

파이썬과 쥬피터 노트북 설치 및 판다스 개요



# 파이썬 설치





1991년 귀도 반 로썸(Guido van Rossum)이 만든 배우기 쉽고 강력한 프로그래밍 언어임

- 효율적인 고수준 데이터 구조
- 효과적인 객체 지향 프로그래밍 접근법
- RAD(Rapid Application Development)

### 파이썬 설치

- ₱ Python 특징
  - Simple and easy to learn
  - Free and open source
  - High-level language
  - Portable
    - ◆ 다양한 플랫폼 지원(Windows, Linux, OS X, etc.)
  - Interpreted
    - ◆ 인터프리터 언어
  - Objected oriented
  - Extensible
  - Embeddable
  - Extensive libraries

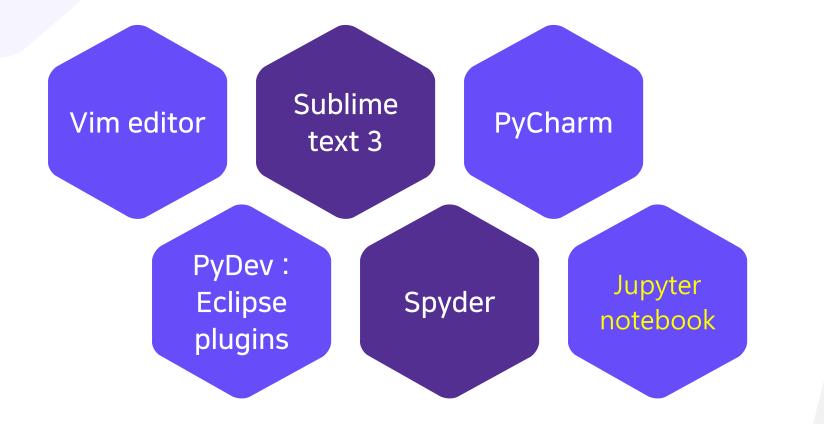
# 파이썬 설치



🕮 Python 개발 환경

Python IDE

- Integrated Development Environment
- 통합개발환경





 과목명
 데이터시각화
 주차명
 02. 데이터 프레임과 Subset Observation
 페이지번호
 02\_02\_02

#### 학습목차

#### 들어가기

- 인트로
- 학습개요

#### 파이썬과 쥬피터 노트북 설치 및 판다스 개요

- 파이썬 설치
- 쥬피터 노트북 설치
- 판다스 개요

#### 시리즈와 데이터프레임

- 판다스 시리즈와 데이터프레임
- 판다스 시리즈
- 판다스 데이터프레임

#### **Subset Observations**

- Subset\_observation(Rows)( 실습)
- Logic 연산(실습)
- Head, Tail 데이터 미리보기(실습)
- 일부 컬럼 기준 데이터 가져오기(실습)

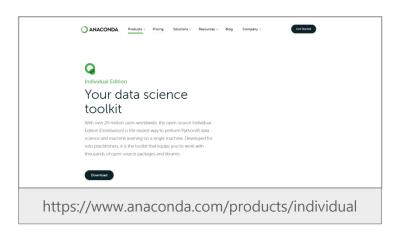
#### 파이썬 공식 홈페이지

이미지를 클릭하여 파이썬 공식 홈페이지를 방문해보세요.



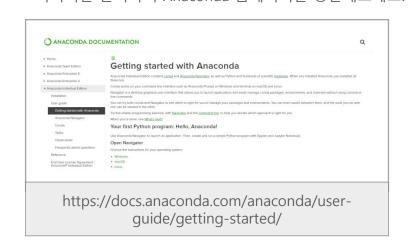
#### Anaconda 홈페이지

이미지를 클릭하여 Anaconda 홈페이지를 방문해보세요.



#### Getting started with Anaconda 페이지

이미지를 클릭하여 Anaconda 홈페이지를 방문해보세요.



