
Data Analysis


Eun Lee


데이터분석 (1)?

- 목표: 판다스를 활용한 실제 데이터 분석을 자신있게 할 수 있게 되는 것.
- 수업 형태: 온라인 - 기본 이론 강의 (목),
대면 프로젝트기반 수업, 피드백, 평가 (화)

- 평가방법

구분	중간고사	기말고사	과제	토론	퀴즈	출결	기타	합계	성적평가기준
평가비율	20%	30%	30%	0%	0%	10%	10%	100%	절대평가
평가내용									

- 
- 과제 및 수업 시간 프로젝트
(발표자료 수업 날짜 내 완성상태)

- 
- 수업참여, 학습 도움

What is pandas?

- ✗ Data analysis library built on top of the Python programming language
- ✗ A robust toolkit for analyzing, filtering, manipulating, aggregating, merging, pivoting, and cleaning data



What is pandas?

- ✗ Pandas is not named after the furry animal. Its name is short after “panel data structures”, a term that refers to “data sets with observations over time”.
- ✗ “Excel for Python” or “Excel on steroids”.



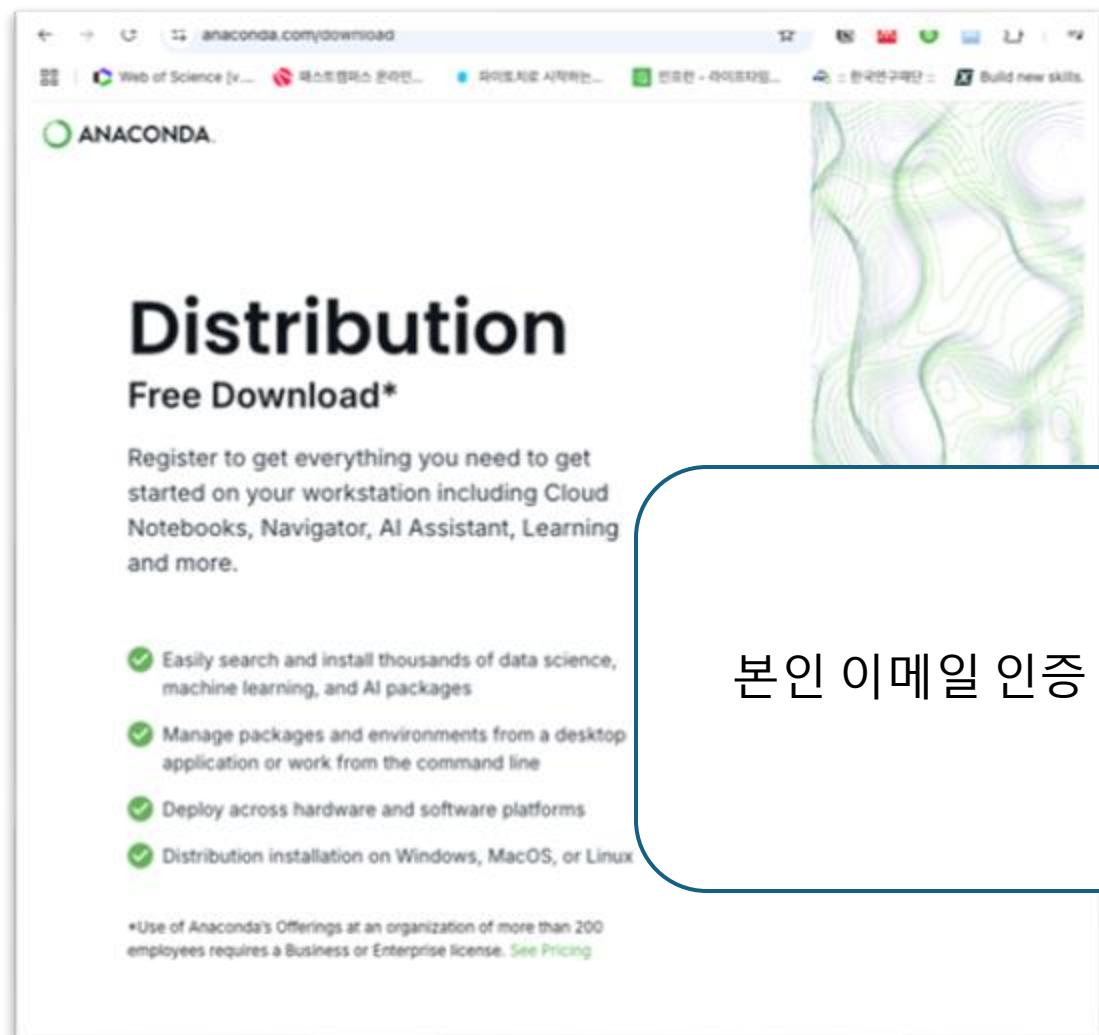
Installing Pandas and Python

- ✖ The easiest setup process is installing the **Anaconda distribution**, a bundle of Python, pandas, and 100+ data analysis libraries

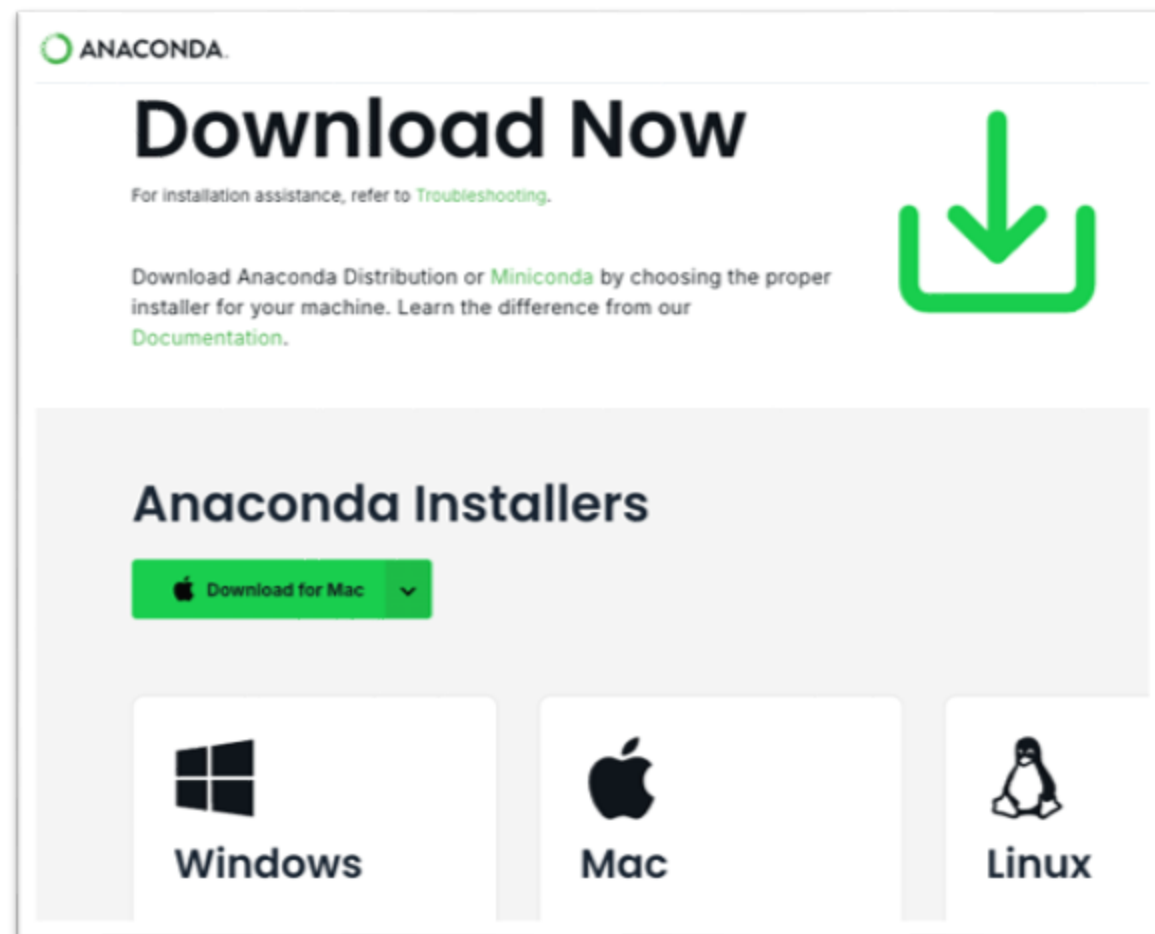
실습실에서는 설치가
어려움. 개인 컴퓨터에
설치 시도하여 사용할
수 있음.

Installing Anaconda for mac OS

- download anaconda: www.anaconda.com/download

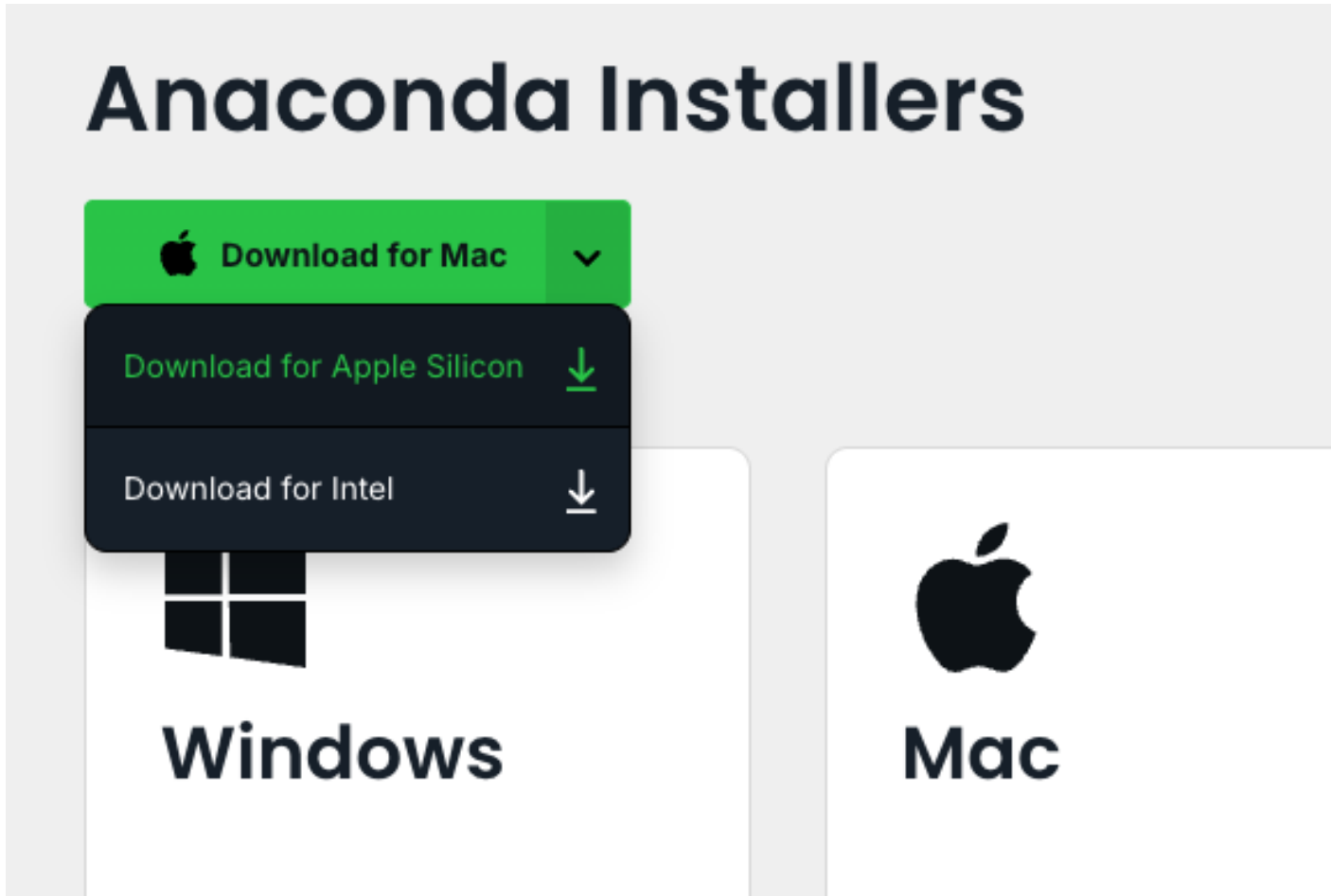


본인 이메일 인증



Installing Anaconda for mac OS

- Mac 칩이 무엇인지에 따라 선택해서 다운로드해야함.



