비정형 빅데이터 분석의 응용과 실습

Week-05. Practice and Assignment

서중원 2020.10.17

1. EasyOCR 활용하기

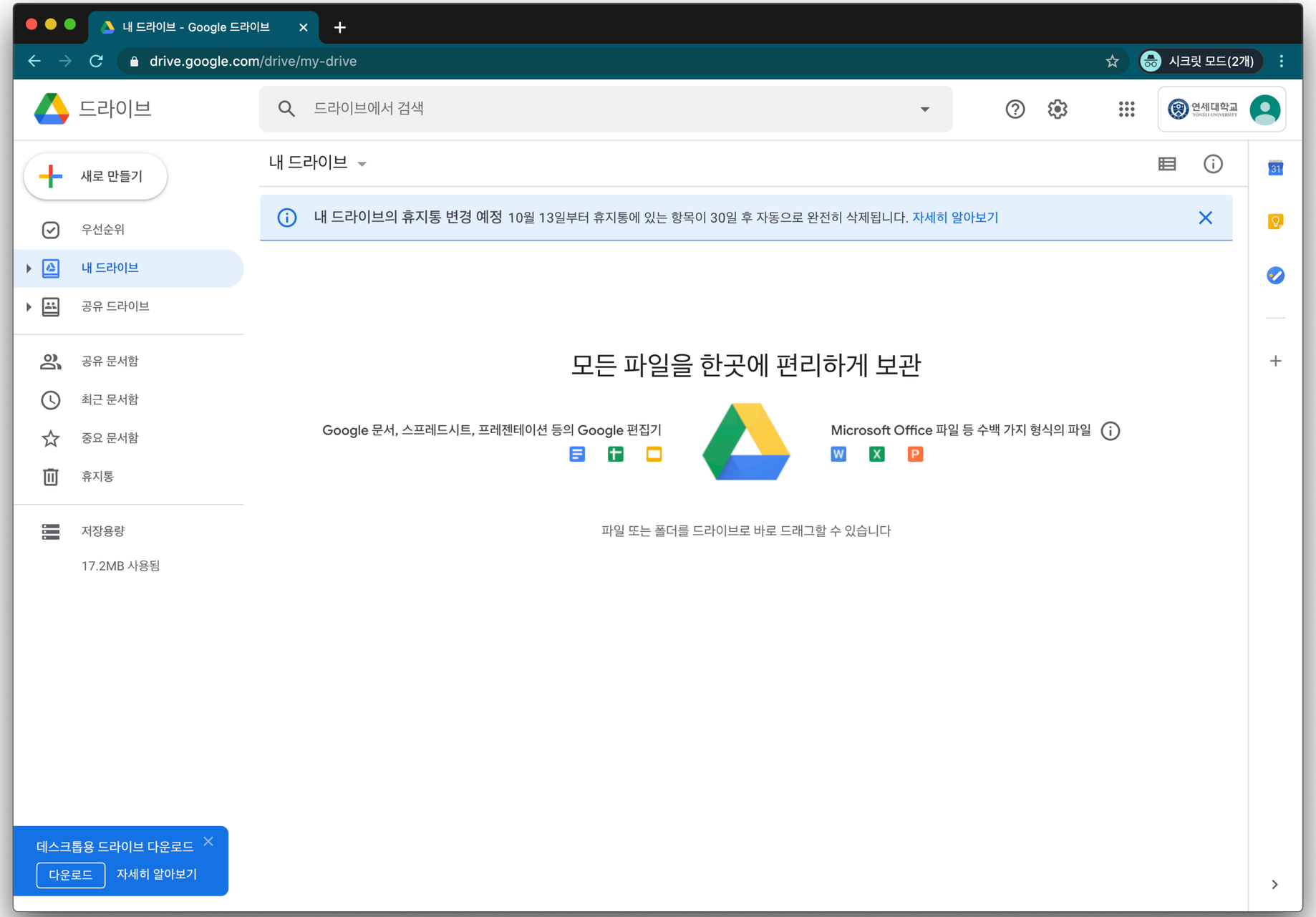
Option1. Pytorch 재설치

!{sys.executable} -m pip install torch==1.6.0+cpu torchvision==0.7.0+cpu -f https://download.pytorch.org/whl/torch_stable.html

