

---

코드랩

AI 인공지능컨택센터(AICC) 웹서비스 개발

2024.04.12.

신동윤

[data4mania@gmail.com](mailto:data4mania@gmail.com)

---

---

강의소통 카톡방...

---

## 빅데이터분석기사(빅분기) 자격소개

---

## 응시자격

### 응시자격

다음 중 하나에 해당하는 사람

1. 대학졸업자등 또는 졸업예정자 (전공 무관)
2. 3년제 전문대학 졸업자등으로서 졸업 후 1년 이상 직장경력이 있는 사람 (전공, 직무분야 무관)
3. 2년제 전문대학 졸업자등으로서 졸업 후 2년 이상 직장경력이 있는 사람 (전공, 직무분야 무관)
4. 기사 등급 이상의 자격을 취득한 사람 (종목 무관)
5. 기사 수준 기술훈련과정 이수자 또는 그 이수예정자 (종목 무관)
6. 산업기사 등급 이상의 자격을 취득한 후 1년 이상 직장경력이 있는 사람 (종목, 직무분야 무관)
7. 산업기사 수준 기술훈련과정 이수자로서 이수 후 2년 이상 직장경력이 있는 사람 (종목, 직무분야 무관)
8. 기능사 등급 이상의 자격을 취득한 후 3년 이상 직장경력이 있는 사람 (종목, 직무분야 무관)
9. 4년 이상 직장경력이 있는 사람 (직무분야 무관)

※ 졸업증명서 및 경력증명서 제출 필요

### 합격기준

필기시험 합격기준	실기시험 합격기준
과목당 100점을 만점으로 1. 전 과목 40점 이상 2. 전 과목 평균 60점 이상	100점을 만점으로 60점 이상

## ● 출제기준(필기)

직무분야	정보통신	중직무분야	정보기술	자격종목	빅데이터 분석기사
○ 직무내용 대용량의 데이터 집합으로부터 유용한 정보를 찾고 결과를 예측하기 위해 목적에 따라 분석기술과 방법론을 기반으로 정형/비정형 대용량 데이터를 구축, 탐색, 분석하고 시각화를 수행하는 업무를 수행한다.					
필기검정방법	객관식	문제수	80	시험시간	120분

## ● 과목별 주요 항목

필기과목명	문제수	주요항목
빅데이터 분석기획	20	빅데이터의 이해
		데이터 분석 계획
		데이터 수집 및 저장 계획
빅데이터 탐색	20	데이터 전처리
		데이터 탐색
		통계기법 이해
빅데이터 모델링	20	분석모형 설계
		분석기법 적용
빅데이터 결과 해석	20	분석모형 평가 및 개선
		분석결과 해석 및 활용

## ● 출제기준(실기)

직무분야	정보통신	중직무분야	정보기술	자격종목	빅데이터 분석기사
○ 직무내용 대용량의 데이터 집합으로부터 유용한 정보를 찾고 결과를 예측하기 위해 목적에 따라 분석기술과 방법론을 기반으로 정형/비정형 대용량 데이터를 구축, 탐색, 분석하고 시각화를 수행하는 업무를 수행한다.					
실기검정방법	통합형(필답형, 작업형)		시험시간	180분	

※ 필기시험 면제기간은 필기합격자 발표일로부터 2년  
(다만, 발표일부터 2년동안 검정이 2회 미만으로 시행된 경우에는 그 다음 회차 필기시험 1회를 면제)

## ● 과목별 주요 항목

실기과목명	주요항목
빅데이터 분석실무	데이터 수집 작업
	데이터 전처리 작업
	데이터 모형 구축 작업
	데이터 모형 평가 작업

- 실무 기반 데이터분석 강의를 통해 수험대비 with 파이썬/R

## • 필기 내용

필기과목명	주요항목	세부항목
빅데이터 분석 기획	빅데이터의 이해	빅데이터 개요 및 활용
		빅데이터 기술 및 제도
	데이터분석 계획	분석방안수립
		분석 작업 계획
	데이터 수집 및 저장 계획	데이터 수집 및 전환 데이터 적재 및 저장
빅데이터 탐색	데이터 전처리	데이터 정제
		분석 변수 처리
	데이터 탐색	데이터 탐색 기초
		고급 데이터 탐색
	통계기법 이해	기술통계 추론통계
빅데이터 모델링	분석모형 설계	분석 절차 수립
		분석 환경 구축
	분석기법 적용	분석기법 고급 분석기법
빅데이터 결과해석	분석모형 평가 및 개선	분석모형 평가
		분석모형 개선
	분석결과 해석 및 활용	분석결과 해석
		분석결과 시각화
		분석결과 활용

## • 실기 내용

실기과목명	주요항목	세부항목
빅데이터 분석 실무	데이터 수집 작업	데이터 수집하기
		데이터 정제하기
	데이터 전처리 작업	데이터 변환하기
		데이터 변환하기
	데이터 모형 구축 작업	분석모형 선택하기
		분석모형 구축하기
	데이터 모형 평가 작업	구축된 모형 평가하기
		분석결과 활용하기

## • ADsP 시험범위

- 빅분기 :