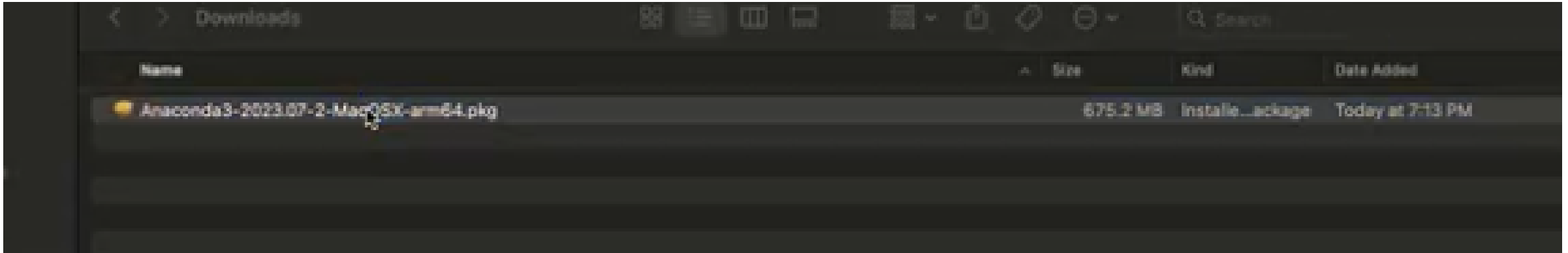
[확인하기]

- 1) 왼쪽 상단 애플 로고 클릭
- 2) 이 맥에 관하여 -> 더 많은 정보
- 3) Chip 에 대한 부분 확인하면 됨.

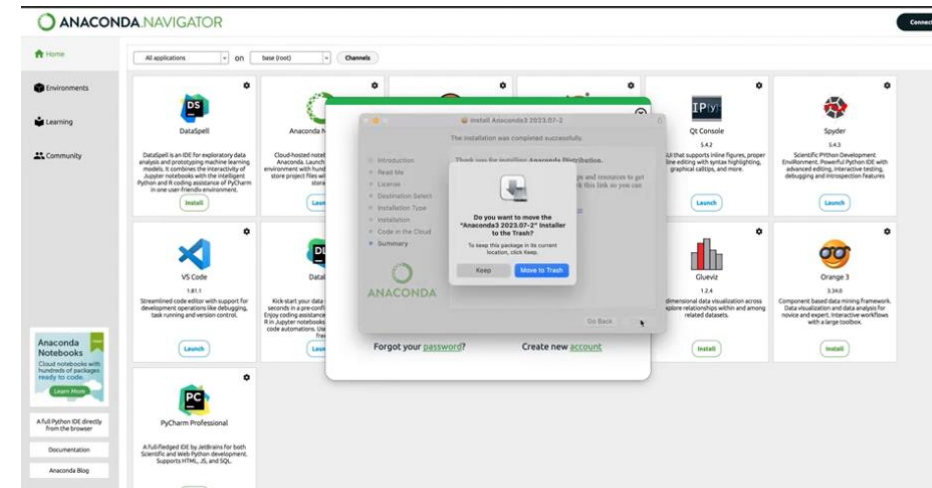


Installing Anaconda for mac OS

- 다운로드가 완료되면, 보통 'Downloads' 폴더에 pkg 가 다운됨.

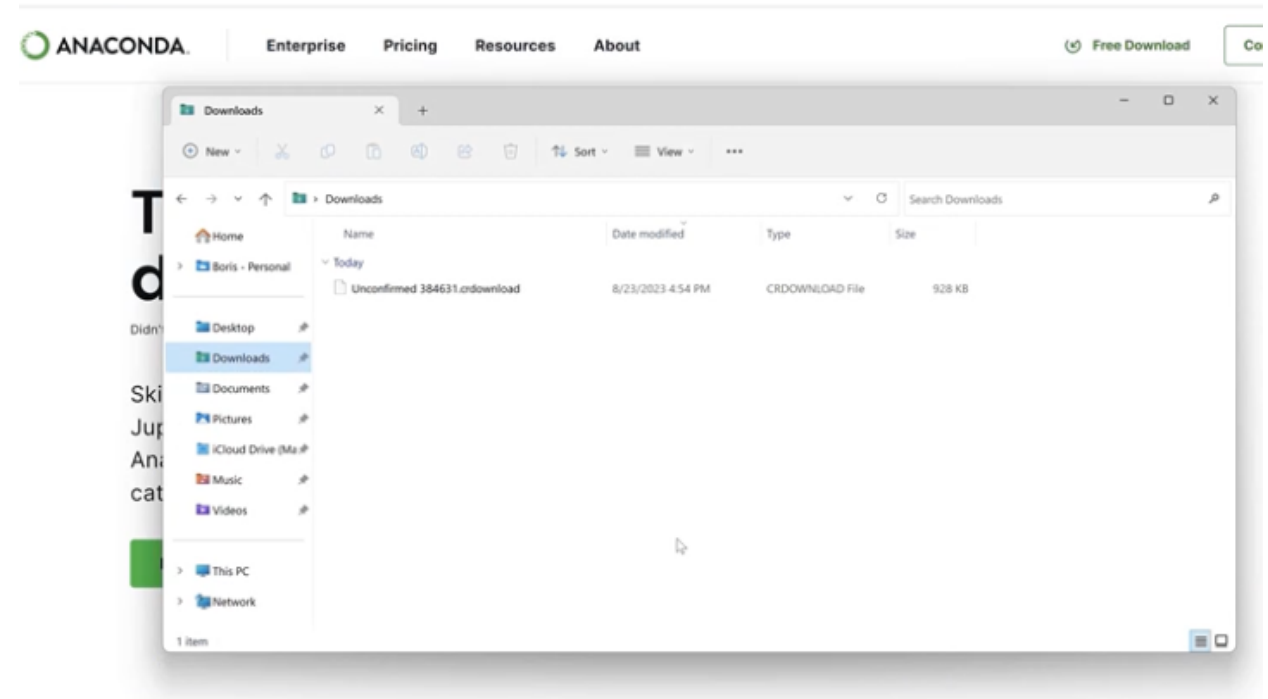
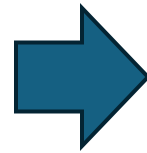
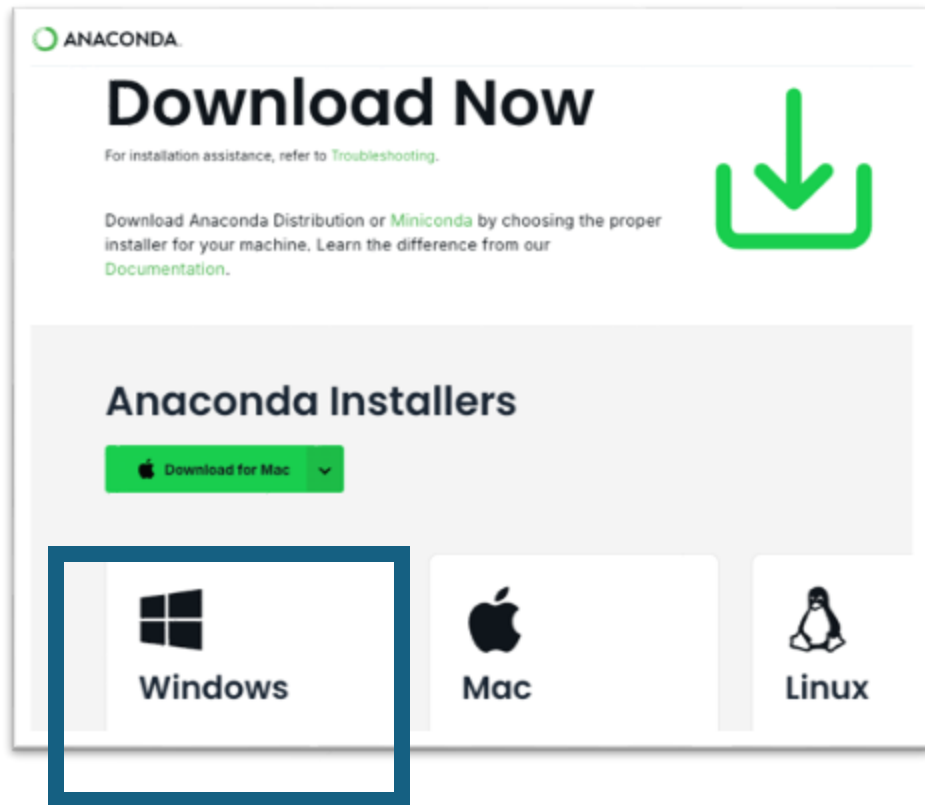


- Allow, continue 클릭하여 설치되도록 함.



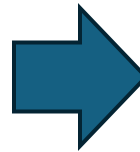
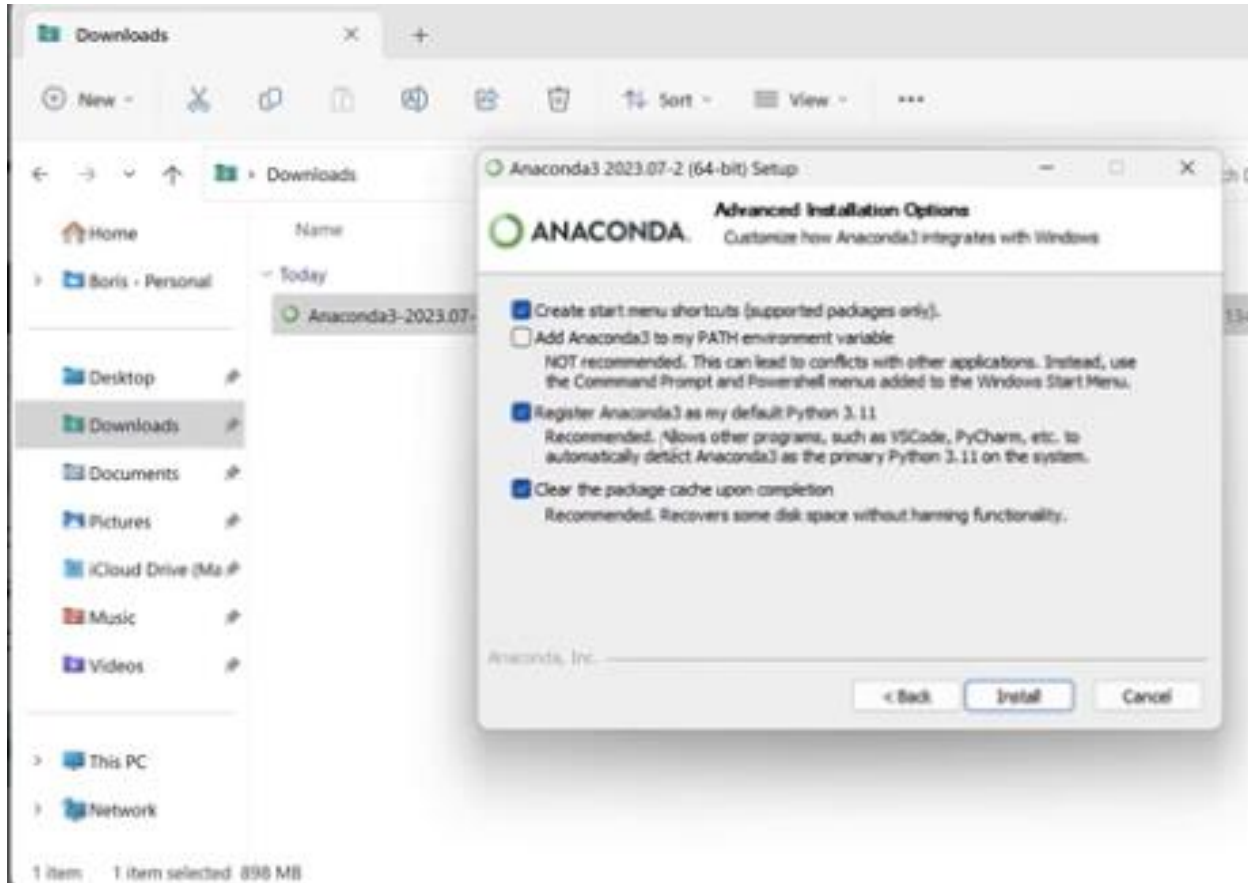
Installing Anaconda for Windows

