

데이터 수집과 분석을 위한 파이썬 개발 환경 구축

학습목표

- 빅데이터 처리를 위한 파이썬 언어의 강점을 설명하고, 파이썬의 개발 환경을 구축할 수 있다.
- 파이썬 프로그래밍에 유용한 개발 툴의
 종류를 파악하고 프로그래밍을 할 수 있다.

학습내용

- 파이썬 소개와 개발 환경 구축 방법
- 파이썬 개발 툴 소개



1 빅데이터 처리를 위한 파이썬



1 파이썬 소개

	 1991년에 첫 버전이 공개된 후 지금까지 많은 인기를 얻고 있는 스크립트 형태의 범용 프로그래밍 언어
	 프로그래머가 원하는 모든 작업을 할 수 있도록 설계한 범용 언어
파이썬	■ 명령형 언어이면서 스크립트 방식 지원- 프로그래밍적인 구현에 적합- 데이터 수집과 처리에 활용할 수 있는 다양한 라이브러리 제공
	 웹 서버 프로그래밍, 데이터 분석, 시스템 자동화, IoT 프로그래밍까지 활용 분야도 다양



1 빅데이터 처리를 위한 파이썬



1 파이썬 소개

파이썬 특징과 장점

- 윈도우, 리눅스, 맥에서 모두 실행할 수 있는 언어
- Free and Open Source 언어
- 인터프리터 언어
- 객체지향 언어
- 배우기 쉽고 간결
- 동적(스크립트) 언어
- loT, 빅데이터, 인공지능, 블록체인 등의 실무 프로그램 구현에 많이 사용되는 고급 라이브러리도 제한 없이 사용 가능



1 빅데이터 처리를 위한 파이썬

- 파이썬 소개
 - 프로그래밍 언어 순위(TIOBE Index for October 2019)

(2019년 10월 기준)

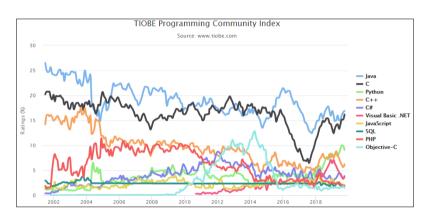
2019. 10월	2018. 10월	변화	프로그래밍 언어	점유율	변화율
1	1		Java	16.884%	-0.92%
2	2		С	14.180%	+0.80%
3	4	٨	Python	9.089%	+1.93%
4	3	V	C++	6.229%	-1.36%
5	6	٨	C#	3.860%	+0.37%

[출처: https://www.tiobe.com/tiobe-index/]



1 빅데이터 처리를 위한 파이썬

- 파이썬 소개
 - 파이썬은 2016년부터 사용량이 가파르게 높아짐
 - ➡ 데이터 분석 분야에서 파이썬이 많이 활용되고 있기 때문
 - 파이썬은 '빠르게 성장하고 있는 언어'



[출처:https://www.tiobe.com/tiobe-index/]



- 1 빅데이터 처리를 위한 파이썬
 - 2 빅데이터 처리 언어로서 파이썬의 장점
 - 1 이해하기 쉽고 유연한 문법으로 좋은 접근성을 가짐
 - 2 빅데이터 처리 언어로서 많은 커뮤니티가 형성되어 있음
 - 3 가독성이 좋고, 간결하며, 스탠다드 라이브러리가 잘 갖춰져 있음
 - 4 데이터 분석 관련 패키지가 최근 몇 년 사이 눈에 띄게 발전하여 NumPy, SciPy, Pandas, Matplotlib 등 데이터 분석 관련 오픈 소스 라이브러리들을 무상으로 사용할 수 있음



2 웹사이트를 통한 설치 방법

- 1 데이터 분석을 위한 파이썬 개발 환경 구축 방법
 - 아래 사이트에서 원하는 버전의 파이썬 설치 프로그램을 다운로드하여 설치 후 필요한 라이브러리와 개발 툴 추가 설치

https://www.python.org/downloads/

■ 아래 사이트에서 Anaconda Distribution 설치

https://www.anaconda.com/distribution/#download-section



2 웹사이트를 통한 설치 방법

Anaconda

Anaconda

■ 데이터 과학 및 머신 러닝을 위한 파이썬과 R 프로그래밍 언어의 'Freemium' 오픈 소스 배포판

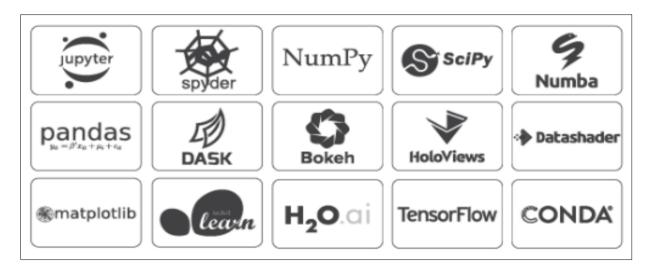
> → 기본적인 기능은 무료로 제공하고 고급 기능은 돈을 받고 판매

- 세계에서 가장 유명한 파이썬 데이터 과학 플랫폼
- 모든 데이터 과학 패키지의 쉬운 설치, 패키지·종속성 및 환경 관리 가능
- CONDA, 파이썬 및 150개가 넘는 과학 패키지와 그 종속성과 함께 제공되는 파이썬 배포판 패키지 및 환경 관리자



2 웹사이트를 통한 설치 방법

- Anaconda
 - 파이썬 개발환경과 수학·과학·데이터 분석 분야에서 필요한 거의 모든 패키지(NumPy, SciPy, Pandas, Matplotlib 등) 포함

