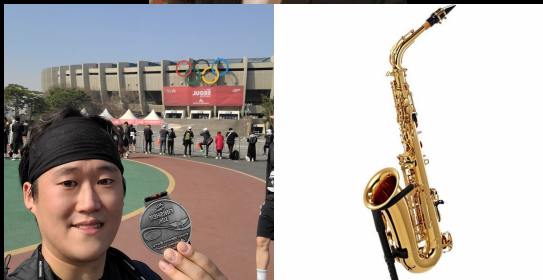


Jupyter Book을 활용해 손쉽게 콘텐츠를 생산하고 공유하자!

안성진 (Tony Ahn)

BACK TO US,
BACK TO PYTHON

발표자 소개



- **Microsoft**
 - Technical Trainer - Data & AI
- **Pseudo Lab (가짜연구소)**
 - 아카데미 빌더
 - 스튜디오 운영

Disclaimer

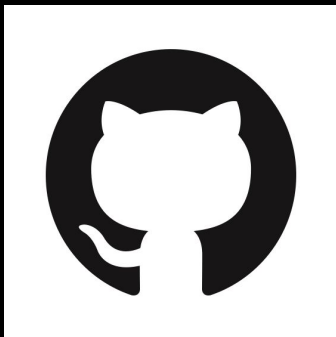
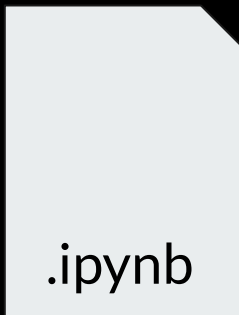
본 발표에서 표현된 의견은 제 개인적인 의견이며, 어떤 식으로든 제 고용주의 견해를 나타내지 않습니다.

The opinions expressed herein are my own personal opinions and do not represent my employer's view in any way.

목차

- 01 - **ipynb**파일 기존 공유 방법의 애로사항
- 02 - **Jupyter Book** 소개 및 특징
- 03 - 콘텐츠 사례
- 04 - **Jupyter Book** 사용법 - 패키지 설치하기
- 05 - **Jupyter Book** 사용법 - 소스 파일 생성
- 06 - **Jupyter Book** 사용법 - **html**파일 빌드
- 07 - **Jupyter Book** 사용법 - 결과물 공유
- 08 - 기타 기능
- 09 - Summary