#### 화면설명

#### [영상 페이지]

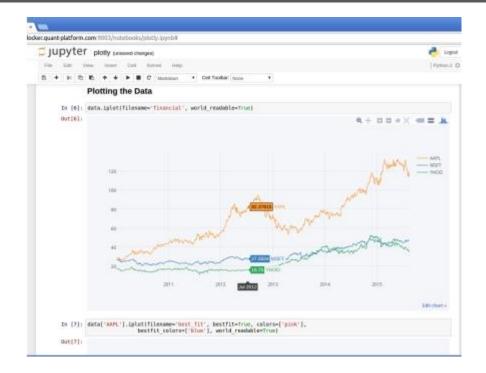
- ■교수님 판서크로마, 전자칠판 동영상 제시
- ■요약 노트 삽입
- ➤ 각 요약 노트는 페이지 메이커의 '게시판 > 글 제출' 컴포넌트로 삽입하며, 게시판별 고유 아이디 부여 (교과목코드(8자리)\_ 개발연도학기(4자리( 연도2자리+학기2자 리))\_주차(2자리)\_생 성순서(1부터 순차적으로))
- ■이미지 클릭 시 홈페이지 이동
- ➤ 홈페이지 url : https://www.python.o rg/
- ➤ 홈페이지 url: https://www.anacond a.com/products/indivi dual
- ➤ 홈페이지 url:

  https://docs.anaconda
  .com/anaconda/userguide/getting-started/
- 파이썬 공식, Anaconda 홈페이지 개발 시 h2로 제목 삽입

# Jupyter notebook

Web application that allow you to create and share documents that contain live code, equations, visualizations and explanatory text

Can do interactive data science via Jupyter notebook!





#### 파이썬 공식 홈페이지에서 다운로드하여 설치

➡ 파이썬 공식 홈페이지 : http://www.python.org

#### 장점

초기 파일의 용량이 적음

#### 단점

라이브러리 설치가 어려움

- 원하는 패키지를 일일이 설치해야 하며, 각 패키지 별
   Dependencies를 잘 고려하여 설치하여야 함
- 윈도우에서 64비트 Python을 사용하는 경우
  - 공식 라이브러리 Repository(PyPI)로부터 직접 설치하는 것이 어려운 경우가 종종 있음

Python 환경 구성하는 방법(2)

#### Anaconda

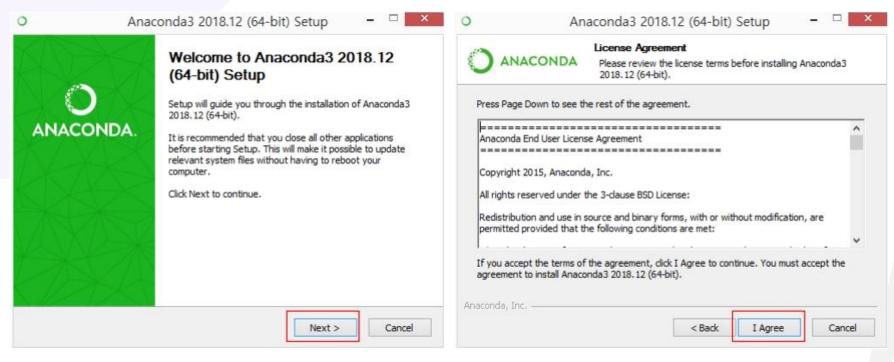
- Open data science platform
- Anaconda is an easy-to-install free package manager, environment manager, Python distribution, and collection of over 720 open source packages.
- Good compatibility with other python packages in PyPI

