

# 2025년 1학기 공공데이터분석

데이터과학부 정진명  
([jmjung@suwon.ac.kr](mailto:jmjung@suwon.ac.kr), 글로벌경상관 918호)

# Contents

- anaconda
- jupyter lab
- Pandas, matplotlib 문제 풀이

# anaconda 개요

# Anaconda

## Anaconda

Python interpreter

+

개발환경 (jupyter lab)

+

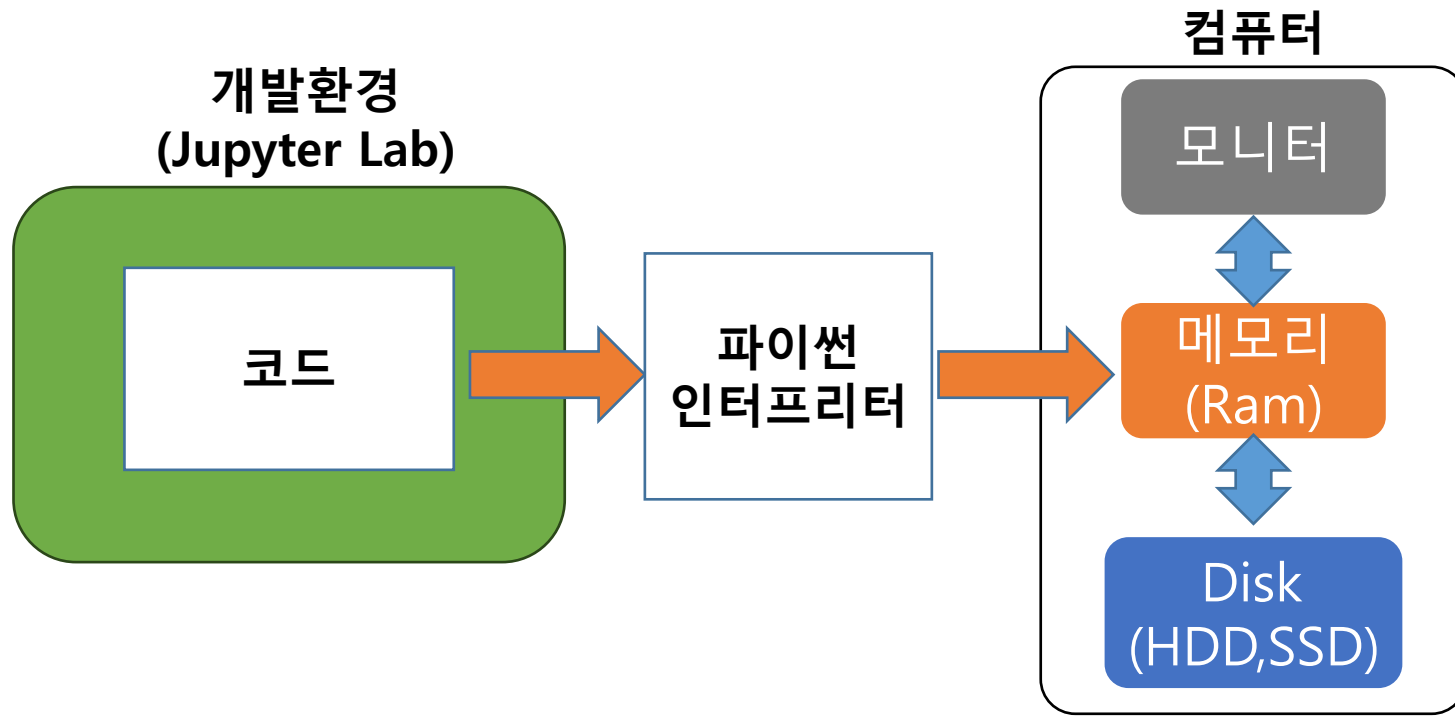
450 여개의 package 보유

(데이터과학에서 주로 사용하는 아래의 6개 package 포함)

name	Summary / License
<a href="#">matplotlib</a>	Publication quality figures in Python / PSF-based
<a href="#">numpy</a>	Array processing for numbers, strings, records, and objects. / BSD 3-Clause
<a href="#">pandas</a>	High-performance, easy-to-use data structures and data analysis tools. / BSD 3-clause
<a href="#">scikit-learn</a>	A set of python modules for machine learning and data mining / BSD 3-Clause
<a href="#">scipy</a>	Scientific Library for Python / BSD 3-Clause
<a href="#">tensorflow</a>	TensorFlow is a machine learning library. / Apache 2.0

# python interpreter

- python interpreter 에는 각 **문법**에 대한 **실행방침**이 명시되어 있음
- 작성한 코드를 **실행하면** python interpreter의 실행 방침에 따라 주로
  - 1) 컴퓨터의 **메모리**에 여러 **변수 (또는 함수)**를 생성
  - 2) 컴퓨터의 **메모리**에 존재하는 **변수**를 업데이트함



\* python 코드의 문법이 업데이트 됨에 따라, **버전**이 증가

# 패키지 (package)

- 특정 기능을 수행하는 모듈(함수)들의 묶음
- 분야 별로 나누어져 있음 (통계, 머신러닝, 시각화, ...)
- 패키지 함수 예제

```
import numpy  
numpy.var(l1)
```

3.44

**numpy Package에 정의된 함수 사용하는 경우**

```
l1=[2,3,4,6,7]  
m1=sum(l1)/len(l1)  
print(m1)  
  
l2=[(ii-m1)*(ii-m1) for ii in l1]  
var=sum(l2)/len(l1)  
print(var)
```

4.4

3.44

**직접 코딩하는 경우**

- 같은 결과 도출
- package 함수 사용이 훨씬 간단
- 패키지에 포함된 함수들을 알고 잘 사용하는 것도 중요한 프로그래밍 능력

# 패키지 (package) 버전

- package별로 구성 함수에 따라 **버전**이 다름

## package A (v1.1)

- func1
- func2
- func3

## package A (v1.2)

- func1
- func2 (수정)
- func3
- func4

# anaconda 설치



# Anaconda 설치 (1/2)

- Anaconda 설치 방법
  - 아래의 site에 가서 본인 컴퓨터 OS bit 에 맞는 (보통 64bit) anaconda를 다운 받은 후 설치 한다.  
(<https://www.anaconda.com/download> )

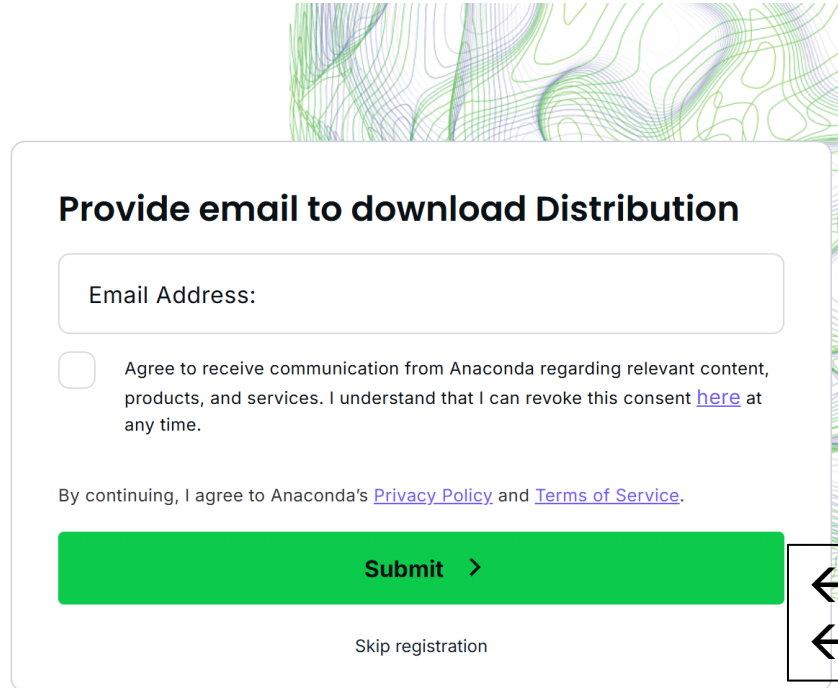
## Distribution

### Free Download\*

Register to get everything you need to get started on your workstation including Cloud Notebooks, Navigator, AI Assistant, Learning and more.