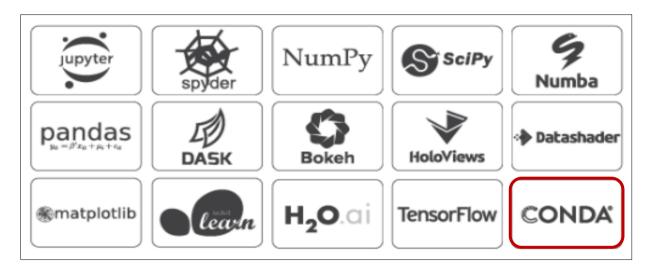


[출처: https://www.anaconda.com/distribution/#download-section]



2 웹사이트를 통한 설치 방법

- Anaconda
 - CONDA라는 응용 프로그램을 제공하여 패키지 및 화경 관리를 편리하게 처리하도록 지원

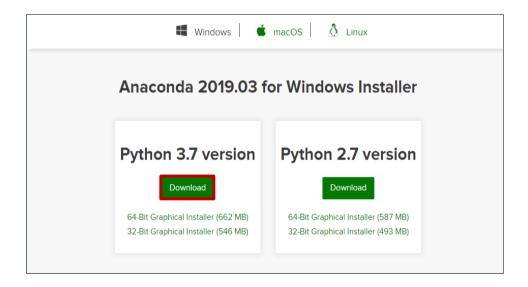


[출처: https://www.anaconda.com/distribution/#download-section]



- 2 웹사이트를 통한 설치 방법
 - ◆ Anaconda 설치
 - 1 다음 URL로 다운로드 사이트를 접속하여 파이썬 다운로드

https://www.anaconda.com/distribution/#downlo ad-section





2 웹사이트를 통한 설치 방법

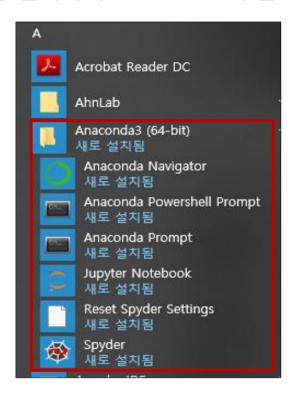
- **③** Anaconda 설치
 - 2 다운로드된 xxx.exe 설치 파일 더블 클릭하여 화면이 출력되면 'Next' 버튼 클릭





2 웹사이트를 통한 설치 방법

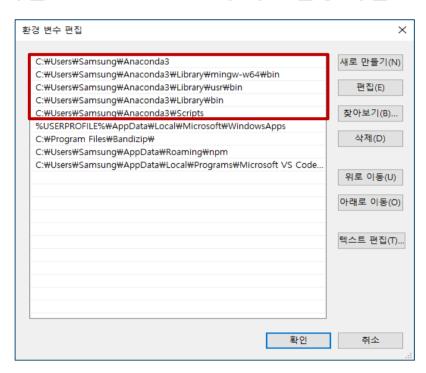
- **③** Anaconda 설치
 - 3 Anaconda 설치를 모두 끝내면, 윈도우 운영체제의 '시작' 메뉴를 선택하여 Anaconda의 설치 상태 확인





2 웹사이트를 통한 설치 방법

- ③ Anaconda 설치
 - 4 윈도우 운영체제의 사용자 환경 변수 path 정보를 출력한 화면으로 Anaconda의 패스 설정 확인

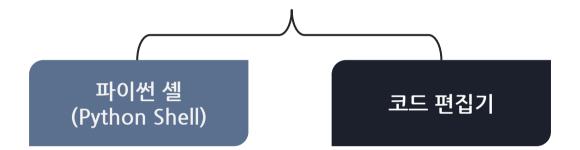




1 IDLE

IDLE

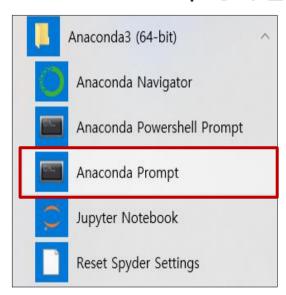
■ Integrated Development Environment의 약자로 '통합 개발 환경'을 뜻함

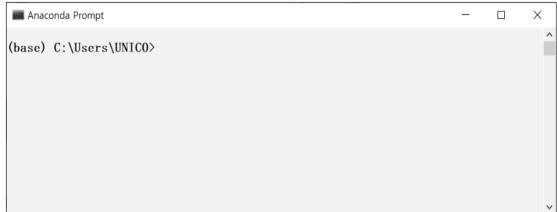




1 IDLE

- 1 Anaconda Prompt 창
 - 1 시작메뉴의 Anaconda3 메뉴를 선택한 후에 Anaconda Prompt라는 서브 메뉴를 클릭하면 Anaconda Prompt창이 출력됨

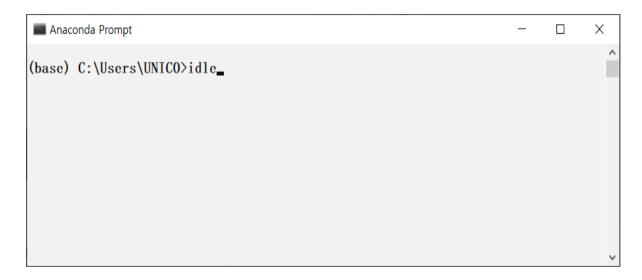






1 IDLE

- 1 Anaconda Prompt 창
 - 2 Anaconda Prompt 창에서 'idle'이라는 명령 입력





1 IDLE

- 2 IDLE 윈도우 창 출력
 - '>>>'는 프롬프트(Prompt)라고 함
 - 프롬프트에 파이썬 문법에 맞는 코드를 작성하고
 엔터키를 누르면 파이썬 인터프리터가 입력된 코드를 해석해 실제 동작 수행