- download anaconda: www.anaconda.com/download



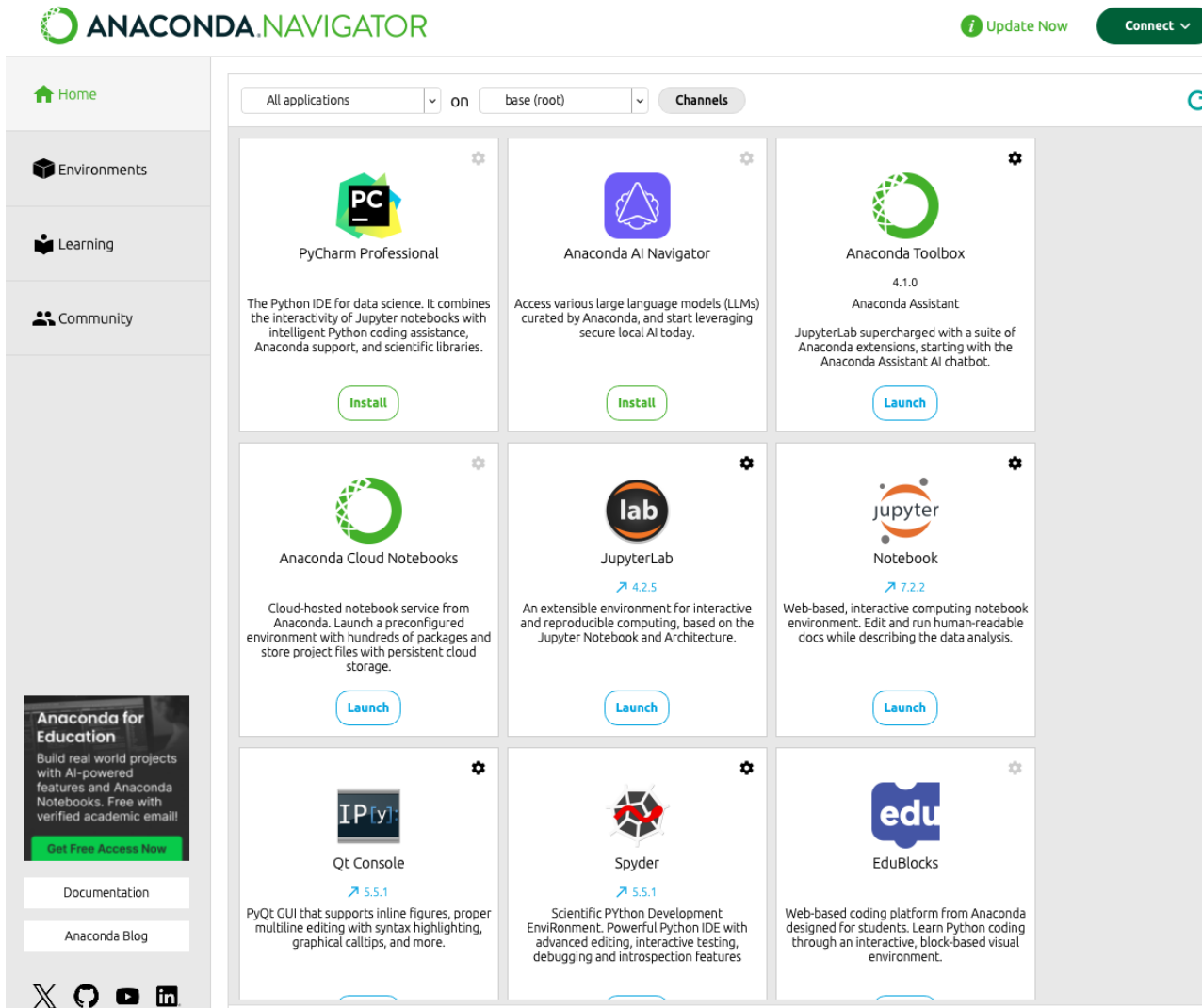
Installing Anaconda for Windows

- Allow, continue 클릭하여 설치되도록 함.
- 아래와 같은 옵션 설정으로 설치가 완료되도록 함.



Anaconda navigator

- 특정 파이썬 버전 에 맞는 라이브러리들을 선택적으로 설치하여 관리 가능.



- 다양한 버전의 파이썬을 쓰는 프로젝트를 수행할 때 활용.

```
(base) eun_m2@ieun-ui-MacBookAir ~ % conda env list

# conda environments:
#
                                /Users/eun_m2/miniconda3
                                /Users/eun_m2/miniconda3/envs/deep1
                                /Users/eun_m2/miniconda3/envs/py310
                                /Users/eun_m2/miniconda3/envs/py311
                                /Users/eun_m2/miniconda3/envs/py38
base
                                * /opt/anaconda3

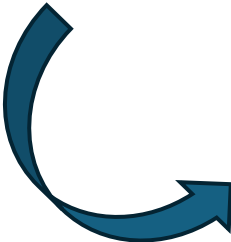
(base) eun_m2@ieun-ui-MacBookAir ~ %
```

Anaconda navigator

```
[(base) eun_m2@ieun-ui-MacBookAir ~ % conda env list

# conda environments:
#
          /Users/eun_m2/miniconda3
          /Users/eun_m2/miniconda3/envs/deepl
          /Users/eun_m2/miniconda3/envs/py310
          /Users/eun_m2/miniconda3/envs/py311
          /Users/eun_m2/miniconda3/envs/py38
base      * /opt/anaconda3

(base) eun_m2@ieun-ui-MacBookAir ~ %
```



ANACONDA.NAVIGATOR

Update Now

Connect

Home

Environments

Learning

Community

Search Environments

base (root)

deepl

miniconda3

py310

py311

py38

Installed

Channels

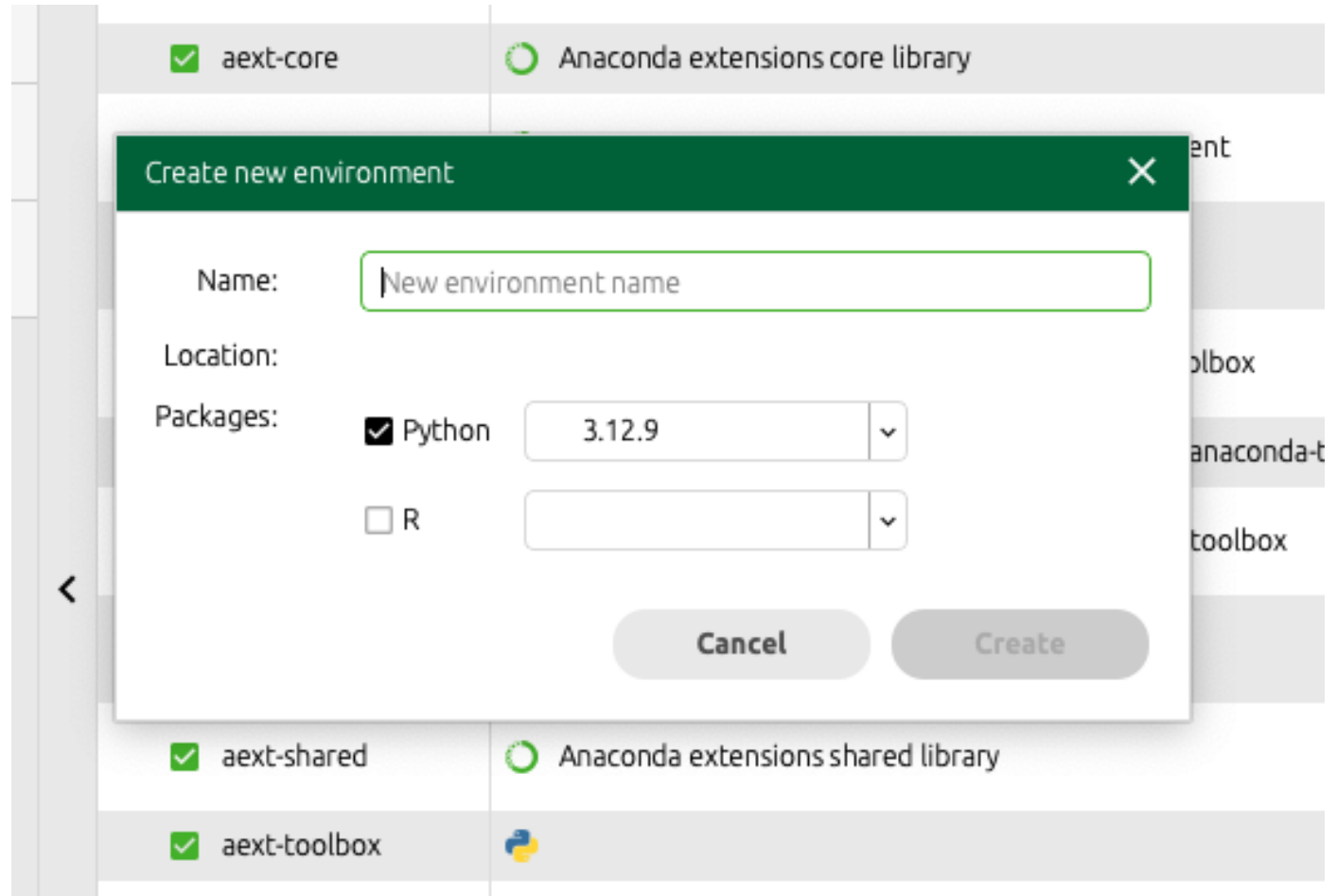
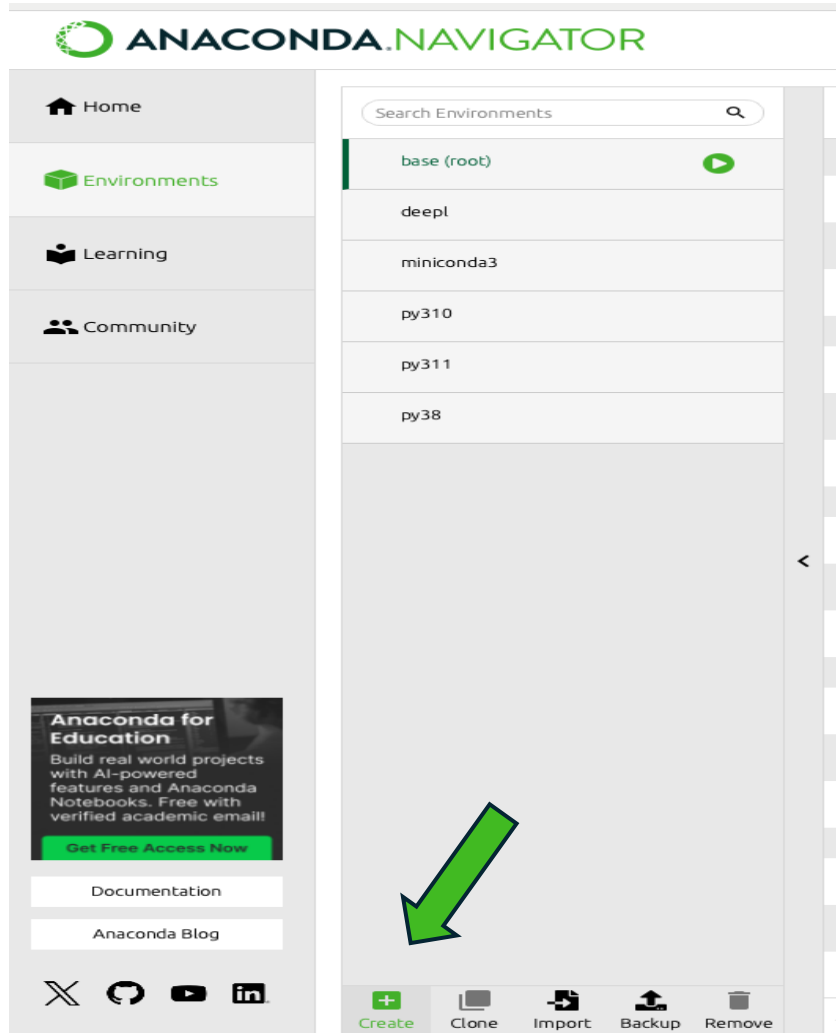
Update index...

