

데이터분석입문

Lecture 02. 데이터 분석 개요 및 실습도구 소개

동양미래대학교
인공지능소프트웨어학과
강 환수

- ❖ 01. 데이터 분석이란?
- ❖ 02. 파이썬이란?
- ❖ 03. 데이터 분석에 필요한 환경 만들기

01. 데이터 분석이란?

02. 파이썬이란?

03. 데이터 분석에 필요한 환경 만들기

❖ 데이터 분석 (Data Analysis)

- 데이터로부터 유용한 정보를 발굴하고 결론 내용을 알리며 의사결정을 지원하는 것을 목표로 데이터를 정리, 변환, 모델링하는 과정 (from Wikipedia)
 - ◆ 특히, 오늘날 비즈니스 부문에서 데이터 분석은 의사 결정을 더 과학적으로 만들어주고 비즈니스를 더 효율적으로 운영할 수 있도록 도와주는 역할을 함



데이터 분석 도구 (Tool)에는 무엇이 있을까?

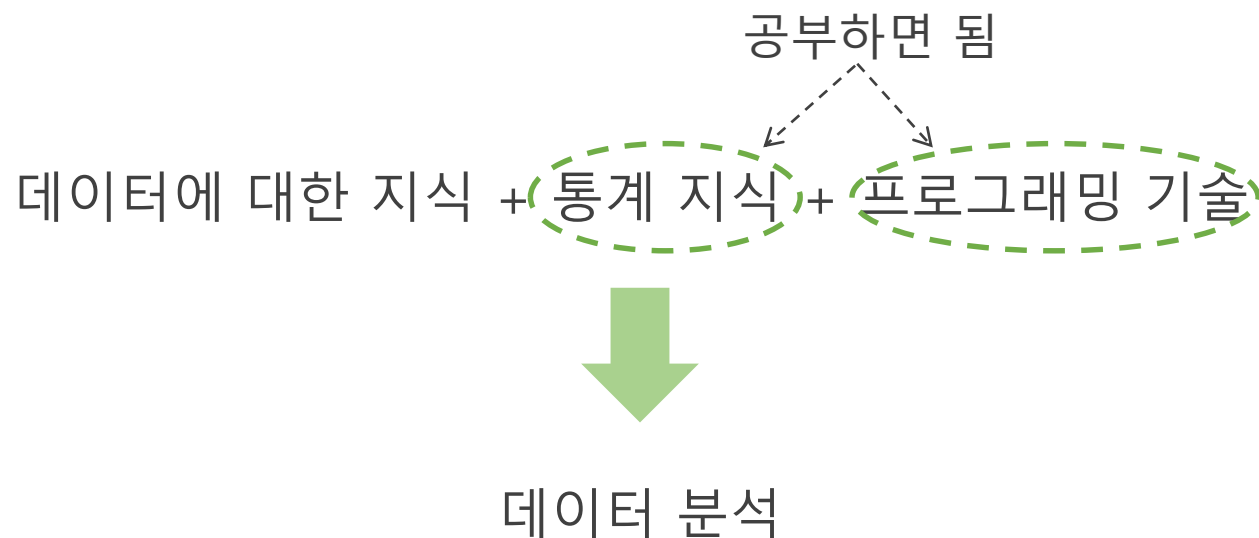
❖ 데이터 분석 도구의 종류

- ① 파이썬 (Python)
- ② R
- ③ 엑셀 (Excel)
- ④ SQL (Structured Query Language, 구조화 질의어)
- ⑤ 태블로 (Tableau)
- ⑥ Power BI(Business Intelligence)
- ⑦ 구글 애널리틱스 (Google Analytics, GA)



❖ 데이터 분석을 하려면?

- ① 데이터에 대한 지식
 - ◆ 데이터를 제공하는 쪽에서 사전 정보를 줄 수도 있고, 안 줄 수도 있음
- ② 통계 지식
 - ◆ 데이터로부터 어떤 정보를 추출해 낼 것인지? 의사결정에 도움이 되는 정보인지 판단
- ③ 프로그래밍 기술
 - ◆ 분석하고자 하는 내용을 프로그램으로 구현하는데 필요



02. 파이썬이란?

01. 데이터 분석이란?

03. 데이터 분석에 필요한 환경 만들기

❖ 파이썬 (Python)

- 1991년에 귀도 반 로섬 (Guido van Rossum)에 의해 개발된 프로그래밍 언어
- 문법이 매우 쉬워서 작성하기에 간단하기 때문에 초보자들이 처음 프로그래밍을 배울 때 추천되는 언어
- 귀도 반 로섬이 즐겨 보던 코미디 <Monty Python's Flying Circus>에서 이름을 따옴



귀도 반 로섬 (Guido van Rossum)

- ✓ 네덜란드의 컴퓨터 프로그래머
- ✓ 2005년부터 2012년 12월까지 구글에서 근무
- ✓ 2013년부터 2019년 10월까지 드롭박스에서 근무
- ✓ 2020년 11월에 Microsoft 입사

[사진출처] [https://en.wikipedia.org/wiki/Python_\(programming_language\)#/media/File:Guido_van_Rossum_OSCON_2006_cropped.png](https://en.wikipedia.org/wiki/Python_(programming_language)#/media/File:Guido_van_Rossum_OSCON_2006_cropped.png)



❖ 파이썬의 활용

- 업무 시간을 단축해 주는 언어로 실무에서 널리 활용되고 있습니다.

- ✓ 엑셀 자동화
- ✓ 파일 처리 자동화
- ✓ 웹 크롤링 자동화
- ✓ 3D 모델링 자동화
- ✓ 기타

- 최근에는 머신러닝, 딥러닝 등 인공지능 개발을 위한 언어로서 각광을 받고 있습니다.

◆ 파이썬으로 접근 가능한 머신러닝 & 딥러닝 라이브러리

- ✓ 사이킷런 (scikit-learn)
- ✓ 텐서플로 (TensorFlow)
- ✓ 파이토치 (PyTorch)
- ✓ 아파치 스파크 MLlib (Apache Spark MLlib)
- ✓ CNTK (Microsoft **C**ognitive **T**oolkit)

❖ 파이썬 버전

- 파이썬 2.0

- ◆ 2000년 10월 16일 배포
- ◆ 2020년 1월 1일 지원 종료

- 파이썬 3.0

- ◆ 2008년 12월 3일 배포
- ◆ 최신 버전: Python 3.11.5 (2023년 8월 24일 기준)

2.X 버전과 3.X 버전은 문법이 많이 다릅니다.

Active Python Releases

For more information visit the [Python Developer's Guide](#).