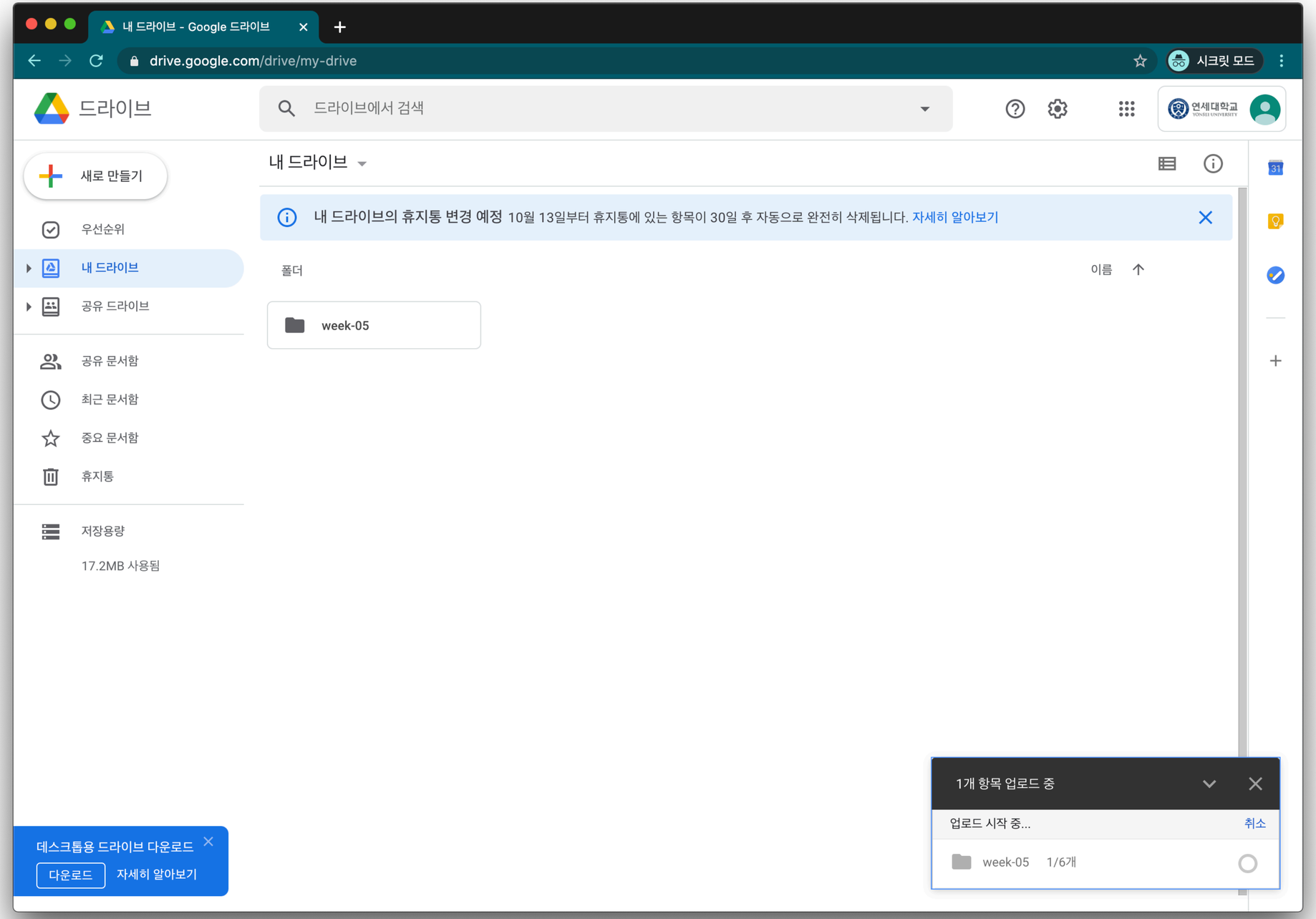
```
import sys
!{sys.executable} -m pip install torch==1.6.0+cpu torchvision==0.7.0+cpu -f https://download.py
!{sys.executable} -m pip install easyocr
```

Option2. Google Colab 활용

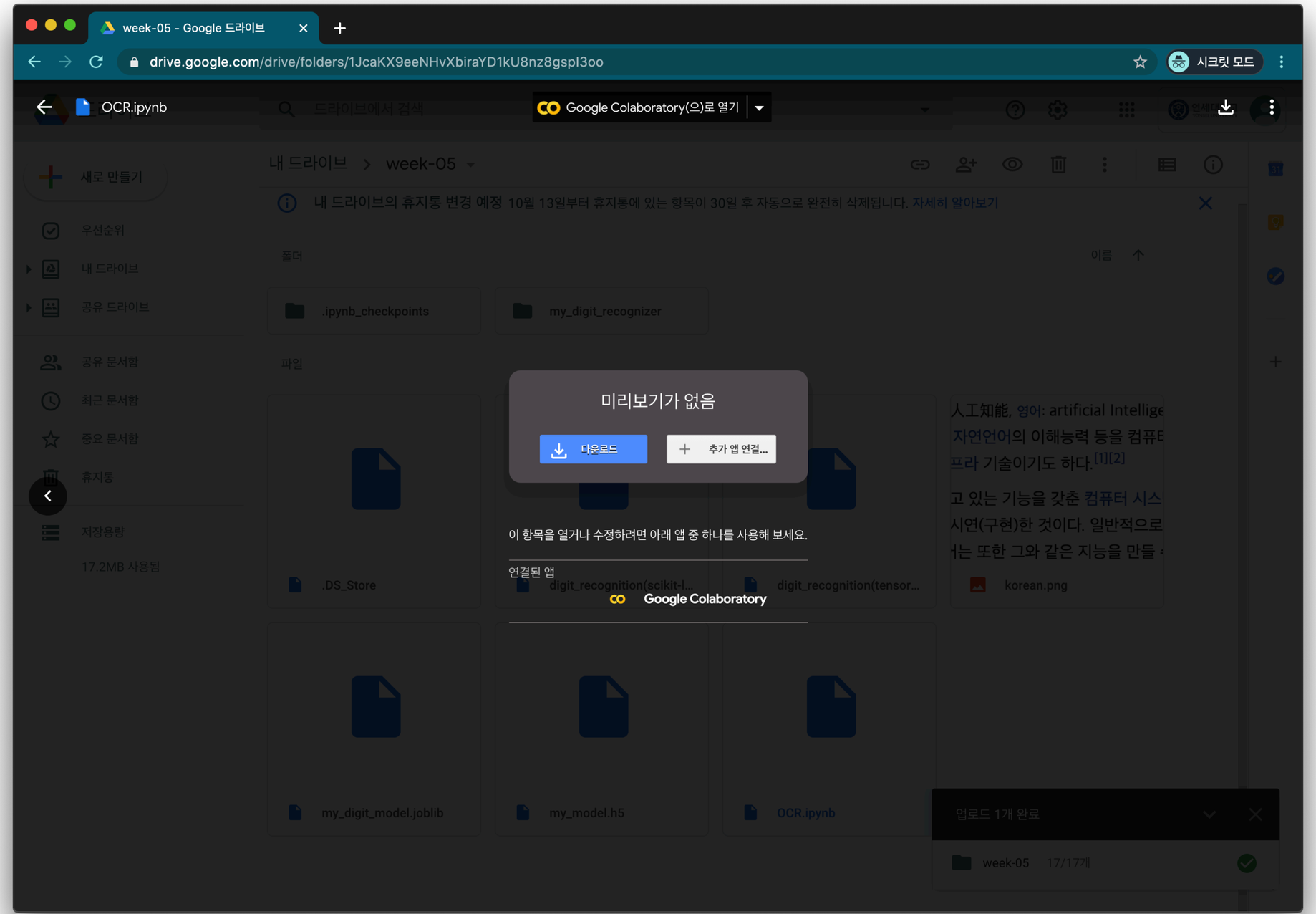
https://drive.google.com/drive/my-drive 접속