Search Packages

Name	T	Description	Version
<input checked="" type="checkbox"/> _anaconda_depends	<input type="radio"/>	Simplifies package management and deployment of anaconda	2024.10
<input checked="" type="checkbox"/> aext-assistant	<input type="radio"/>	Anaconda extensions assistant library	4.1.0
<input checked="" type="checkbox"/> aext-assistant-server	<input type="radio"/>	Anaconda extensions assistant server	4.1.0
<input checked="" type="checkbox"/> aext-core	<input type="radio"/>	Anaconda extensions core library	4.1.0
<input checked="" type="checkbox"/> aext-core-server	<input type="radio"/>	Anaconda toolbox backend lib core server component	4.1.0
<input checked="" type="checkbox"/> aext-panels	<input type="radio"/>	The aext-panels component of anaconda-toolbox	4.1.0
<input checked="" type="checkbox"/> aext-panels-server	<input type="radio"/>	The aext-panels-server component of anaconda-toolbox	4.1.0
<input checked="" type="checkbox"/> aext-project-filebrowser-server	<input type="radio"/>	The aext-project-filebrowser-server component of anaconda-toolbox	4.1.0
<input checked="" type="checkbox"/> aext-share-notebook	<input type="radio"/>	The aext-share-notebook component of anaconda-toolbox	4.1.0
<input checked="" type="checkbox"/> aext-share-notebook-server	<input type="radio"/>	Anaconda extensions share notebook server	4.1.0

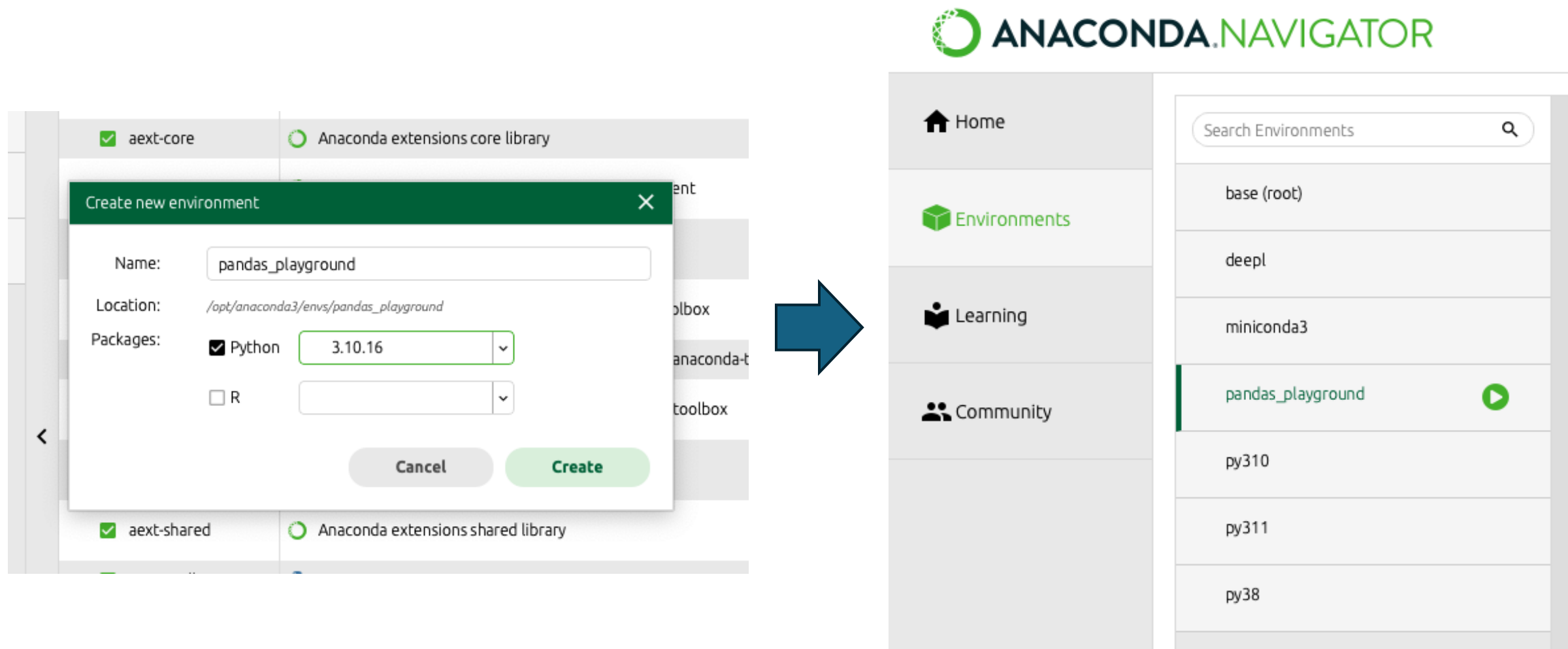
Anaconda navigator

- 최신 버전이 아닌, 이전 버전으로 작업하고자하면 새로운 환경 **create**



Anaconda navigator

- 설정한 파이썬 버전으로 새로운 환경이나타남.



Anaconda navigator

- 'All' 로 하면 해당 환경에서 사용가능한 모든 라이브러리가 보임.
- Search -> pandas 검색해서 Apply

Search Environments

All Channels Update index...

Name	T	Description
base (root)		
deepl		
miniconda3		
pandas_playground		
py310		
py311		
py38		

Name	T	Description
<input type="checkbox"/> _anaconda_core...	🟢	Simplifies package management and deployment of anacond
<input type="checkbox"/> _anaconda_core	🟢	Simplifies package management and deployment of anacond
<input type="checkbox"/> _anaconda_depends	🟢	Simplifies package management and deployment of anacond
<input type="checkbox"/> _go_select	🟢	The golang select package.
<input type="checkbox"/> _ipyw_jlab_nb_ex...	🟢	A configuration metapackage for enabling anaconda-bundlec
<input type="checkbox"/> _libarchive_static...	🟢	A static build of libarchive containing only conda-related part
<input type="checkbox"/> _mutex_mxnet	🟢	Mutex package to pin a variant of mxnet conda package
<input type="checkbox"/> _ppc64le-microarch- level	🟢	Meta package to build conda recipes with microarchitecture l
<input type="checkbox"/> _py-xgboost-mutex	🟢	Scalable, portable and distributed gradient boosting (gbdt, gl machine, hadoop, spark, flink and dataflow



pandas_playground

py310

py311

py38

<input type="checkbox"/> geopandas	🟢 Geographic pandas extensions
<input type="checkbox"/> geopandas-base	🟢 Geographic pandas extensions
<input type="checkbox"/> modin	🟢 Speed up your pandas workflows by changing a single line of code
<input type="checkbox"/> modin-all	🟢 Speed up your pandas workflows by changing a single line of code
<input type="checkbox"/> modin-core	🟢 Speed up your pandas workflows by changing a single line of code
<input type="checkbox"/> modin-dask	🟢 Speed up your pandas workflows by changing a single line of code
<input type="checkbox"/> modin-ray	🟢 Speed up your pandas workflows by changing a single line of code
<input type="checkbox"/> pandarallel	🟢 An easy to use library to speed up computation (by parallelizing on multi cpus) with pandas.
<input checked="" type="checkbox"/> pandas	🟢 High-performance, easy-to-use data structures and data analysis tools.
<input type="checkbox"/> pandas-datareader	🟢 Up to date remote data access for pandas, works for multiple versions

<input checked="" type="checkbox"/> pandas	🟢 High-performance, easy-to-use data structures and data analysis tools.	2.2.3
<input type="checkbox"/> pandas-datareader	🟢 Up to date remote data access for pandas, works for multiple versions of pandas	0.10.0
<input type="checkbox"/> pandas-profiling	🟢 Generate profile report for pandas dataframe	3.6.3
<input type="checkbox"/> pandas-stubs	🟢 Collection of pandas stub files	2.1.4.2.
<input type="checkbox"/> pandasql	🟢 Sqldf for pandas	0.7.3
<input type="checkbox"/> qgrid	🟢 Pandas dataframe viewer for jupyter notebook	1.1.1
<input type="checkbox"/> qpd	🟢 Query pandas using sql	0.4.4
<input type="checkbox"/> sklearn-pandas	🟢 Pandas integration with sklearn	2.2.0
<input type="checkbox"/> spatialpandas	🟢 Pandas extension arrays for spatial/geometric operations	0.4.10

22 packages available matching "pandas" 1 package selected

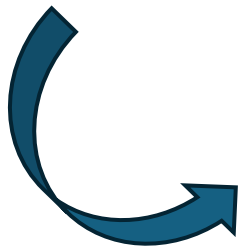
Apply Clear

Anaconda navigator

- Anaconda Navigator에서 특정 패키지를 설치할 때 **Apply** 버튼을 누르면, 선택한 패키지가 현재 활성화된 가상 환경에 설치되거나 업데이트됨.

Apply 버튼을 누르면

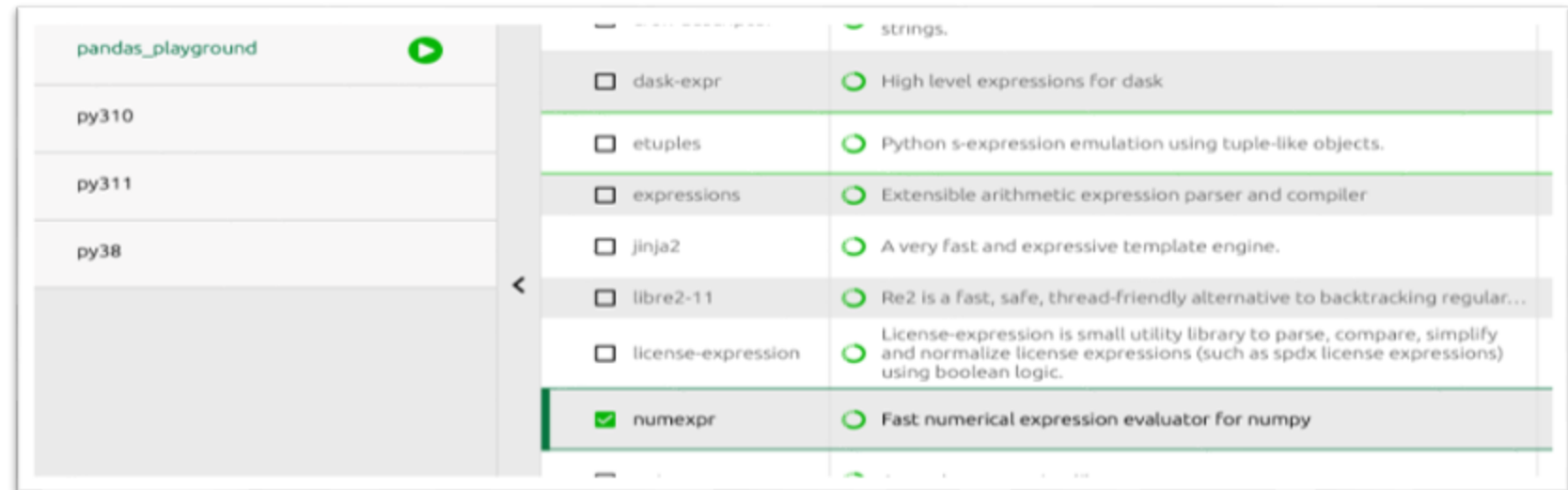
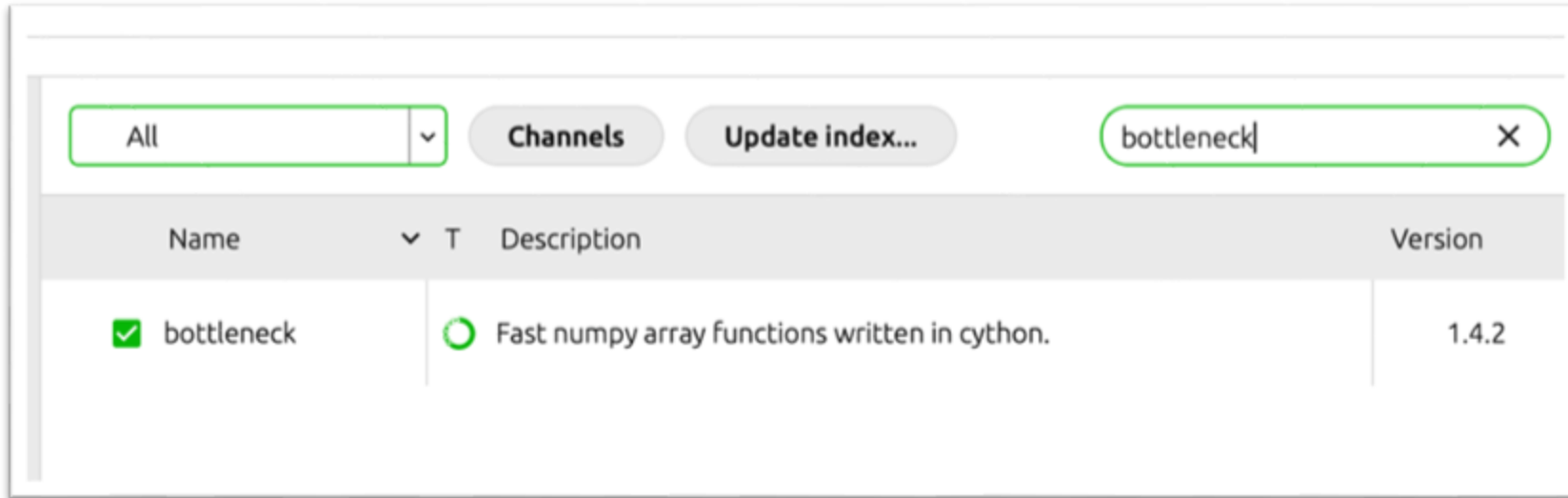
- `conda install package_name` 과 같은 명령어가 실행됨.
- 필요한 종속성(Dependencies)이 자동으로 설치됨.




```
(base) eun_m2@ieun-ui-MacBookAir ~ % conda activate pandas_playground
(pandas_playground) eun_m2@ieun-ui-MacBookAir ~ %
(pandas_playground) eun_m2@ieun-ui-MacBookAir ~ %
(pandas_playground) eun_m2@ieun-ui-MacBookAir ~ %
(pandas_playground) eun_m2@ieun-ui-MacBookAir ~ %
(pandas_playground) eun_m2@ieun-ui-MacBookAir ~ %
(pandas_playground) eun_m2@ieun-ui-MacBookAir ~ % conda install pandas
Retrieving notices: done
Channels:
- defaults
Platform: osx-arm64
Collecting package metadata (repodata.json): done
Solving environment: done
```