빅분기 이론 =

ADsP +

데이터 수집/전  
환/적재/저장 +  
분석결과 시각화

과목명	주요항목	세부항목
데이터 이해	데이터의 이해	<ul style="list-style-type: none"> <li>•데이터와 정보</li> <li>•데이터베이스의 정의와 특징</li> <li>•데이터베이스 활용</li> </ul>
	데이터의 가치와 미래	<ul style="list-style-type: none"> <li>•빅데이터의 이해</li> <li>•빅데이터의 가치와 영향</li> <li>•비즈니스 모델</li> <li>•위기 요인과 통제 방안</li> <li>•미래의 빅데이터</li> </ul>
	가치 창조를 위한 데이터 사이언스와 전략 인사이트	<ul style="list-style-type: none"> <li>•빅데이터분석과 전략 인사이트</li> <li>•전략 인사이트 도출을 위한 필요 역량</li> <li>•빅데이터 그리고 데이터 사이언스의 미래</li> </ul>
데이터분석 기획	데이터분석 기획의 이해	<ul style="list-style-type: none"> <li>•분석 기획 방향성 도출</li> <li>•분석 방법론</li> <li>•분석 과제 발굴</li> <li>•분석 프로젝트 관리 방안</li> </ul>
	분석 마스터 플랜	<ul style="list-style-type: none"> <li>•마스터 플랜 수립</li> <li>•분석 거버넌스 체계 수립</li> </ul>
데이터분석	R기초와 데이터 마트	<ul style="list-style-type: none"> <li>•R기초</li> <li>•데이터 마트</li> <li>•결측값 처리와 이상값 검색</li> </ul>
	통계분석	<ul style="list-style-type: none"> <li>•통계학 개론</li> <li>•기초 통계분석</li> <li>•다변량 분석</li> <li>•시계열 예측</li> </ul>
	정형 데이터 마이닝	<ul style="list-style-type: none"> <li>•데이터 마이닝 개요</li> <li>•분류분석(Classification)</li> <li>•군집분석(Clustering)</li> <li>•연관분석(Association Analysis)</li> </ul>

<https://url.kr/4iub18>



---

**Week01, Week02**

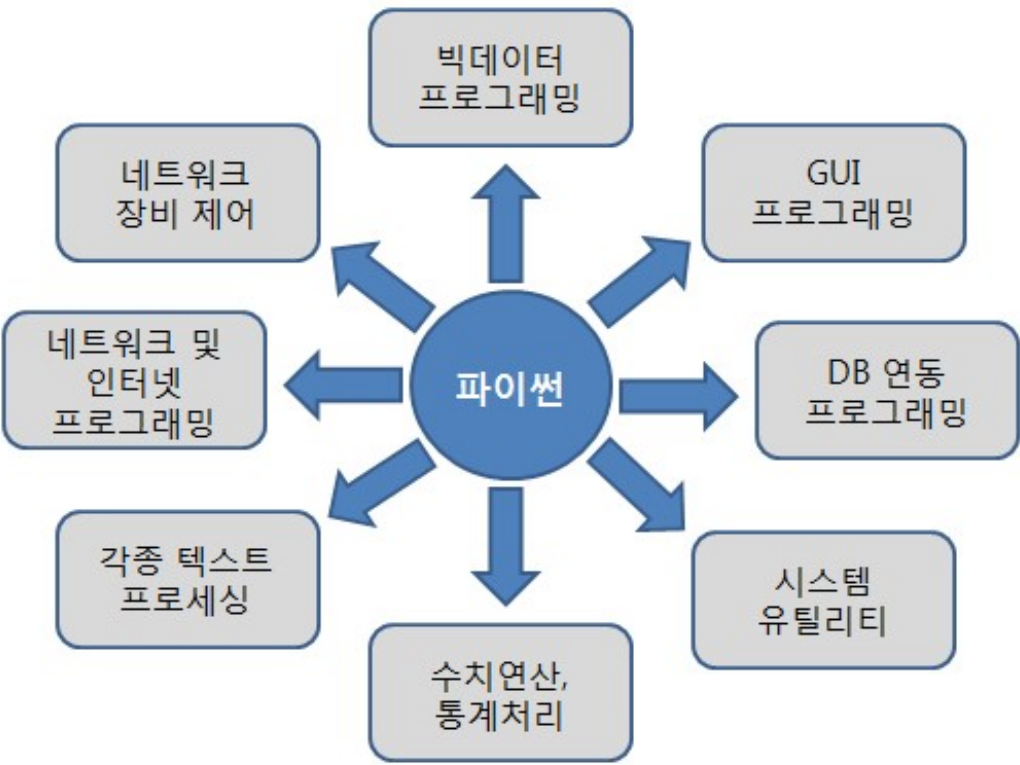
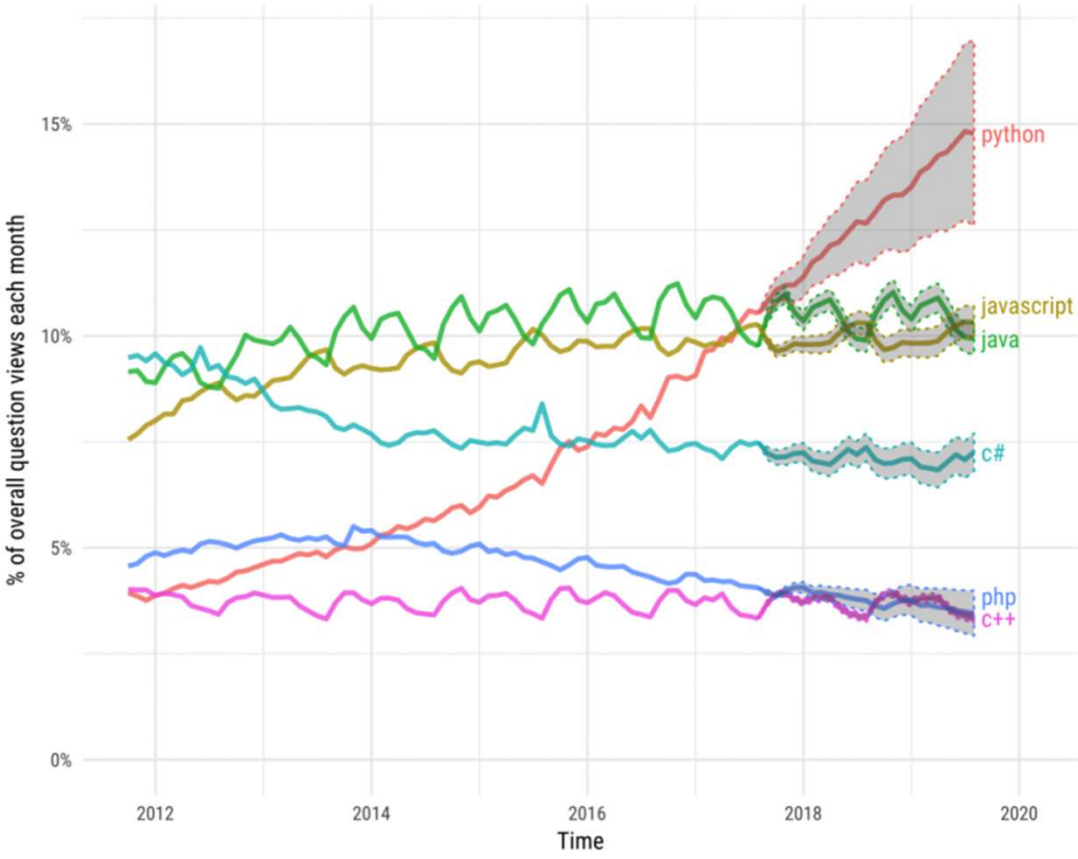
**파이썬 설치 및 기초문법, 판다스**

---

# 파이썬 소개

## Projections of future traffic for major programming languages

Future traffic is predicted with an STL model, along with an 80% prediction interval.