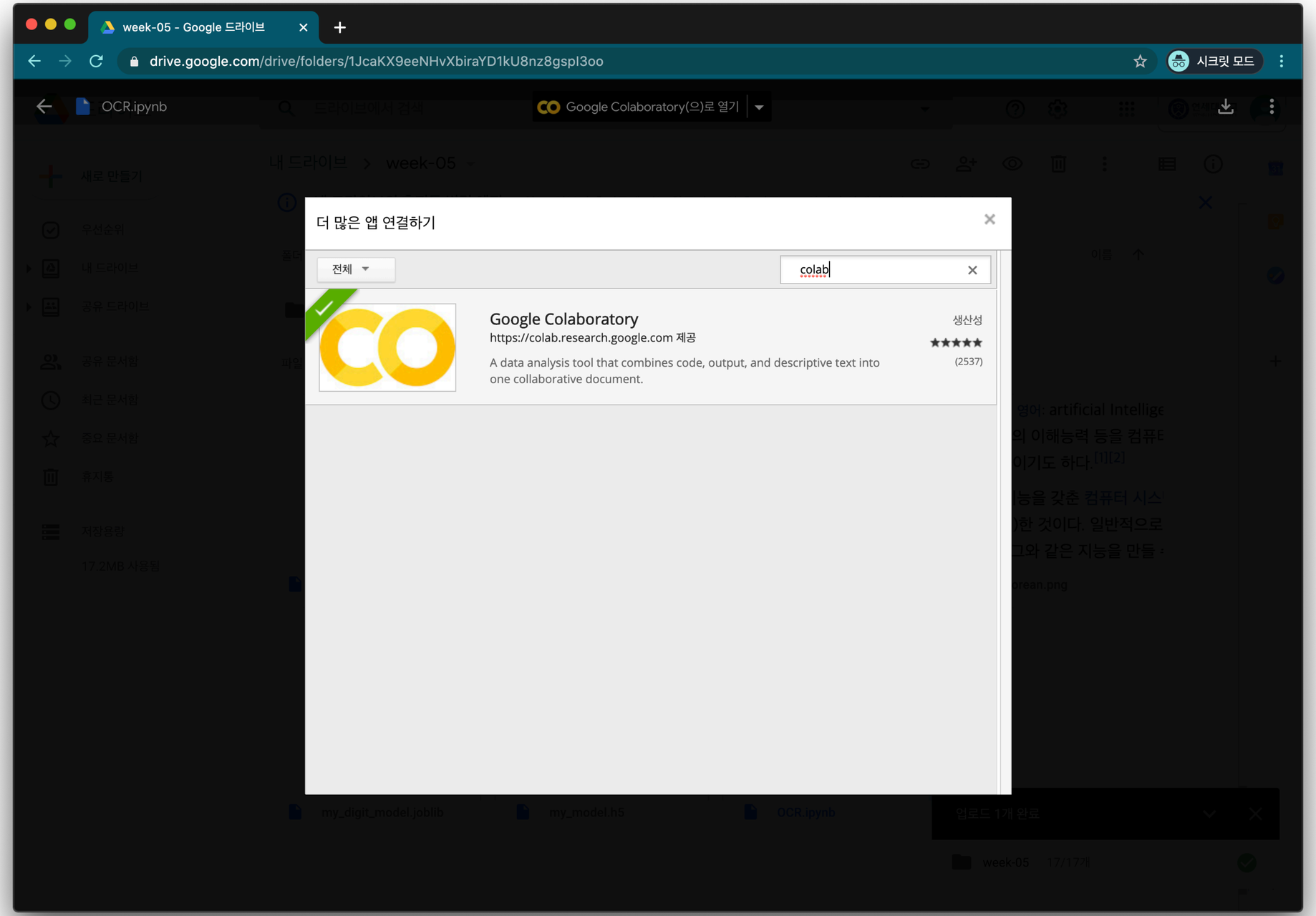
Practice의 week-05 업로드!