- ✓ Easily search and install thousands of data science, machine learning, and AI packages
- ✓ Manage packages and environments from a desktop application or work from the command line
- ✓ Deploy across hardware and software platforms
- ✓ Distribution installation on Windows, MacOS, or Linux

\*Use of Anaconda's Offerings at an organization of more than 200 employees requires a Business or Enterprise license. [See Pricing](#)



The screenshot shows a registration form titled "Provide email to download Distribution". It includes an "Email Address:" input field, a checkbox for agreeing to communication, and a green "Submit >" button. Below the button is a link for "Skip registration".

**Provide email to download Distribution**

Email Address:

☐ Agree to receive communication from Anaconda regarding relevant content, products, and services. I understand that I can revoke this consent [here](#) at any time.

By continuing, I agree to Anaconda's [Privacy Policy](#) and [Terms of Service](#).

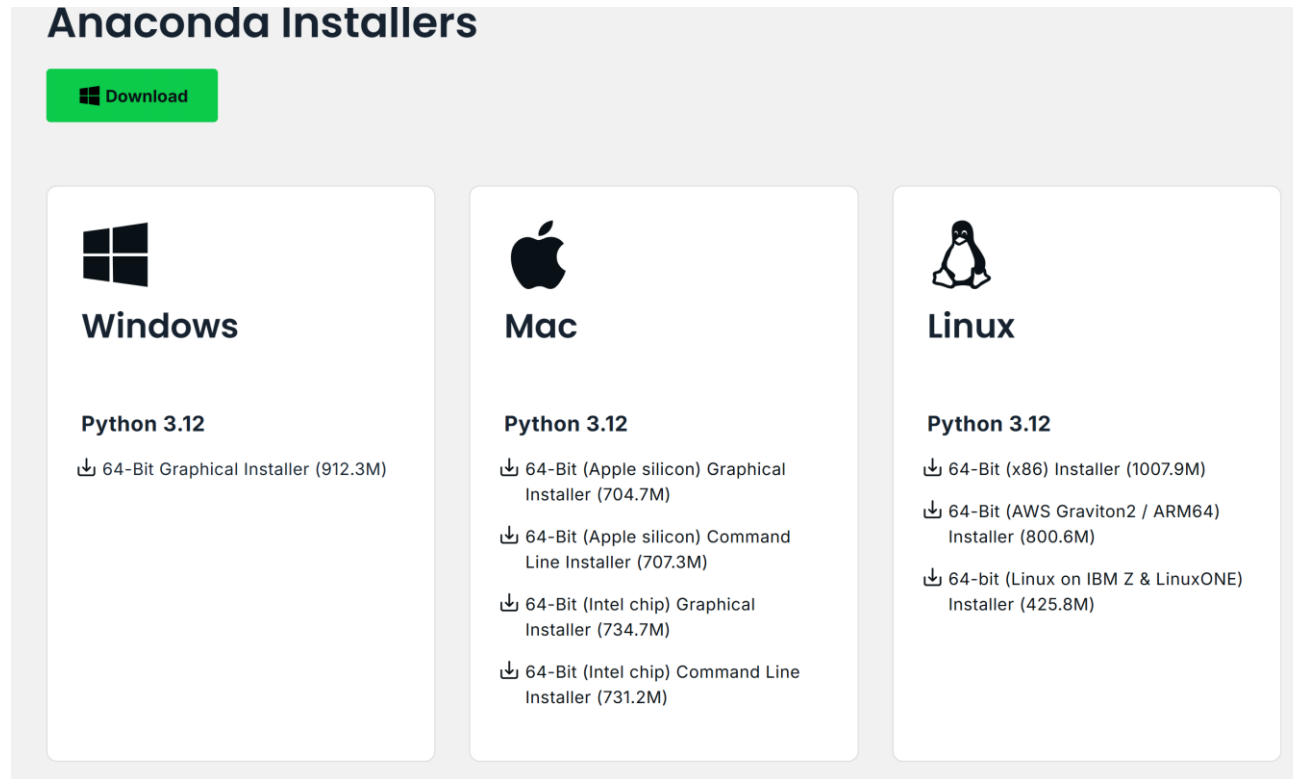
**Submit >**

[Skip registration](#)

← 위에 이메일 입력 후 submit 또는  
← 그냥 skip

# Anaconda 설치 (2/2)

- 본인 컴퓨터 OS에 맞는 anaconda를 다운 받은 후 설치 한다.



- 다운받은 후 실행하고, 변경사항 없이 다음 (next)를 여러 번 누르면 설치 됨 (소요시간 약 15분 이내, 컴퓨터 사양에 따라 다를 수 있음)

# Anaconda uninstall 방법

## ■ uninstall 방법

아래 경로의 **Anaconda3** 폴더에 가서 **Uninstall-Anaconda3** 실행



내 PC > 로컬 디스크 (C:) > 사용자 > **jmjung** > Anaconda3  
(또는 Users)

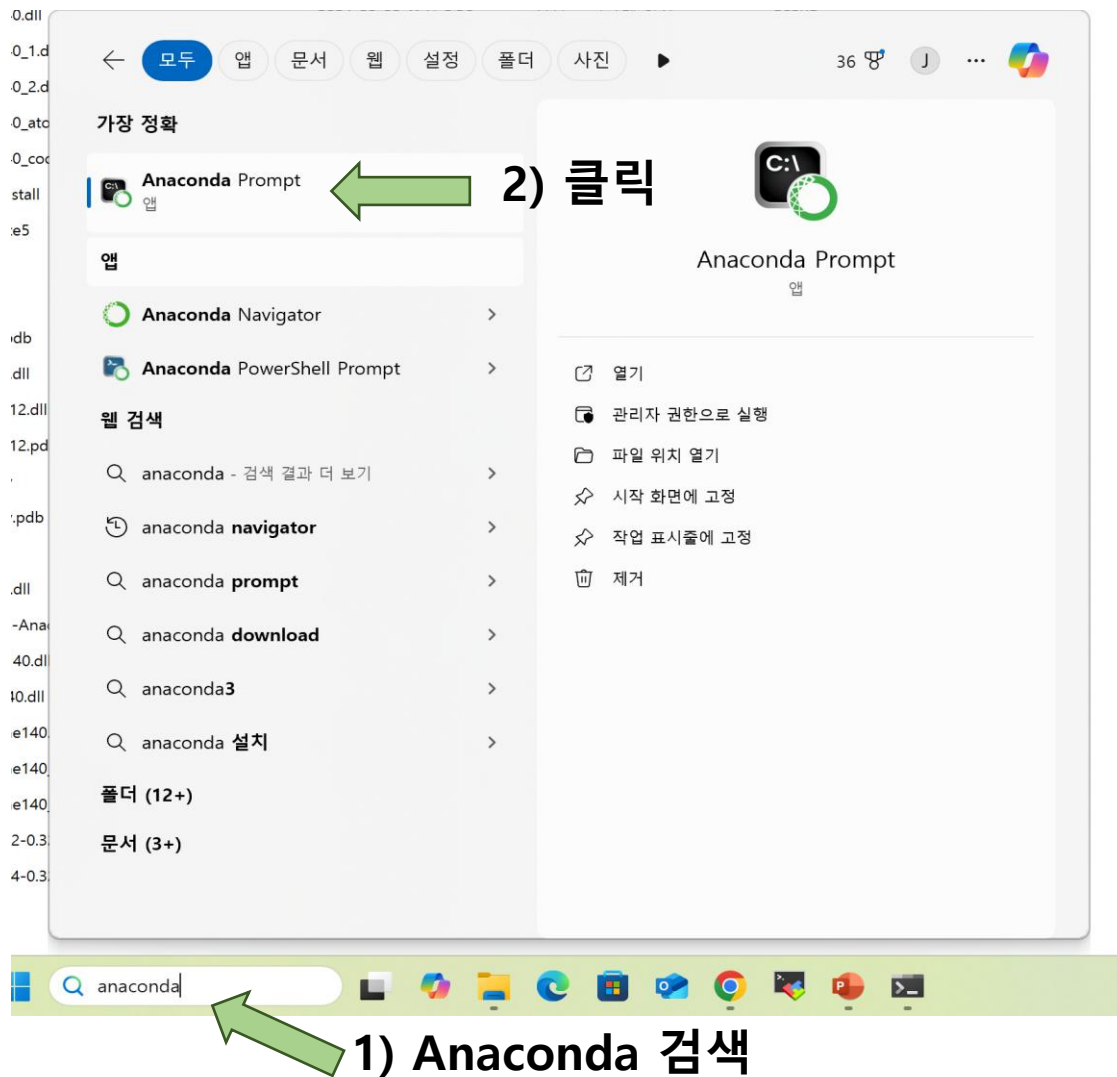
이 **jmjung** <user name>은 컴퓨터마다 다른  
컴퓨터 로그인 화면에 나오는 이름