# 파이썬(Python)이란?



- 1990년 암스테르담의 귀도 반 로섬(Guido Van Rossum)에 의해 만들어진 인터프리터 언어
- 사전적인 뜻은 고대 신화 속의 파르나수스 산의 동굴에 살던 큰 뱀으로, 아폴로가 델파이에서 파이썬을 퇴치했다는 ...
- 어디에서 많이 사용하는가?
  - 구글에서 만들어진 소프트웨어의 50%이상이 파이썬으로 만들어졌다고 함
  - Dropbox(파일 동기화 서비스), Django(파이썬 웹 프레임워크) 등

# 파이썬 특징

- 문법이 쉬워 빠르게 학습할 수 있다
  - 파이썬을 공부한지 단 하루만에 자신이 원하는 프로그램을 작성할 수 있었다고 함 (프로그래밍 경험이 있었음)
- 간결하다
  - Perl은 하나의 일을 하기 위한 방법이 100가지
  - 파이썬은 가장 좋은 방법 1가지를 선호
- 강력하다
  - 대부분의 모든 일을 할 수 있음
    - 시스템 프로그래밍, 하드웨어 제어, 복잡하고 많은 반복연산에는 어울리지 않음
  - 파이썬으로 안되면 C, C++로 만들어서 쉽게 파이썬에 포함시킬 수 있음
- 무료이다
- 개발속도가 빠르다
  - C, C++, Java 등으로 한달 이상 걸릴 것을 1주일 안에 개발 가능
  - Python을 이용하여 빠르게 Prototype 시스템을 구축 후, 검증 되면 이를 C, C++, Java 등으로 구현 → 전체 개발 시간을 단축!

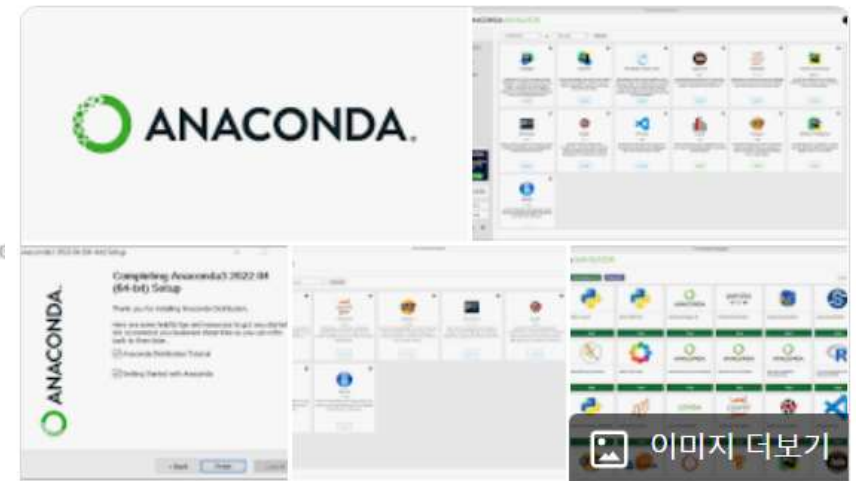
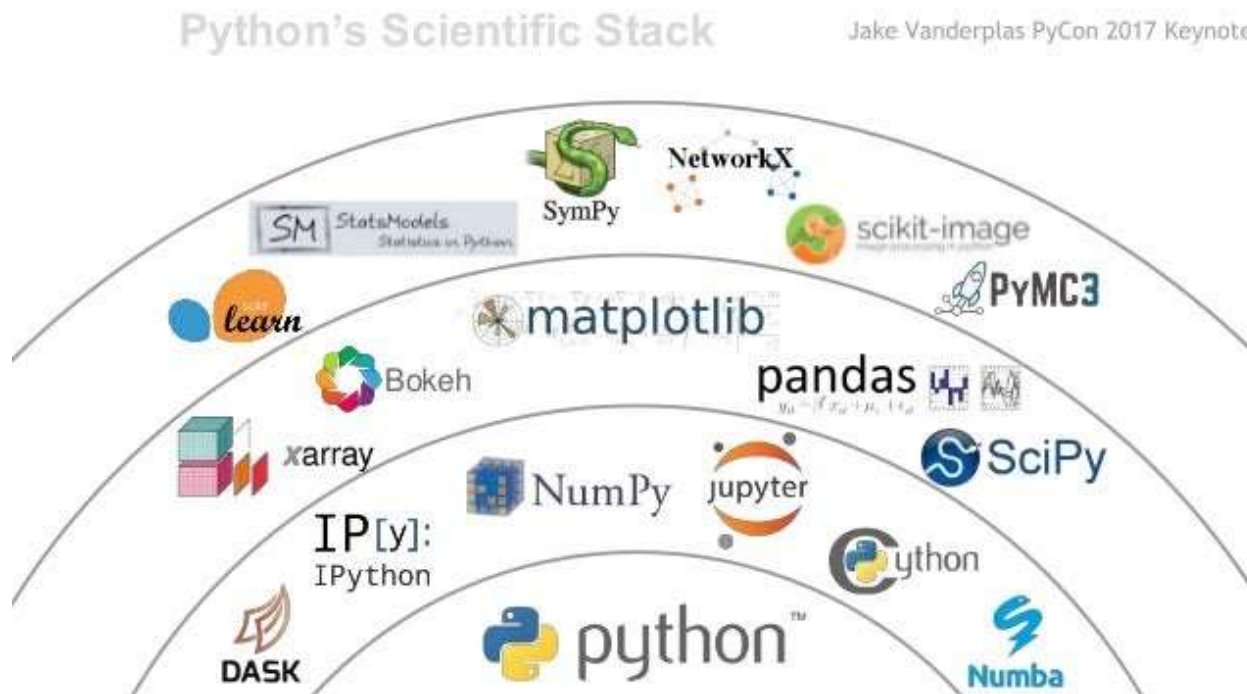