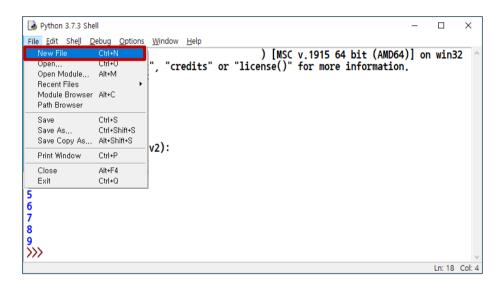


- 1 IDLE
 - ③ IDLE 창에서 파이썬 코드 실행
 - 1 코드를 작성하고 엔터키를 누를 때마다 해당 코드를 수행시키는 방식으로 대화식 프로그래밍 가능



1 IDLE

- ③ IDLE 창에서 파이썬 코드 실행
 - 2 파이썬 코드를 입력하는 코드 단위 수행, 파이썬 파일 생성 후 수행도 가능
 - File 메뉴의 New File 메뉴 선택



 소스 편집기 창이 독립적으로 출력되며, 하나의 모듈로 구성하려는 파이썬 코드 작성

```
*untitled* - - X

File Edit Format Run Options Window Help

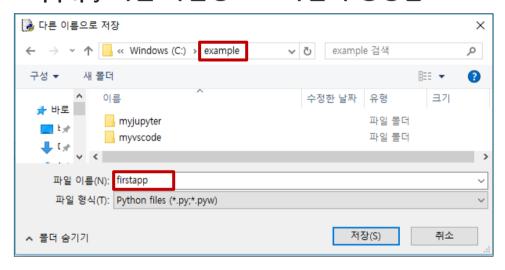
v1 = 5
v2 = 10
print(v1+v2)
print(v1*v2)
for i in range(v1, v2):
    print(i)

Ln: 6 Col: 12
```

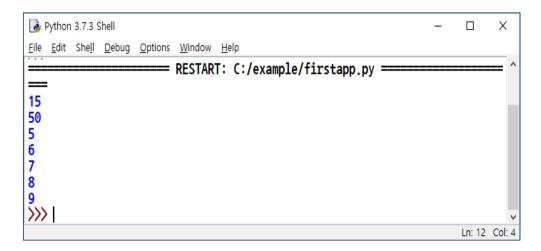


1 IDLE

- ③ IDLE 창에서 파이썬 코드 실행
 - 2 파이썬 코드를 입력하는 코드 단위 수행, 파이썬 파일 생성 후 수행도 가능
 - example 폴더에 firstapp 명칭으로 저장
 - firstapp.py 라는 파일명으로 파일이 생성됨



■ Run 메뉴의 Run Module을 선택하면 수행 결과가 IDLE 창에 출력





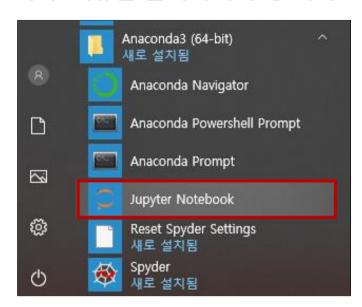
Z Jupyter Notebook

Jupyter Notebook 크롬과 같은 웹 브라우저에서 코드를 작성하고 실행이 가능한 툴



2 Jupyter Notebook

- ◆ Jupyter Notebook 실행
 - 1 Anaconda에는 Jupyter Notebook이 포함되어 있으므로 '시작' 메뉴를 클릭하여 수행 시작



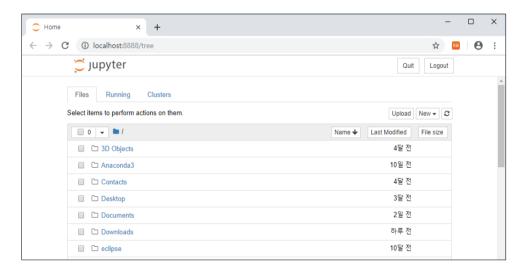
2 Jupyter Notebook 실행을 시작하는 윈도우 출력

```
💆 Jupyter Notebook
                                                                                                                 П
I 04:20:36.008 NotebookApp]
                             JupyterLab application directory is C:\Users\Samsung\Anaconda3\share\jupyter\lab
 I 04:20:36.011 NotebookApp]
                             Serving notebooks from local directory: C:\Users\Samsung
I 04:20:36.011 NotebookApp
                             The Jupyter Notebook is running at:
I 04:20:36.011 NotebookApp
                             http://localhost:8888/?token=eae5c62cf28c90d84d70bda78d8eae467b66e81e2761f715
[I 04:20:36.012 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 04:20:36.130 NotebookApp]
    To access the notebook, open this file in a browser:
        file:///C:/Users/Samsung/AppData/Roaming/jupyter/runtime/nbserver-16064-open.html
   Or copy and paste one of these URLs:
        http://localhost:8888/?token=eae5c62cf28c90d84d70bda78d8eae467b66e81e2761f715
```



2 Jupyter Notebook

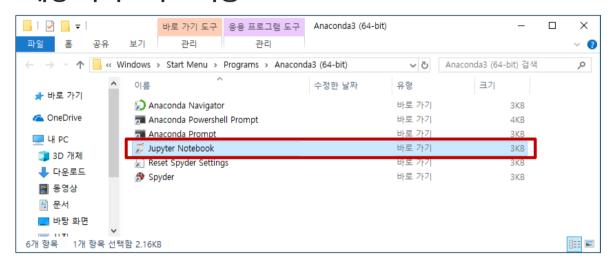
- ② Jupyter Notebook의 작업 페이지
 - 1 브라우저가 기동되고 Jupyter Notebook의 작업 페이지 출력
 - Jupyter Notebook 디폴트로 인식하는 시작 디렉토리가 시스템의 사용자 디렉토리이므로 c:/example/myjupyter로 변경 필요