🕮 Anaconda 다운로드



- https://www.anaconda.com/products/individual
  - 3.7, 64bit 버전 다운로드

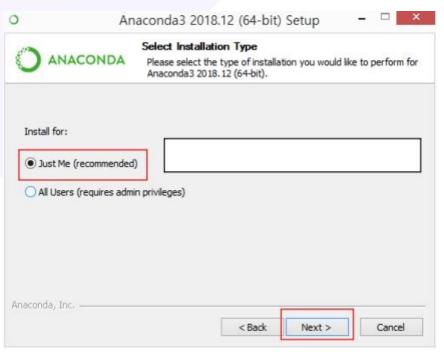
#### ♣ Anaconda 설치



➡ 다운로드 받은 실행파일 시작

설치하는 과정은 Windows나 OS X나 거의 똑같으며, 오히려 OS X가 더 단순함

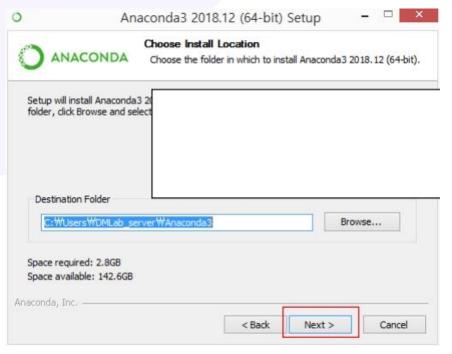
### ♣ Anaconda 설치



■ 반드시 "Just Me"를 선택

OS X에서는 이 과정이 없으며, 사용자 계정에 로그인하는 화면이 나타남

### Anaconda 설치

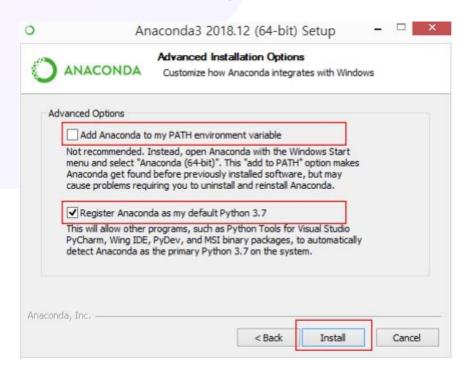


- 기본값
  - 자신의 사용자계정 경로 ₩Anaconda3
- 주의
  - 경로에 한글이 없도록 할 것
  - 만약 사용자이름이 한글이 예경로를 변경해야 함 예상C:₩Anaconda3