01 - ipynb파일 기존 공유 방법의 애로사항



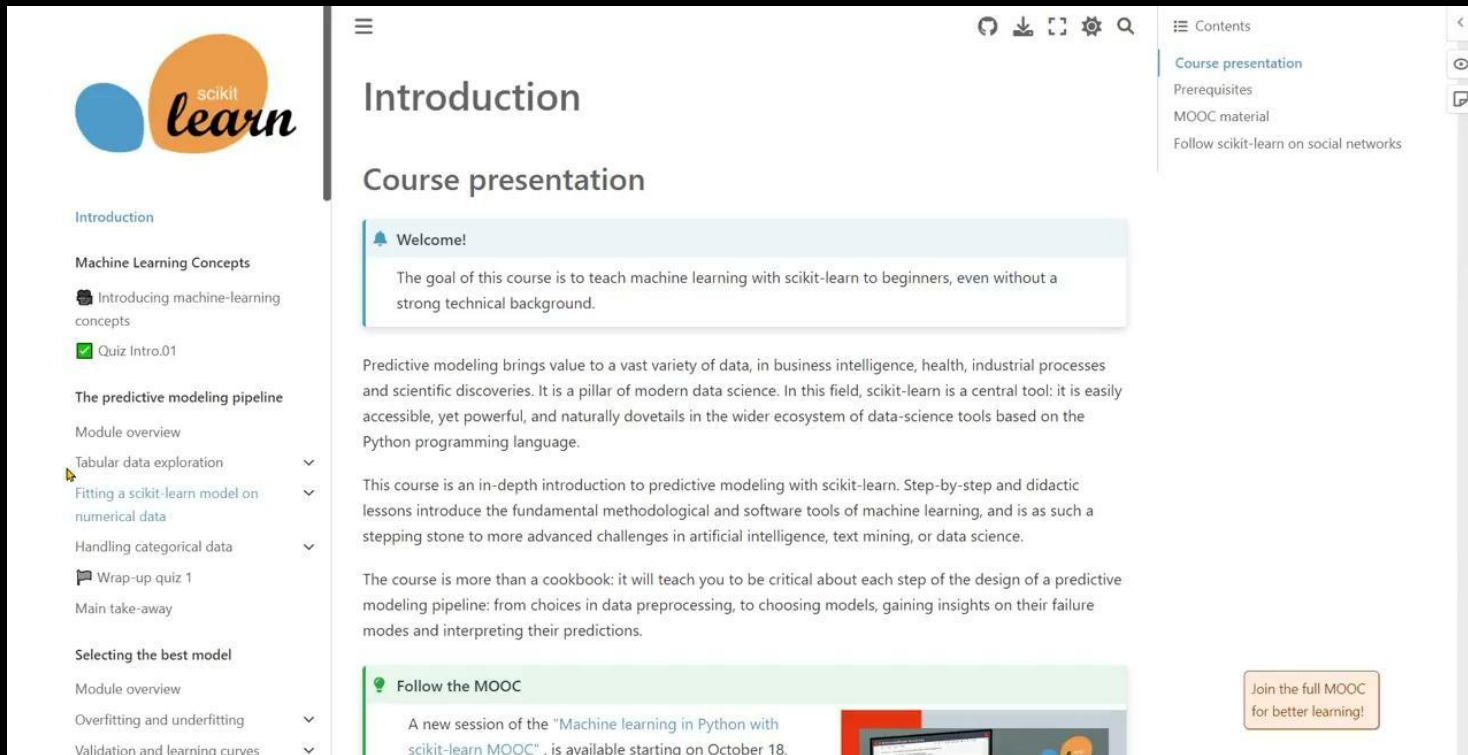
- **.ipynb** 작업물을 커뮤니티에 공유 할 때 **GitHub**을 주로 활용
- **GitHub**에 공유된 **ipynb**파일에 접근 시, 과거에는 렌더링에 시간 소요 및 실패 하는 경우 종종 발생 (최근에는 개선됨)
- **GitHub**은 개발자에게 친숙하고 효율적인 환경이지만, 비개발자에게 상대적으로 낯설고 친화적이지 않은 **UI** 환경

02 - Jupyter Book 소개 및 특징



- .md 및 .ipynb 파일을 활용하여 출판물 수준(publication-quality)의 책과 문서를 만들게 해주는 오픈 소스 도구
- 파일들을 접근성이 높은 **html** 파일로 변환
- **GitHub Pages**와 자연스럽게 연동되어 손쉽게 웹페이지 생성
- **Project Jupyter**의 핵심 기여자들이 모인 **Executable Book Project** 커뮤니티에서 만든 산출물

03 - 콘텐츠 사례 영상자료 1



The screenshot displays the 'Introduction' page of the scikit-learn MOOC. The page is divided into three main sections: a left sidebar, a central content area, and a right sidebar.


Left Sidebar: Features the scikit-learn logo at the top. Below it, the 'Introduction' section is highlighted. Under 'Machine Learning Concepts', there is a link to 'Introducing machine-learning concepts' and a 'Quiz Intro.01' with a green checkmark. The 'The predictive modeling pipeline' section includes links for 'Module overview', 'Tabular data exploration', 'Fitting a scikit-learn model on numerical data', 'Handling categorical data', 'Wrap-up quiz 1', and 'Main take-away'. The 'Selecting the best model' section includes links for 'Module overview', 'Overfitting and underfitting', and 'Validation and learning curves'.

Central Content Area: The main heading is 'Introduction'. Below it is the 'Course presentation' section. A 'Welcome!' message states: 'The goal of this course is to teach machine learning with scikit-learn to beginners, even without a strong technical background.' The text continues: 'Predictive modeling brings value to a vast variety of data, in business intelligence, health, industrial processes and scientific discoveries. It is a pillar of modern data science. In this field, scikit-learn is a central tool: it is easily accessible, yet powerful, and naturally dovetails in the wider ecosystem of data-science tools based on the Python programming language.' Another paragraph states: 'This course is an in-depth introduction to predictive modeling with scikit-learn. Step-by-step and didactic lessons introduce the fundamental methodological and software tools of machine learning, and is as such a stepping stone to more advanced challenges in artificial intelligence, text mining, or data science.' A final paragraph notes: 'The course is more than a cookbook: it will teach you to be critical about each step of the design of a predictive modeling pipeline: from choices in data preprocessing, to choosing models, gaining insights on their failure modes and interpreting their predictions.'