Anaconda navigator

- Search -> bottleneck, numexpr 검색해서 Apply




Jupyter Lab


 ANACONDA.NAVIGATOR





[Home](#)
[Environments](#)
[Learning](#)
[Community](#)


All applications on pandas_playground Channels



ORACLE
Cloud Infrastructure
Oracle Data Science Service
OCI Data Science offers a machine learning platform to build, train, manage, and deploy your machine learning models on the cloud with your favorite open-source tools
[Launch](#)


PyScript
Code and share Python in the Browser. A vibrant community of makers, builders, and hackers building the next frontier of Python-powered web applications.
[Launch](#)


PythonAnywhere
Host, run, and code Python in the cloud! Get started for free.
[Launch](#)


Glueviz
1.2.4
Multidimensional data visualization across files. Explore relationships within and among related datasets.
[Install](#)


JupyterLab
4.3.4
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Archicore.
[Install](#)


jupyter
Notebook
7.3.2
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
[Install](#)

Jupyter Lab 에서 쓸 수업 자료를 가져오기

- Github 를 써보자: https://github.com/eunlee-netsci/Data_analysis1.git

Data_analysis1 / Class_materials /

eunlee-netsci Add files via upload

Name

Last commit message

..

01 - Installation and Setup.ipynb

Add files via upload

material_readme.md

Data_analysis1 / Class_materials / 01 - Installation and Setup.ipynb

eunlee-netsci Add files via upload

Preview

Code

Blame

189 lines (189 loc) · 6.74 KB

Introduction to pandas

- **Pandas** is a data analysis library built on top of the Python program

Raw file content

Download

⌘ shift s

Copy path

⌘ shift .

Copy permalink

⌘ shift ,

View options

✓ Show code folding buttons

Jupyter Lab 에서 쓸 수업 자료를 가져오기

- Git clone >> git pull 을 쓰면, 새롭게 업데이트 부분만 업데이트하면서 폴더 최신으로 유지 가능

The image illustrates the process of cloning a GitHub repository for use in Jupyter Lab. It consists of three main components:

- GitHub Repository Page:** Shows the repository 'Data_analysis1' by 'eunlee-netsci'. The 'Code' button is highlighted, and the 'Clone' dropdown menu is open, showing the 'HTTPS' option selected. A green arrow points to the 'Clone' button.
- Terminal Window:** Displays the command `git clone https://github.com/eunlee-netsci/Data_analysis1.git` and its output, which shows the repository being cloned into 'Data_analysis1'. A green arrow points to the terminal window.
- File Explorer:** Shows the local file system structure, including the 'Data_analysis1' directory. A green arrow points to the 'Data_analysis1' directory.

Jupyter Lab

ANACONDA.NAVIGATOR

Home

Environments

Learning

Community

All applications

on

pandas_playground

Channels



PyCharm Professional

The Python IDE for data science. It combines the interactivity of Jupyter notebooks with intelligent Python coding assistance, Anaconda support, and scientific libraries.

Install



Anaconda AI Navigator

Access various large language models (LLMs) curated by Anaconda, and start leveraging secure local AI today.

Install



Anaconda Cloud Notebooks

Cloud-hosted notebook service from Anaconda. Launch a preconfigured environment with hundreds of packages and store project files with persistent cloud storage.

Launch



JupyterLab

4.3.4

An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.

Launch

Anaconda for Education

localhost:8888/lab/tree/Dropbox/5_Teaching/2025_1_수업준비/DS1

File Edit View Run Kernel Tabs Settings Help

/ Dropbox / 5_Teaching / 2025_1_수업준비 / DS1 /

Name	Modified	Size
01 - Installation and Se...	last yr.	6.7 KB

Launcher

Dropbox/5_Teaching/2025_1_수업준비/DS1

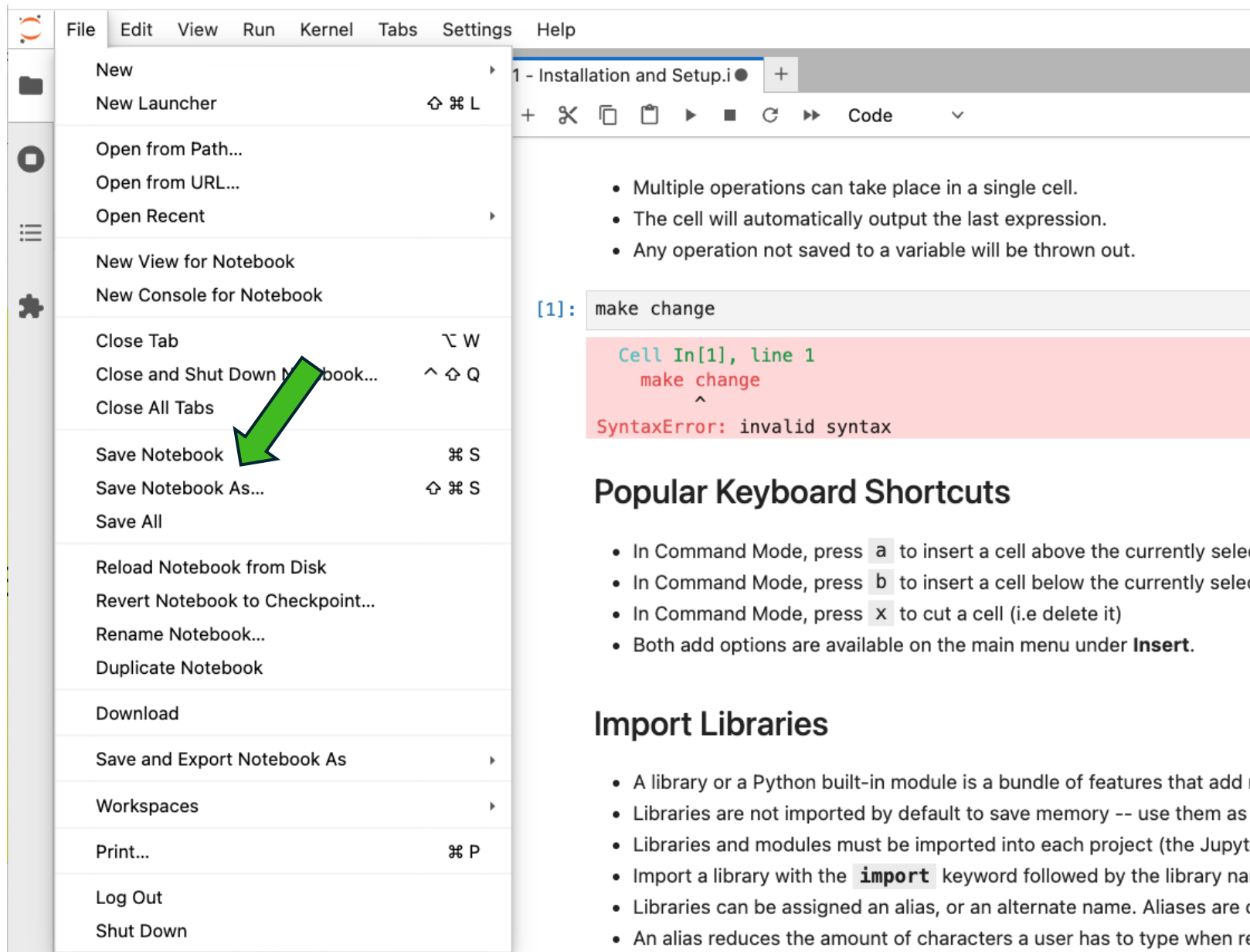
Notebook

Python 3 (ipykernel)

Console

Python 3 (ipykernel)

Jupyter Lab: 반드시 저장하도록 하자.



The screenshot displays the Jupyter Lab interface. On the left, the 'File' menu is open, showing various options. A green arrow points to the 'Save Notebook' option, which is highlighted. The main area shows a code cell with the text 'make change'. Below the code, a red error message is displayed: 'SyntaxError: invalid syntax'. The error message is highlighted in red, and a green arrow points to the word 'make' in the code, indicating the source of the error.

File Edit View Run Kernel Tabs Settings Help

- New
- New Launcher ⌘ L
- Open from Path...
- Open from URL...
- Open Recent
- New View for Notebook
- New Console for Notebook
- Close Tab ⌘ W
- Close and Shut Down Notebook... ^ ⌘ Q
- Close All Tabs
- Save Notebook ⌘ S
- Save Notebook As... ⌘ S
- Save All
- Reload Notebook from Disk
- Revert Notebook to Checkpoint...
- Rename Notebook...
- Duplicate Notebook
- Download
- Save and Export Notebook As
- Workspaces
- Print... ⌘ P
- Log Out
- Shut Down

1 - Installation and Setup.i +

+ ✂ 📄 📌 ▶ ■ ↺ ▶▶ Code ▾

- Multiple operations can take place in a single cell.
- The cell will automatically output the last expression.
- Any operation not saved to a variable will be thrown out.