OS X에서의 기본 경로

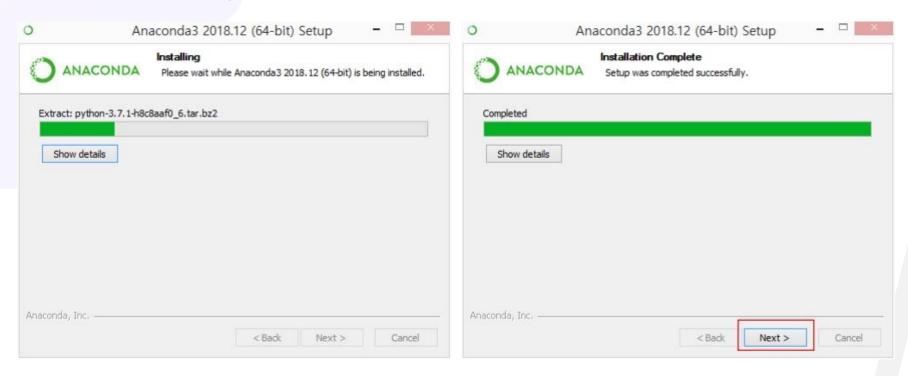
/home/자신의 사용자계정 경로/Anaconda3

### ♣ Anaconda 설치

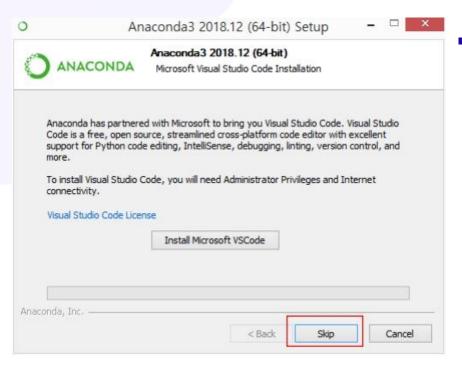


OS X에서는 이 과정이 없음

#### ♣ Anaconda 설치

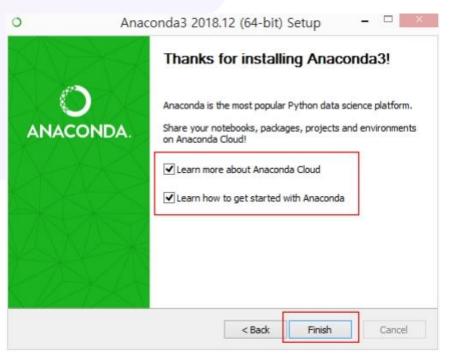


### ♣ Anaconda 설치



- MS Visual Studio Code (텍스트 편집기) 설치를 원하는 경우
  - Install Microsoft VSCode 클릭

#### ♣ Anaconda 설치



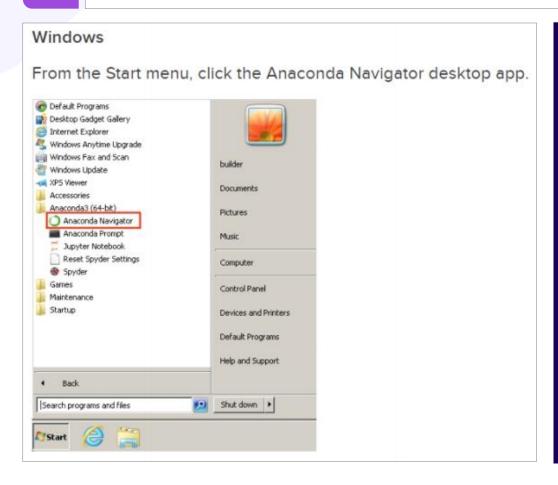
• 관련 내용을 보고 싶으면 체크

Anaconda 및 Jupyter Notebook 실행

#### Anaconda 실행

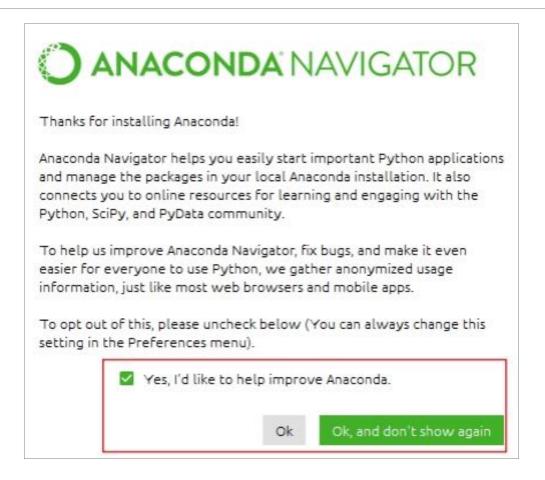
- → 다운로드 : http://docs.anaconda.com/anaconda/userguide/getting-started
- Anaconda Navigator를 활용하여 실행하는 방법
- Anaconda Prompt(for Windows)
- Terminal(for Linux, OS X)을 활용하여 실행하는 방법

- Anaconda 및 Jupyter Notebook 실행
  - Anaconda Navigator를 이용한 실행
    - 1 윈도우 시작 메뉴에서 "Anaconda Navigator"를 검색하여 실행