# 파이썬으로 무엇을 할 수 있나?

- 시스템 유틸리티 - 운영체제의 시스템 명령어 지원
- GUI(Graphic User Interface) 프로그램
  - Tkinter, wxPython, PyQt, PyGTK
- 웹 프로그래밍 - 많은 라이브러리 제공
- C/C++과의 결합
  - C/C++로 만든 프로그램을 파이썬에서 호출, 반대로 가능
- 수치연산 프로그래밍 - **Numpy 이용 (C로 작성 됨)**
- 데이터베이스 프로그래밍 - 오라클, MySQL 등의 도구 제공
- 파이썬으로 할 수 없는 일
  - 대단히 빠른 속도를 요구하거나 하드웨어를 직접 건드려야 하는 프로그램



# 파이썬 패키지 vs 아나콘다 (Anaconda)



## 아나콘다 (Anaconda)

파이썬 배포판

아나콘다는 패키지 관리와 디플로이를 단순케 할 목적으로 과학 계산을 위한 파이썬과 R 프로그래밍 언어의 자유-오픈 소스 배포판이다. 패키지 버전들은 패키지 관리 시스템 conda를 통해 관리된다. [위키백과](#)


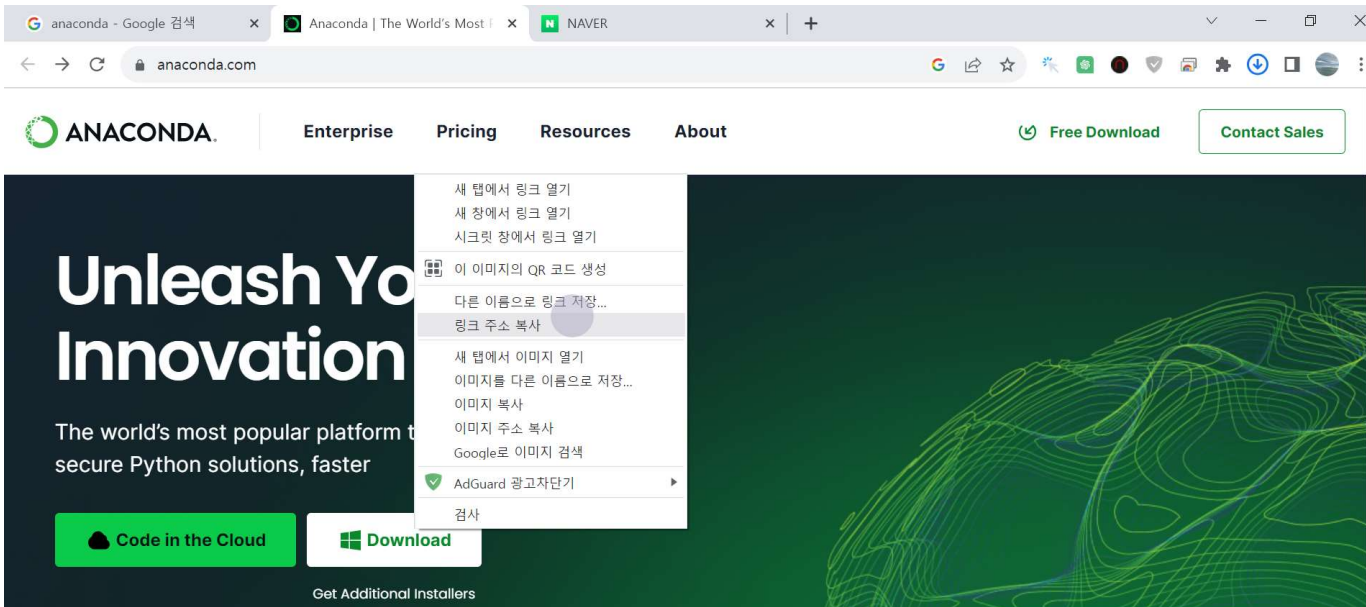
프로그래밍 언어: 파이썬

개발: Continuum analytics

최초 출시일: 2012년 7월 17일

# Old Version 다운로드 <https://repo.anaconda.com/archive/>

- 강의목적 안정버전  
2022-05
- 마지막 tf\_gpu 포함버전  
2023-03

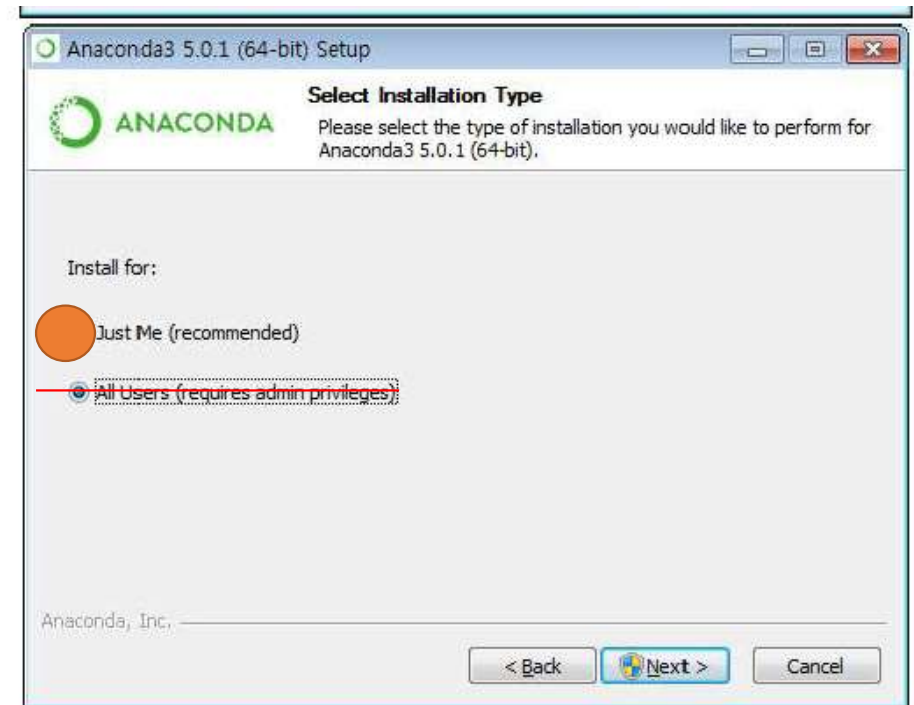


<a href="#">Anaconda3-2022.05-MacOSX-x86_64.sh</a>	584.0M	2022-05-10 13:22:01	1a10c06660ebe1204e538b4e9d810142441af9dfd74b077eee2761ec6e675f39
<a href="#">Anaconda3-2022.05-Linux-s390x.sh</a>	279.8M	2022-05-10 13:22:01	c14415df69e439acd7458737a84a45c6067376cbec2fccf5e2393f9837760ea7
<a href="#">Anaconda3-2022.05-Linux-ppc64le.sh</a>	367.3M	2022-05-10 13:22:01	a50bf5bd26b5c5a2c24028c1aff6da2fa4d4586ca43ae3acd7fffb9b50d7f282