Python version	Maintenance status	First released	End of support	Release schedule
3.12	prerelease	2023-10-02 (planned)	2028-10	PEP 693
3.11	bugfix	2022-10-24	2027-10	PEP 664
3.10	security	2021-10-04	2026-10	PEP 619
3.9	security	2020-10-05	2025-10	PEP 596
3.8	security	2019-10-14	2024-10	PEP 569

[사진출처] <https://www.python.org/downloads/>

❖ 파이썬 개발 환경 (1/4)

- ① 명령줄 인터페이스 (Command Line Interface, CLI)

- ◆ 터미널

- 짧은 파이썬 코드가 잘 작동하는지 즉흥적으로 실험해 볼 수 있음

- ◆ IDLE (Integrated Development and Learning Environment)

- 파이썬 기본 탑재 콘솔

- 직접 명령을 입력하는 것 외에도 파이썬 스크립트를 직접 편집할 수 있음

❖ 파이썬 개발 환경 (2/4)

- ② 통합 개발 환경 (Integrated Development Environment, IDE)

- ◆ PyCharm

- JetBrains사에서 개발한 IDE로, 코드 컴플리션 기능이 우수하고 파이썬의 장점인 각종 프레임워크 지원이 가장 뛰어남
- Community Edition은 무료
- Professional Edition은 유료

- ◆ Spyder

- 오픈 소스 IDE로, PyCharm Professional Edition 같은 IDE를 사용하기 힘든 학생이나 개인 개발자가 사용



[왼쪽 사진출처] https://en.wikipedia.org/wiki/Visual_Studio_Code#/media/File:Visual_Studio_Code_1.35_icon.svg

[중앙 사진출처] https://en.wikipedia.org/wiki/PyCharm#/media/File:PyCharm_Icon.svg

[오른쪽 사진출처] [https://en.wikipedia.org/wiki/Spyder_\(software\)#/media/File:Spyder_logo.svg](https://en.wikipedia.org/wiki/Spyder_(software)#/media/File:Spyder_logo.svg)

❖ 파이썬 개발 환경 (3/4)

- ③ 코드 편집기

- ◆ Notepad++

- ◆ Sublime Text

- ◆ Atom

- ◆ 메모장

- ◆ Visual Studio Code

- Microsoft사에서 개발한 IDE로 전세계적으로 가장 인기 있는 개발 환경 중에 하나임

개인 취향이나 프로젝트 성격에 따라 IDE가 아닌 가볍고 빠른 코드 편집기를 더 선호하기도 합니다



❖ 파이썬 개발 환경 (4/4)

- ④ 노트북 편집기

- ◆ IPython

- ◆ Jupyter Notebook

- ◆ JupyterLab

- ◆ Google Colaboratory

- ◆ Kaggle

셀(Cell)을 만들어서 셀 안에 코드를 작성해 놓고,
실행시키면서 셀 별로 수행 결과를 확인을 할 수 있습니다.

코딩 → 결과 확인 → 코드 수정 → 결과 확인 등의
반복적인 작업에 유용

- Vs code 에서도 노트북 편집이 가능

03. 데이터 분석에 필요한 환경 만들기

- 01. 데이터 분석이란?
- 02. 파이썬이란?

❖ 아나콘다 (Anaconda)

- 과학 연구 및 머신러닝 분야에 적합한 파이썬 및 R 언어의 패키지/의존성 관리 및 배포를 편리하게 해주는 조건부 무료 오픈소스 SW

❖ 주피터 노트북 (Jupyter Notebook)

- 파이썬 프로그래밍 언어로 코드를 작성하고 실행하는 개발 환경 제공
- 일부 코드의 실행 결과를 확인할 수 있고 문서화하여 다른 사람과 공유 가능
- 아나콘다를 설치하면 주피터 노트북을 바로 사용할 수 있음

❖ 웹 브라우저

- Microsoft Edge
- Chrome

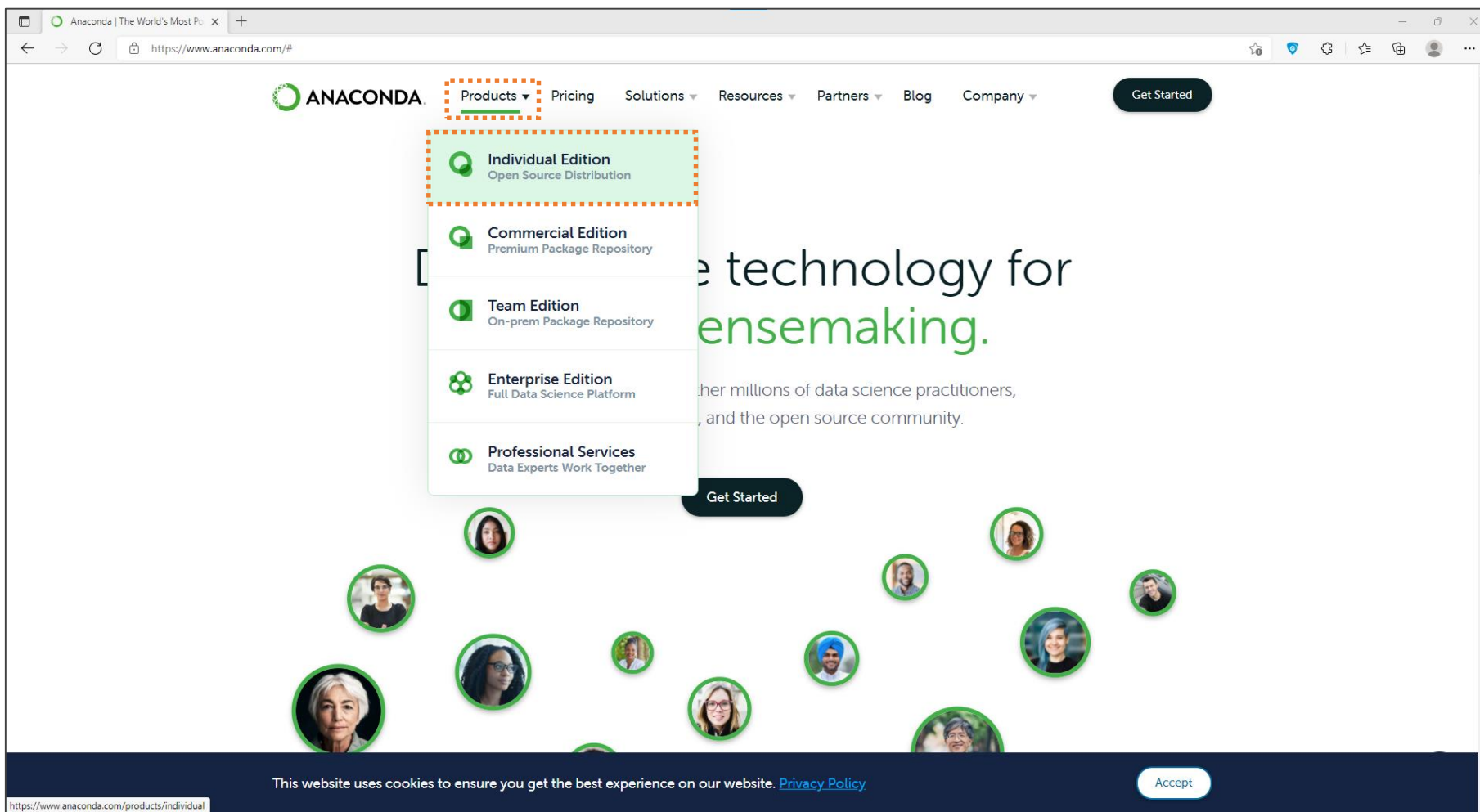


[왼쪽 사진출처] [https://en.wikipedia.org/wiki/Microsoft_Edge#/media/File:Microsoft_Edge_logo_\(2019\).svg](https://en.wikipedia.org/wiki/Microsoft_Edge#/media/File:Microsoft_Edge_logo_(2019).svg)

[오른쪽 사진출처] [https://en.wikipedia.org/wiki/Google_Chrome#/media/File:Google_Chrome_icon_\(February_2022\).svg](https://en.wikipedia.org/wiki/Google_Chrome#/media/File:Google_Chrome_icon_(February_2022).svg)