Right Sidebar: Contains a 'Contents' menu with links to 'Course presentation', 'Prerequisites', 'MOOC material', and 'Follow scikit-learn on social networks'.

Bottom Section: A green banner titled 'Follow the MOOC' contains the text: 'A new session of the "Machine learning in Python with scikit-learn MOOC", is available starting on October 18.' To the right of this banner is a small image of a laptop displaying the course content. A separate orange box on the right side of the page says: 'Join the full MOOC for better learning!'.

03 - 콘텐츠 사례 영상자료 2



PseudoLab Tutorial Book

Search this book...

PyTorch를 활용한 딥러닝 튜토리얼
(Deep Learning Tutorials with PyTorch)

객체 탐지(OBJECT DETECTION)

의료용 마스크 탐지 모델 구축

1. 객체 탐지 소개
2. 데이터 탐색
3. 데이터 전처리
4. RetinaNet
5. Faster R-CNN
6. 참고 문헌

시계열 분석(TIME SERIES)

코로나 확진자 수 예측 모델 구축

1. Time Series 소개
2. 데이터 탐색
3. 데이터 전처리

```
print('-----train start-----')
for epoch in range(num_epochs):
    start = time.time()
    model.train()
    i = 0
    epoch_loss = 0
    for imgs, annotations in data_loader:
        i += 1
        imgs = list(img.to(device) for img in imgs)
        annotations = [(k: v.to(device) for k, v in t.items()) for t in annotations]
        loss_dict = model(imgs, annotations)
        losses = sum(loss for loss in loss_dict.values())

        optimizer.zero_grad()
        losses.backward()
        optimizer.step()
        epoch_loss += losses
    print(f'epoch : {epoch+1}, Loss : {epoch_loss}, time : {time.time() - start}')
```

```
-----train start-----
epoch : 1, Loss : 77.14759063720703, time : 252.42370867729187
epoch : 2, Loss : 48.91315460205078, time : 263.22984743118286
epoch : 3, Loss : 43.18947982788086, time : 264.4591932296753
epoch : 4, Loss : 36.07373046875, time : 265.2568733692169
epoch : 5, Loss : 31.8864688873291, time : 265.57766008377075
epoch : 6, Loss : 31.76308250427246, time : 265.0076003074646
epoch : 7, Loss : 31.24744415283203, time : 265.16882514953613
epoch : 8, Loss : 29.340274810791016, time : 265.73448038101196
epoch : 9, Loss : 25.922008514404297, time : 267.91367626190186
epoch : 10, Loss : 23.59230613708496, time : 266.9004054069519
```

학습시킨 가중치를 저장하고 싶다면, `torch.save` 를 이용하여 저장해두고 나중에 언제든지 불러와 사용할 수 있습니다.

```
torch.save(model.state_dict(), f'model_{num_epochs}.pt')
```

```
model.load_state_dict(torch.load(f'model_{num_epochs}.pt'))
```

Contents

- 5.1 데이터 불러오기
- 5.2 데이터 분리
- 5.3 데이터셋 클래스 정의
- 5.4 모델 불러오기
- 5.5 전이 학습
- 5.6 예측

03 - 콘텐츠 사례 영상자료 3

Python으로 하는 인과추론

Search this book **Ctrl** + **K**

Causal Inference for The Brave and True

인과추론 Part I

- 01 - Introduction To Causality
- 02 - Randomised Experiments
- 03 - Stats Review: The Most Dangerous Equation
- 04 - Graphical Causal Models
- 05 - The Unreasonable Effectiveness of Linear Regression
- 06 - Grouped and Dummy Regression
- 07 - Beyond Confounders
- 08 - Instrumental Variables
- 09 - Non Compliance and LATE
- 10 - Matching
- 11 - Propensity Score
- 12 - Doubly Robust Estimation
- 13 - Difference-in-Differences

☰

🔍 🔄 ⬇

Causal Inference for The Brave and True

안녕하세요. 가짜연구소 빌더 신진수입니다. 가짜연구소 인과추론팀은 한국어 자료가 많지 않은 인과추론을 많은 분들이 쉽게 접하실 수 있도록 기여하고자 합니다!