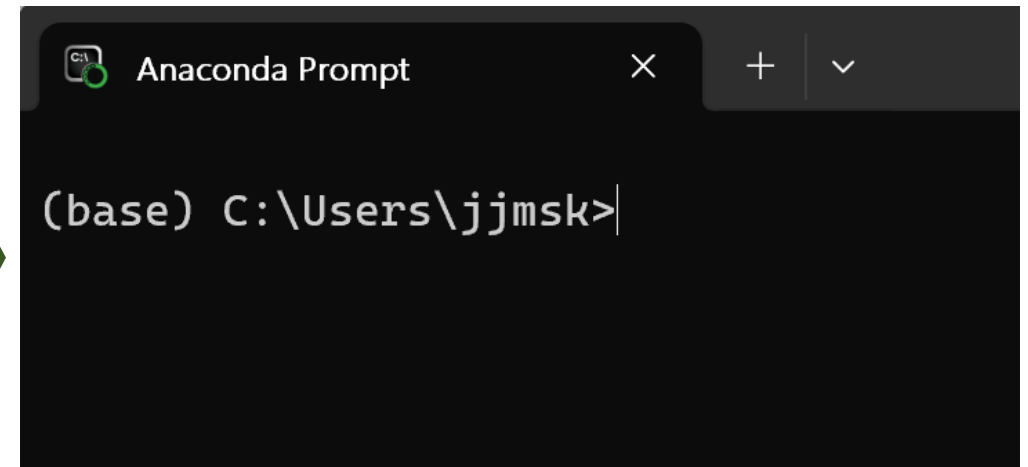
qt.conf	CONF 파일	1KB
ucrtbase.dll	응용 프로그램 확장	993KB
<b>Uninstall-Anaconda3</b>	응용 프로그램	304KB
vccorlib140.dll	응용 프로그램 확장	358KB

설치된 package 버전 확인

# anaconda prompt 실행



## anaconda 환경



# 설치된 interpreter, package 버전 확인

- interpreter: `python --version`
- package: `conda list <package name>`
  - **package name:**  
*numpy, pandas, matplotlib, seaborn, folium ...*

- python interpreter: 3.12.7
- pandas: 2.2.2
- numpy: 1.26.4
- matplotlib: 3.9.2
- seaborn: 0.13.2
- folium: 없음

- 버전이 조금 달라도 상관없음
- 수업 제공 코드가 오류 나거나, 또는 실행 결과가 다른 경우 알려주기 바람

```
Anaconda Prompt
(base) C:\Users\jjmsk>python --version
Python 3.12.7

(base) C:\Users\jjmsk>conda list pandas
# packages in environment at C:\Users\jjmsk\anaconda3:
#
# Name                        Version                        Build      Channel
pandas                        2.2.2                        py312h0158946_0

(base) C:\Users\jjmsk>conda list numpy
# packages in environment at C:\Users\jjmsk\anaconda3:
#
# Name                        Version                        Build      Channel
numpy                        1.26.4                        py312hfd52020_0
numpy-base                   1.26.4                        py312h4dde369_0
numpydoc                      1.7.0                        py312haa95532_0

(base) C:\Users\jjmsk>conda list matplotlib
# packages in environment at C:\Users\jjmsk\anaconda3:
#
# Name                        Version                        Build      Channel
matplotlib                    3.9.2                        py312haa95532_0
matplotlib-base               3.9.2                        py312hbd63d0_0
matplotlib-inline              0.1.6                        py312haa95532_0

(base) C:\Users\jjmsk>conda list seaborn
# packages in environment at C:\Users\jjmsk\anaconda3:
#
# Name                        Version                        Build      Channel
seaborn                       0.13.2                       py312haa95532_0

(base) C:\Users\jjmsk>conda list folium
# packages in environment at C:\Users\jjmsk\anaconda3:
#
# Name                        Version                        Build      Channel

(base) C:\Users\jjmsk>
```

# 새로운 package 설치

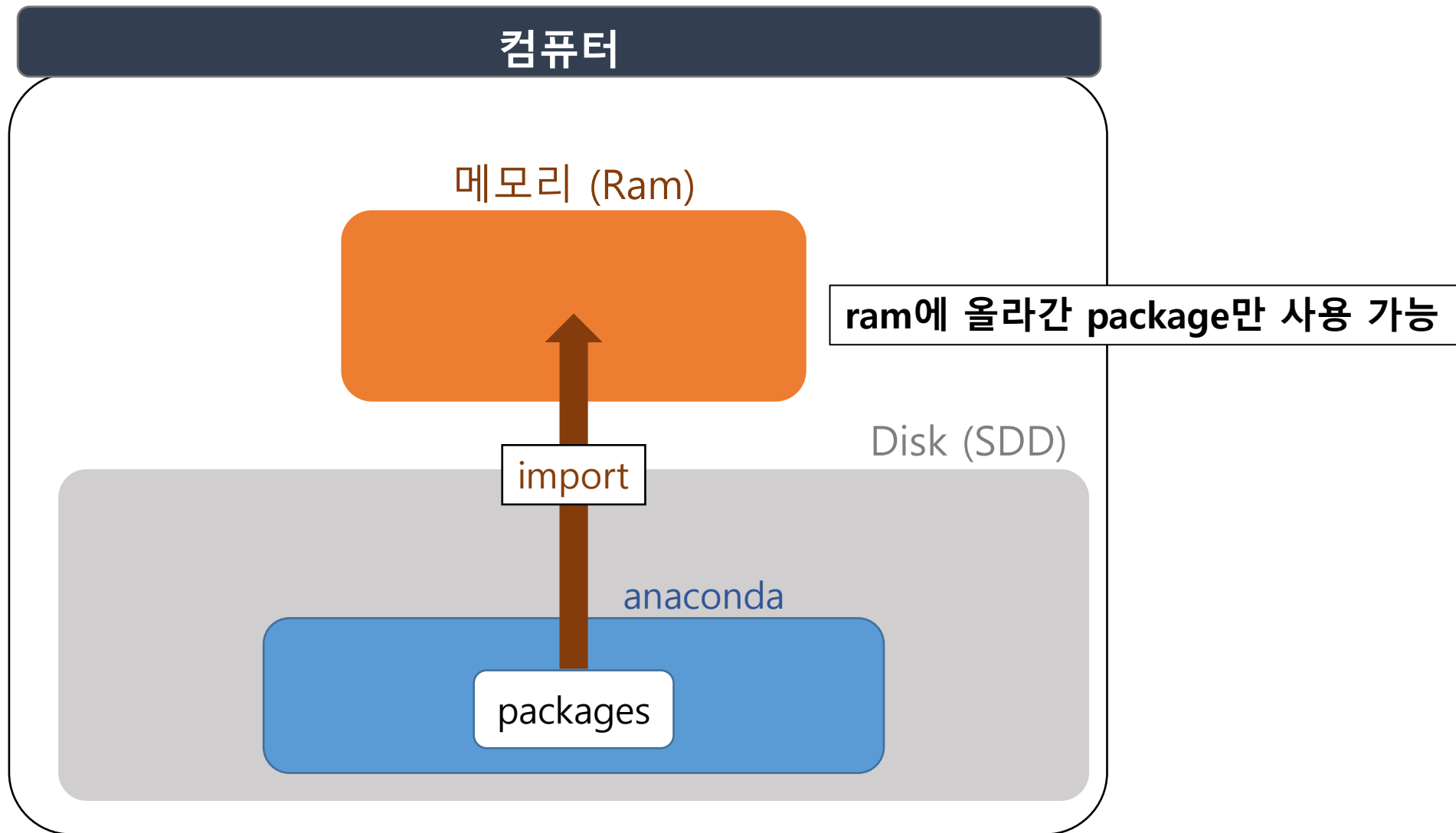
```
(base) C:\Users\jjmsk>pip install folium
Collecting folium
  Downloading folium-0.19.5-py2.py3-none-any.whl.metadata
Collecting branca>=0.6.0 (from folium)
  Downloading branca-0.8.1-py3-none-any.whl.metadata
Requirement already satisfied: Jinja2>=2.9 in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from folium)
Requirement already satisfied: numpy in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from folium)
Requirement already satisfied: requests in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from folium)
Requirement already satisfied: xyzservices in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from folium)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from Jinja2>=2.9->folium)
Requirement already satisfied: charset-normalizer<4,>=2.0 in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from requests->folium)
Requirement already satisfied: idna<4,>=2.5 in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from requests->folium)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from requests->folium)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\jjmsk\anaconda3\envs\base\lib\site-packages (from requests->folium)
Downloading folium-0.19.5-py2.py3-none-any.whl (110 kB)
Downloading branca-0.8.1-py3-none-any.whl (26 kB)
Installing collected packages: branca, folium
Successfully installed branca-0.8.1 folium-0.19.5

(base) C:\Users\jjmsk>conda list folium
# packages in environment at C:\Users\jjmsk\anaconda3:
#
# Name                               Version    Bu
folium                               0.19.5     pyp
```