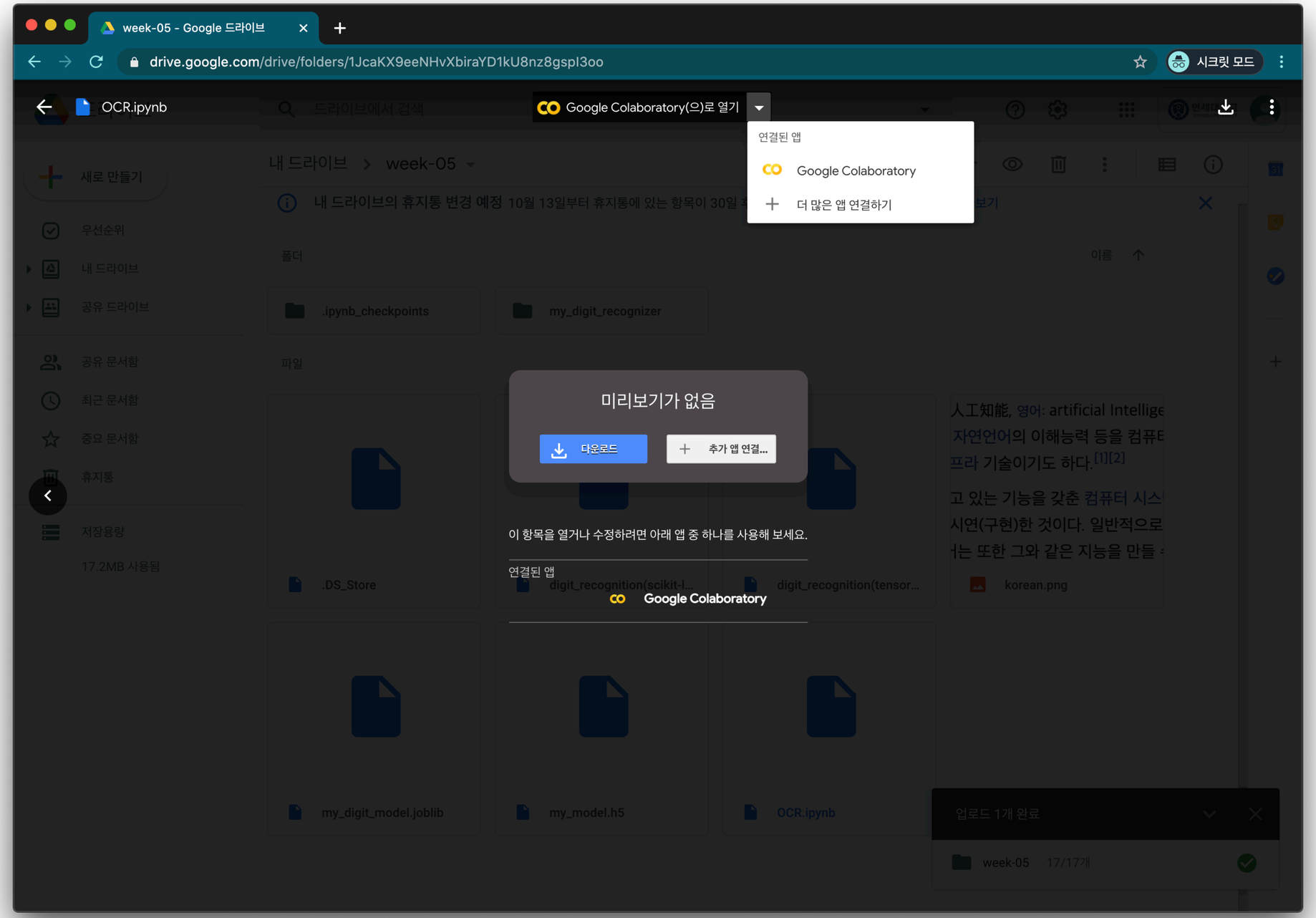
OCR.ipynb 클릭후 추가 앱 연결 클릭!