책 번역 및 도입

이 책은 Matheus Fature (Nubank Data Scientist)의 [Causal Inference for The Brave and True](#)을 한국어로 번역한 자료입니다. 이 책은 인과추론에 대한 기본 개념과 Python 실습, 나아가 최신 사례까지 모두 다루고 있습니다. 인과추론에 대한 이론과 실습자료가 많지 않은 상황에서 이 책은 인과추론 처음 접하거나 인과추론에 대한 이해를 바탕으로 사례를 접하고 싶은 분 모두를 만족시킬 수 있는 책이라고 생각합니다. 한국에서 인과추론을 공부하시는 분들께 많은 도움이 되었으면 좋겠습니다.

가짜연구소 Causal Inference 팀은 전문 번역가가 아닌, 데이터분석가/과학자/연구자로 구성 되어 있습니다. 따라서, 한국어 번역 과정에서 이상하거나 어색한 부분이 발생할 수 있어요. 해당 부분에 대한 여러분의 피드백을 부탁드립니다. 또한, 한국어 표현이 익숙하지 않은 부분에 대해서는 영어로 그대로 표현했습니다. 해당 부분 이해 부탁드립니다. 여러분의 관심은 꼭 스타로 표현해주세요!

마지막으로, 이 책의 한국어 번역에 동의해 주신 Matheus Fature에게 감사의 말씀을 전합니다. 그리고, 번역과 실습자료를 만들기 위해 같이 고생해준 Causal Inference 팀 멤버들에게도 진심으로 감사드립니다!

이 책은 Joshua Angrist, Alberto Abadie, Christopher Walters, Miguel Hernan 및 Jamie Robins 등을 포함한 많은 연구자들의 계량 경제학 수업과 바탕으로 하고 있으며, 대부분의 아이디어는 전미경제학회(American

☰ On this page

- 책 번역 및 도입
- 자료 이슈 및 관련 질문

03 - 콘텐츠 사례

강의 교보재

The screenshot displays the scikit-learn MOOC interface. On the left is a sidebar with navigation links: Introduction, Machine Learning Concepts (with a sub-link for Quiz Intro.01), The predictive modeling pipeline, and Selecting the best model. The main content area is titled 'Exercise M4.01' and describes the goal of understanding linear model parametrization and quantifying fitting accuracy. It lists prerequisites like 'Data loading' and 'Model definition'. A code block shows a pandas script for loading the 'penguins' dataset and defining a linear model function.

<https://inria.github.io/scikit-learn-mooc/>

BACK TO US,
BACK TO PYTHON

PYCON KOREA 2023

학습 내용 공유

The screenshot shows the PseudoLab Tutorial Book interface. It features a search bar and a list of topics including Object Detection, Time Series, and Generative Modeling. The 'Generative Modeling' section is highlighted, showing a diagram of a GAN architecture. The diagram illustrates the flow from a 'D-dimensional noise vector' through a 'Generator Network' to produce 'Fake Images', which are then evaluated by a 'Discriminator Network' against 'Real Images' to produce 'Predicted Labels'. Below the diagram, there is a list of references and a mathematical formula for the GAN Value Function.

1.2. GAN 모델

1.2.1. 모델 구조

GAN의 아키텍처를 도식화하면 다음과 같습니다.

D-dimensional noise vector → Generator Network → Fake Images → Discriminator Network → Predicted Labels

• 그림 1-2 GAN의 아키텍처 (출처: Hamed Alqahtani, 2019. An Analysis Of Evaluation Metric Of GANs)

생성기(G)는 실제 데이터와 비슷한 데이터를 만들어내도록 학습되며, 판별기(D)는 실제 데이터와 G가 생성한 가짜 데이터를 구별하도록 학습됩니다. GAN의 목적함수는 다음과 같은데, 이는 게임 이론 타입의 목적함수로 G와 D 2명의 플레이어의 싸움 면서 서로 균형점(nash equilibrium)을 찾아가도록 하는 방식입니다.