- pip install <package name>
  - package name: *folium*, *geopandas*, ...

- 
- python interpreter: 3.12.7
  - pandas: 2.2.2
  - numpy: 1.26.4
  - matplotlib: 3.9.2
  - seaborn: 0.13.2
  - folium: 0.19.5

# import package





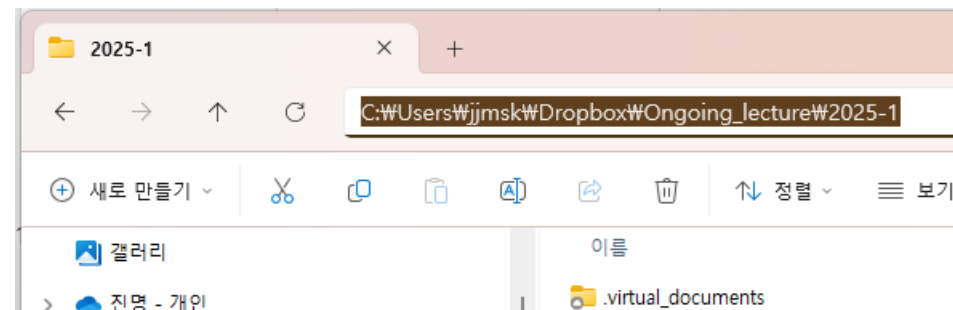
jupyter lab

# Jupyter lab 실행방법 1

## 1. Anaconda prompt 접속

## 2. (옵션) 작업할 폴더로 이동:

- 1) window 탐색기에서 해당 폴더로 이동 후 주소복사
- 2) *cd* (change directory) 명령 사용



```
(base) C:\Users\jjmsk>cd C:\Users\jjmsk\Dropbox\Ongoing_lecture\2025-1\파이썬 데이터 분석\pd_2w
```

```
(base) C:\Users\jjmsk\Dropbox\Ongoing_lecture\2025-1\파이썬 데이터 분석\pd_2w>jupyter lab
```

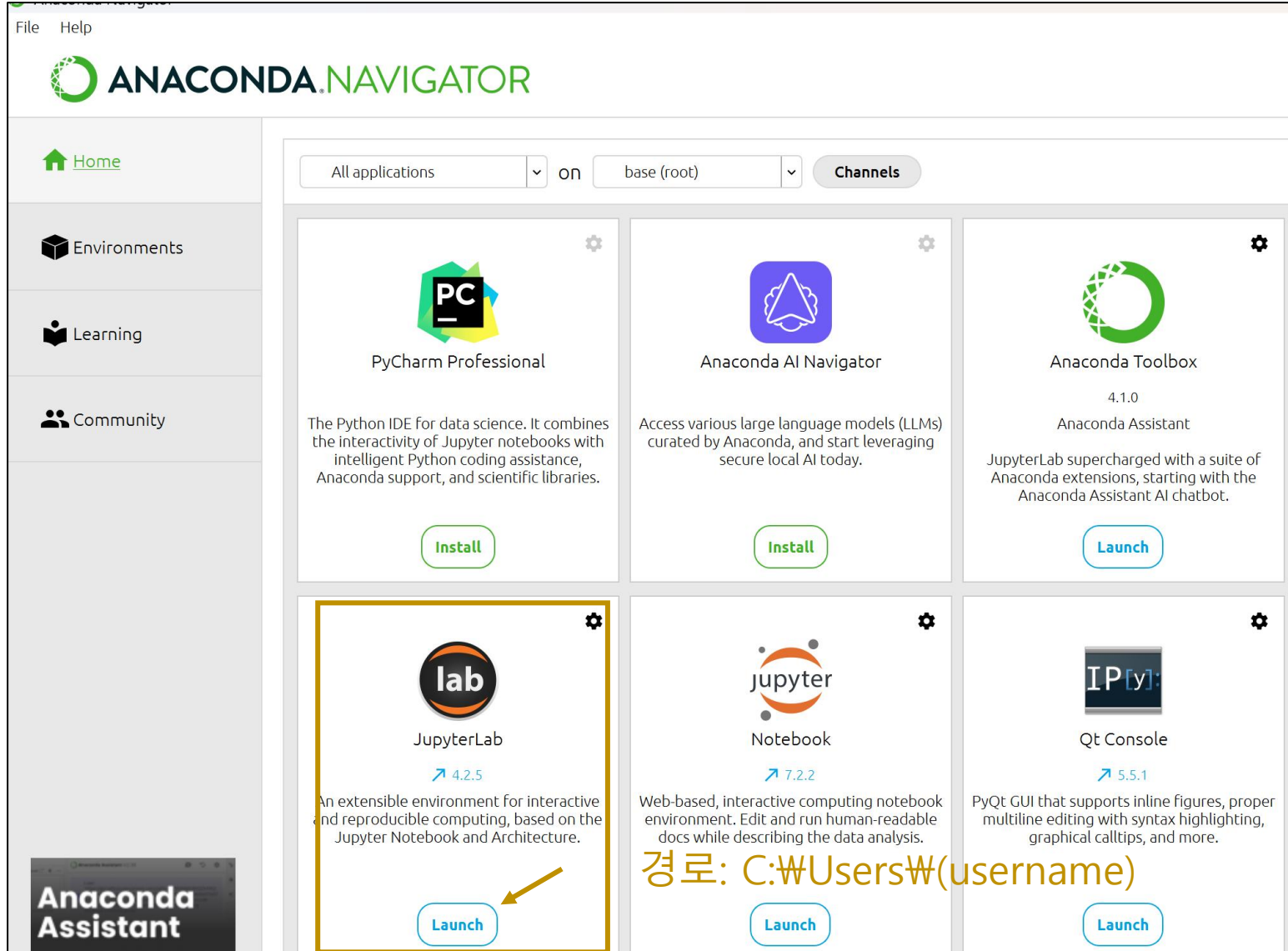
## 3. jupyter lab 실행

- 2번 없이 실행 시 경로: C:\Users\W(username)