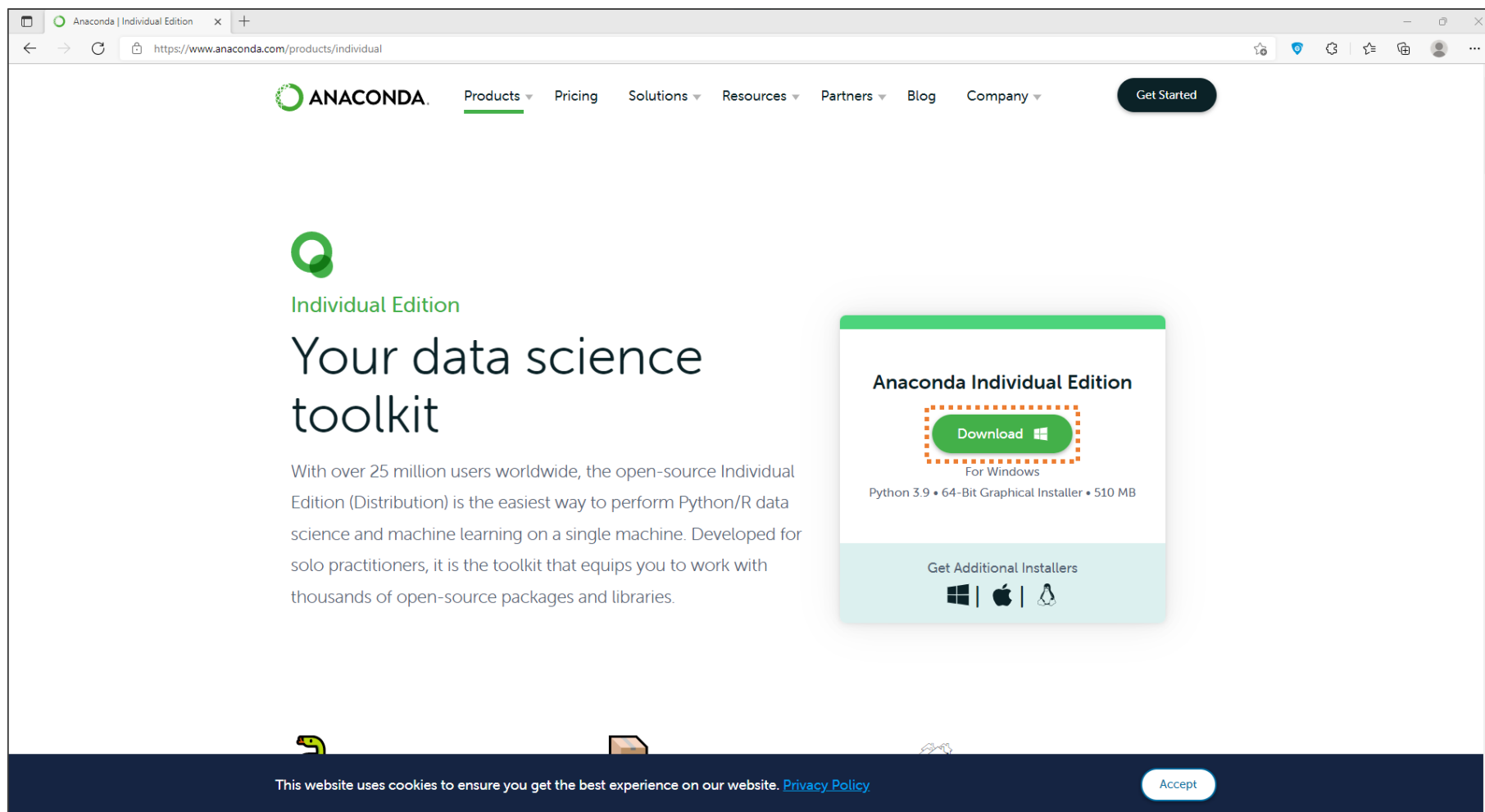
❖ ① 아나콘다 설치하기 (1/10)

- 아나콘다 사이트 (<https://www.anaconda.com/>)에 접속하여, Products → Individual Edition을 클릭



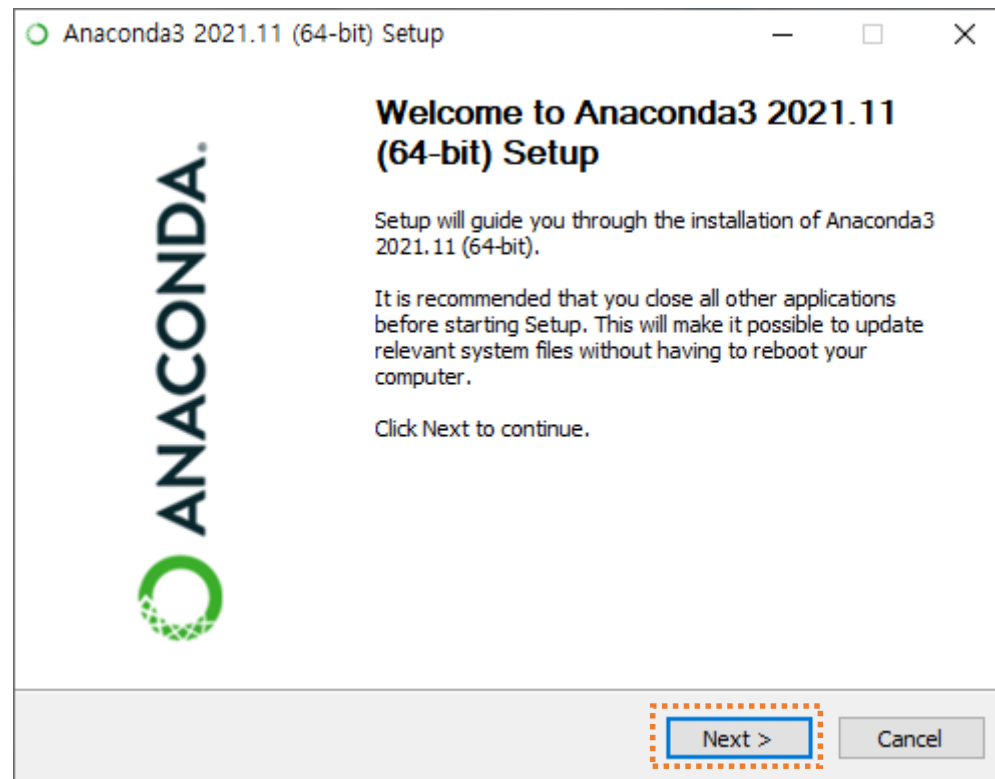
❖ ① 아나콘다 설치하기 (2/10)