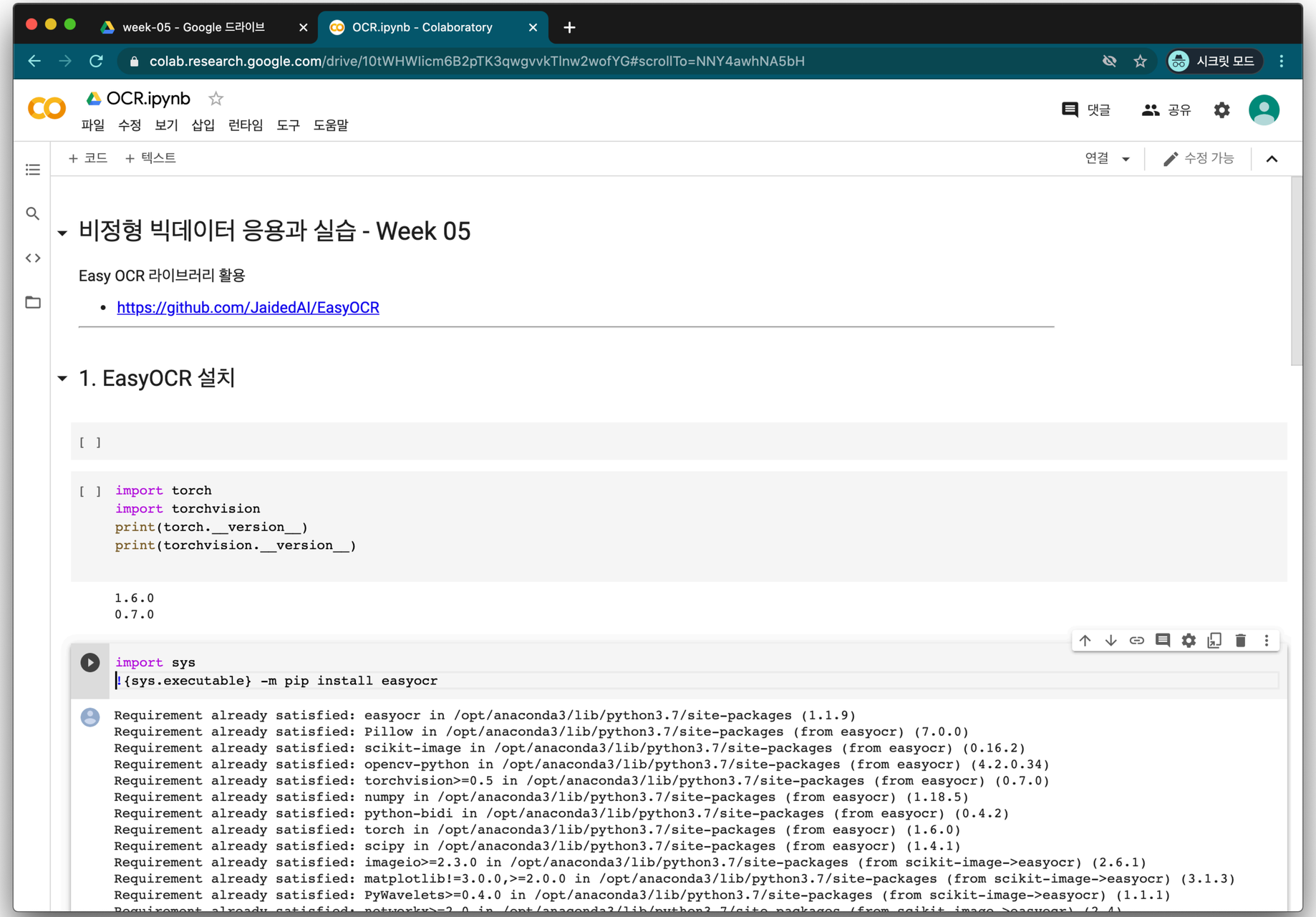
Colab 검색후 연결/설치!



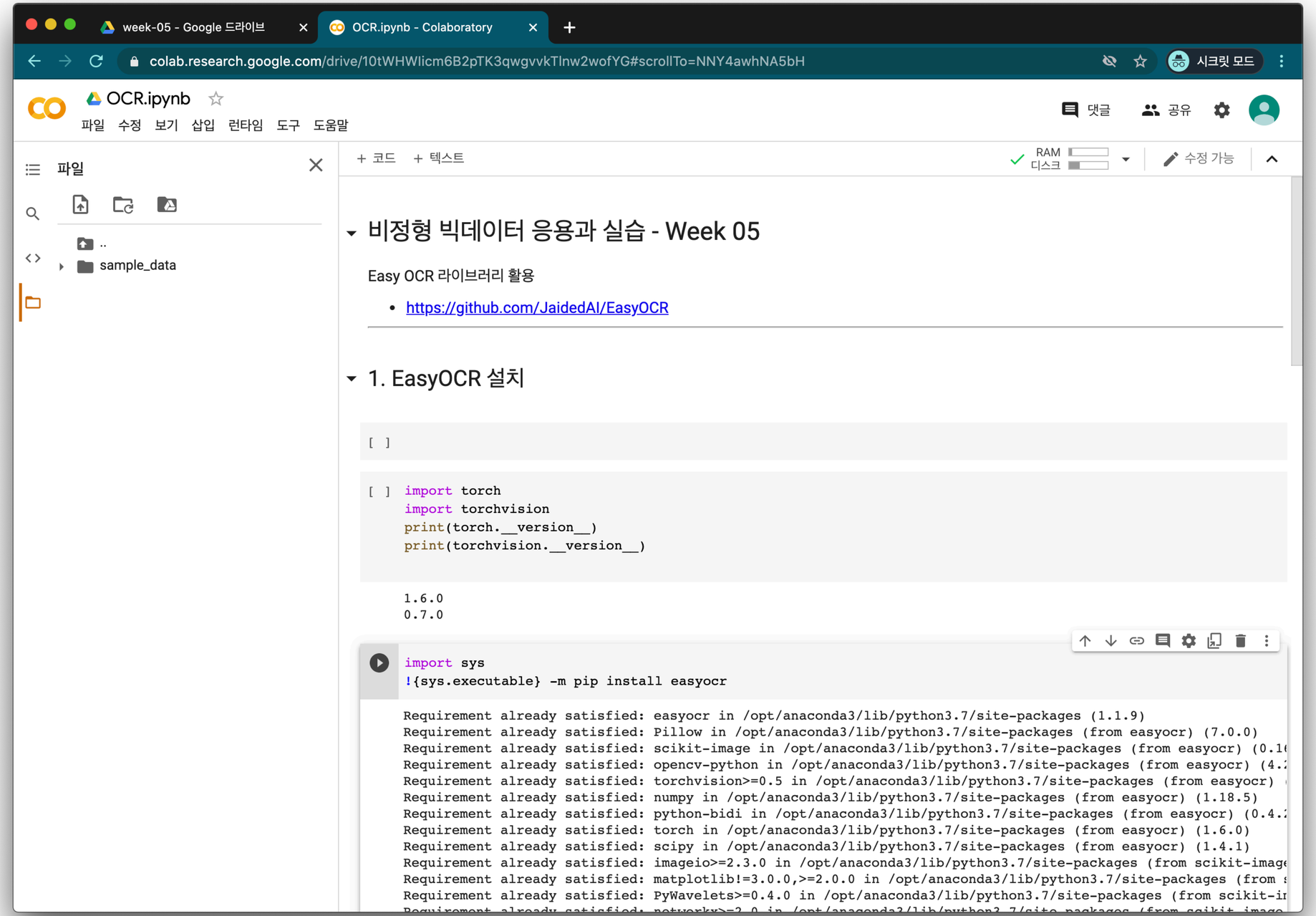
다시 OCR.ipynb 쪽으로 와서, Google Colaboratory로 열기!