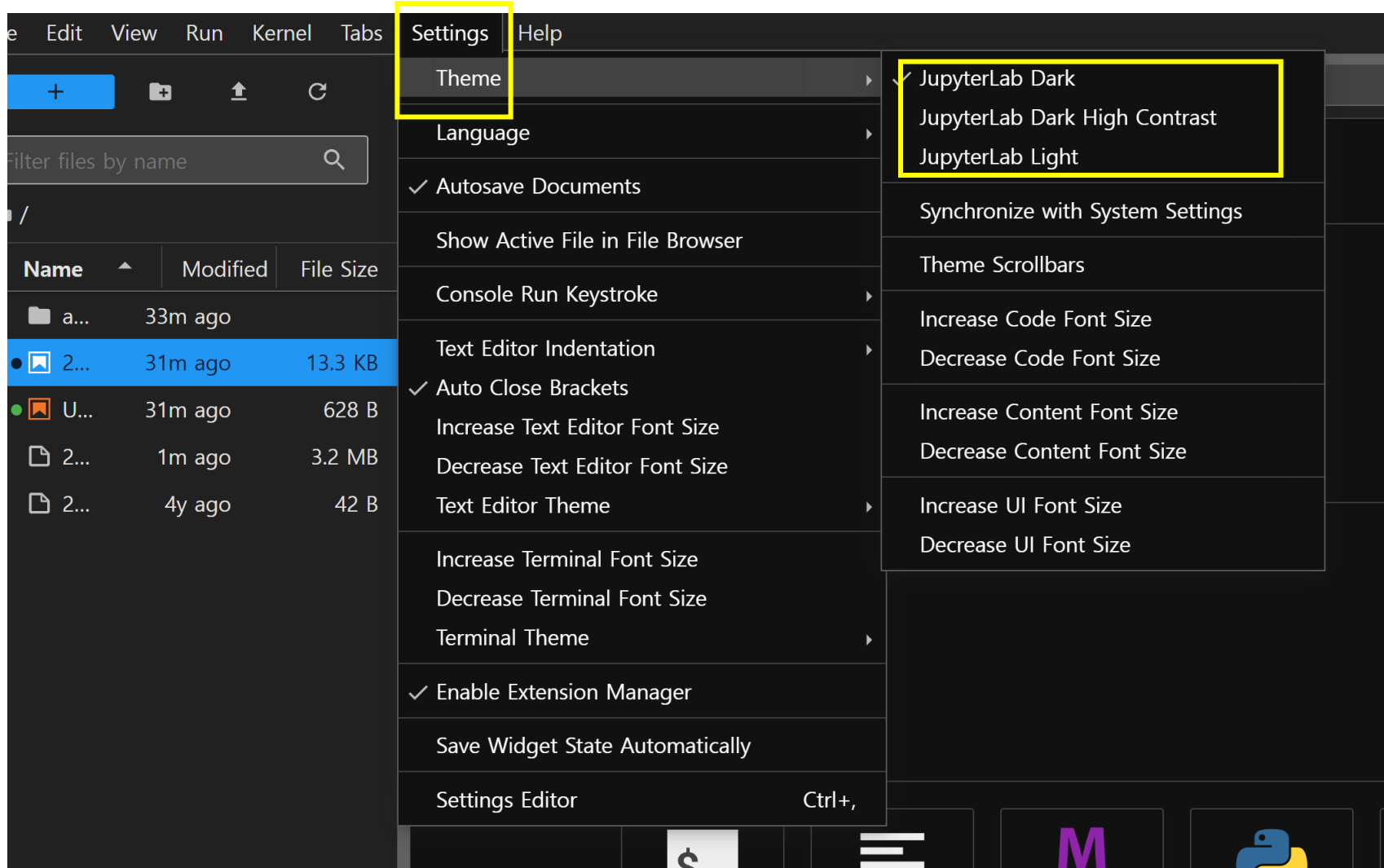
- ❖ anaconda prompt에서 작업할 폴더(ipynb파일이 있는폴더)로 이동 후에 jupyter lab 실행하면 편리함.
- ❖ Jupyter lab 실행 후, 실행한 폴더의 하위폴더들은 이동이 가능하지만 상위폴더로는 이동하지 못함

# Jupyter lab 실행방법 2

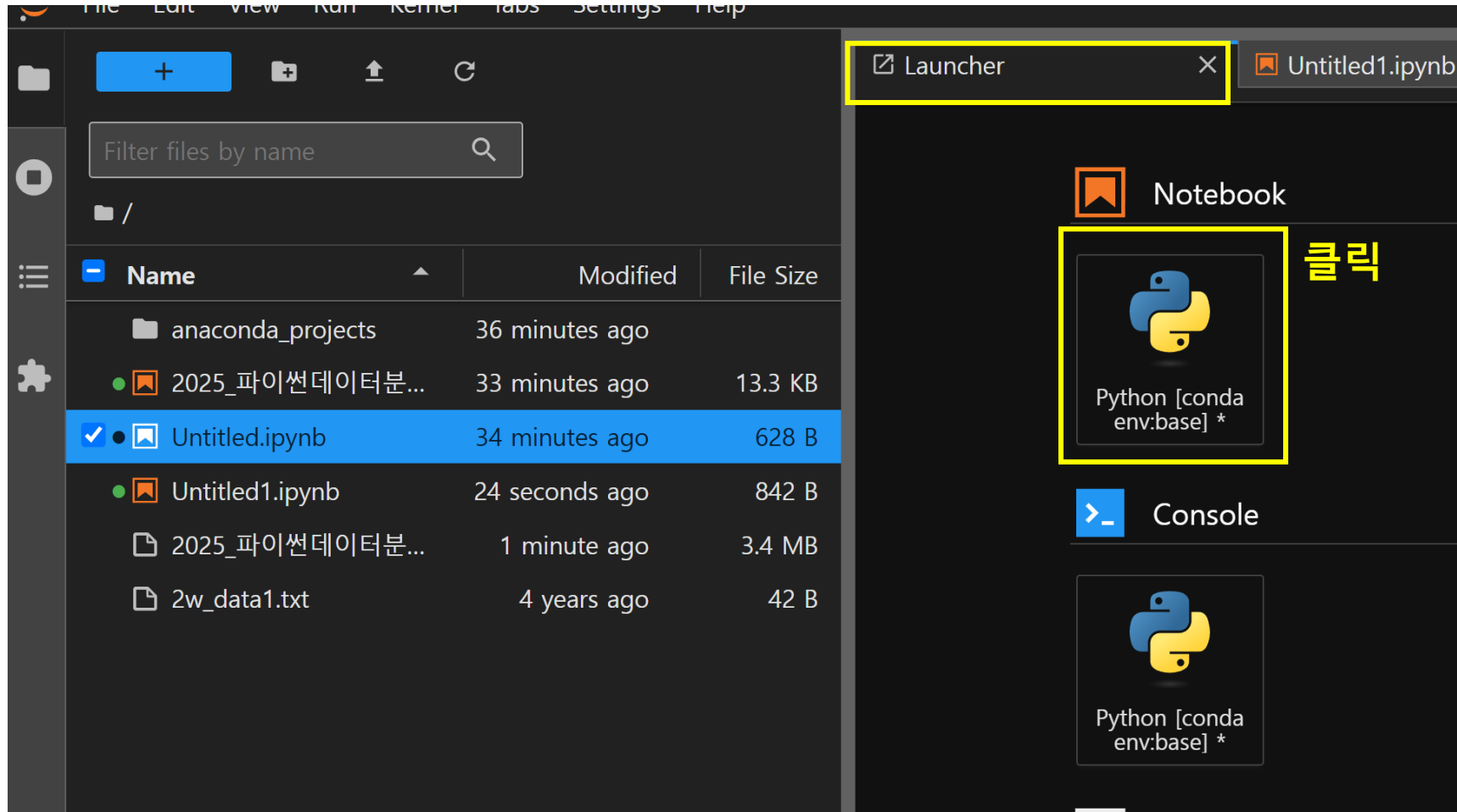
- Anaconda Navigator



# Jupyter lab 배경설정



# Jupyter lab ipynb 열기 (새로운 파일)



# Jupyter lab ipynb 열기 (기존 파일)

The screenshot displays the Jupyter Lab interface. On the left, the file browser shows a list of files in the 'anaconda\_projects' directory. The file '2025\_파이썬데이터분석...' is selected, highlighted with a yellow box, and labeled '클릭' (Click). The main area shows the code editor for 'Untitled1.ipynb'. The code is in Korean and checks the versions of Python, NumPy, and Pandas. The output shows the versions: Python 3.12.7, NumPy 1.26.4, and Pandas 2.2.2.