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{z} \sim p_{\text{noise}}(\mathbf{z})} [\log D(\mathbf{z})] + \mathbb{E}_{\mathbf{z} \sim p_G(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))]$$

• 그림 1-3 GAN Value Function (출처: Ian J Goodfellow, 2014. Generative Adversarial Nets)

<https://pseudo-lab.github.io/Tutorial-Book/>

03 - 콘텐츠 사례

패키지 사용법

The screenshot shows the InterpretML API documentation for the `LinearRegression` class. The left sidebar contains a navigation menu with categories like 'Quick Overview', 'Algorithms', 'Glassbox Models', 'More EBMs', 'Framework', and 'Deployment Guide'. The main content area displays the class signature, a description of the wrapper, initialization details, and parameters for `explain_global` and `explain`.

```
class interpret.glassbox.LinearRegression(feature_names=None,
feature_types=None, linear_class=<class
'sklearn.linear_model._base.LinearRegression'>, **kwargs)
Linear regression.

Currently wrapper around linear models in scikit-learn: scikit-learn/scikit-learn

Initializes class.

Parameters::
  • feature_names – List of feature names.
  • feature_types – List of feature types.
  • linear_class – A scikit-learn linear class.
  • **kwargs – Kwargs pass to linear class at initialization time.

explain_global(name=None)
Provides global explanation for model.

Parameters::
  name – User-defined explanation name.

Returns::
  An explanation object, visualizing feature-value pairs as horizontal bar
  chart.
```

<https://interpret.ml/docs/intro.html>

BACK TO US,
BACK TO PYTHON

PYCON KOREA 2023

연구 내용 공유

The screenshot shows the 'The Turing Way' project illustration. It features a stylized tree-like structure with a person walking on a path. The path is labeled 'WELCOME TO THE TURING WAY'. The tree branches out into 'CHAPTERS' and 'GUIDE' sections. The 'GUIDE' section includes 'GUIDE I: REPRODUCIBLE RESEARCH!', 'GUIDE II', and 'GUIDE III'. The 'CHAPTERS' section includes 'X', 'Y', and 'Z'. The illustration is signed 'Scriberia'.

Fig. 2 The Turing Way project illustration by Scriberia. Original version on Zenodo.
<http://doi.org/10.5281/zenodo.3695300>.



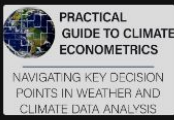

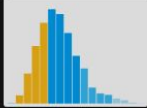

Although *The Turing Way* receives support and funding from *The Alan Turing Institute*, the project is designed to be a global collaboration. We have contributions from across the UK, and from India, Mexico, Australia, USA, and many European countries. Chapters have been written, reviewed and curated by members of research institutes and universities, government departments, and industry. We are committed to creating a space where people with diverse expertise and lived experiences can share their knowledge with others to allow us all to use data science to improve the world.







<https://the-turing-way.netlify.app/>

03 - 콘텐츠 사례 - Gallery of Jupyter Books

Gallery of Jupyter Books

This is a gallery of Jupyter Books built from across the community. If you'd like to add your book to this list, simply [add an entry to this gallery.yml file](#) and open a Pull Request to add it.

<p>PyTorch를 활용한 딥러닝 튜토리얼 (Deep Learning Tutorials with PyTorch)</p>  <p>website repo Stars: 35</p>	<p>Computational Mechanics</p>  <p>website repo Stars: 24</p>	<p>Practical Guide to Climate Econometrics</p>  <p>website repo Stars: 24</p>
<p>Argo Online School</p>  <p>website repo Stars: 12</p>	<p>UW Data Visualization Curriculum</p>  <p>website repo Stars: 126</p>	<p>Quantitative Economics with Python</p>  <p>website repo Stars: 66</p>

<p>The Data Science Interview Book</p>  <p>website repo Stars: 26</p>	<p>Classifying 19th Century British Library books using Crowdsourcing and Machine Learning</p>  <p>website repo Stars: 10</p>	<p>단국대 2020 캐글 보내기 (DKU Kaggle Class)</p>  <p>website repo Stars: 40</p>
<p>Open Source in Environmental Sustainability</p>  <p>website repo Stars: 37</p>	<p>Python Programming for Economics and Finance</p>  <p>website repo Stars: 25</p>	<p>The Data-Sitters Club</p>  <p>website repo Stars: 12</p>

04 - Jupyter Book 만들기 - 패키지 설치하기

```
pip install -U jupyter-book
```

```
jupyter-book --help
```

Usage: jupyter-book [OPTIONS] COMMAND [ARGS]...