<a href="#">Anaconda3-2022.05-Linux-x86_64.sh</a>	658.8M	2022-05-10 13:22:00	a7c0afe862f6ea19a596801fc138bde0463abcbce1b753e8d5c474b506a2db2d
<a href="#">Anaconda3-2022.05-Linux-aarch64.sh</a>	567.6M	2022-05-10 13:22:00	dc6bb4eab3996e0658f8bc4bbd229c18f55269badd74acc36d9e23143268b795
<a href="#">Anaconda3-2022.05-Windows-x86.exe</a>	487.8M	2022-05-10 13:21:59	cd8c688349bcd1f429e3b383620fb0d19f52be0f765b2eae78d63b41aefb2e73
<a href="#">Anaconda3-2021.11-Windows-x86_64.exe</a>	510.3M	2021-11-17 12:08:45	1b3d593d1deb22b835be5c68897075e0fc9dea240ab4191c55674aba259a78ff
<a href="#">Anaconda3-2021.11-Windows-x86.exe</a>	404.1M	2021-11-17 12:08:45	dc0746dded06cc480328c20b73369803ce98df1971bda669d93859e02c1c0664

## 아나콘다 설치와 쥬피터 노트북(웹브라우저)에서 파이썬 코딩

- ① Anaconda(아나콘다) - Python 기반의 데이터 분석에 필요한 오픈소스를 모아놓은 개발 플랫폼이다.
  - 가상환경 관리자 제공 / - 패키지 관리자 제공
- ② 아나콘다에 포함된 수준 높은 패키지 관리자를 통해서 파이썬의 효율성을 극대화 시켜 활용할 수 있습니다.
- ③ 아나콘다의 또 다른 장점은 가상환경 관리자를 각 프로젝트별 개발 환경을 효율적으로 구성할 수 있다

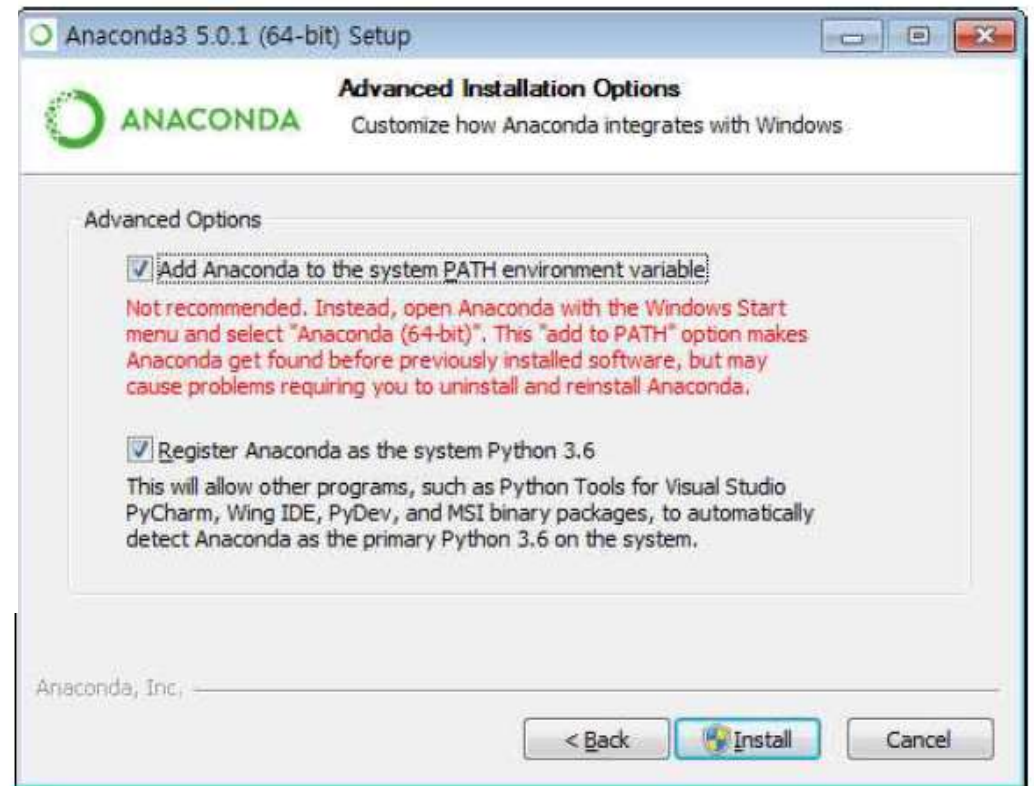
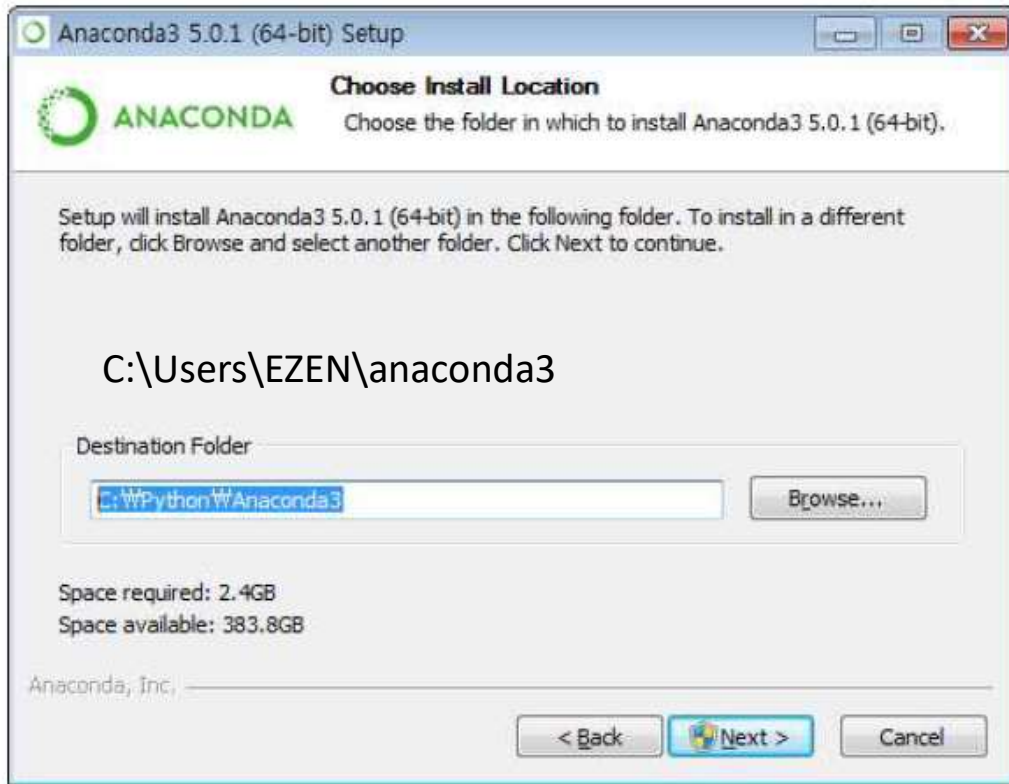
- 1) 크롬 브라우저 실행. (익스플로러는 윈도우 설치파일 다운 안됨)
- 2) <https://www.anaconda.com/download/> 링크로 이동 설치파일 다운로드  
(64bit or 32bit 개발환경에 맞는 버전다운)  
(Anaconda3-5.0.1-Windows-x86\_64 , Anaconda3-5.0.1-Windows-x86)
- 3) 설치 파일 실행