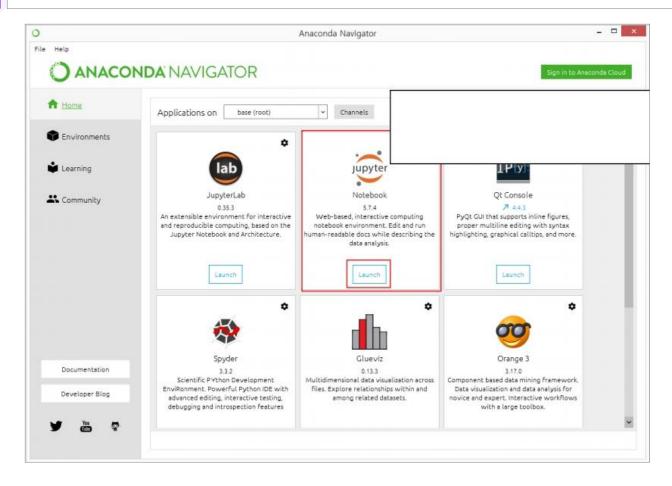




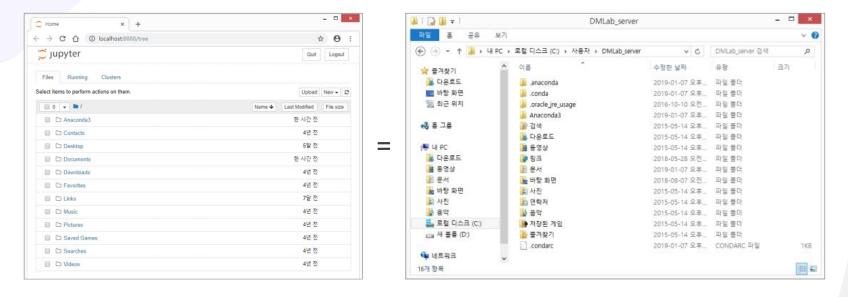
- Anaconda 및 Jupyter Notebook 실행
  - Anaconda Navigator를 이용한 실행
    - 2 각자 원하는 옵션으로 체크 및 실행



- ♠ Anaconda 및 Jupyter Notebook 실행
  - Anaconda Navigator를 이용한 실행
    - 2 각자 원하는 옵션으로 체크 및 실행

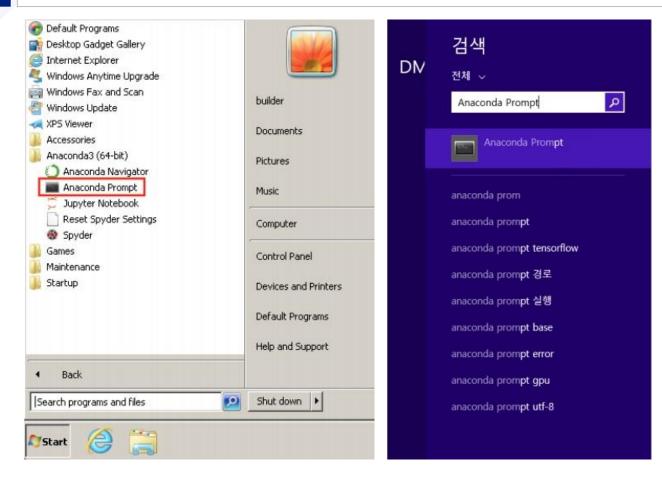


- Anaconda 및 Jupyter Notebook 실행
  - Anaconda Navigator를 이용한 실행
    - 3 Jupyter notebook 실행 완료



이 경로는 "C:₩Users₩ user\_name"임 향후 실습파일을 이 폴더에 넣어두고 사용할 예정임

- Anaconda 및 Jupyter Notebook 실행
  - Anaconda Prompt(for Windows)를 이용한 실행
    - 윈도우 시작 메뉴에서 "Anaconda Prompt"를 검색하여 실행



- Anaconda 및 Jupyter Notebook 실행
  - Anaconda Prompt(for Windows)를 이용한 실행
    - jupyter notebook 을 타이핑 후 ENTER KEY 누르고 난 뒤, 현재 프롬프트가 있는 경로에서 'Jupyter notebook' 입력 후 실행



Linux, OS X에서도 터미널 Terminal)을 실행하여 똑같이 Jupyter notebook을 실행할 수 있음



 과목명
 데이터시각화
 주차명
 02. 데이터 프레임과 Subset Observation
 페이지번호
 02\_03\_02

#### 학습목차

#### 들어가기

- 인트로
- 학습개요

#### 파이썬과 쥬피터 노트북 설치 및 판다스 개요

- 파이썬 설치
- 쥬피터 노트북 설치
- 판다스 개요

#### 시리즈와 데이터프레임

- 판다스 시리즈와 데이터프레임
- 판다스 시리즈
- 판다스 데이터프레임

#### **Subset Observations**

- Subset\_observation(Rows)( 실습)
- Logic 연산(실습)
- Head, Tail 데이터 미리보기(실습)
- 일부 컬럼 기준 데이터 가져오기(실습)

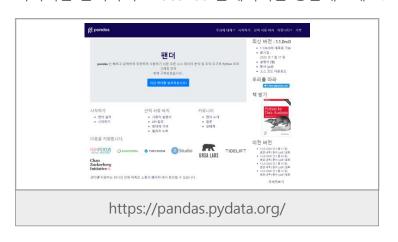
#### 판다스 실습 홈페이지

이미지를 클릭하여 판다스 실습 홈페이지를 방문해보세요.