- Download 버튼 클릭



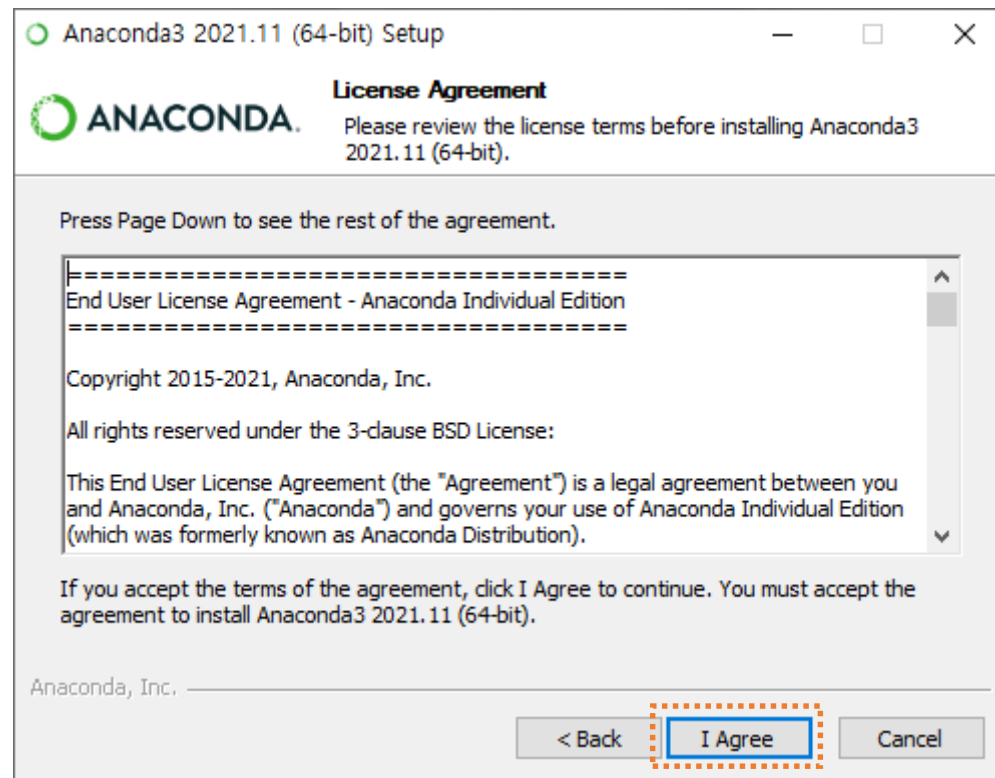
❖ ① 아나콘다 설치하기 (3/10)

- 내려받은 파일을 실행해 다음 화면이 나오면 Next 버튼을 클릭하여 설치를 진행



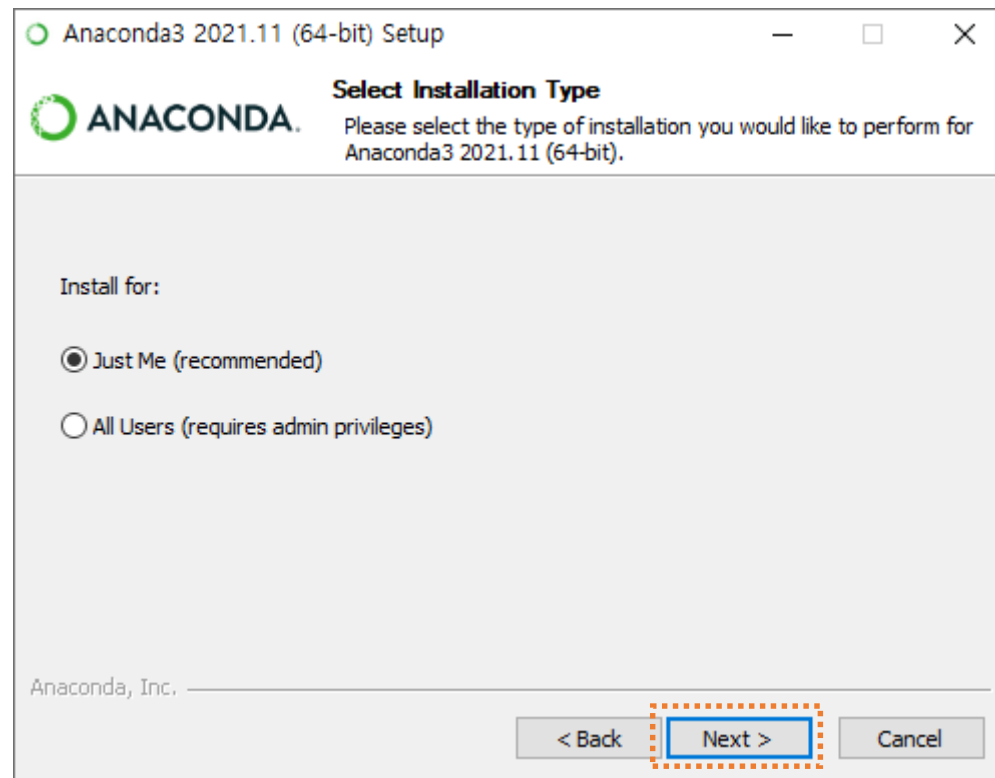
❖ ① 아나콘다 설치하기 (4/10)

- 라이선스 동의 화면이 나오면 I Agree 버튼을 클릭



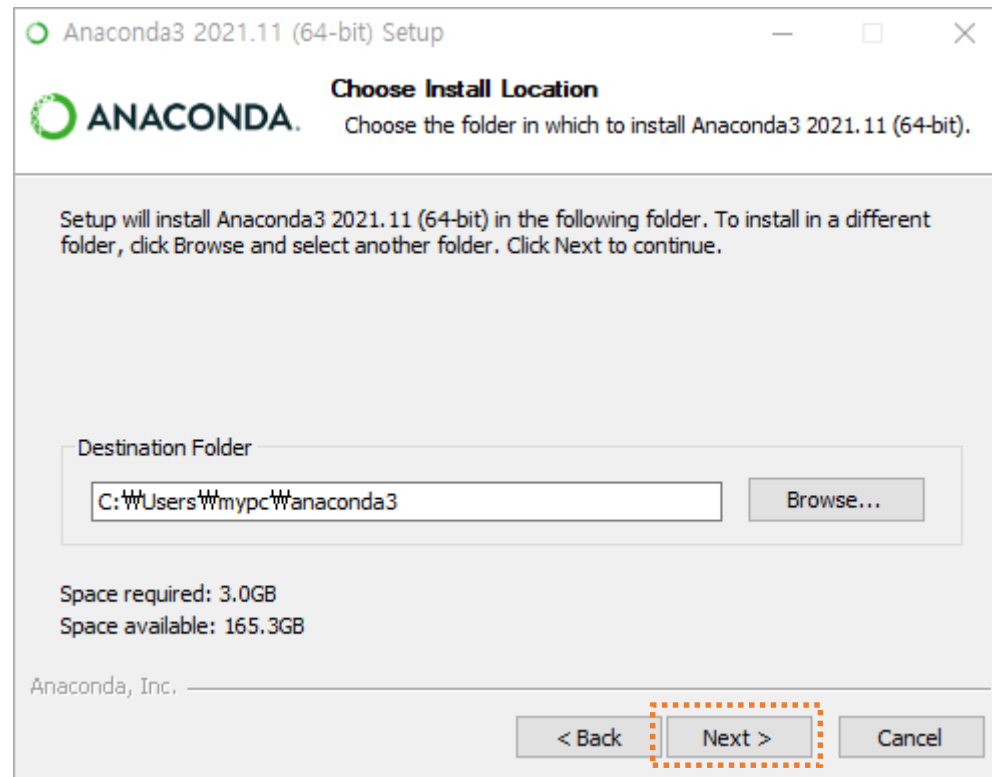
❖ ① 아나콘다 설치하기 (5/10)