Just Me 선택



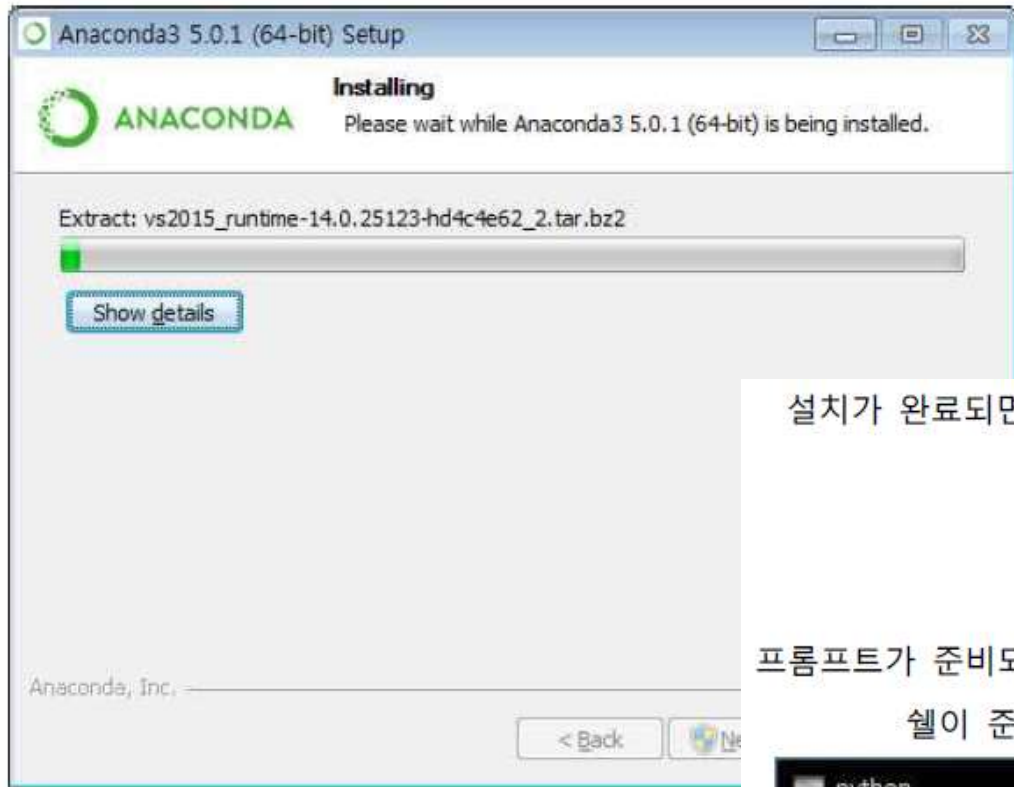
## 아나콘다 설치와 쥬피터 노트북(웹브라우저)에서 파이썬 코딩



Options는 2개 모두 체크 선택

아나콘다의 경로를 환경 변수의 PATH에 추가할지 설정하는 화면입니다.

## 아나콘다 설치와 쥬피터 노트북(웹브라우저)에서 파이썬 코딩



설치가 완료되면 아래 아이콘을 클릭하거나 없다면, 검색에서 Anaconda Prompt 를 찾아 실행합니다.