#### 판다스 코드 연습 홈페이지

이미지를 클릭하여 Anaconda 홈페이지를 방문해보세요.



#### 화면설명

#### [영상 페이지]

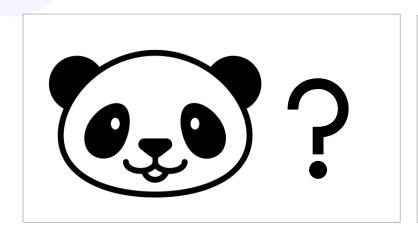
- ■교수님 판서크로마, 전자칠판 동영상 제시
- ■요약 노트 삽입
- ➤ 각 요약 노트는 페이지 메이커의 '게시판 > 글 제출' 컴포넌트로 삽입하며, 게시판별 고유 아이디 부여 (교과목코드(8자리)\_ 개발연도학기(4자리( 연도2자리+학기2자 리))\_주차(2자리)\_생 성순서(1부터 순차적으로))
- ■이미지 클릭 시 홈페이지 이동
- ➤ 홈페이지 url: https://dataitgirls2.git hub.io/10minutes2pan das/
- ■이미지 클릭 시 홈페이지 이동
- ➤ 홈페이지 url : https://pandas.pydata .org/
- 판다스 실습, 코드 연습 홈페이지 개발 시 h2로 제목 삽입

# 판다스 개요

學 판다스(Pandas)?



Pandas is a software library written for the Python programming language for data manipulation and analysis.





### 판다스 개요

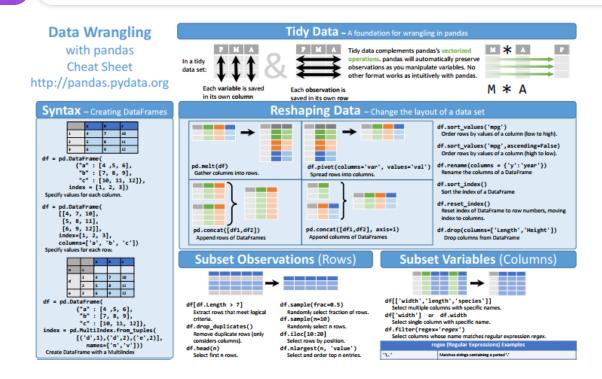
☞ 판다스 10분 완성

실습

https://dataitgirls2.github.io/10minutes2pandas/

코드 연습

- Data Wrangling with pandas Cheat Sheet
- http://pandas.pydata.org







# 시리즈와 데이터프레임



- ☞ 판다스 시리즈와 데이터프레임의 이해
  - 판다스 자료구조

#### 01 목적

- 서로 다른 형식을 갖는 여러 종류의 데이터를 컴퓨터가 이해할 수 있도록 동일한 형식을 갖는 구조로 통합할 필요가 있음
- **판다스의 일차적인 목적은 형식적으로** 서로 다른 여러 가지 유형의 데이터를 공통의 포맷으로 정리하는 것임