- 설치 타입을 묻는 창이 나오면, Just Me (recommended)를 선택하고 Next 버튼을 클릭



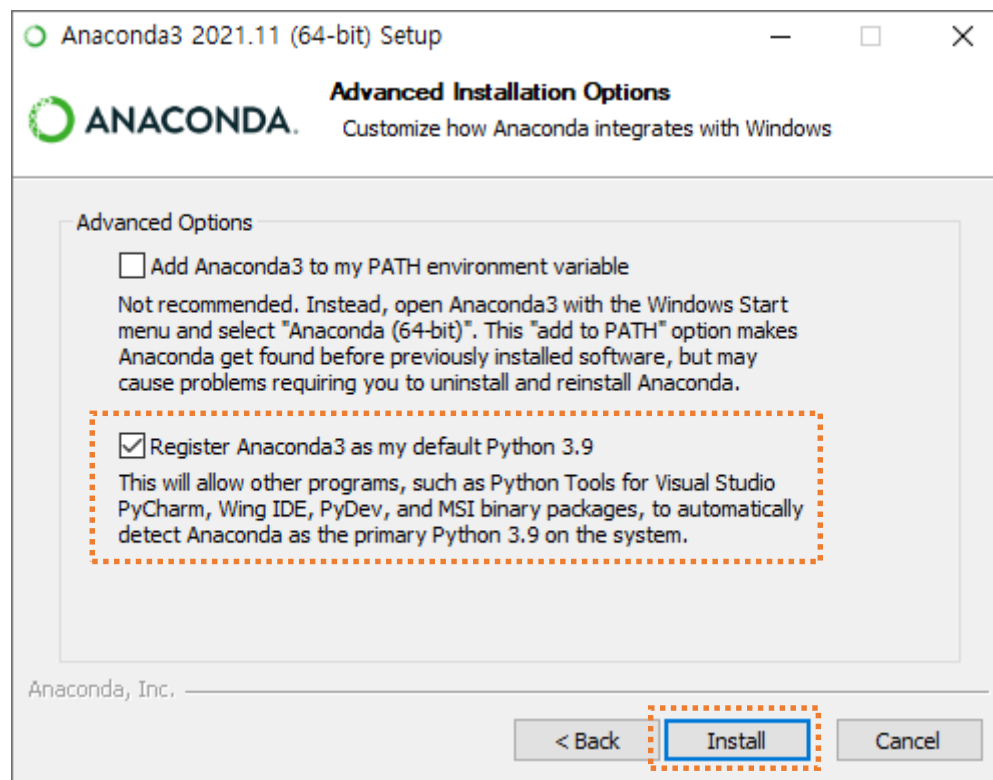
❖ ① 아나콘다 설치하기 (6/10)

- 설치 경로를 묻는 창이 나오면 기본값으로 두거나 원하는 경로를 지정하고 Next 버튼 클릭



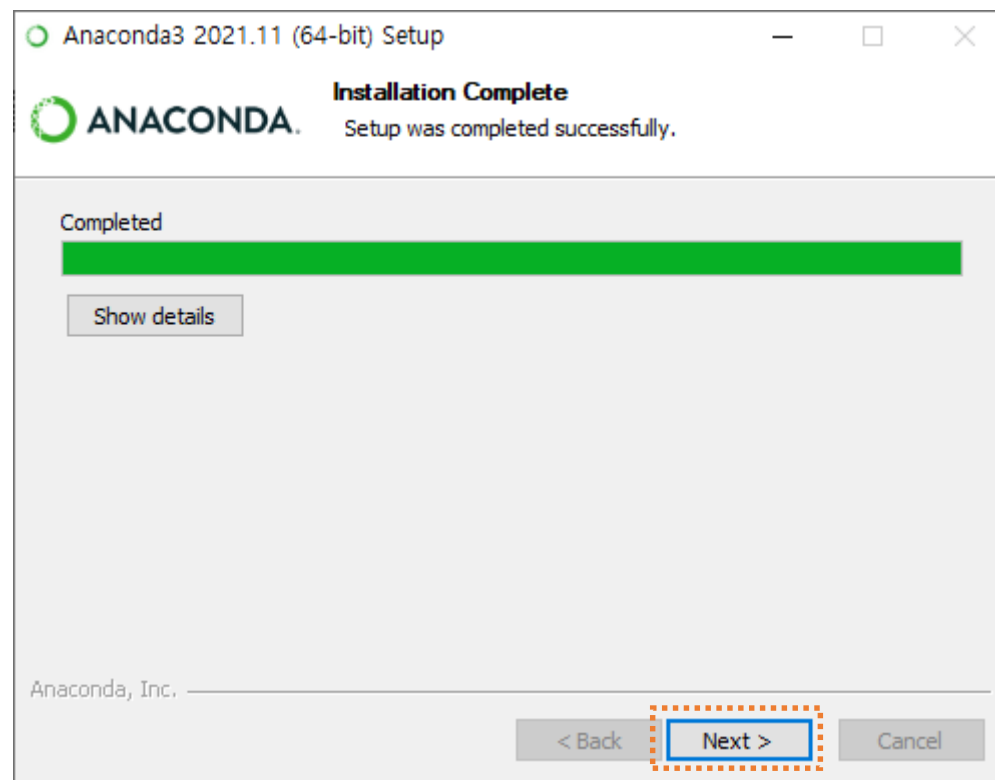
❖ ① 아나콘다 설치하기 (7/10)

- 옵션을 설정하는 화면이 나오면 기본값으로 두고 Install 버튼을 클릭



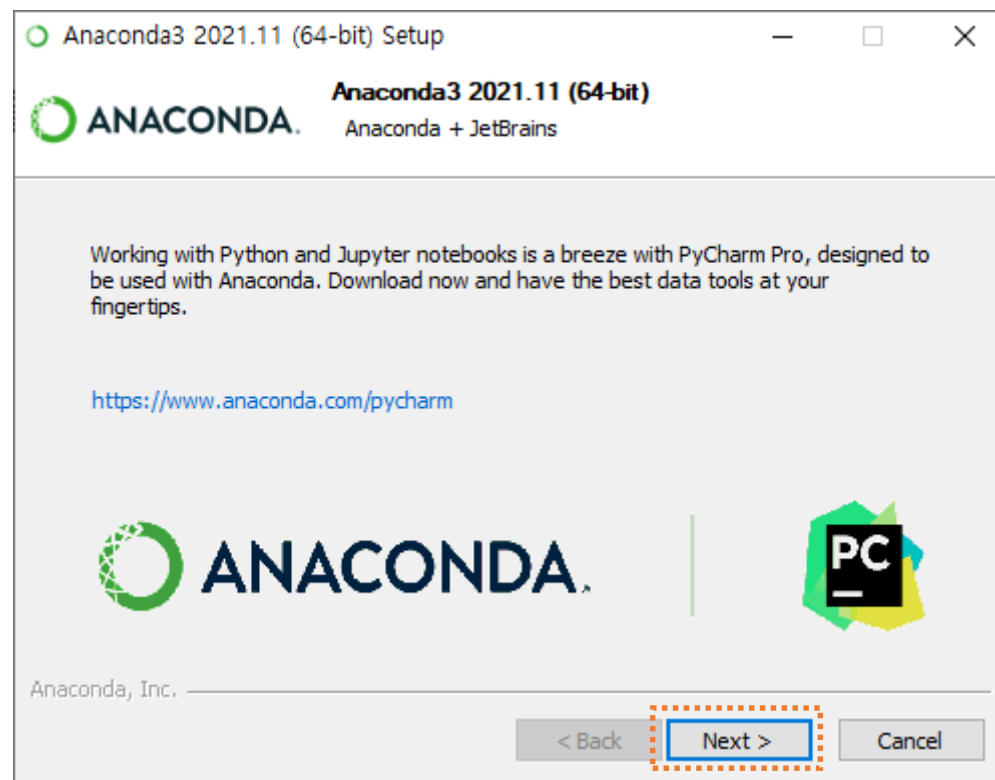
❖ ① 아나콘다 설치하기 (8/10)