File browser table:

Name	Modified	File Size
anaconda_projects	37 minutes ago	
2025_파이썬데이터분석...	35 minutes ago	13.3 KB
Untitled.ipynb	35 minutes ago	628 B
Untitled1.ipynb	1 minute ago	842 B
2025_파이썬데이터분석...	3 minutes ago	3.4 MB
2w_data1.txt	4 years ago	42 B

Code editor content:

```
[6]: import sys
print('python', sys.version)

import numpy as np
print('numpy', np.__version__)

import pandas as pd
print('pandas', pd.__version__)

# 결과 확인을 용이하게 하기 위한 코드
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = 'all'

python 3.12.7 | packaged by Anaconda, Inc. | (main, Oct 4 2024, 13:17:27) [MSC v.1929 64 bit (AMD64)]
numpy 1.26.4
pandas 2.2.2

[14]: a=10
b=20
a
b
```

# markdown으로 작성한 목차 확인

The screenshot shows the JupyterLab interface. On the left, the sidebar contains a menu icon (three horizontal lines) which is highlighted with a yellow box. A yellow callout box labeled '클릭' (Click) points to this icon, containing a list of items: '1. 버전 체크', '2. 문제', and '3. 풀이'. The main area displays a code cell with the following Python code:

```
[6]: import sys
      print('python', sys.version)

      import numpy as np
      print('numpy', np.__version__)

      import pandas as pd
```

# Highlight selected word

- 현재 Jupyter lab에는 highlight selected word 기능이 없음
- find (ctrl + f)로 구현 가능
  - 찾기를 원하는 변수 더블클릭 후 ctrl+f 수행

49

```
# 실습2

## 방법1
n=15
k=3

l1 = list(range(n))
l1
l2 = l1[:k]
l2
sum(l2)

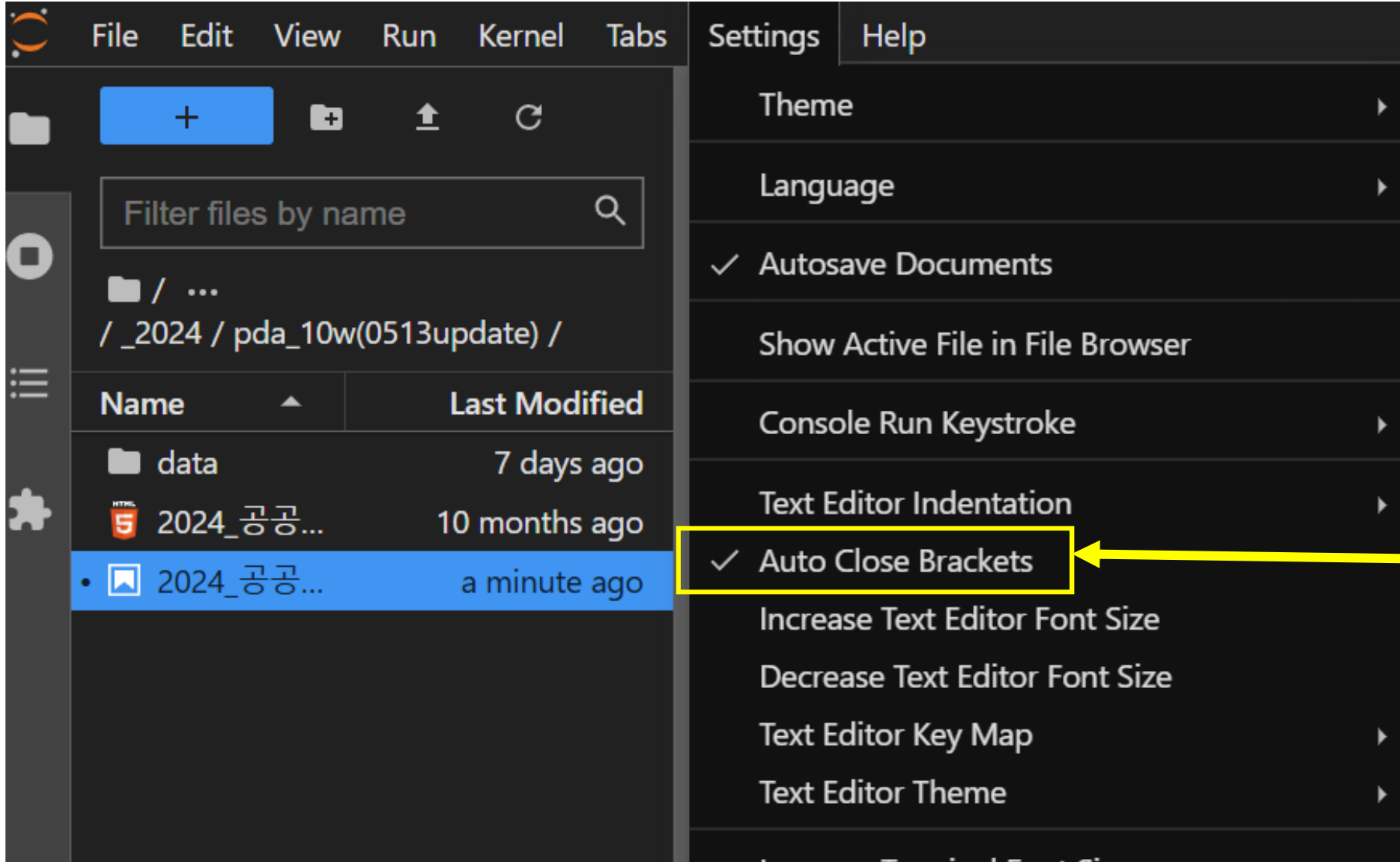
## 방법2
s1=0
for ii in range(n):
    if ii%k==0:
```

l2 A<sup>a</sup> ab .\* 2/3

- replace
- 대소문자
- Match whole word (exact matching)
- 2024.1 월 이후 버전에만 있음
- Anaconda prompt에서 *conda update jupyterlab* 으로 버전 업데이트 하면 사용 가능



# Auto close brackets



여는 괄호 작성시 닫는  
괄호 생성해 줌

# Jupyter lab 두 가지 모드

## 1. edit mode – cell 안 상태

- enter를 치거나 cell 안을 클릭하면 edit mode로 바뀜
- 코드 작성시에 사용

## 2. commend mode – cell 밖 상태

- ESC를 치거나 cell 밖을 클릭하면 commend mode로 바뀜
- commend mode에서는 타이핑을 해도 코드 작성이 되지않고, 미리 예약된 단축키가 작동함
- 코드 작성 외에 코드 관리에 편리한 작업을 수행

# 코드 실행 방법

## ■ 코드 실행 명령어

```
[17]: a=10  
      b=20  
      a  
      b  
  
[17]: 10  
[17]: 20
```

- cell 단위로 실행 → Interpreter에서 코드를 실행함 (그 전에는 X)
- 실행 명령어 **ctrl + enter**, **shift + enter**, **alt + enter**

- cell의 위치가 아닌 실행 순서에 의하여 코드가 실행 됨
- 셀을 블록으로 지정후 실행하면 블록지정된 셀 이 순차적으로 모두 실행됨
- 코드 실행 시 에러 종류
  - **구문오류**: 실행하기 전, interpreter 문법에 맞는지 체크해서 맞지 않으면 에러 발생, 코드 전체가 실행이 되지 않음
  - **실행오류**: 실행하는 중, 실행이 불가능하여 생기는 에러, 실행오류 발생하기 전까지의 코드는 실행됨

# 주요 단축키

## 주요 단축키 (command mode – cell 밖 상태)

- 아래 셀이랑 합치기 : shift + M
- 셀 내려두기 : x      셀 copy : c      셀 붙여넣기 : v or shift + v
- 셀 지우기 : d, d      셀 지우기 취소 : z
- Markdown 변환 : m      Code로 변환 : y
- Code line: l (소문자 L)
- Find and replace: f (해당 셀에만 적용)
- Find: ctrl+f (web page 기능)

## 주요 단축키 (edit mode – cell 안 상태)

- 툴팁 표시하기 : Shift + Tab (사용 함수 설명)
- 셀 나누기 : ctrl + shift + -
- 들여쓰기: tab, shift+tab
- 변수나 함수 작성시 tab: 자동완성기능
- 입력 취소: ctrl + z
- 입력 취소 되돌리기 : ctrl + y
- 주석처리: ctrl + / (블록지정 후에도 가능)

### ■ 단축키 모음

<http://intellegibilisverum.tistory.com/entry/IPython-notebook-%EB%8B%A8%EC%B6%95%ED%82%A4-%EB%AA%A8%EC%9D%8C>

# Magic command

해당 명령어들은 현재 메모리에 할당되어 있는 변수들의 정보를 보여줍니다.