프롬프트가 준비되면 python 명령을 실행해서 파이썬 셸(python shell)로 들어갑니다. 파이썬 셸이 준비되면 아래 명령을 통해 파이썬이 잘 작동하는지 확인해봅니다.

```
python
(C:\Users\jaewo\Anaconda3) C:\Users\jaewo>python
Python 3.6.1 [Anaconda 4.4.0 (64-bit)] (default, May 11 2017, 13:25:24) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> print("hello world")
hello world
>>>
```

---

## [ 기본 명령어 ]

### #아나콘다 버전 확인

conda --version

### #아나콘다 정보 보기

conda info

### #아나콘다 업데이트

conda update conda

### #아나콘다 가상환경 생성

conda create --name(-n) 가상환경명 설치할 패키지

#예) 파이썬 3.5 버전 설치 & test 이름으로 가상환경 생성

conda create --name test python=3.5 <=>

conda create --n test python=3.5

### # 아나콘다 가상환경 목록보기

conda info --envs <=> conda info -e

### # 아나콘다 개발환경 활성화 하기

activate <개발환경 이름>

### # 아나콘다 개발환경 제거

conda remove --name <개발환경 이름> --all

### #패키지 설치

#예)conda install 패키지명

conda install simplejson

### #패키지 리스트 확인

conda list

### #패키지 삭제

#예)conda remove --name test --all

conda remove --name 가상환경명 --all <=> conda remove -n 가상환경명 --all

### #아나콘다 클린(인덱스 캐시, 잠긴 파일, 사용하지 않는 패키지, 소스 캐시 등을 삭제)

conda clean -all <=> conda clean -a

- 
- VS Code 사용법 :  
<https://demun.github.io/vscode-tutorial/>
  - 쥬피터 노트북 단축키 :  
<https://speedspeed.tistory.com/92>
  - jupytertext 설치 및 사용 :  
<https://limitt.tistory.com/200>
  - 코랩 단축키 :  
<https://surfonmedia.tistory.com/1>

# 직접 해보는 데이터분석 ... 이후는 실습과 함께 ...

The image shows the Kaggle website interface. On the left is a sidebar with navigation links: 'kaggle' logo, 'Create', 'Home', 'Competitions' (highlighted), 'Datasets', 'Code', and 'Discussions'. The main content area features a search bar at the top. Below it is a large banner for the 'Titanic - Machine Learning from Disaster' competition. The banner includes the text 'GettingStarted Prediction Competition', the title 'Titanic - Machine Learning from Disaster', and the subtitle 'Start here! Predict survival on the Titanic and get familiar with ML basics'. It also shows the Kaggle logo, '13,988 teams', and 'Ongoing' status. At the bottom of the banner is a navigation bar with links: 'Overview' (underlined), 'Data', 'Code', 'Discussion', 'Leaderboard', 'Rules', and 'Team'. To the right of this bar are links for 'My Submissions' and a 'Submit Predictions' button.

[Titanic - Machine Learning from Disaster | Kaggle](#)

# Google Colab 실습환경

- 사전확인 사항 : 크롬 설치여부 및 구글 아이디
- 실습자료 주소 : <https://url.kr/2lghb4> 에서 <전체 다운로드> 및 압축해제

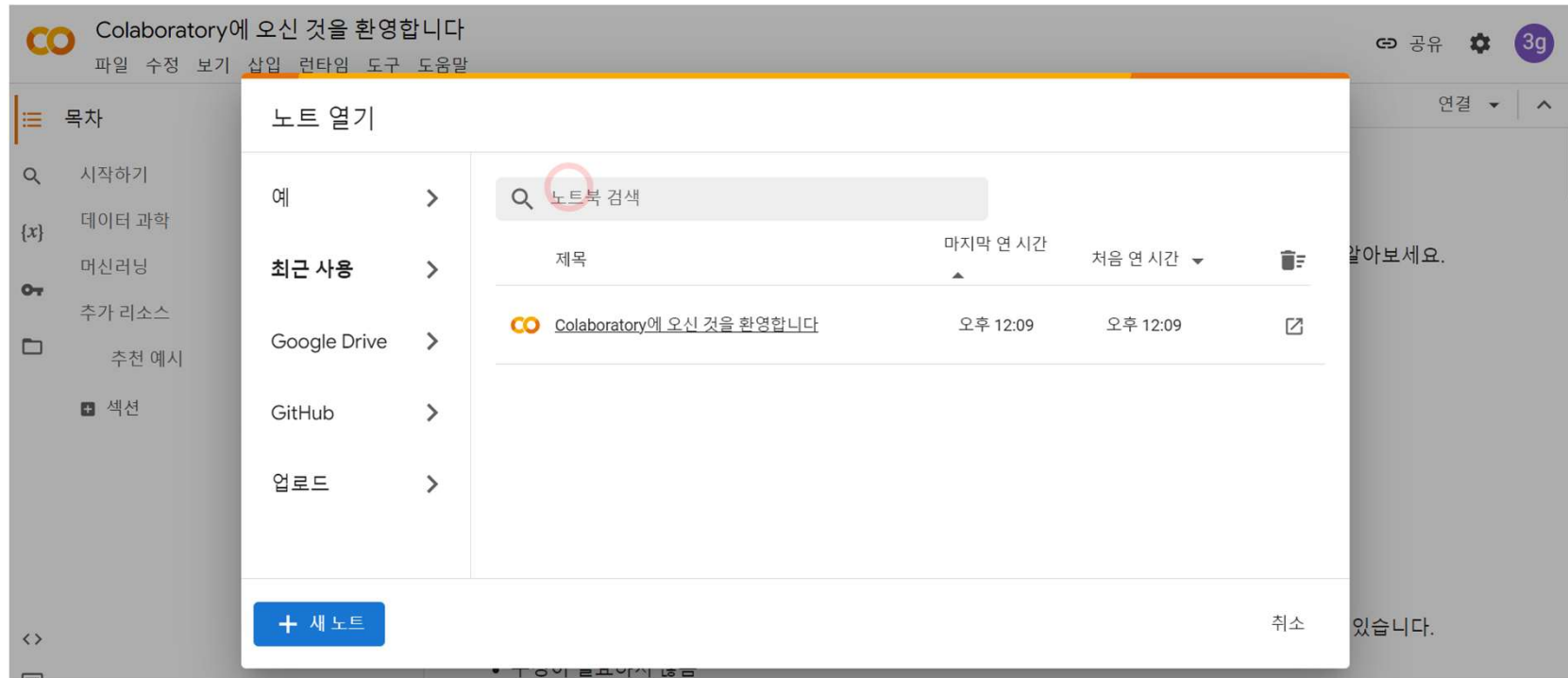
강의자료

[전체 다운로드](#)