Z Jupyter Notebook

- 🛂 Jupyter Notebook의 작업 페이지
 - 2 Jupyter Notebook의 시작 디렉토리 변경
 - 시작 메뉴에서 Jupyter Notebook 선택
 - Jupyter Notebook의 설치 디렉토리 파악 후 해당 디렉토리로 이동



■ Jupyter Notebook 파일의 속성 정보에 들어가 대상 정보 수정



파이썬 <u>개발 툴 소개</u>



2 Jupyter Notebook

- Jupyter Notebook의 작업 페이지
 - 2 Jupyter Notebook의 시작 디렉토리 변경
 - cmd 창에서 다음 명령을 수행시켜 Jupyter Notebook의 설정 파일 생성

jupyter notebook --generate-config

```
Microsoft Windows [Version 10.0.17134.706]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Samsung\jupyter notebook --generate-config
Writing default config to: C:\Users\Samsung\.jupyter\jupyter_notebook_config.py

C:\Users\Samsung\cd .jupyter

C:\Users\Samsung\.jupyter>
```

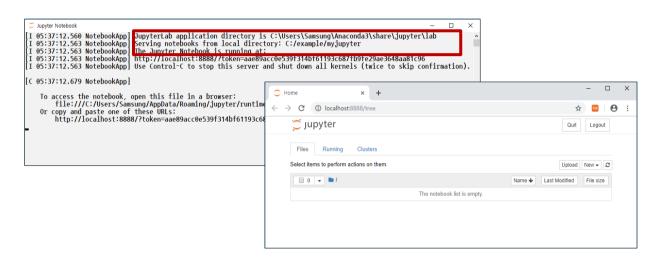
jupyter_notebook_config.py 파일에서 notebook_dir 항목을 찾아서 주석을 해제하고 'C:/example/myjupyter' 작성

The directory to use for notebooks and kernels.
c.NotebookApp.notebook_dir = 'C:/example/myjupyter'



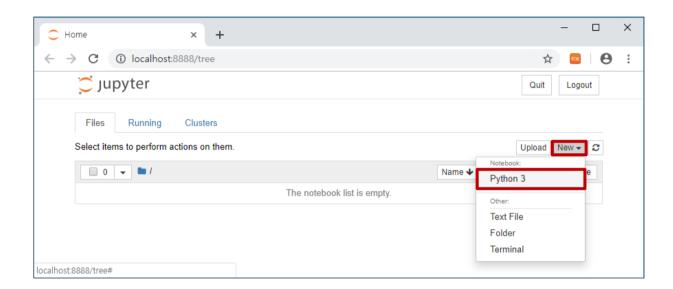
Z Jupyter Notebook

- **②** Jupyter Notebook의 작업 페이지
 - 4 Jupyter Notebook을 재기동하면 시작 디렉토리가 변경된 것을 볼 수 있음

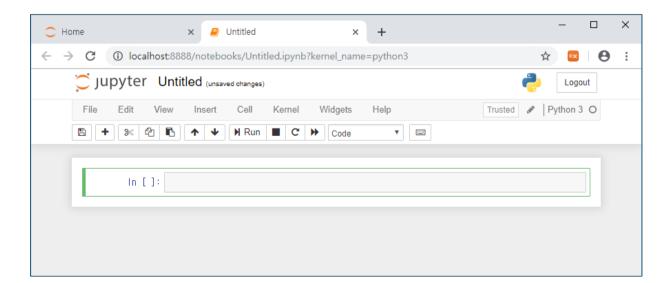




- 2 Jupyter Notebook
 - ③ 파이썬 코드 테스트
 - 1 우측에 있는 'New'를 누르고 Python 3 클릭

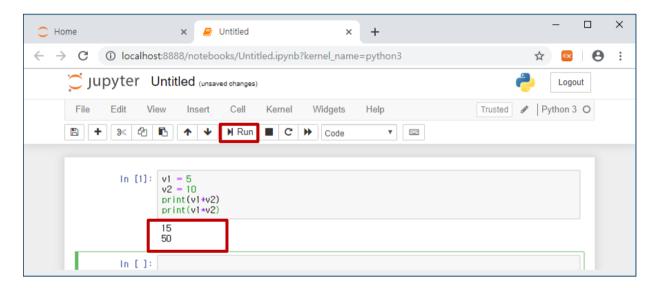


2 Untitled라는 이름의 새로운 창이 출력되고 파이썬 코드를 입력할 수 있는 박스 출력

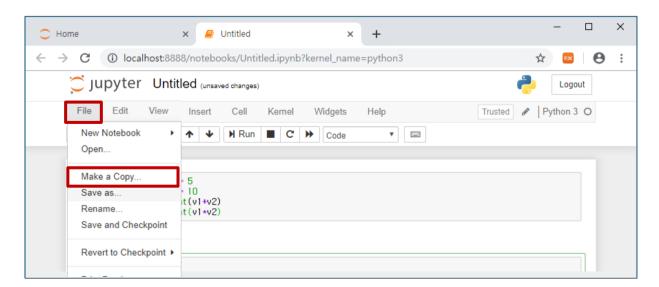




- 2 Jupyter Notebook
 - ❸ 파이썬 코드 테스트
 - 3 박스 안에 적당한 파이썬 코드를 작성하고 위에 있는 'Run' 버튼을 클릭하면 박스 아래에 수행 결과 출력