[1]: make change

Cell In[1], line 1
make change
^
SyntaxError: invalid syntax

Popular Keyboard Shortcuts

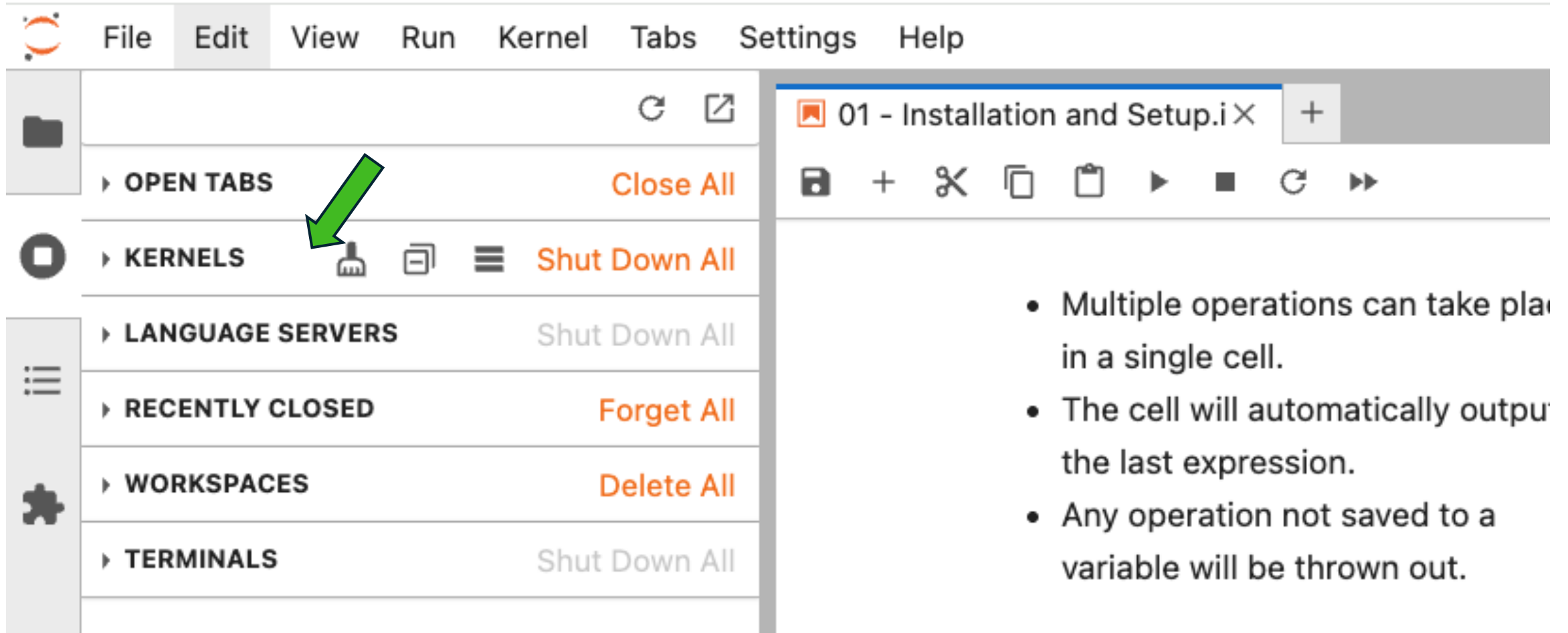
- In Command Mode, press **a** to insert a cell above the currently selected cell.
- In Command Mode, press **b** to insert a cell below the currently selected cell.
- In Command Mode, press **x** to cut a cell (i.e delete it)
- Both add options are available on the main menu under **Insert**.

Import Libraries

- A library or a Python built-in module is a bundle of features that add new functionality to your program.
- Libraries are not imported by default to save memory -- use them as you need them.
- Libraries and modules must be imported into each project (the Jupyter Notebook).
- Import a library with the **import** keyword followed by the library name.
- Libraries can be assigned an alias, or an alternate name. Aliases are often used to shorten the name of a library.
- An alias reduces the amount of characters a user has to type when referring to a library.

Jupyter Lab: 커널 정리

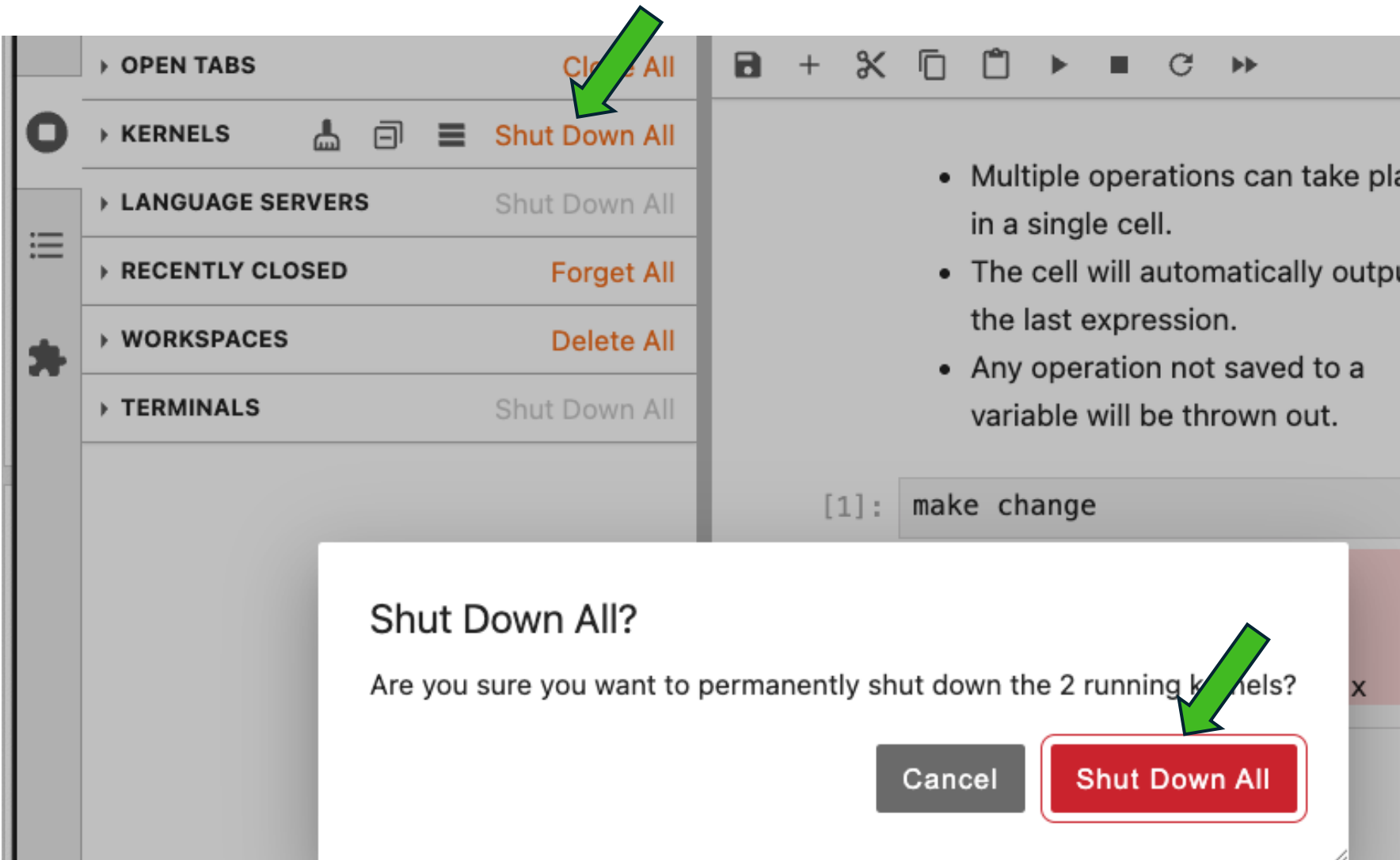
- 노트북을 열때마다 파이썬 서버가 명령 수행 대기를 하게 됨.
- 해당 서버는 제한된 컴퓨팅 자원을 사용하는 것이라 조절이 필요함.



The screenshot displays the Jupyter Lab user interface. The top menu bar includes 'File', 'Edit', 'View', 'Run', 'Kernel', 'Tabs', 'Settings', and 'Help'. The left sidebar contains several sections: 'OPEN TABS' with a 'Close All' button, 'Kernels' with a 'Shut Down All' button (highlighted by a green arrow), 'LANGUAGE SERVERS' with a 'Shut Down All' button, 'RECENTLY CLOSED' with a 'Forget All' button, 'WORKSPACES' with a 'Delete All' button, and 'TERMINALS' with a 'Shut Down All' button. The right pane shows a code editor with a single cell titled '01 - Installation and Setup.i'.

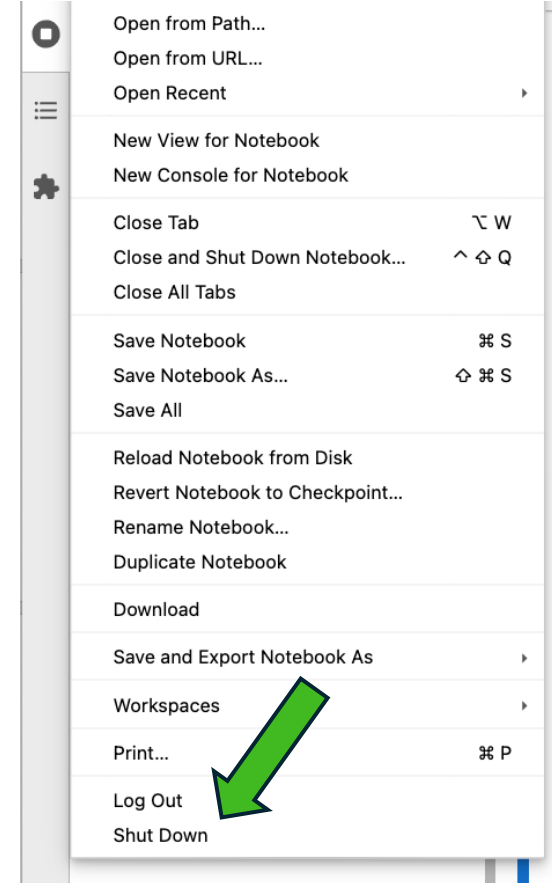
- Multiple operations can take place in a single cell.
- The cell will automatically output the last expression.
- Any operation not saved to a variable will be thrown out.

Jupyter Lab: 커널 정리



The screenshot shows the Jupyter Lab interface. On the left sidebar, the 'Kernels' tab is selected, and the 'Shut Down All' button is highlighted with a green arrow. In the center, a modal dialog titled 'Shut Down All?' is displayed, asking 'Are you sure you want to permanently shut down the 2 running kernels?'. The dialog has two buttons: 'Cancel' and 'Shut Down All'. The 'Shut Down All' button is highlighted with a red border and a green arrow. The background shows a code editor with a cell containing the text '[1]: make change' and a list of bullet points:

- Multiple operations can take place in a single cell.
- The cell will automatically output the last expression.
- Any operation not saved to a variable will be thrown out.

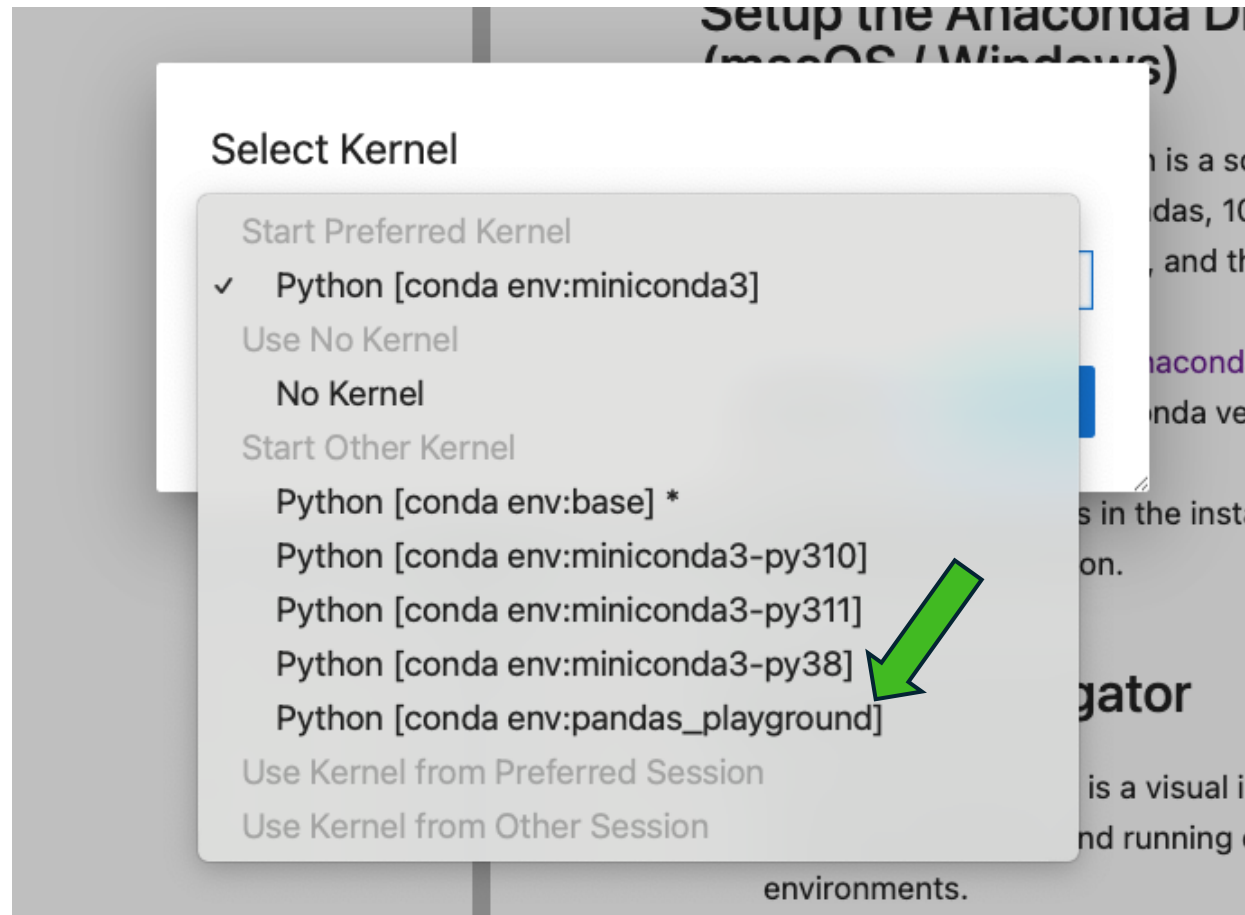


The screenshot shows the 'File' menu in Jupyter Lab. The 'Shut Down' option is highlighted with a green arrow. The menu items are as follows:

- Open from Path...
- Open from URL...
- Open Recent
- New View for Notebook
- New Console for Notebook
- Close Tab (⌘ W)
- Close and Shut Down Notebook... (⇧ ⌘ Q)
- Close All Tabs
- Save Notebook (⌘ S)
- Save Notebook As... (⇧ ⌘ S)
- Save All
- Reload Notebook from Disk
- Revert Notebook to Checkpoint...
- Rename Notebook...
- Duplicate Notebook
- Download
- Save and Export Notebook As
- Workspaces
- Print... (⌘ P)
- Log Out
- Shut Down

Jupyter notebook 사용법

- 아나콘다 네비게이터에서 jupyter lab 실행
- 커널 설정 요청하면 새롭게 만든 커널 (파이썬 환경) 선택하여 실행



Jupyter notebook 사용법

- 새로운 파일 만들 때, 첫 화면에서 사용할 환경 설정 필요.