<code>%who</code>	변수들을 출력합니다.
<code>%who_ls</code>	변수들을 리스트의 형태로 출력합니다.
<code>%whos</code>	변수들을 변수명과 유형, 그리고 데이터를 포함하여 출력합니다.

```
import numpy as np
```

```
a = 1234
b = "hello"
c = 1e-3
```

```
def foo(a, b):
    print(a * b)
```

```
%who
```

```
a      b      c      foo      np      plt
```

```
%who_ls
```

```
['a', 'b', 'c', 'foo', 'np', 'plt']
```

```
%whos
```

Variable	Type	Data/Info
----------	------	-----------

a	int	1234
b	str	hello
c	float	0.001
foo	function	<function foo at 0x00000296BED309D0>
np	module	<module 'numpy' from 'C:\>ges\numpy\__init__.py'>
plt	module	<module 'matplotlib.pyplot' from 'C:\>matplotlib\pyplot.py'>

# Magic command

%%time	코드 블록(셀)의 실행 시간을 측정합니다.
%%timeit	코드 블록(셀)을 n회 반복 실행하여 평균 실행 시간을 측정합니다.

```
%%time  
sum(range(10000000))  
sum(range(1000000))
```

CPU times: total: 266 ms  
Wall time: 257 ms

499999500000

```
%%timeit  
sum(range(10000000))  
sum(range(1000000))
```

262 ms  $\pm$  1.77 ms per loop (mean  $\pm$  std. dev. of 7 runs, 1 loop each)

# Anaconda prompt 명령어 사용 (pwd)

## 현재 파일 경로 확인

```
In [6]: | pwd
```

```
'C:\\Users\\jjm\\Dropbox\\Ongoing_lecture\\2024_1\\공공데이터분석'
```



pandas 복습

# 실습1

1. 아래와 같은 series sr1을 만드시오
  - Index: a,b,c,d,e
  - Data: 0~14까지의 random value 5개
2. sr1에 3을 곱하고 1을 더한 sr2를 만들고,  
sr2 중에 data가 짝수인 데이터만 구하기
3. sr1에서 두번째로 큰 data의 값과 index를 구하시오

## 실습2

- pandas\_data1.txt를 읽어서 다음을 출력하시오 ('이름'을 index로 선택)
  1. 처음부터 '김열' 학생까지의 데이터 (한글파일 read\_table 시: encoding='cp949')
  2. 최하나학생의 기말점수
  3. 마지막 5명 학생의 출석, 과제 점수

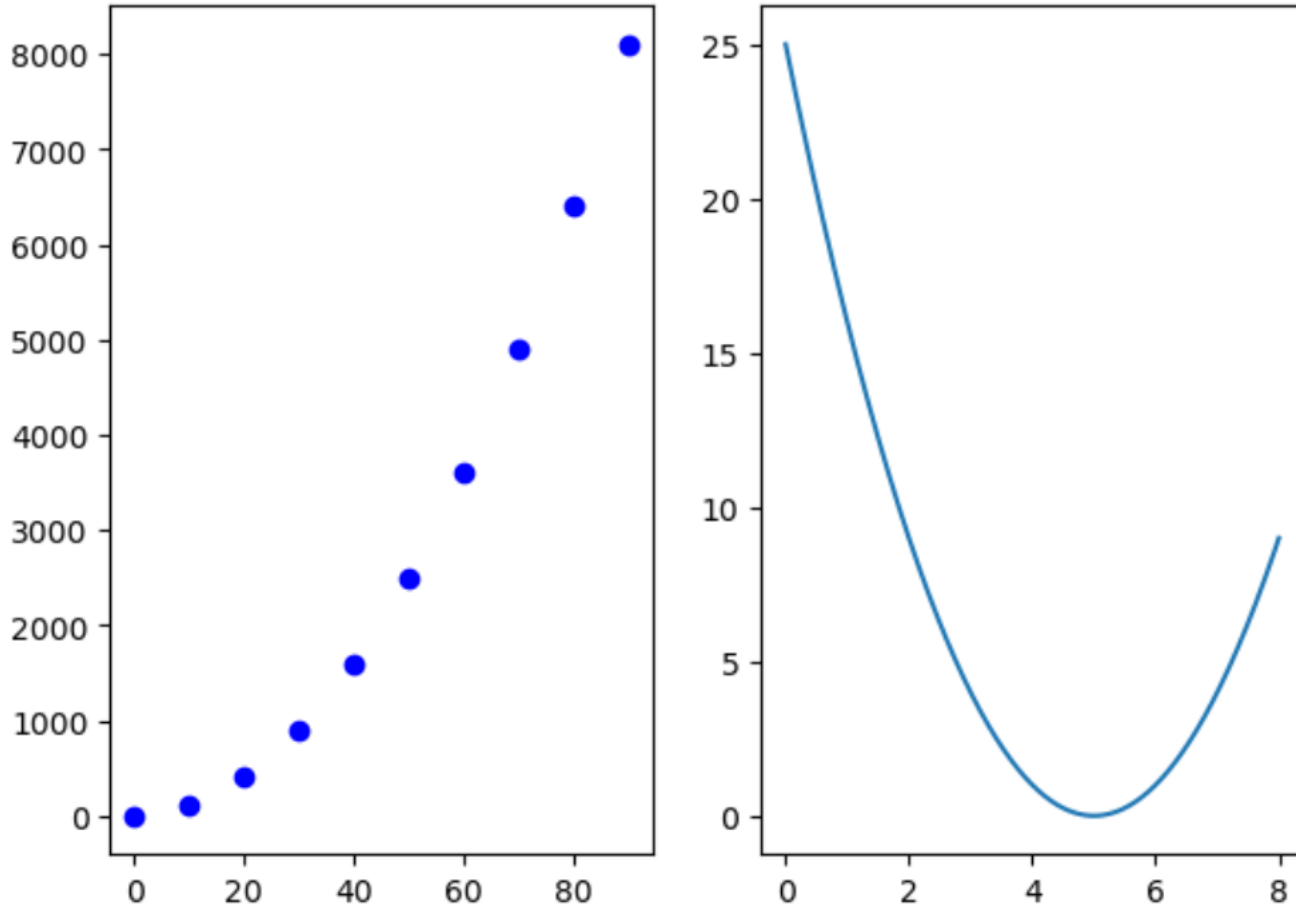
## 실습3

- pandas\_data3.txt를 읽어서 다음을 수행하시오
  1. final이 40점 이상인 학생의 수
  2. 1학년 학생의 기말 점수. 그리고 그 점수들의 평균
  3. proj가 100점인 학생의 mid, final, att 점수
  4. att이 5점 이상 이고 proj가 90점 이상인 학생 name
  5. "mid\_grade" column 만들어, mid 점수가 25점 이상 A, 20점 이상 B, 나머지는 C를 넣으시오
  6. gender 별 기말점수의 평균값은?