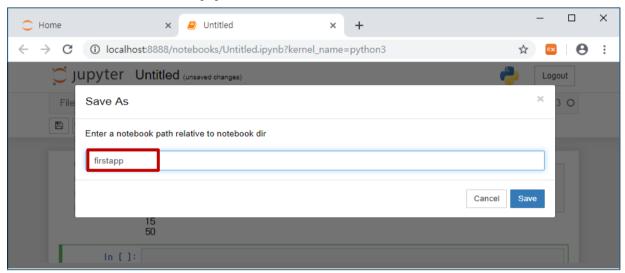


4 작성한 파이썬 코드를 파일로 저장하려면 File 메뉴를 선택하고 'Save as…' 메뉴 클릭





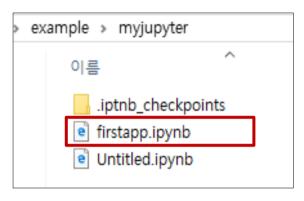
- 2 Jupyter Notebook
 - ③ 파이썬 코드 테스트
 - 5 파일명 입력
 - 여기서는 firstapp 입력

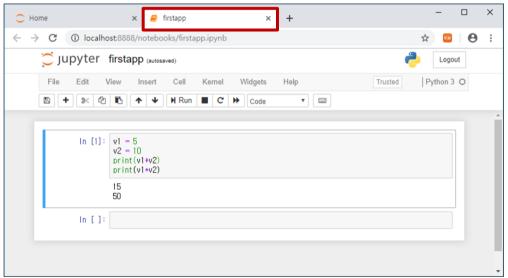




2 Jupyter Notebook

- ③ 파이썬 코드 테스트
 - 6 firstapp.ipynb라는 파일 생성, 창의 이름도 firstapp으로 변경됨

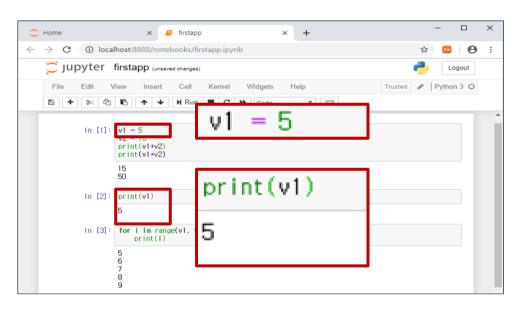






Z Jupyter Notebook

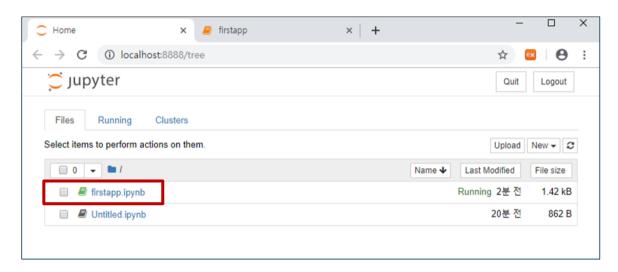
- 4 파이썬 소스코드 출력
 - 1 파이썬 소스코드를 입력하는 각각의 박스들은 하나의 파일 내에서 계속 이어짐
 - 위의 박스에서 선언한 v1 변수를 아래 박스들에서도 사용 가능





2 Jupyter Notebook

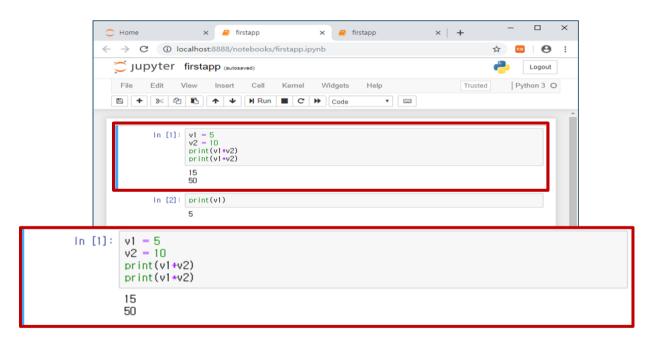
- 4 파이썬 소스코드 출력
 - 2 Jupyter Notebook를 재가동하고 http://localhost:8888/tree으로 요청
 - 시작 디렉토리에 저장된 firstapp.ipynb 파일이 리스트로 출력되는 것을 볼 수 있음





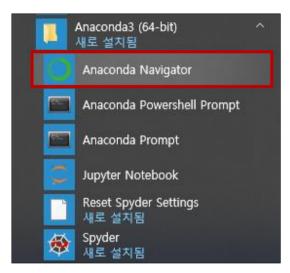
2 Jupyter Notebook

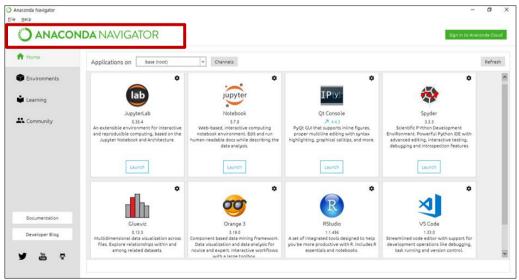
- 4 파이썬 소스코드 출력
 - 3 리스트에서 firstapp.ipynb 파일을 클릭하면 새로운 창이 출력
 - 마지막으로 저장된 소스코드의 내용이 출력되는 것을 볼 수 있음