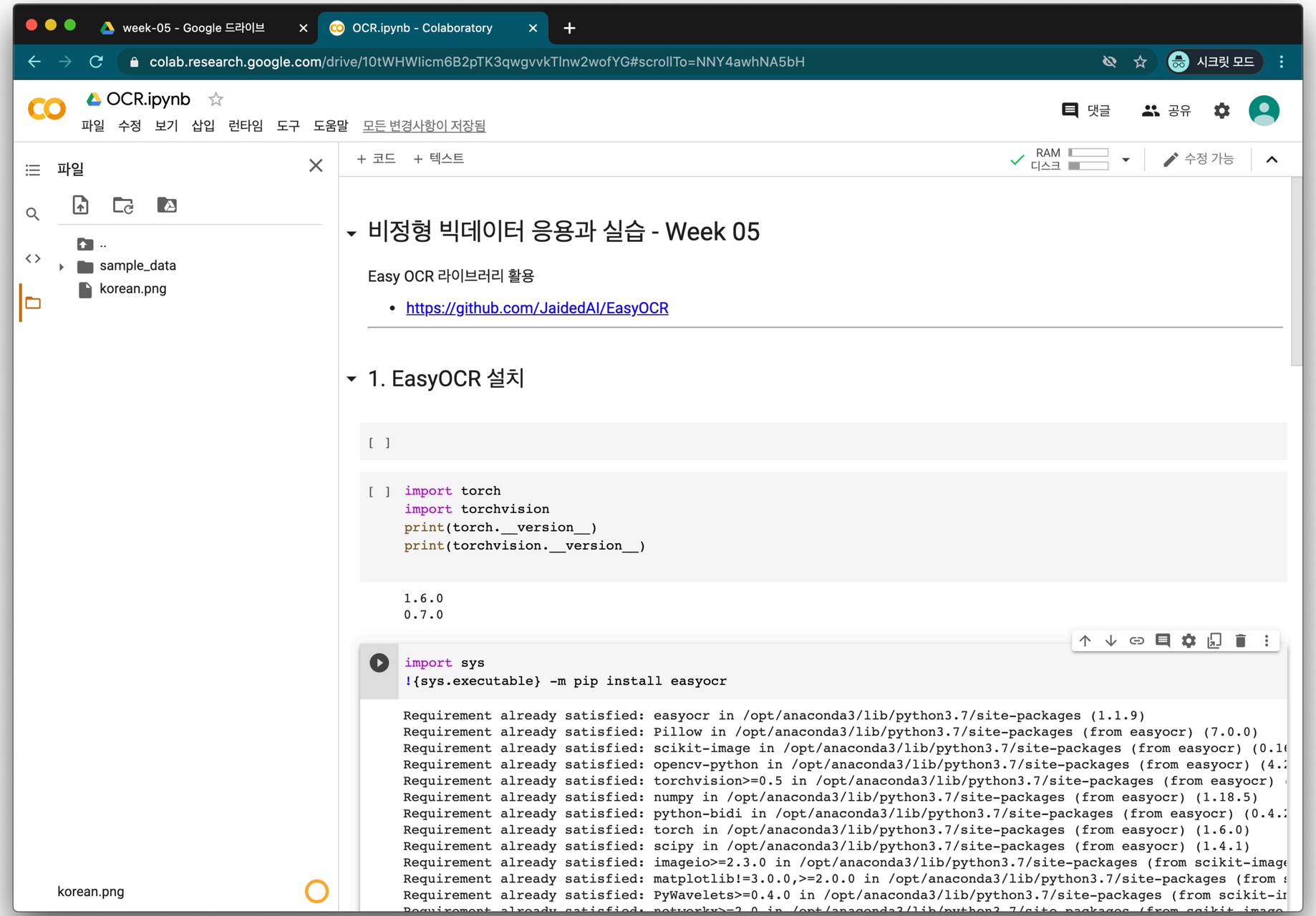
Google Colaboratory로 열린상태, 연결 클릭!