- 설치에는 몇 분 정도 소요되며, 다음과 같이 설치가 진행되면 Next 버튼을 클릭



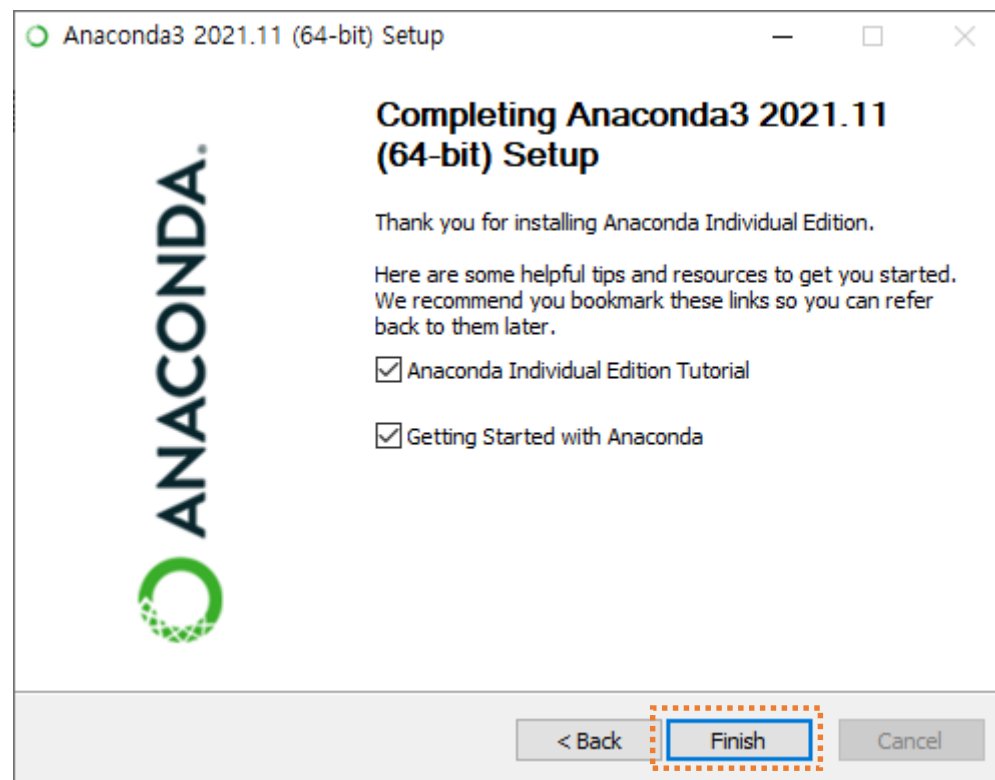
❖ ① 아나콘다 설치하기 (9/10)

- Next 버튼을 클릭



❖ ① 아나콘다 설치하기 (10/10)

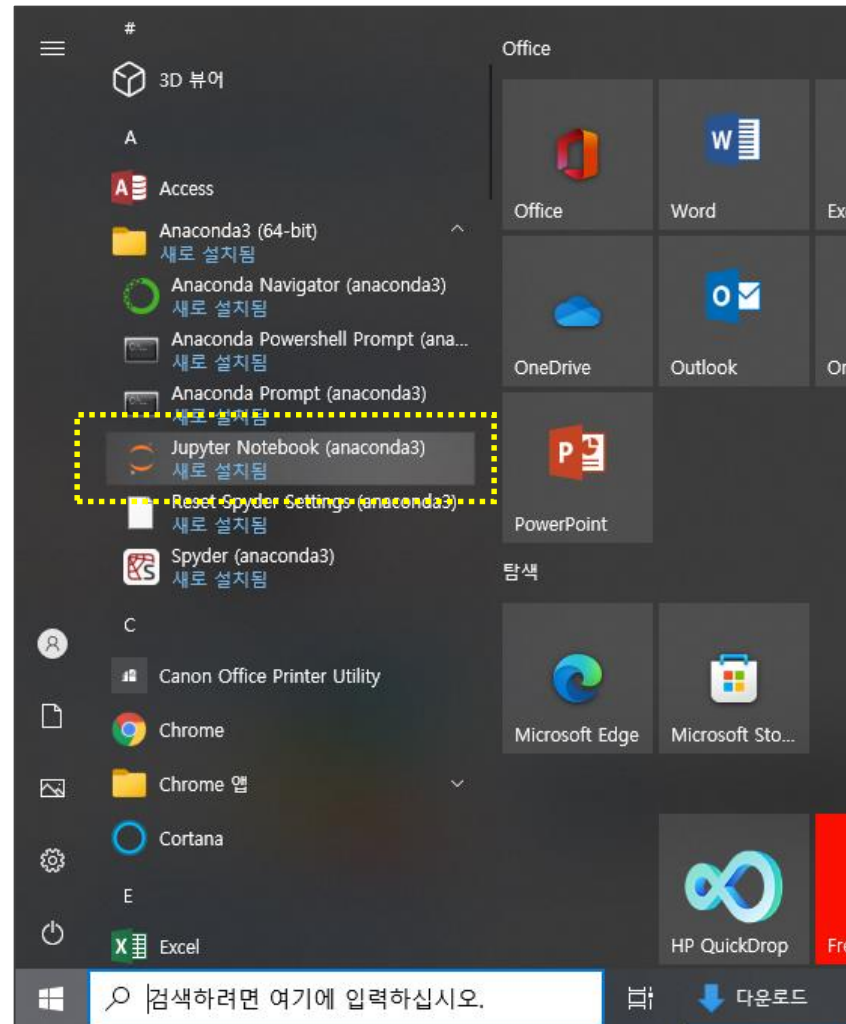
- 설치를 마쳤으면 Finish 버튼을 클릭



❖ ② 주피터 노트북 (Jupyter Notebook) 실행하기 (1/4)

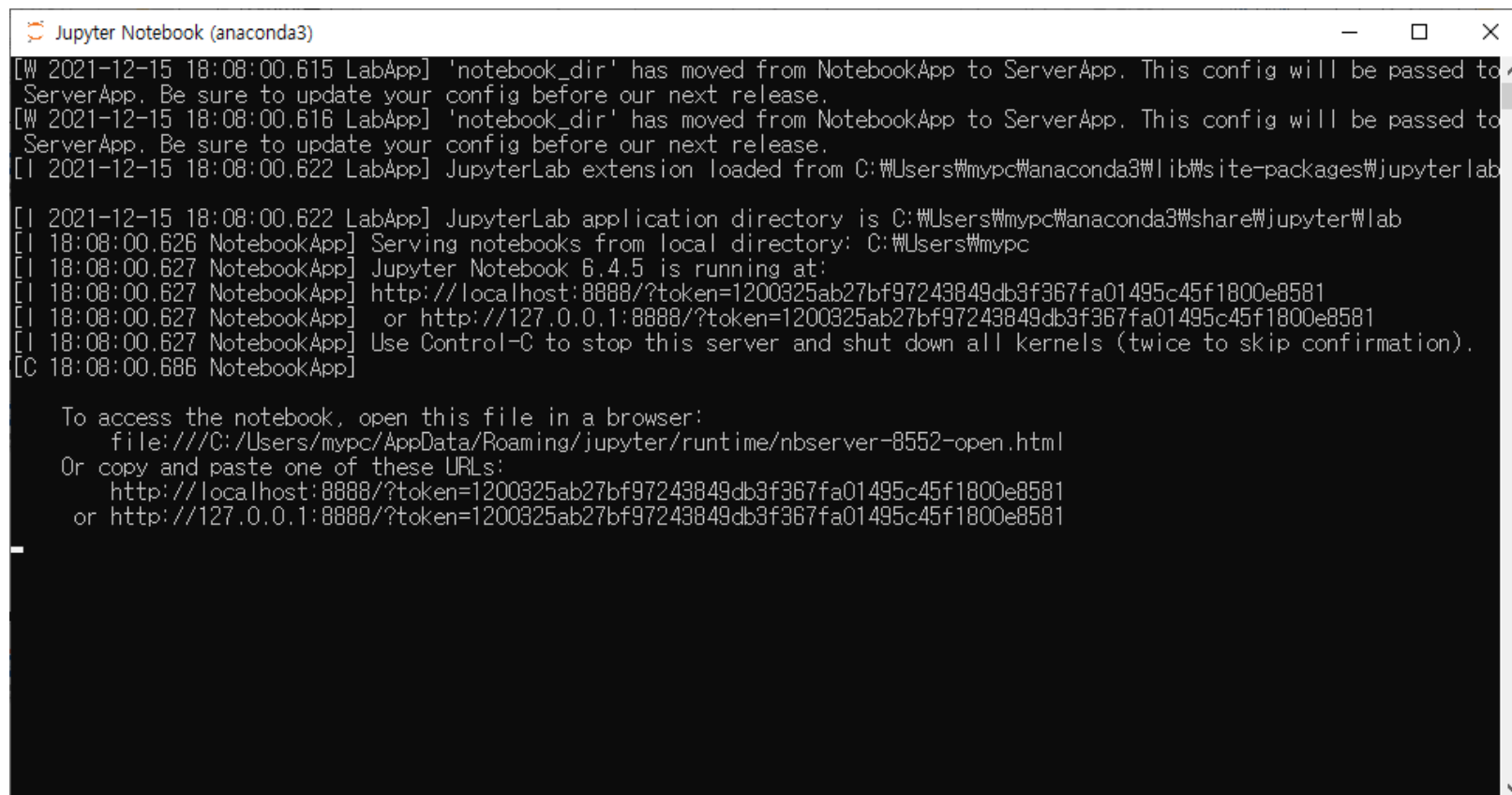
- 작업 표시줄의 윈도우 키를 눌러서 Anaconda3 폴더의 Jupyter Notebook 아이콘을 클릭