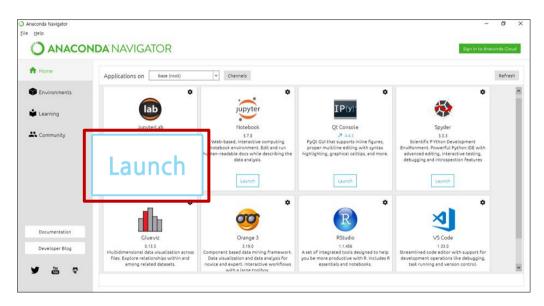
- 3 VS Code
 - 1 VS Code 설치
 - 1 시작메뉴에서 Anaconda Navigator를 선택하면 ANACONDA NAVIGATOR 창 출력



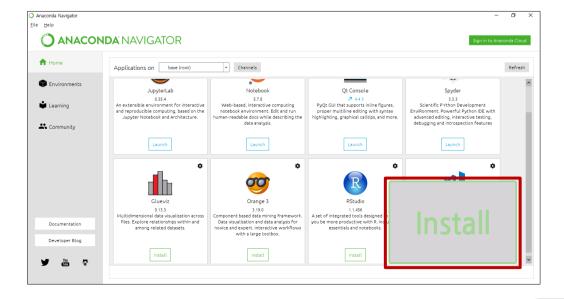




- 3 VS Code
 - ◆ VS Code 설치
 - 2 Anaconda가 설치되면서 함께 설치된 툴들은 'Launch' 버튼과 함께 출력

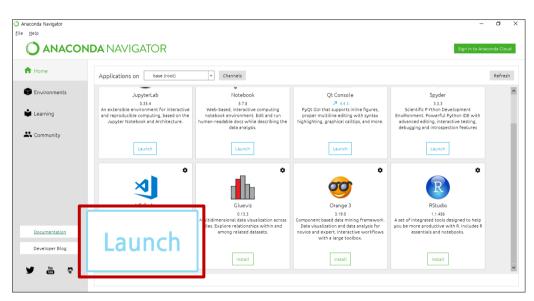


3 VS Code처럼 추가 설치 과정을 진행해야 하는 툴들은 'Install' 버튼과 함께 출력



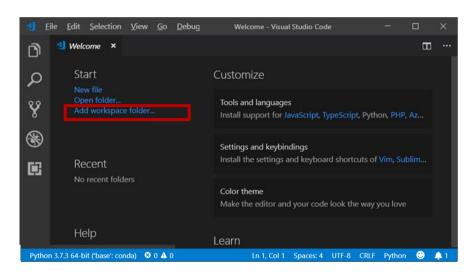


- 3 VS Code
 - ◆ VS Code 설치
 - 4 VS Code 설치가 완료되면 'Install' 버튼에서 'Launch' 버튼으로 변경

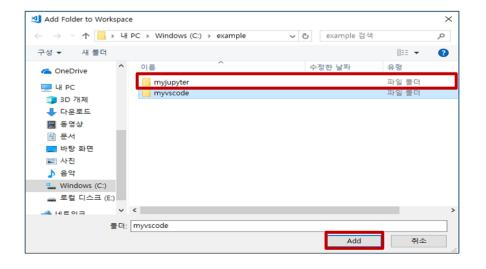




- 3 VS Code
 - **2** VS Code 실행
 - 1 'Launch' 버튼을 클릭하여 VS Code를 실행시킨 후 Welcome 화면에서 좌측의 'Add workspace folder...' 클릭

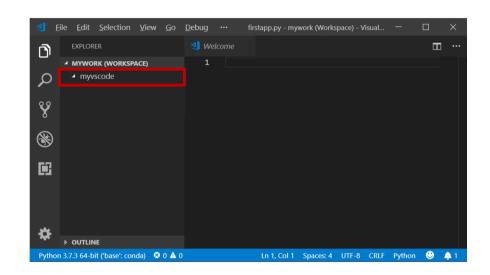


2 example 폴더에 미리 만들어 놓은 myvscode 폴더를 선택하고 'Add' 버튼 클릭

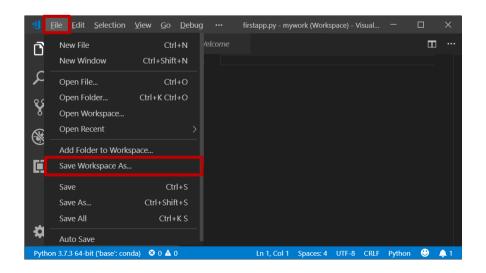




- 3 VS Code
 - 🛂 VS Code 실행
 - 3 myvscode라는 폴더가 WORKSPACE 폴더에 추가된 것을 볼 수 있음

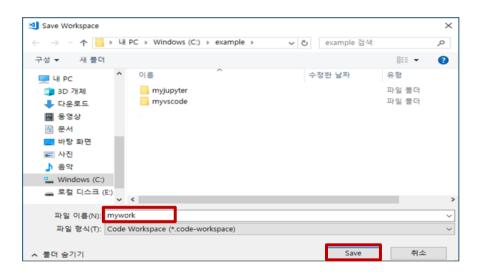


4 WORKSPACE명인 UNTITLED를 변경하기 위해 File 메뉴의 'Save Workspace As...' 메뉴 선택

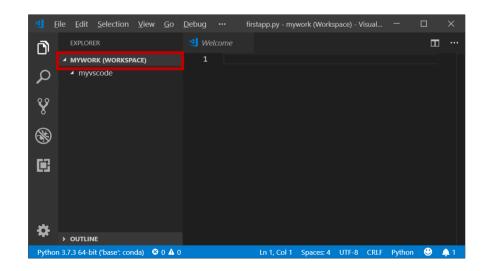




- 3 VS Code
 - **2** VS Code 실행
 - 5 파일 다이얼로그(Dialog)가 출력되면 example 폴더 아래에 mywork라는 명칭으로 Code Workspace가 생성되도록 입력 후 'Save' 버튼 클릭