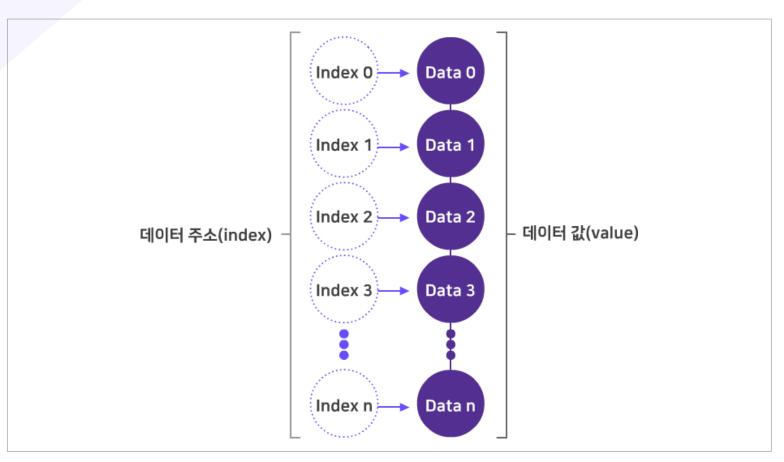
- 판다스 시리즈와 데이터프레임의 이해
  - 판다스 자료구조

#### 02 종류

- 판다스는 시리즈(Series)와 데이터프레임(DataFrame)이라는 구조화된 데이터 형식을 제공함
- 서로 다른 종류의 데이터를 한곳에 담는 그릇(컨테이너)이 됨
  - **다만,** 시리즈는 1차원 배열이고, 데이터프레임이 2차원 배열이라는 점에서 차이**가 있음**
- 특히, 행과 열로 이루어진 2차원 구조의 데이터프레임은 데이터 분석 실무에서 자주 사용됨

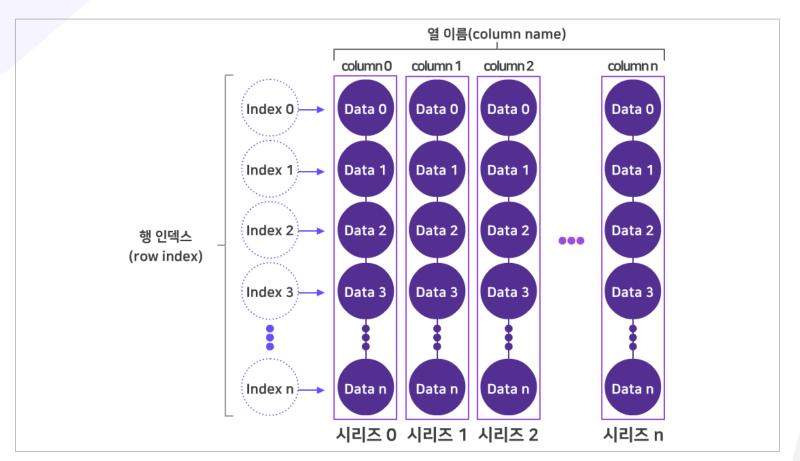
- ☞ 판다스 시리즈와 데이터프레임의 이해
  - 판다스 자료구조





- 판다스 시리즈와 데이터프레임의 이해
  - 판다스 자료구조







### 판다스 시리즈

- ❷ 시리즈
  - 01 딕셔너리와 시리즈의 구조가 비슷하기 때문에, 딕셔너리를 시리즈로 변환하는 방법을 많이 사용함
  - 02 판다스 내장 함수인 Series()를 이용하고, 딕셔너리를 함수의 매개변수(인자)로 전달함

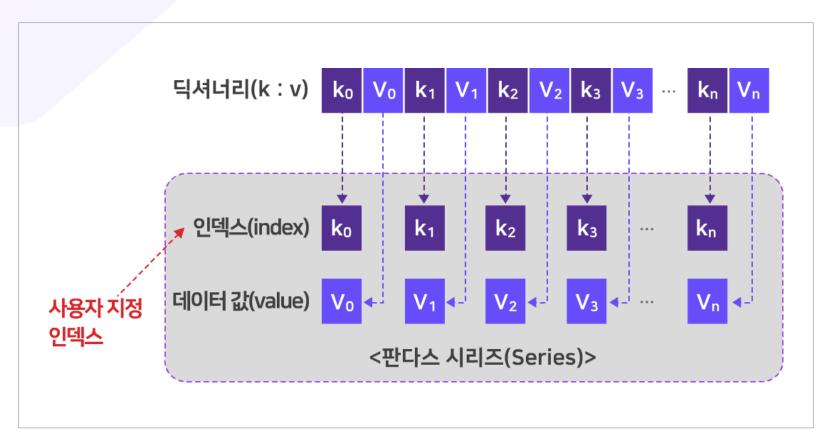
딕셔너리 → 시리즈 변환: pandas.Series ( 딕셔너리 )