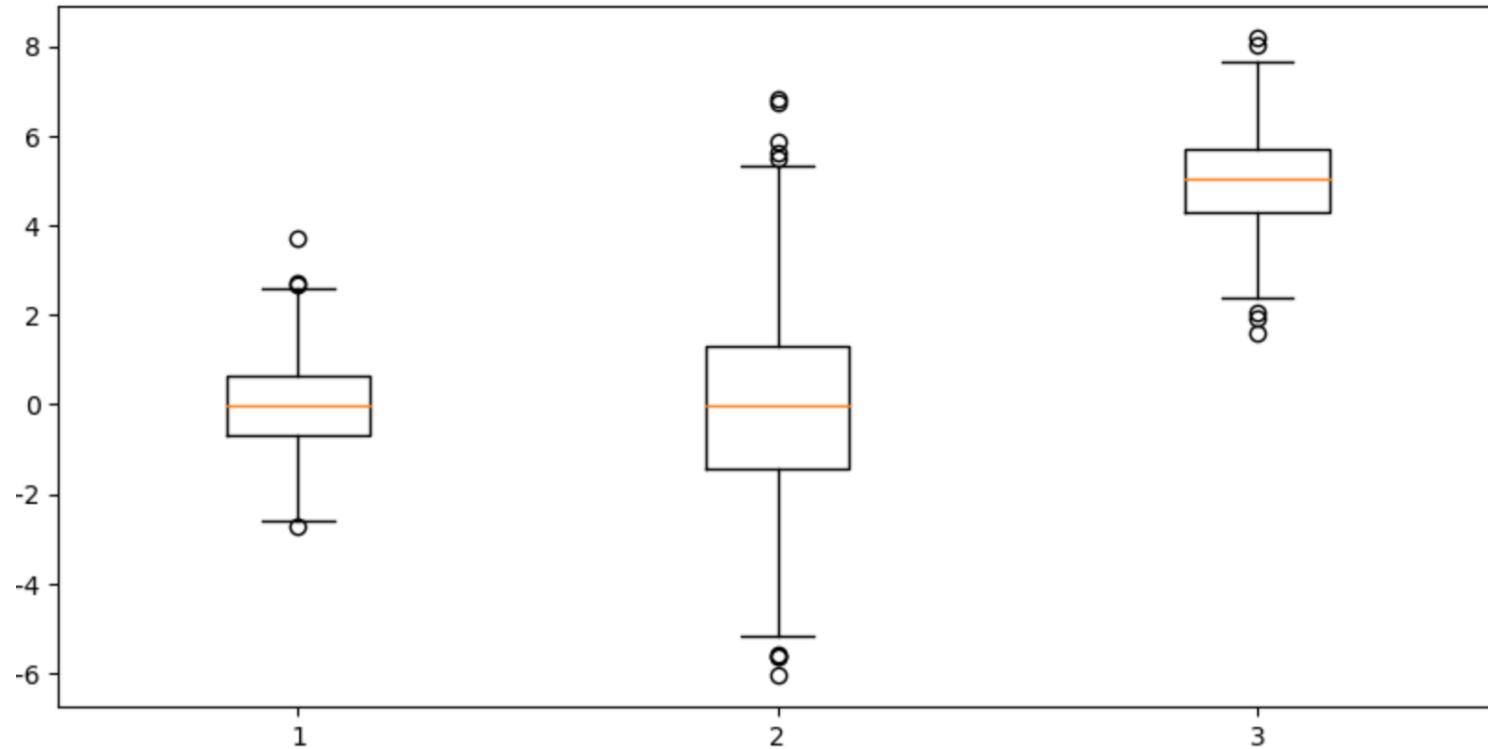
# matplotlib 복습

# 실습 1

- $Y=X^2$ ,  $Y=(X-5)^2$  두 개 함수를 하나의 figure에 나란히 그리시오



## 실습 2



위 그림과 같이 3개의 data에 각각의 boxplot을 그리시오

- data 1) 평균이 0, 표준편차 1의 정규분포에서 샘플링한 1000개의 값
- data 2) 평균이 0, 표준편차 2의 정규분포에서 샘플링한 1000개의 값
- data 3) 평균이 5, 표준편차 1의 정규분포에서 샘플링한 1000개의 값

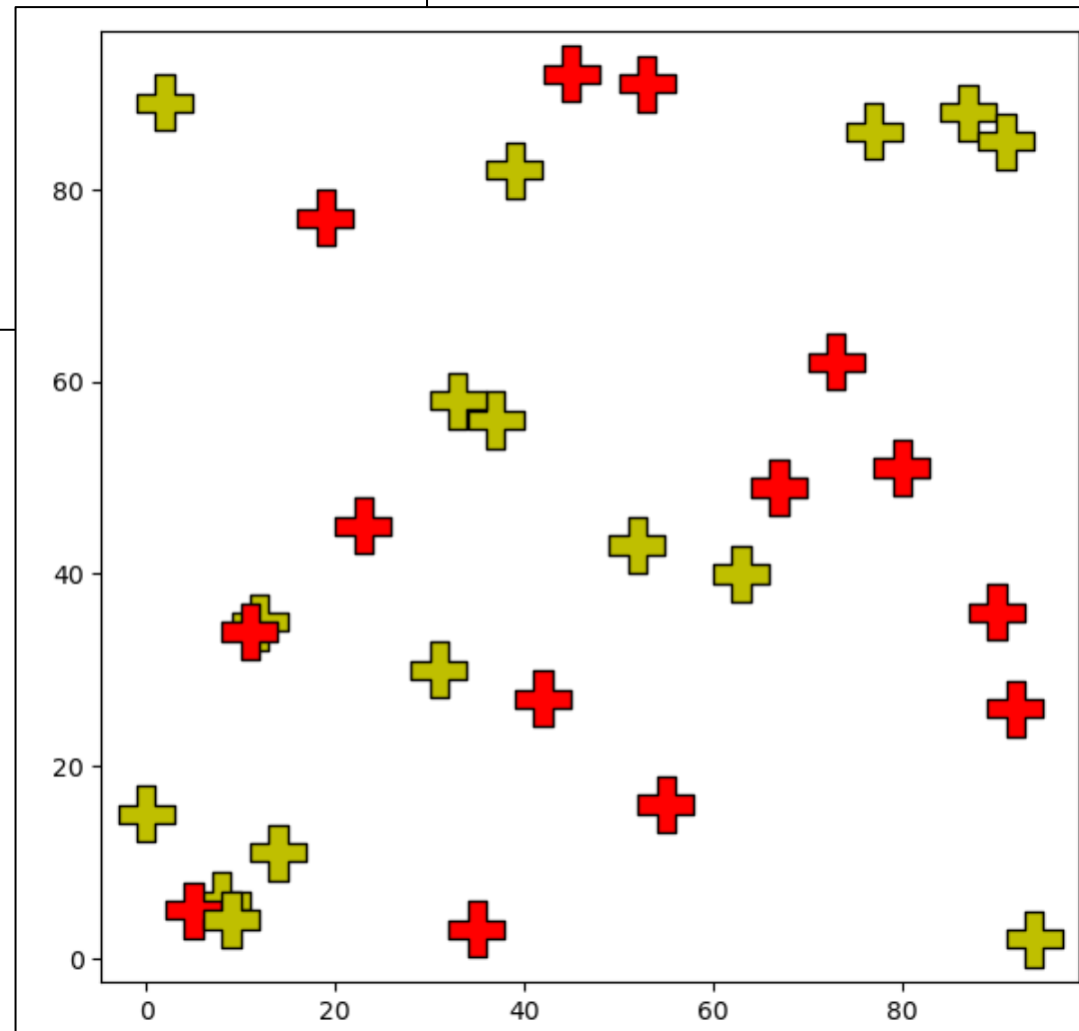
## 실습3

'xy\_price.txt' 파일이 있다. (x,y는 좌표, price는 해당 위치의 부동산 가격)

'xy\_price.txt' 파일을 읽어서, 해당 위치의 부동산 가격이

- 20 미만이면 노란색,
- 20 이상이면 빨간색

으로 나타내시오





End