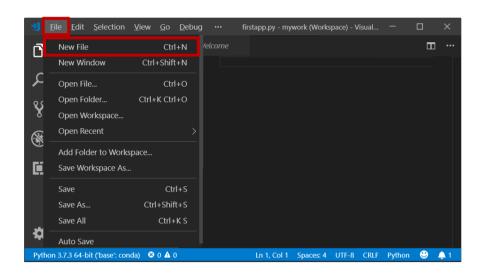


6 WORKSPACE명이 MYWORK로 변경된 것 확인

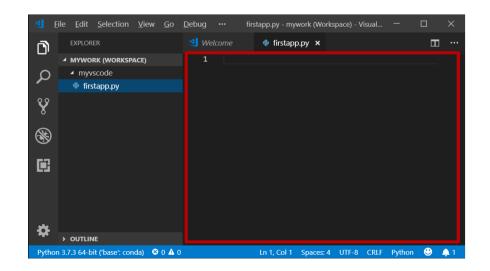




- 3 VS Code
 - ❖ VS Code로 파이썬 코드 수행
 - 1 파이썬 코드를 작성하여 수행하기 위해 myvscode 폴더 선택 후, 오른쪽 버튼을 클릭하여 서브메뉴가 출력되면 'New File' 메뉴 선택

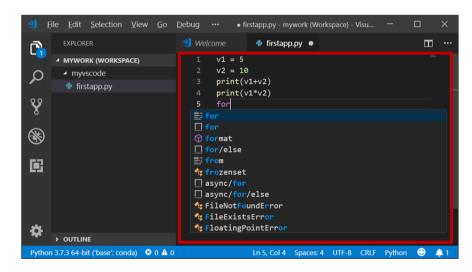


2 파일명으로 firstapp.py를 입력하면, 오른편에 코드 편집기 출력

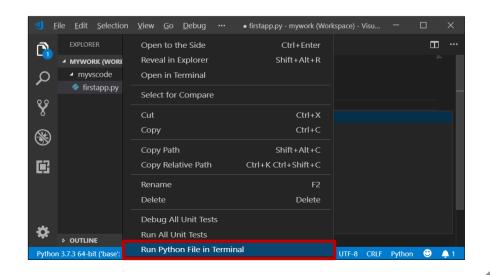




- 3 VS Code
 - ₃ VS Code로 파이썬 코드 수행
 - 3 수행하려는 파이썬 코드 입력
 - 입력하는 동안 함수 호출 방법 뿐만 아니라 파이썬 구문에 대해서도 Code Assistant를 지원하므로 코드 작성에 많은 도움이 됨

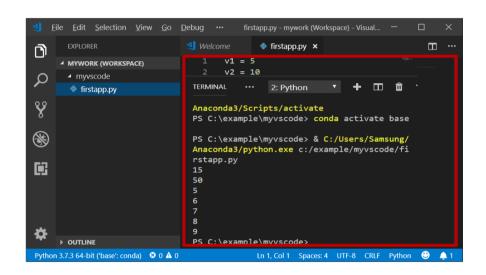


4 코드 작성이 종료되면 파일명을 선택한 후에 'Run Python File in Terminal'을 클릭하여 수행





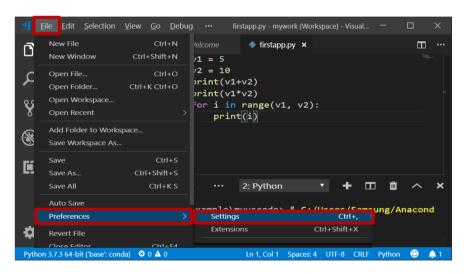
- 3 VS Code
 - ₃ VS Code로 파이썬 코드 수행
 - 5 오른쪽 하단에 Terminal 뷰가 출력되고 작성한 파이썬 코드 수행 결과 출력



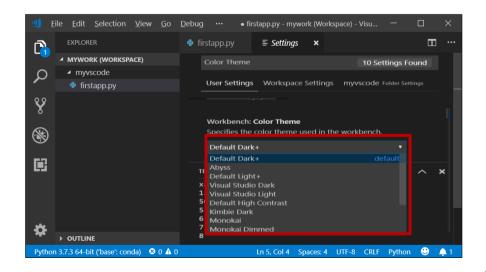


3 VS Code

- 4 VS Code 테마 변경
 - 1 VS Code 툴의 기본 컬러 테마(Color Theme)는 Dark이므로 검은 바탕으로 출력
 - 흰 바탕으로 변경하고자 하면 File 메뉴의 'Preferences' 메뉴와 'Settings' 메뉴를 차례대로 선택

