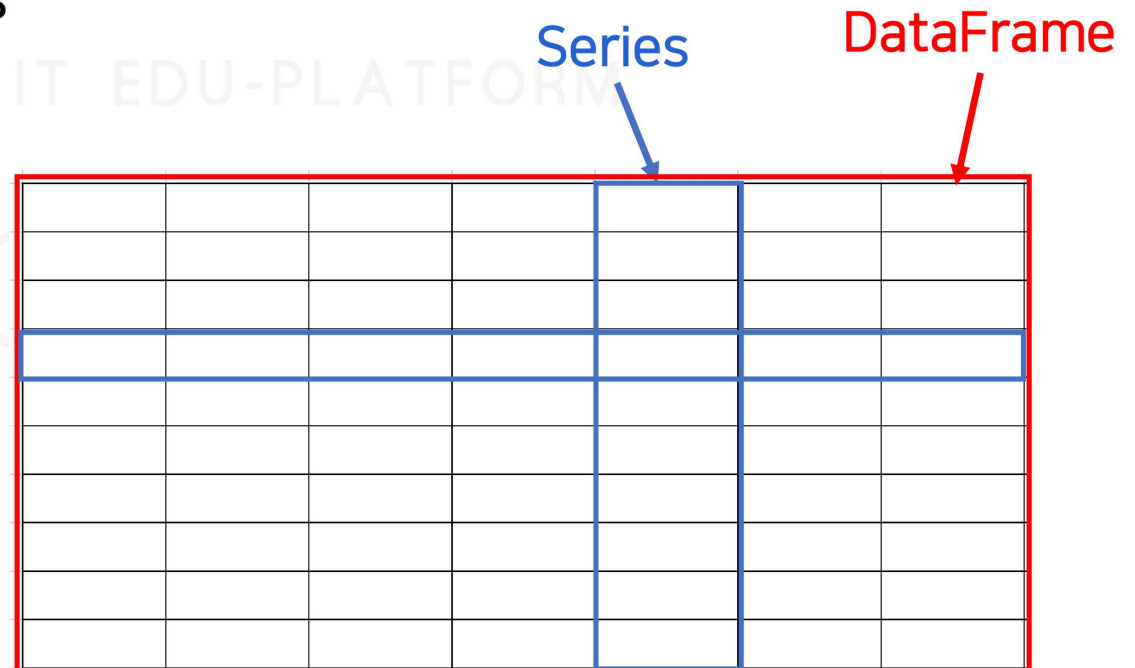
| survived | pclass | sex | age | sibsp | parch | fare | embarked | class | who | adult | male | deck | embark | town | alive | alone |
|----------|--------|--------|------|-------|-------|------|----------|-------|-----|-------|------|------|--------|------|-------|-------|
| 0 | 3 | male | 22.0 | 1 | | | | | | | | | | | | |
| 1 | 1 | female | 38.0 | 1 | | | | | | | | | | | | |

Iris 품종 예측 데이터셋

자전거 수요 예측 데이터셋

| time | season | holiday | workingday | weather | temp | atemp | humidity | windspeed | casual | registered | count |
|-------|--------|---------|------------|---------|------|--------|----------|-----------|--------|------------|-------|
| 00:00 | 1 | 0 | 0 | 1 | 9.84 | 14.395 | 81 | 0.0000 | 3 | 13 | 16 |
| 00:00 | 1 | 0 | 0 | 1 | 9.02 | 13.635 | 80 | 0.0000 | 8 | 32 | 40 |
| 00:00 | 1 | 0 | 0 | 1 | 9.02 | 13.635 | 80 | 0.0000 | 5 | 27 | 32 |
| 00:00 | 1 | 0 | 0 | 1 | 9.84 | 14.395 | 75 | 0.0000 | 3 | 10 | 13 |
| 00:00 | 1 | 0 | 0 | 1 | 9.84 | 14.395 | 75 | 0.0000 | 0 | 1 | 1 |
| 00:00 | 1 | 0 | 0 | 2 | 9.84 | 12.880 | 75 | 6.0032 | 0 | 1 | 1 |
| 00:00 | 1 | 0 | 0 | 1 | 9.02 | 13.635 | 80 | 0.0000 | 2 | 0 | 2 |
| | | | | | | 6.3 | | | 3.3 | | 6.0 |

- 자료구조란 데이터를 저장하는 방식 또는 구조를 의미함
 - Python : 리스트, 튜플, 딕셔너리 등
 - NumPy : ndarray
 - Pandas 는?



| Pandas 를 배우다는 것은?

- Pandas 라는 데이터 분석을 위한 도구의 사용법을 배우는 과정
- 분석하고자 하는 데이터를 Series, DataFrame 에 저장
- 자료구조를 필요에 따라 추가, 수정, 삭제, 전처리 할 수 있고,
- 자료구조에 저장된 데이터를 조회, 정렬, 연산할 수 있고,
- 다양한 데이터들을 원하는 뷰로 결합하여 활용할 수 있는
- 2차원 데이터로부터 의미를 도출하는 전과정에 필수적인 라이브러리를 학습하는 과정

- Pandas 실습 환경 (Pandas 이미 설치되어 있음)



- Pandas 라이브러리 설치
 - > conda install pandas
 - > pip install pandas