Build and manage books with Jupyter.

Options:

--version Show the version and exit.
-h, --help Show this message and exit.

Commands:

build Convert your book's or page's content to HTML or a PDF.
clean Empty the _build directory except jupyter_cache.
config Inspect your _config.yml file.
create Create a Jupyter Book template that you can customize.
myst Manipulate MyST markdown files.
toc Command-line for sphinx-external-toc.

05 - Jupyter Book 만들기 - 소스 파일 생성

```
mybookname/  
├── _config.yml  
├── _toc.yml  
├── intro.md  
├── logo.png  
├── markdown-notebooks.md  
├── markdown.md  
├── notebooks.ipynb  
├── references.bib  
└── requirements.txt
```

- 콘텐츠
 - .md 파일
 - .ipynb 파일
- 콘텐츠 목차 정의
 - _toc.yml
- 환경 설정
 - _config.yml

05 - Jupyter Book 만들기 - 소스 파일 생성 - `_toc.yml` 파일

```
format: jb-book
root: index
chapters:
- file: path/to/chapter1
- file: path/to/chapter2
```

- **format** - 해당 `_toc.yml` 파일이 어떤 구조를 가지게 되는지 명시 (**jb-book**, **jb-article**)
- **root** - 랜딩 페이지로 사용할 파일 경로
- **chapters** - 단원 구분자
- **file** - 단원 내에 들어갈 콘텐츠가 담긴 파일, **root**에서 부터 상대적인 경로 사용하여 명시
 - **root** 및 **file**은 확장자명 사용 안함

05 - Jupyter Book 만들기 - 소스 파일 생성 - `_config.yml` 파일

```
# In _config.yml
title: My sample book
author: The Jupyter Book Community
logo: logo.png
execute:
  execute_notebooks: force
```

- **title:** 전자책의 제목
- **author:** 전자책의 저자
- **logo:** 로고 이미지 파일로 향하는 상대경로
- **execute:** 코드 실행 여부 관련 설정

06 - Jupyter Book 만들기 - html파일 빌드

```
jupyter-book build mybookname/
```

```
mybookname
├── _build
│   └── html
│       ├── _images
│       ├── _static
│       ├── index.html
│       ├── intro.html
│       └── ...
```

- **build** 명령어를 통해 콘텐츠 파일들을 **html**으로 변환
- 결과물은 **_build** 폴더에 위치

06 - Jupyter Book 만들기 - html파일 빌드

```
08/11/2023 04:02 AM <DIR> .
08/11/2023 04:02 AM <DIR> ..
08/11/2023 03:51 AM      263 intro.md
08/11/2023 03:50 AM    21,697 logo.png
03/12/2023 12:09 AM     1,787 markdown-notebooks.md
03/12/2023 12:09 AM     3,378 notebooks.ipynb
03/12/2023 12:09 AM         30 requirements.txt
08/11/2023 04:02 AM      929 _config.yml
08/11/2023 03:55 AM      163 _toc.yml
      7 File(s)      28,247 bytes
      2 Dir(s) 13,711,011,840 bytes free

(jb-dev-15) C:\Users\WAhnSungJin\Github\sample-book>
```

07 - Jupyter Book 만들기 - 결과물 공유



```
ghp-import -n -p -f _build/html
```

- 변환된 **html** 파일들을 **GitHub Pages** 서비스를 활용해 무료로 웹 상에 배포
 - **GitHub** 레포 필요
- **ghp-import** 패키지를 활용해 명령어 한 줄만으로 손쉽게 웹페이지 생성
- 추후 **GitHub Actions**와 연동하여 자동화 가능