윈도우 키



❖ ② 주피터 노트북 (Jupyter Notebook) 실행하기 (2/4)

- 검은 창이 뜬 후 자동으로 브라우저가 켜질 때까지 기다립니다.
- 아래 검은 창을 끄면 주피터 노트북이 작동되지 않으니 실습 할 때는 항상 켜 두세요.

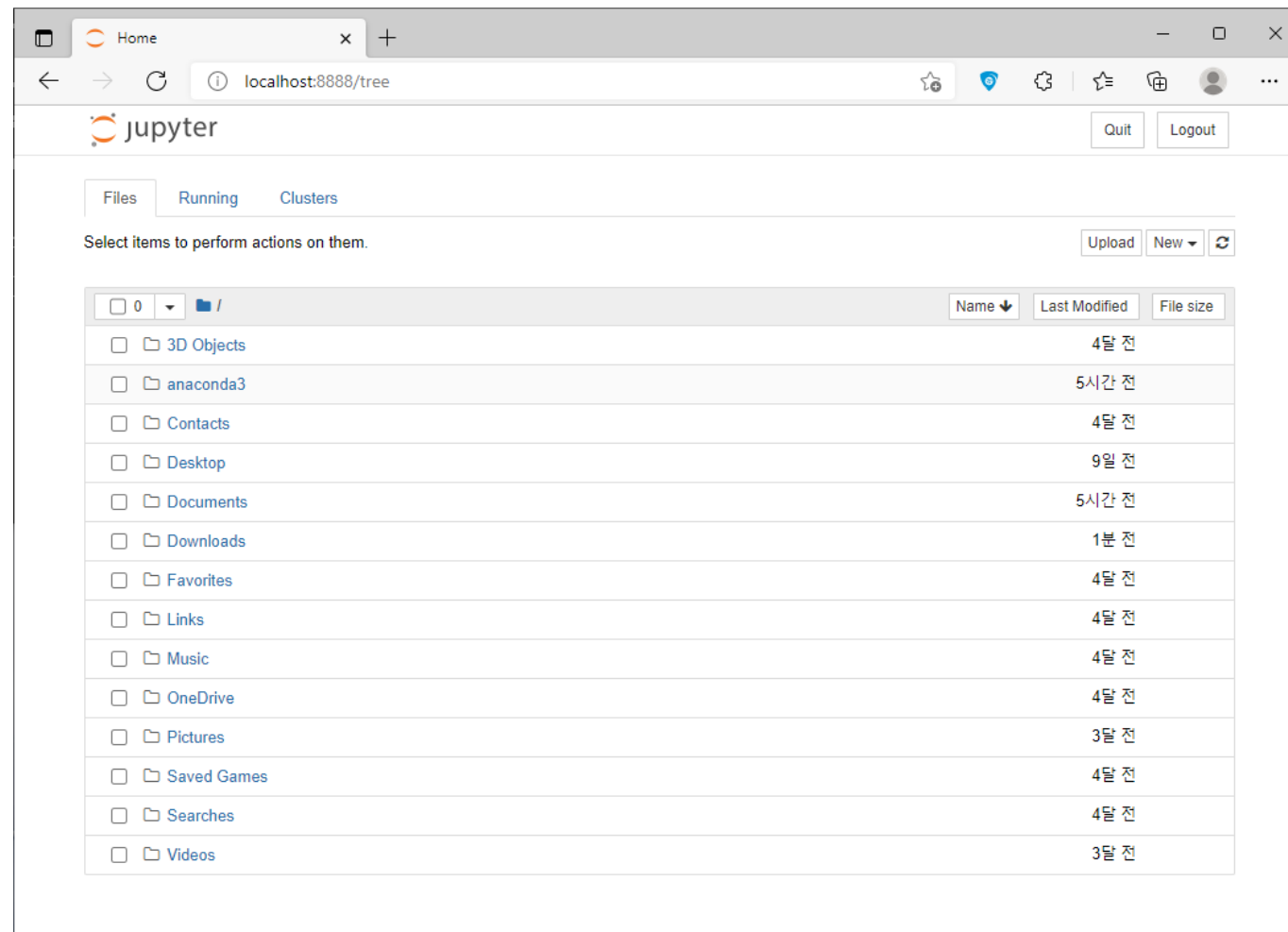


```
Jupyter Notebook (anaconda3)
[W 2021-12-15 18:08:00.615 LabApp] 'notebook_dir' has moved from NotebookApp to ServerApp. This config will be passed to
ServerApp. Be sure to update your config before our next release.
[W 2021-12-15 18:08:00.616 LabApp] 'notebook_dir' has moved from NotebookApp to ServerApp. This config will be passed to
ServerApp. Be sure to update your config before our next release.
[I 2021-12-15 18:08:00.622 LabApp] JupyterLab extension loaded from C:\Users\mypc\anaconda3\lib\site-packages\jupyterlab
[I 2021-12-15 18:08:00.622 LabApp] JupyterLab application directory is C:\Users\mypc\anaconda3\share\jupyter\lab
[I 18:08:00.626 NotebookApp] Serving notebooks from local directory: C:\Users\mypc
[I 18:08:00.627 NotebookApp] Jupyter Notebook 6.4.5 is running at:
[I 18:08:00.627 NotebookApp] http://localhost:8888/?token=1200325ab27bf97243849db3f367fa01495c45f1800e8581
[I 18:08:00.627 NotebookApp] or http://127.0.0.1:8888/?token=1200325ab27bf97243849db3f367fa01495c45f1800e8581
[I 18:08:00.627 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 18:08:00.686 NotebookApp]

To access the notebook, open this file in a browser:
    file:///C:/Users/mypc/AppData/Roaming/jupyter/runtime/nbserver-8552-open.html
Or copy and paste one of these URLs:
    http://localhost:8888/?token=1200325ab27bf97243849db3f367fa01495c45f1800e8581
    or http://127.0.0.1:8888/?token=1200325ab27bf97243849db3f367fa01495c45f1800e8581
```