Dropbox/5_Teaching/2025_1_수업준비/Data_analysis1/Class_materials



Notebook



Python [conda
env:miniconda3
]



Python [conda
env:base] *



Python [conda
env:miniconda3
-py310]



Python [conda
env:miniconda3
-py311]



Python [conda
env:miniconda3
-py38]

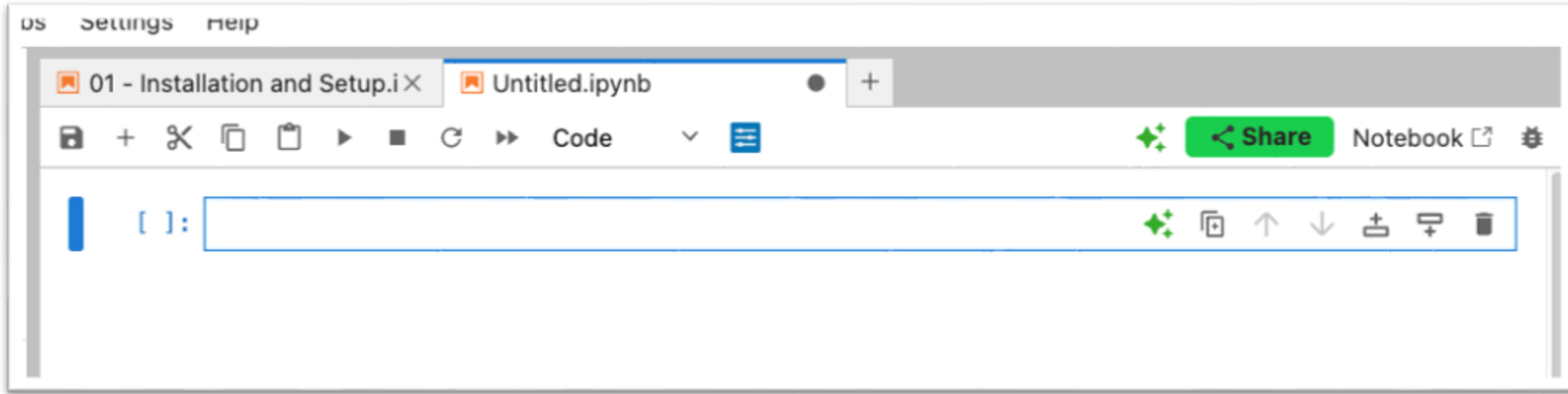


Python [conda
env:pandas_pla
yground]



Jupyter notebook 사용법

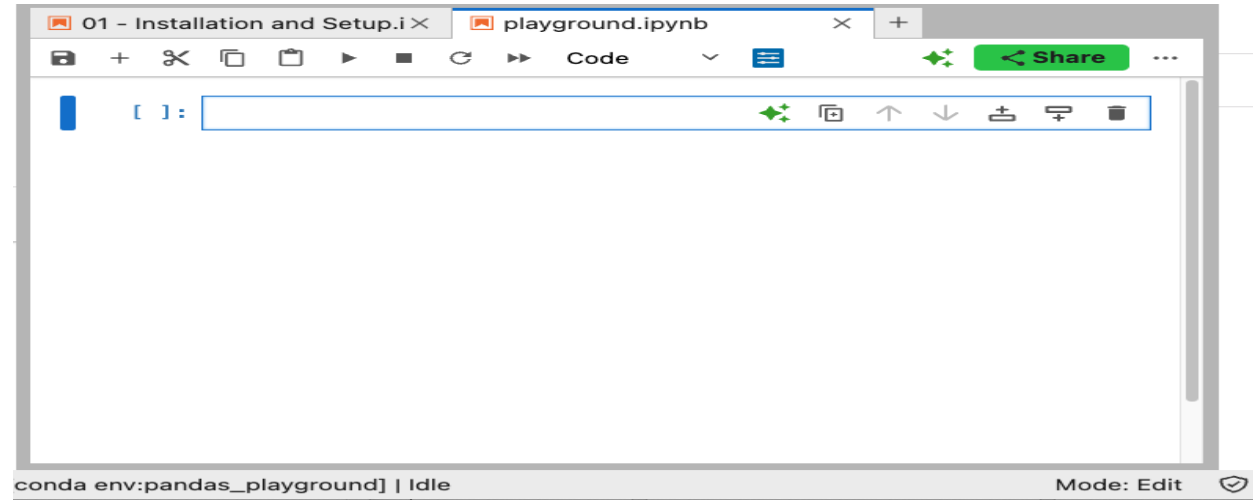
- 해당 커널을 사용하여 쥬피터노트북 형태로 코드 작성 가능



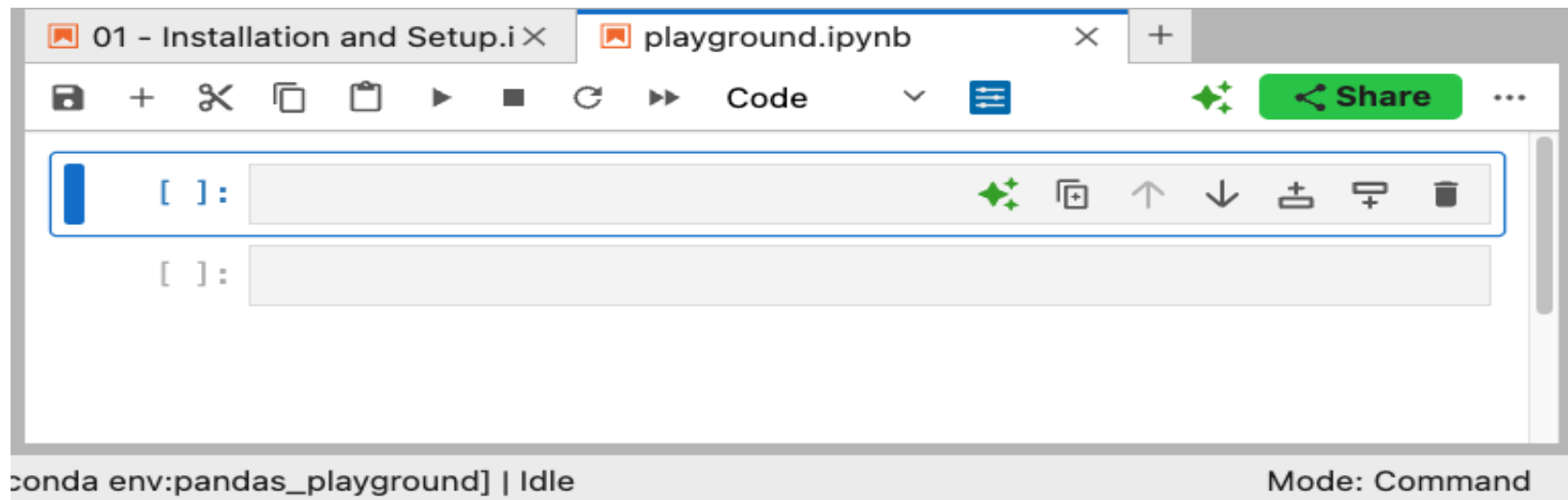
- 쥬피터 노트북은 여러 셀의 집합으로 이해할 수 있음.
 - 셀은 <코드> 를 위한 셀, <텍스트> 나 <이미지> 를 위한 셀 등이 될 수 있음.
 - 코드를 위한 셀의 경우 아웃풋을 바로 다음에 보여주어 편리함.

Jupyter notebook 사용법

- 쥬피터노트북의 Mode:
 - Edit Mode:



- Command Mode: shortcut 을 사용할 수 있는 상태. ESC 를 누르면 활성화

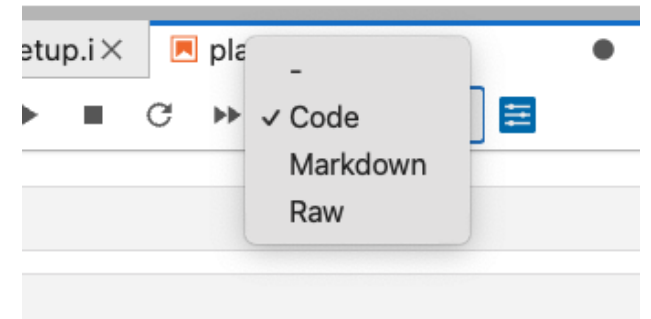
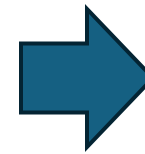
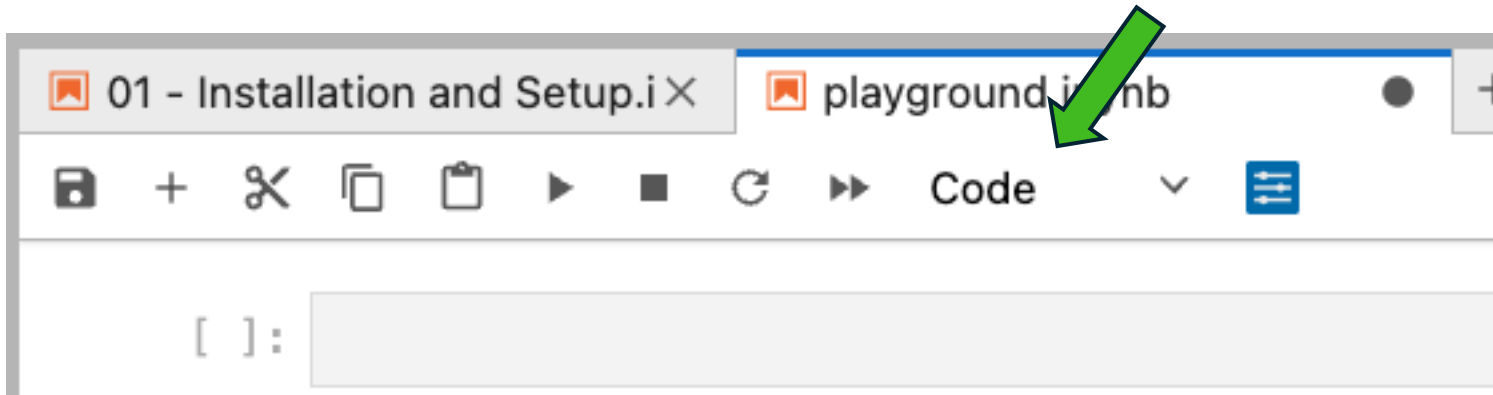


Jupyter notebook 사용법

- 단축키 (Command mode 에서 사용 가능)
 - 'a' 누르면: 해당 셀 앞에 추가 셀 형성
 - 'b' 누르면: 해당 셀 뒤에 추가 셀 형성
 - 툴바에서 선택해도 됨.



- 셀의 성격을 결정할 수 있음.