07 - Jupyter Book 만들기 - 결과물 공유

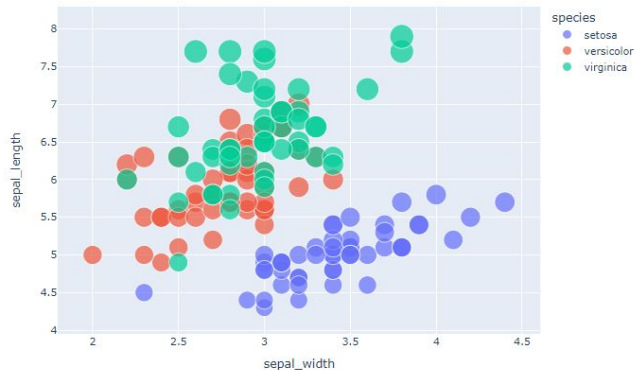
```
08/11/2023 04:23 AM <DIR> .
08/11/2023 04:23 AM <DIR> ..
08/11/2023 04:22 AM      3,247 .gitignore
08/11/2023 03:51 AM      263 intro.md
08/11/2023 04:22 AM     1,096 LICENSE
08/11/2023 03:50 AM    21,697 logo.png
03/12/2023 12:09 AM     1,787 markdown-notebooks.md
03/12/2023 12:09 AM     3,378 notebooks.ipynb
03/12/2023 12:09 AM        30 requirements.txt
08/11/2023 04:23 AM <DIR> _build
08/11/2023 04:02 AM      929 _config.yml
08/11/2023 03:55 AM      163 _toc.yml
      9 File(s)      32,590 bytes
      3 Dir(s)  15,281,266,688 bytes free

(jb-dev-15) C:\Users\WAhnSungJin\Github\PyConKR2023_session>
```

08 - 기타 기능

```
import plotly.io as pio
import plotly.express as px
import plotly.offline as py

df = px.data.iris()
fig = px.scatter(df, x="sepal_width", y="sepal_length", color="species", size="sepal_length")
fig
```



Contents ▾

Plotly

Plotly is another interactive plotting library that provides a high-level API for visualization. See the [Plotly JupyterLab documentation](#) to get started with Plotly in the notebook.

Below is some example output.

Important

For these plots to show, it may be necessary to load `require.js`, in your `_config.yml`:

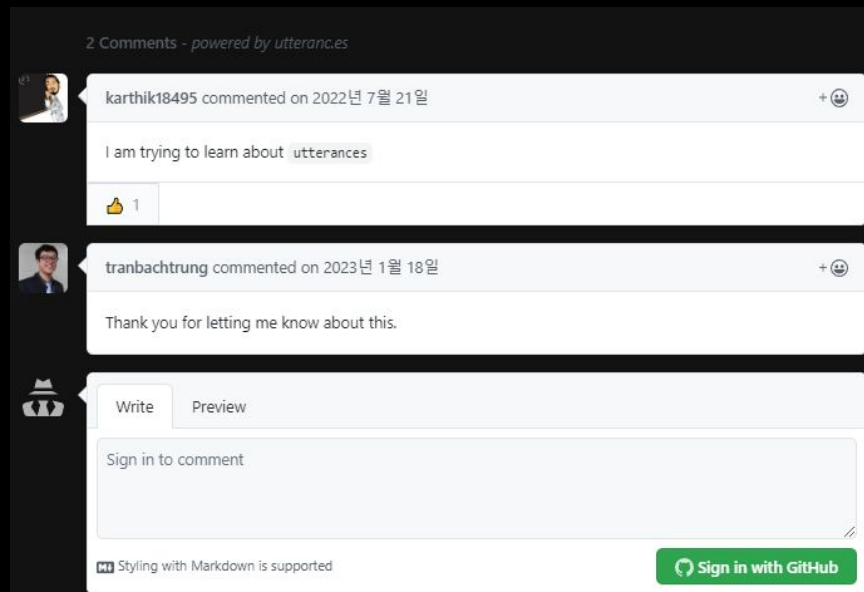
```
sphinx:
  config:
    html_js_files:
      - https://cdnjs.cloudflare.com/ajax/libs/require.js/2.3.4/require.min.js
```

Important

Including plotly plots in a Jupyter Book page is *currently not compatible* with the `dollarmath` syntax extension (mathematical notation written between two `"$"` characters). If you are trying to include both plotly plots and mathematical notation within the same page, and finding that plotly plots are not being rendered, this may be the cause. Try removing all use of the `dollarmath` syntax within the page and rebuilding the book.