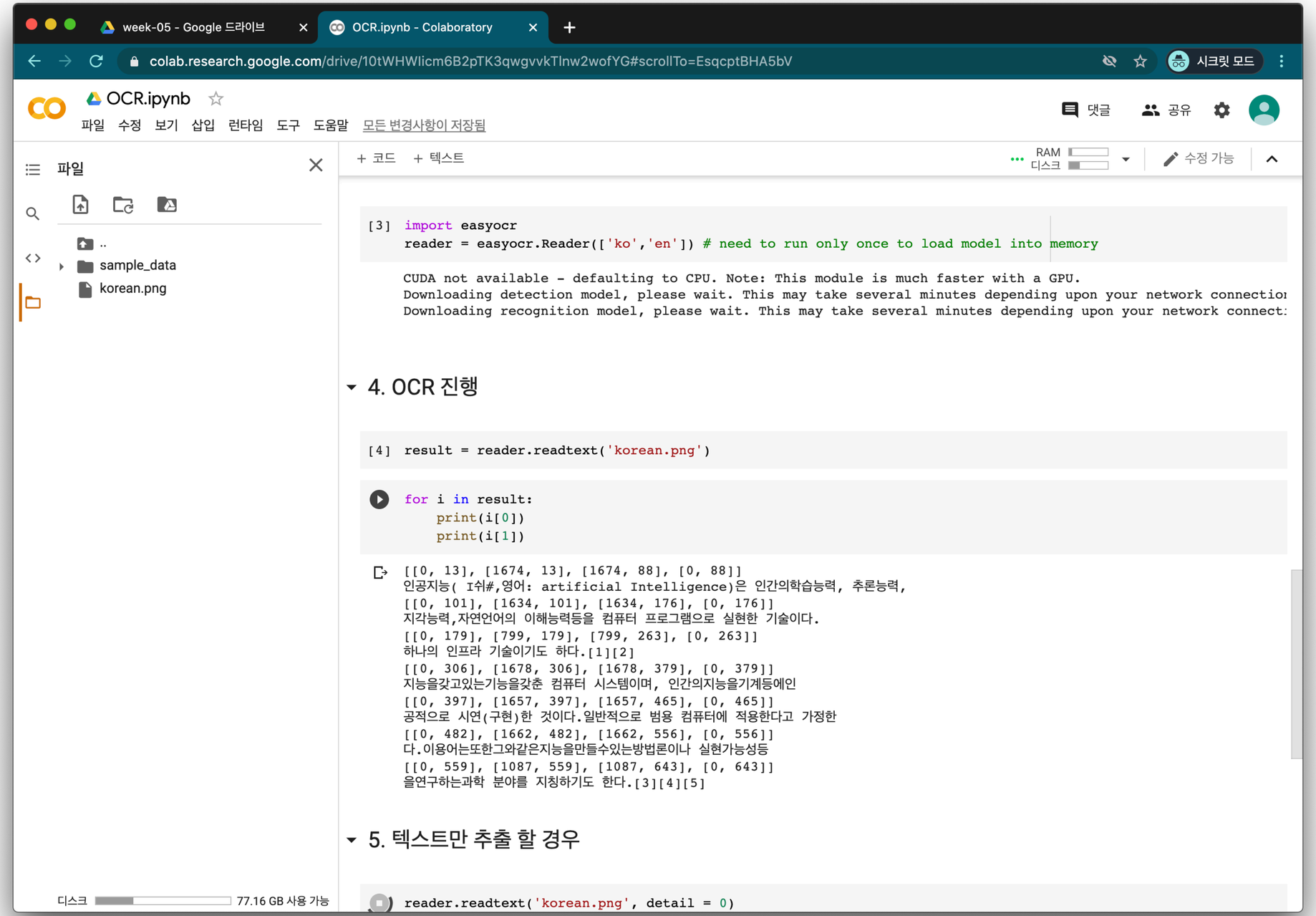
좌측에 폴더 클릭후, local에 있는 korean.png 업로드!



전체 실행!


week-05 - Google 드라이브
OCR.ipynb - Colaboratory
colab.research.google.com/drive/10tWHWIicm6B2pTK3qwgvvkTlnw2wofYG#scrollTo=NNY4awhNA5bH
시크릿 모드
OCR.ipynb
파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨
댓글
공유
파일
+ 코드 + 텍스트
RAM
디스크
수정 가능
..
sample_data
korean.png
비정형 빅데이터 응용과 실습 - Week 05
Easy OCR 라이브러리 활용
https://github.com/JaidedAI/EasyOCR
1. EasyOCR 설치
[ ]
[ ] import torch
import torchvision
print(torch.__version__)
print(torchvision.__version__)
1.6.0
0.7.0
import sys
!{sys.executable} -m pip install easyocr
Requirement already satisfied: easyocr in /opt/anaconda3/lib/python3.7/site-packages (1.1.9)
Requirement already satisfied: Pillow in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (7.0.0)
Requirement already satisfied: scikit-image in /opt/anaconda3/lib/python3.7/site-packages (from easyocr)
Requirement already satisfied: opencv-python in /opt/anaconda3/lib/python3.7/site-packages (from easyocr)
Requirement already satisfied: torchvision>=0.5 in /opt/anaconda3/lib/python3.7/site-packages (from easyocr)
Requirement already satisfied: numpy in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.18.5)
Requirement already satisfied: python-bidi in /opt/anaconda3/lib/python3.7/site-packages (from easyocr)
Requirement already satisfied: torch in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.6.0)
Requirement already satisfied: scipy in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.4.1)
Requirement already satisfied: imageio>=2.3.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image
Requirement already satisfied: matplotlib!=3.0.0,>=2.0.0 in /opt/anaconda3/lib/python3.7/site-packages (from
Requirement already satisfied: PyWavelets>=0.4.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-im
korean.png