덕셔너리의 키(k)는 시리즈의 인덱스에 대응하고, 딕셔너리의 각 키에 매칭되는 값(v)이 시리즈의 데이터 값(원소)로 변환함

# 판다스 시리즈



【 딕셔너리 ➡ 시리즈 변환 】





### 판다스 데이터프레임

- 學 판다스 데이터프레임과 시리즈의 이해
  - 데이터프레임

2차원 배열이며, R의 데이터프레임에서 유래함

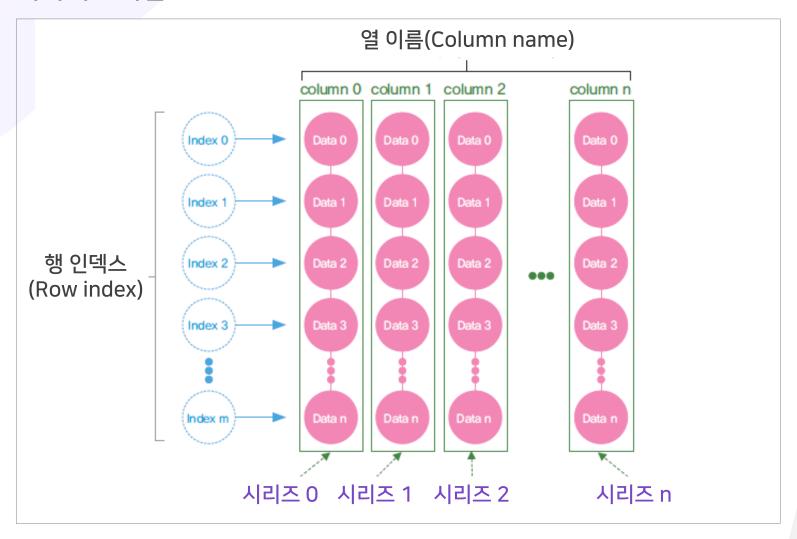
데이터프레임의 열은 시리즈 객체임

시리즈를 열벡터(Vector)라고 하면, 데이터프레임은 여러 개의 열벡터들이 같은 행 인덱스를 기준으로 줄지어 결합된 2차원 벡터 또는 행렬(Matrix)임

데이터프레임은 행과 열을 나타내기 위해 행 인덱스(Row index)와 열 이름(Column name 또는 Column label)으로 구분함

# 판다스 데이터프레임

- ☞ 판다스 데이터프레임과 시리즈의 이해
  - 데이터프레임



### 판다스 데이터프레임

- 판다스 데이터프레임과 시리즈의 이해
  - 데이터프레임

【주식 종목 리스트】

열 •

	종목 코드	회사 이름	액면가	총 주식수
행 ←	005930	삼성전자	100원	5,970백만 주
	017670	SK텔레콤	500원	81백만 주
	005380	현대자동차	5000원	214백만 주

- 데이터프레임의 각 열은 공통의 속성을 갖는 일련의 데이터를 나타냄
- 각 행은 개별 관측대상에 대한 다양한 속성 데이터들의 모음인 레코드(Record)임



<b>과목명</b> 데이터시각화	주차명 01. 데이터 프레임과 Subset Observation 페이지번호	05_01_02
학습목차	학습평가	화면설명
<b>평가하기</b> - 학습평가	학습한 내용을 바탕으로 다음 문제를 풀어 보세요.	[학습평가 페이지] ■ 페이지 퀴즈 컴포넌트 사용하여 페이지 개발
	번호     문제     정답     해설       파이선 프로그래밍 언어의 특징이 <b>아닌</b> 것은?     ① 객체지향 언어이다.       1     ② 다양한 플랫폼을 지원한다.     ③ 기계어의 일종으로 저급언어이다.       ④ 인터프리터 언어이다.	기 이 이 이 때에 에 에 기 기 교
	파이선 프로그램을 위해 만들어진 소프트웨어 라이브러리로서 데이터 2 조작과 분석에 사용되는 것은? 판다스 판다스에 대한 설명이다.	
	판다스의 자료구조 중 데이터프레임에 관한 설명으로 <b>틀린</b> 것은?  ① 서로 다른 여러가지 유형의 데이터를 공통의 포맷으로 정리하는 데이터 형식이다. ② 1차원 배열이다. ③ 행과 열로 이루어진 2차원 구조이다. ④ R의 데이터프레임에서 유래했다	