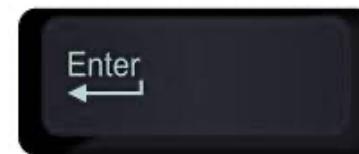
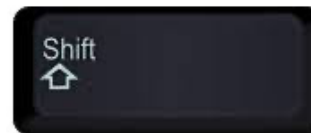
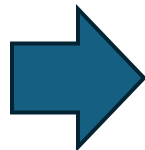


Jupyter notebook 사용법

- '코드'셀 실행하기

```
[3]: 1+1
```

```
[3]: 2
```



- 셀은 실행된 순서대로 실행된다. (위에서 아래로가 아니라, 실행한 순서!)

```
[7]: a = 1
```

```
[9]: a = a + 2
```

```
[11]: print(a)
```

3

VS.

```
[7]: a = 1
```

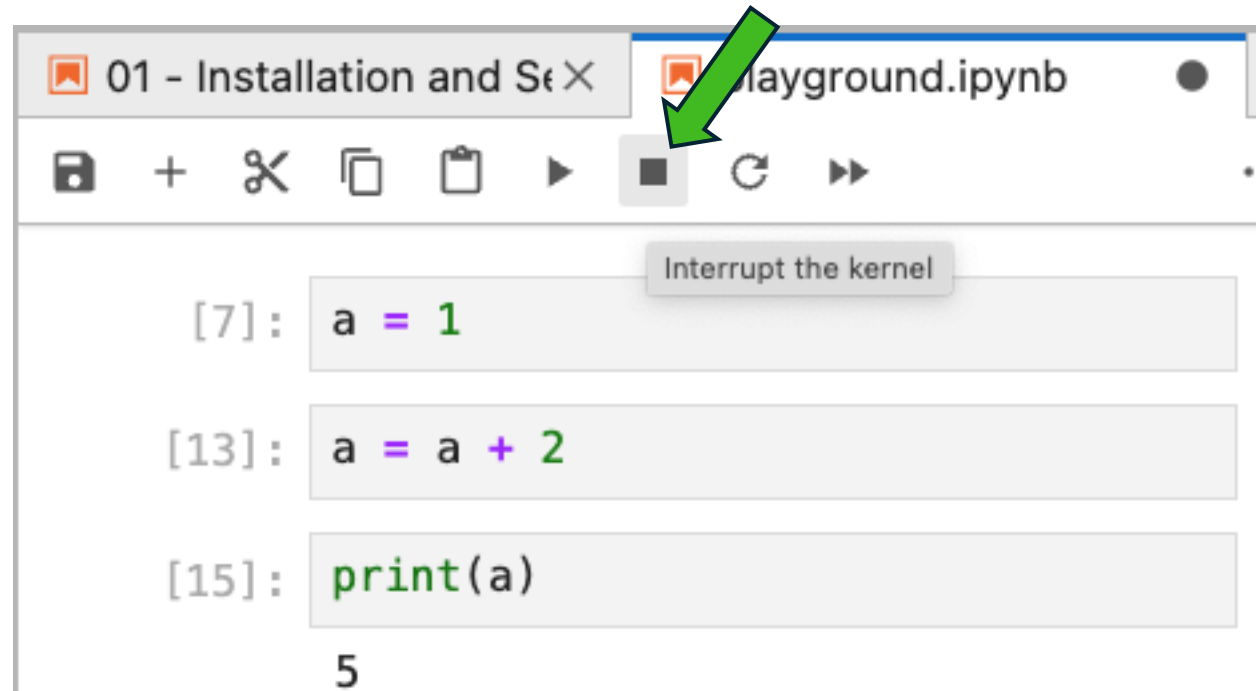
```
[13]: a = a + 2
```

```
[15]: print(a)
```

5

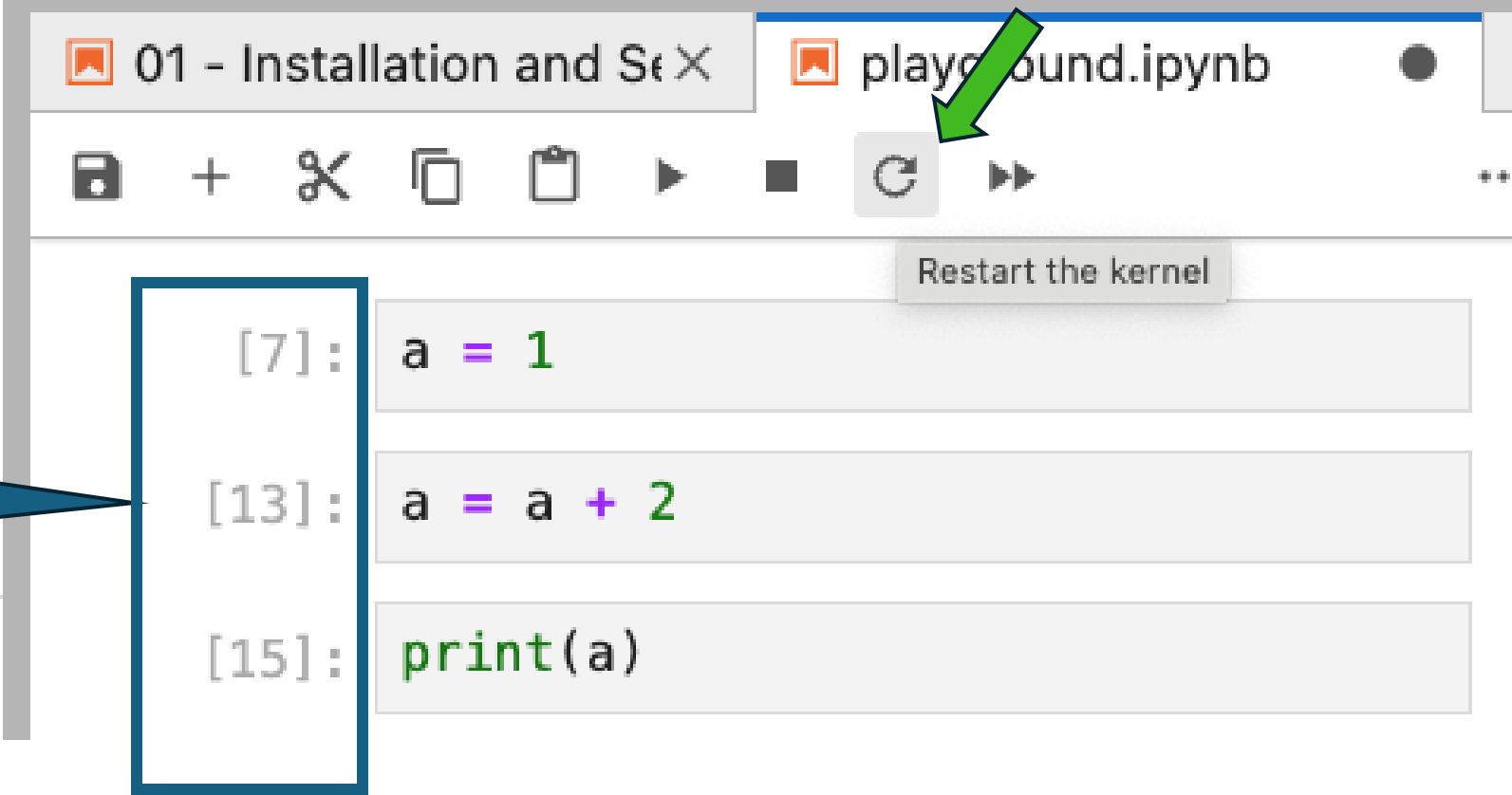
Jupyter notebook 사용법

- 실행 중이지만 멈춰야하는 상황에 멈출 수 있음.



Jupyter notebook 사용법

- 커널 다시 시작: restart kernel



The screenshot shows a Jupyter Notebook window with two tabs: '01 - Installation and Se...' and 'playground.ipynb'. The 'playground.ipynb' tab is active. Below the tabs is a toolbar with icons for saving, adding, deleting, copying, pasting, running, and restarting the kernel. A green arrow points to the 'Restart the kernel' button, which has a tooltip that says 'Restart the kernel'. Below the toolbar are three code cells. The first cell is highlighted with a blue box and contains the code: `[7]: a = 1`. The second cell contains the code: `[13]: a = a + 2`. The third cell contains the code: `[15]: print(a)`. A blue speech bubble with the text '상대적인 실행 순서' (Relative execution order) points to the first cell.

상대적인 실행
순서

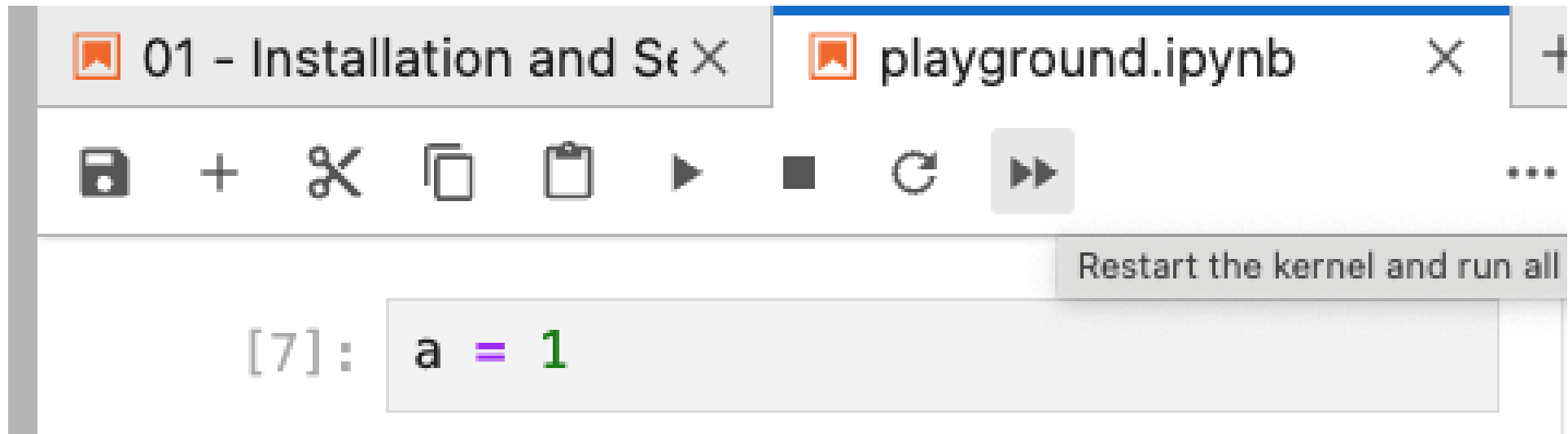
```
[7]: a = 1
```

```
[13]: a = a + 2
```

```
[15]: print(a)
```

Jupyter notebook 사용법

- 커널 다시 시작: restart the kernel
 - 초기 상태로 만들 때
 - 커널 (= 세팅한 환경) 에 새로 설치한 모듈 혹은 환경 적용 필요할 때
 - Restart the kernel + run all from top



Jupyter notebook 사용법

- Shift + enter vs. command (or ctrl) + enter
 - 셀 이동의 차이.
- 셀 안에 여러 줄의 코드를 쓸 수 있음.

```
[25]: a = a + 2  
      10 + 8
```

```
[25]: 18
```



여러 줄의 코드 중 마지막 실행 결과만 보여줌.
위의 코드도 실행은 되었음.
결과가 안 보인다고 해서 실행이 안된 것이 아님.

Jupyter notebook 사용법

- Jupyter notebook 에서 쓸 라이브러리를 갖고 와야함: `import [library 이름]`

```
[2]: import pandas
```

```
[ ]: |
```

- 모듈 이름을 줄이거나, 모듈 중 사용할 함수만 따로 불러올 수 있음.
 - 설명할 수 있는 사람?

```
[4]: import pandas as pd
```


과제 1: Python crash (팀과제)

- 총 17문제, 핵심적으로 알아야하는 내용 포함. -> 다음 시간 테스트

Python Crash Course

다음 코드 셀들을 모두 실행시키고, 관련 질문에 맞는 코드를 작성하시오.

Comments

- A **comment** is a line of text ignored by Python.
- Use a hashtag/octothorpe (#) to create a comment.

1. 다음 계산값이 모두 출력되도록 cell 2 에 수정하여 작성하시오.

```
# cell 1
3 + 3
5 + 5
10 + 10 # Solve really complex mathematical expression using Python
```

과제2 (개인과제) – 토요일 자정까지.

- Ipython 설명 번역해서 업로드