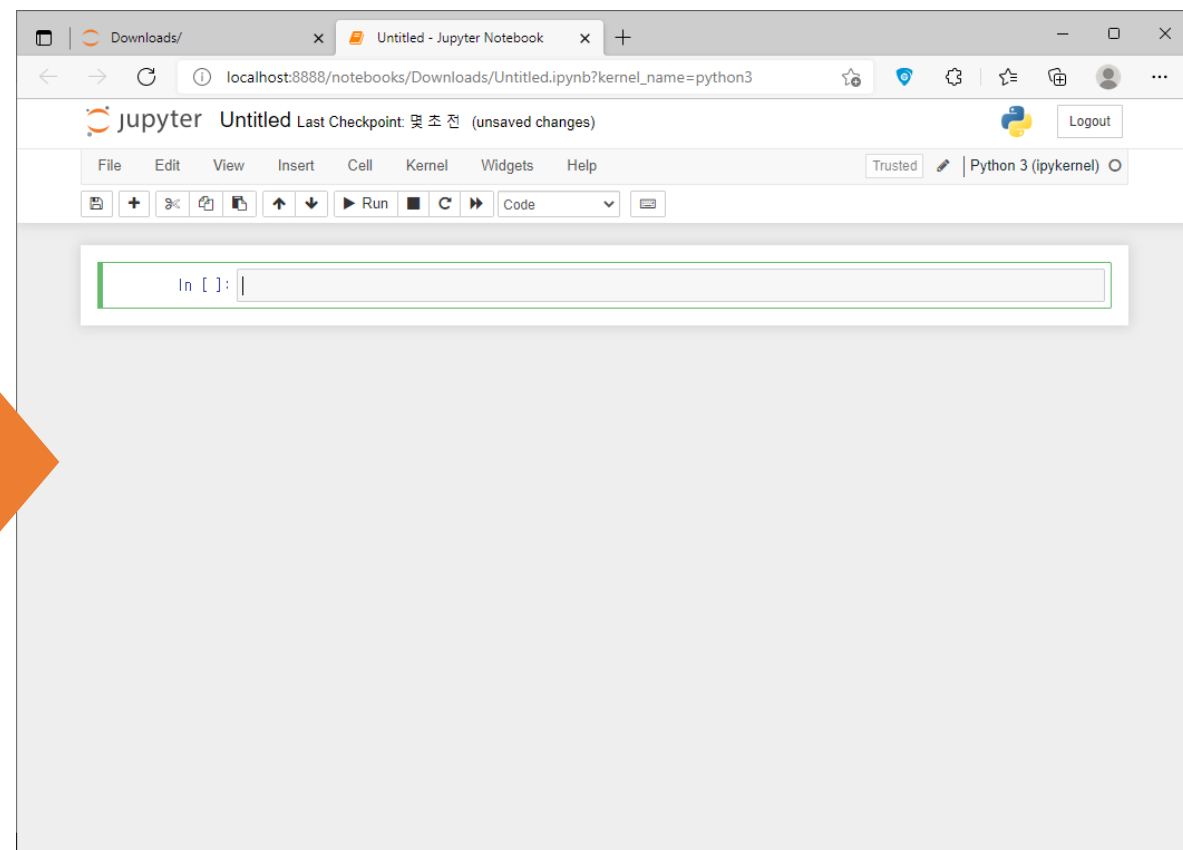
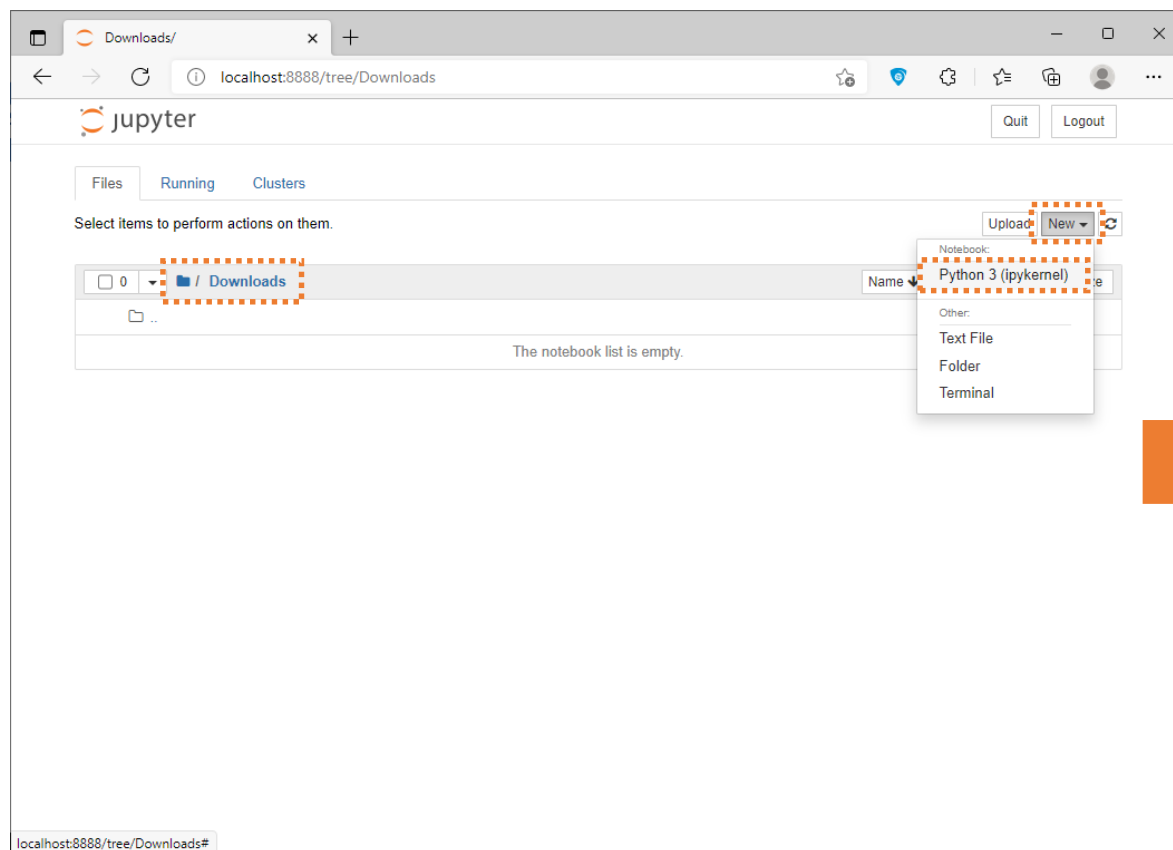
❖ ② 주피터 노트북 (Jupyter Notebook) 실행하기 (3/4)

- 브라우저가 켜지면 아래와 같은 화면이 나옵니다.



❖ ② 주피터 노트북 (Jupyter Notebook) 실행하기 (4/4)

- 브라우저에서 Downloads 폴더를 선택한 후 오른쪽에 있는 New → Python 3를 클릭하여 새 파이썬 노트북을 생성합니다.



- ❖ 01. 데이터 분석이란?
- ❖ 02. 파이썬이란?
- ❖ 03. 데이터 분석에 필요한 환경 만들기

THANK YOU!

Q & A

- Name: 강환수
- Office: 동양미래대학교 2호관 706호 (02-2610-1941)
- E-mail: hsknag@dongyang.ac.kr
- Homepage: <https://github.com/ai7dnn/2023-DA>