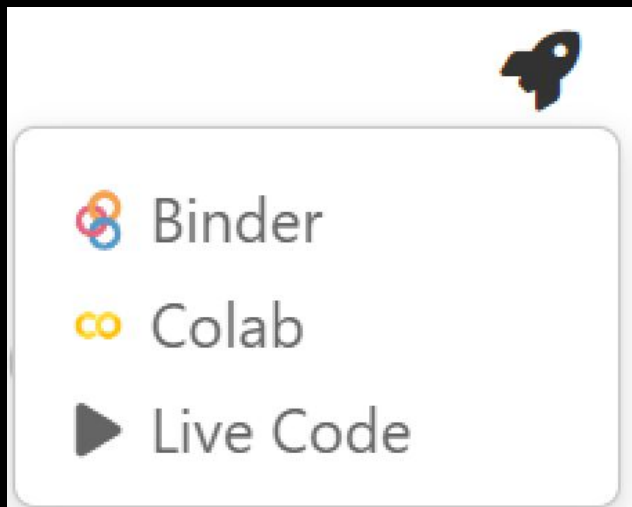
Plotly uses `renderers` kinds of information plot. Experiment with the output you want

08 - 기타 기능



```
html:
  comments:
    utterances:
      repo: "github-org/github-repo"
```

08 - 기타 기능



⚠ The latest release refactored our HTML, so double-check your custom CSS rules! ⚠

☰ 🔍 ⚙️ 📄 🔄 📄 🔍

Make your code cells executable

❌❌❌❌ Launching from mybinder.org: failed

This section describes how to bring interactivity to your book. This lets users run code and see outputs *without leaving the page*. Interactivity is provided by a kernel running on the public [MyBinder](#) service.

For an example, click the 🚀 -> [Live Code](#) button above on this page, and run the code below.


```
import numpy as np
import matplotlib.pyplot as plt
plt.ion()

x = np.arange(500)
y = np.random.randn(500)

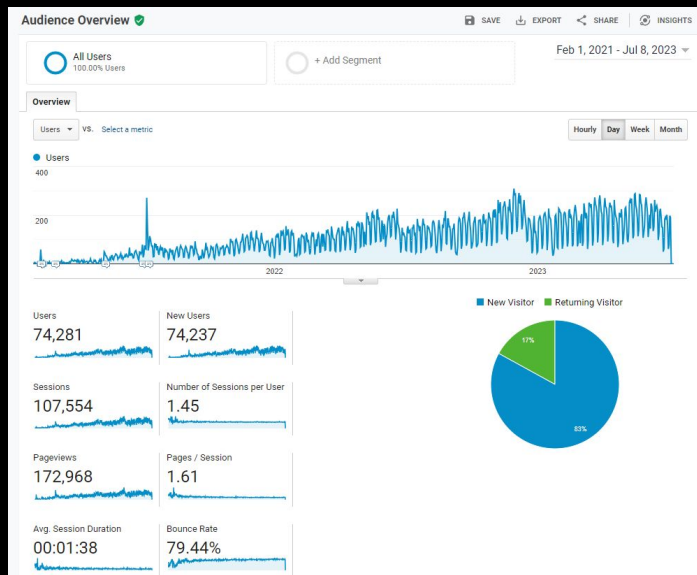
fig, ax = plt.subplots()
ax.scatter(x, y, c=y, s=x)
```

run restart restart & run all

<matplotlib.collections.PathCollection at 0x7ffa354bd640>



08 - 기타 기능



```
html:  
google_analytics_id: G-XXXXXXX
```


09 - Summary

- Jupyter Book을 활용하면 데이터 분석가/과학자 및 교육자 분들께서 손쉽게 연구 결과물을 접근성이 높은 형태로 변환하여 온라인 상에 공유가 가능
- 콘텐츠 변환 및 배포는 Jupyter Book으로 손쉽게 가능하기 때문에, 실무자는 콘텐츠 내용 자체에 집중 가능
- 나도 멋진 콘텐츠 만들어서 국내 오픈 소스 커뮤니티에 기여해보자!

Q&A

**BACK TO US,
BACK TO PYTHON**

감사합니다

**BACK TO US,
BACK TO PYTHON**