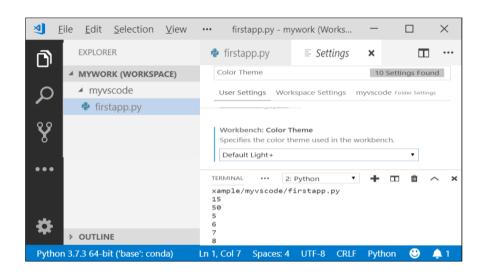


2 세팅하려는 항목으로 Color Theme를 입력하면 선택할 수 있는 테마 리스트 출력

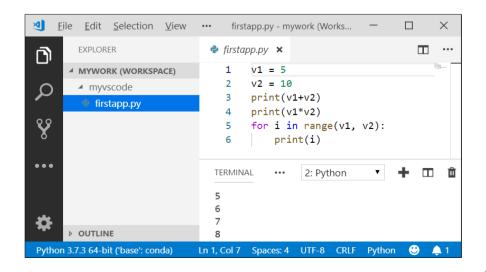




- 3 VS Code
 - 4 VS Code 테마 변경
 - 3 컬러 테마를 Default Dark에서 Default Light로 변경



- 4 코드 컬러 테마를 Default Light로 변경한 후의 코드 편집기 화면
 - 검은 바탕과 흰 바탕을 비교해 보고 원하는 것 선택 가능





3 VS Code

- 4 VS Code 테마 변경
 - 5 VS Code가 디폴트(Default)로 내장하고 있는 코드 테마에서 원하는 것을 찾지 못한 경우, 아래 사이트에서 확장 테마에 대한 정보를 얻고 추가 설치 가능

https://code.visualstudio.com/docs/getstarted/themes

6 Dracula 테마 설치를 위해 아래 사이트에 방문 후 하단에 있는 'Visual Studio Code' 클릭

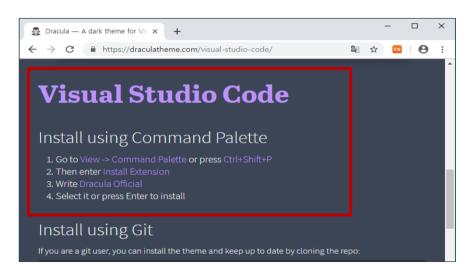
https://draculatheme.com



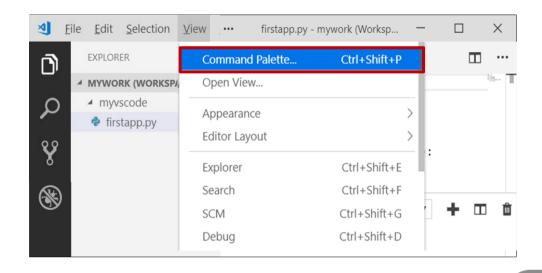




- 3 VS Code
 - 4 VS Code 테마 변경
 - 7 Visual Studio Code에서 설치 방법 확인 후, Command Palette를 이용하여 설치하는 방법으로 설치 진행

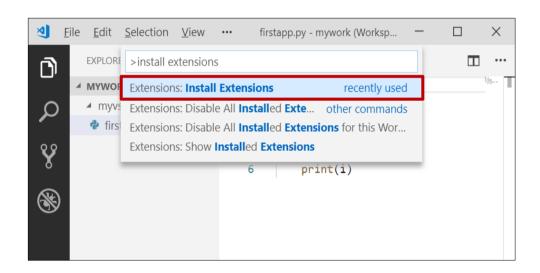


8 View 메뉴에서 'Command Palette...' 메뉴 선택

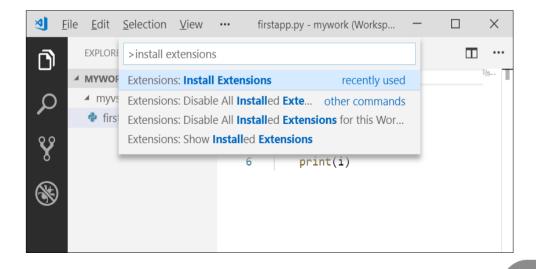




- 3 VS Code
 - 4 VS Code 테마 변경
 - 9 Install Extensions를 입력하여 'Extensions: Install Extensions' 선택

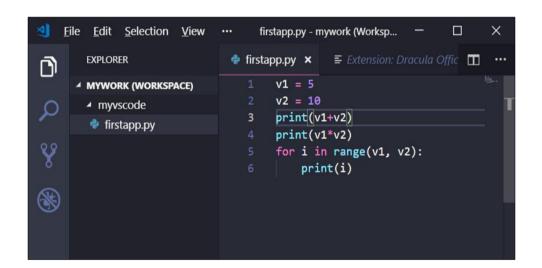


10 왼편에 출력된 Extensions: Install Extensions 뷰에서 dra까지만 입력하면 제일 위에 Dracula Official 항목이 출력되고 'Install' 버튼을 클릭하여 설치 진행





- 3 VS Code
 - 4 VS Code 테마 변경
 - 11 설치가 완료된 후 출력한 VS Code의 편집기 화면
 - 다양한 컬러 테마가 지원되므로 원하는 컬러 테마를 설치하여 사용 가능



학습정리

1. 파이썬 소개와 개발 환경 구축 방법



- 파이썬은 이해하기 쉽고 유연한 문법을 가지고 있으며 가독성이 좋고, 간결하며, 스탠다드 라이브러리가 잘 갖춰져 있음
- Anaconda는 파이썬 실행 환경을 포함하므로 파이썬 개발 환경을 따로 설치하지 않아도 됨
- Anaconda는 수학·과학·데이터 분석 분야에서 필요한 거의 모든 패키지들(Numpy, SciPy, Pandas, Matplotlib 등)이 포함되어 있으며 데이터 분석을 위해서 파이썬 개발 환경을 준비할 때 Anaconda가 가장 많이 선택되고 있음

학습정리



- 파이썬 프로그램을 개발하는데 있어서 선택할 수 있는 개발 툴은 IDLE, Jupyter Notebook, VS Code 등 다양함
- 파이썬이 자체적으로 지원하는 개발 툴은 IDLE이며 Anaconda에서 Jupyter Notebook을 포함하고 제공함
- VS Code는 Anaconda Navigator에서 확장 프로그램으로 설치하여 사용 가능함