이름 	소유자	마지막...	파일 크기
 data	 소유자 숨겨짐	오전 11:58	—
 data_titanic	 소유자 숨겨짐	오전 11:58	—
 output	 소유자 숨겨짐	오전 11:58	—
 BD-Python-Part1-01_파이썬.ipynb 	 소유자 숨겨짐	2023. 11. 8.	122KB
 BD-Python-Part1-02_판다스.ipynb 	 소유자 숨겨짐	2023. 11. 8.	60KB
 BD-Python-Part1-03_시각화.ipynb 	 소유자 숨겨짐	2023. 11. 8.	67KB
 BD-Python-Part1-04_타이타닉초급.ipynb 	 소유자 숨겨짐	2023. 11. 8.	13KB
 BD-Python-Python_Final_Review.ipynb 	 소유자 숨겨짐	2023. 11. 8.	49KB
 BD-Python-Python_Pandas_Review.ipynb 	 소유자 숨겨짐	2023. 11. 8.	37KB

- 구글 코랩 <https://colab.research.google.com/?hl=ko> 이동 및 로그인

# Google Colab 실습환경



- <+새노트> 선택
- <파일> - <노트업로드> - <다운로드한 파일 선택>
- <런타임> - <모두 실행>

---

## Week01 : 파이썬 기초문법

- 사칙연산 계산기
- 등호 / 부등호 연산자
- 변수 Variables
- Boolean
- 문자열
- 리스트 List
- 제어문 : if ~ elif ~ else  
condition
- 반복문 for loop
- 함수 Function

### 파이썬 자료형 종류 리뷰

1. 숫자형
  - 1, 2, 3, 0, -1, -5, 3.14
2. Boolean
  - True, False
3. 문자열
  - "Hello", "메가IT", '빅데이터'
4. 리스트 List
  - [], [1,3,5], ["홍", "길", "동"]
5. 튜플 Tuple
  - (), (1,3,5)
6. 사전 Dictionary
  - {key1 : value1}, {"이름":"홍길동", "성별":"남자"}
7. 집합 Set
  - {}, set([1,3,5]), set("Hello")



---

## Week02 : 판다스 및 시각화

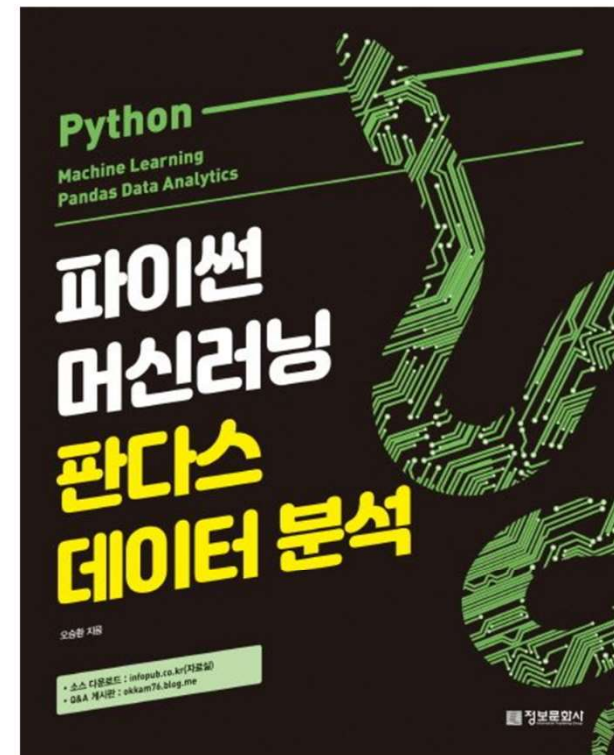
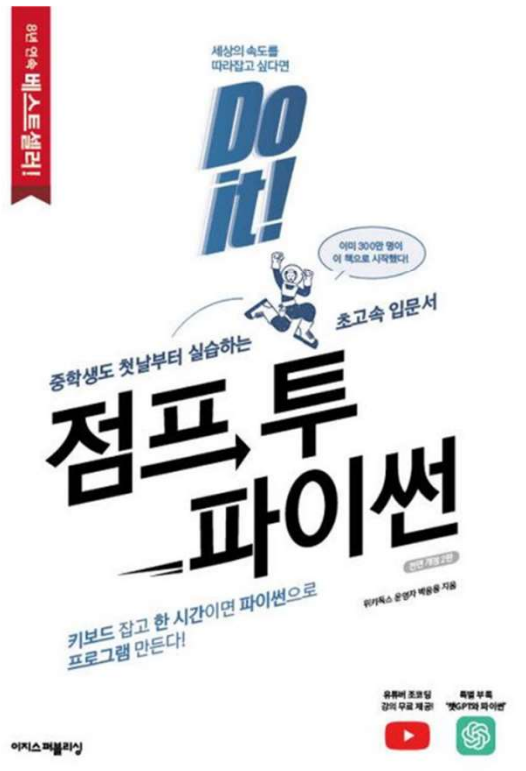
- 판다스(Pandas) ?
- Series : 1차원 구조
- DataFrame : 2차원의 구조 형태
- 파일 읽어오기
- 기본 내장 기능
- 행렬
- 색인(인덱스)
- pivot\_table
- 파일 저장하기
- 데이터 시각화 소개
- 시각화 패키지 사용법
- CountPlot
- BarPlot
- DistPlot
- BoxPlot
- ScatterPlot

[점프 투 파이썬] 온라인 서적  
<https://wikidocs.net/book/1>  
[점프 투 파이썬] 깃허브  
<https://github.com/pahkey/jump2python>

## 도서추천 : 파이썬 및 판다스

중학생도 첫날부터 실습하는 초고속 입문서

판다스 활용과 데이터 분석을 동시에!



---

파이썬으로 가능한 일들...

[3] 파이썬 학습이 늦지 않은 이유

# 당신의 전문성



여러분만이 가지고 있는 노하우, 전문성이 분명 있을 거예요

## 말로 (논리적으로) 코딩하자