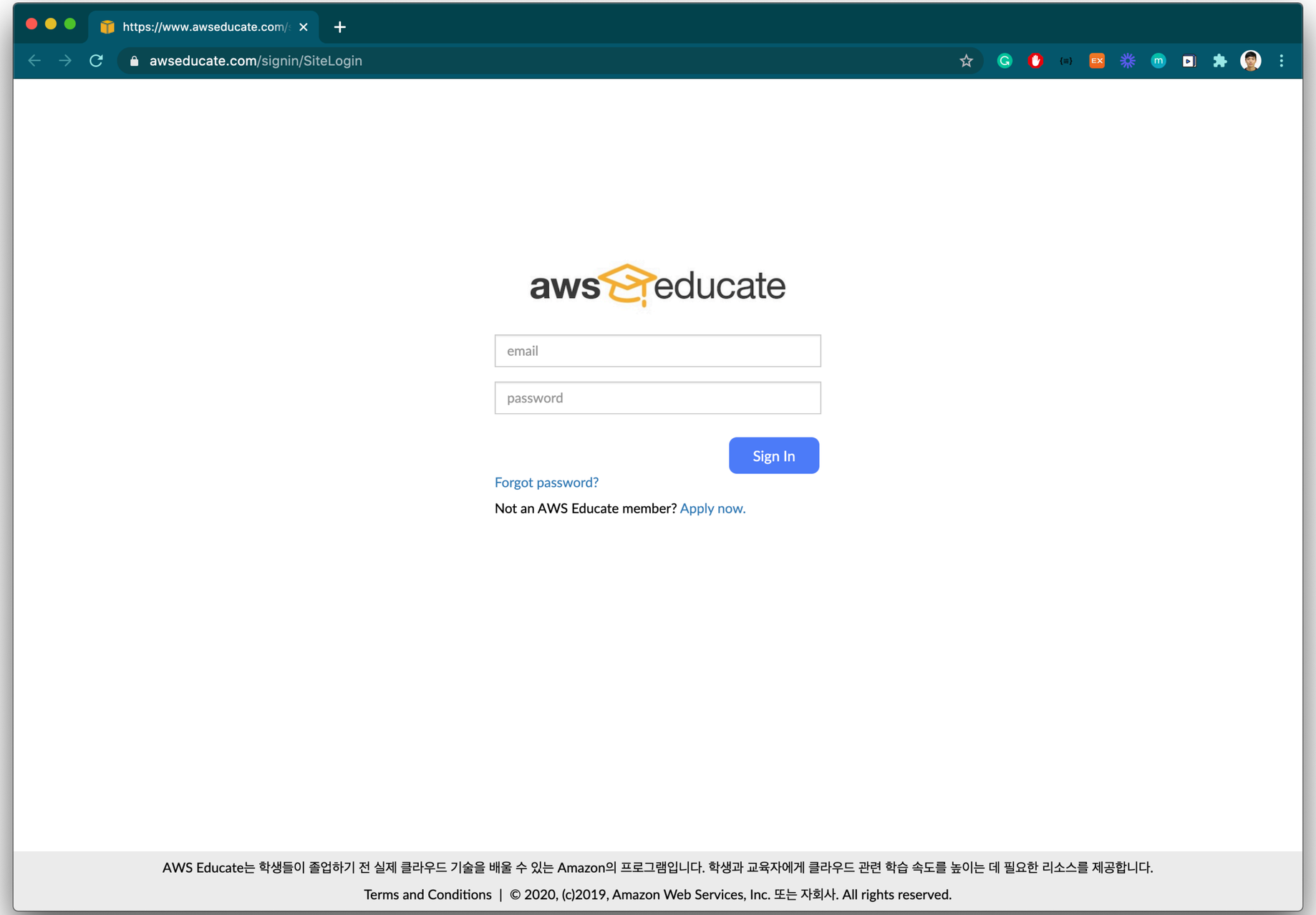
성공!


week-05 - Google 드라이브
OCR.ipynb - Colaboratory
colab.research.google.com/drive/10tWHWIicm6B2pTK3qwgvvkTlnw2wofYG#scrollTo=EsqcptBHA5bV
시크릿 모드
OCR.ipynb
파일 수정 보기 삽입 런타임 도구 도움말 모든 변경사항이 저장됨
댓글
공유
파일
..
sample_data
korean.png
+ 코드 + 텍스트
RAM
디스크
수정 가능
[3] import easyocr
reader = easyocr.Reader(['ko','en']) # need to run only once to load model into memory
CUDA not available - defaulting to CPU. Note: This module is much faster with a GPU.
Downloading detection model, please wait. This may take several minutes depending upon your network connectio
Downloading recognition model, please wait. This may take several minutes depending upon your network connect
4. OCR 진행
[4] result = reader.readtext('korean.png')
for i in result:
print(i[0])
print(i[1])
[[0, 13], [1674, 13], [1674, 88], [0, 88]]
인공지능(I쉬#,영어: artificial Intelligence)은 인간의학습능력, 추론능력,
[[0, 101], [1634, 101], [1634, 176], [0, 176]]
지각능력,자연언어의 이해능력등을 컴퓨터 프로그램으로 실현한 기술이다.
[[0, 179], [799, 179], [799, 263], [0, 263]]
하나의 인프라 기술이기도 하다.[1][2]
[[0, 306], [1678, 306], [1678, 379], [0, 379]]
지능을갖고있는기능을갖춘 컴퓨터 시스템이며, 인간의지능을기계등에인
[[0, 397], [1657, 397], [1657, 465], [0, 465]]
공적으로 시연(구현)한 것이다.일반적으로 범용 컴퓨터에 적용한다고 가정한
[[0, 482], [1662, 482], [1662, 556], [0, 556]]
다.이용어는또한그와같은지능을만들수있는방법론이나 실현가능성등
[[0, 559], [1087, 559], [1087, 643], [0, 643]]
을연구하는과학 분야를 지칭하기도 한다.[3][4][5]
5. 텍스트만 추출 할 경우
reader.readtext('korean.png', detail = 0)
디스크
77.16 GB 사용 가능

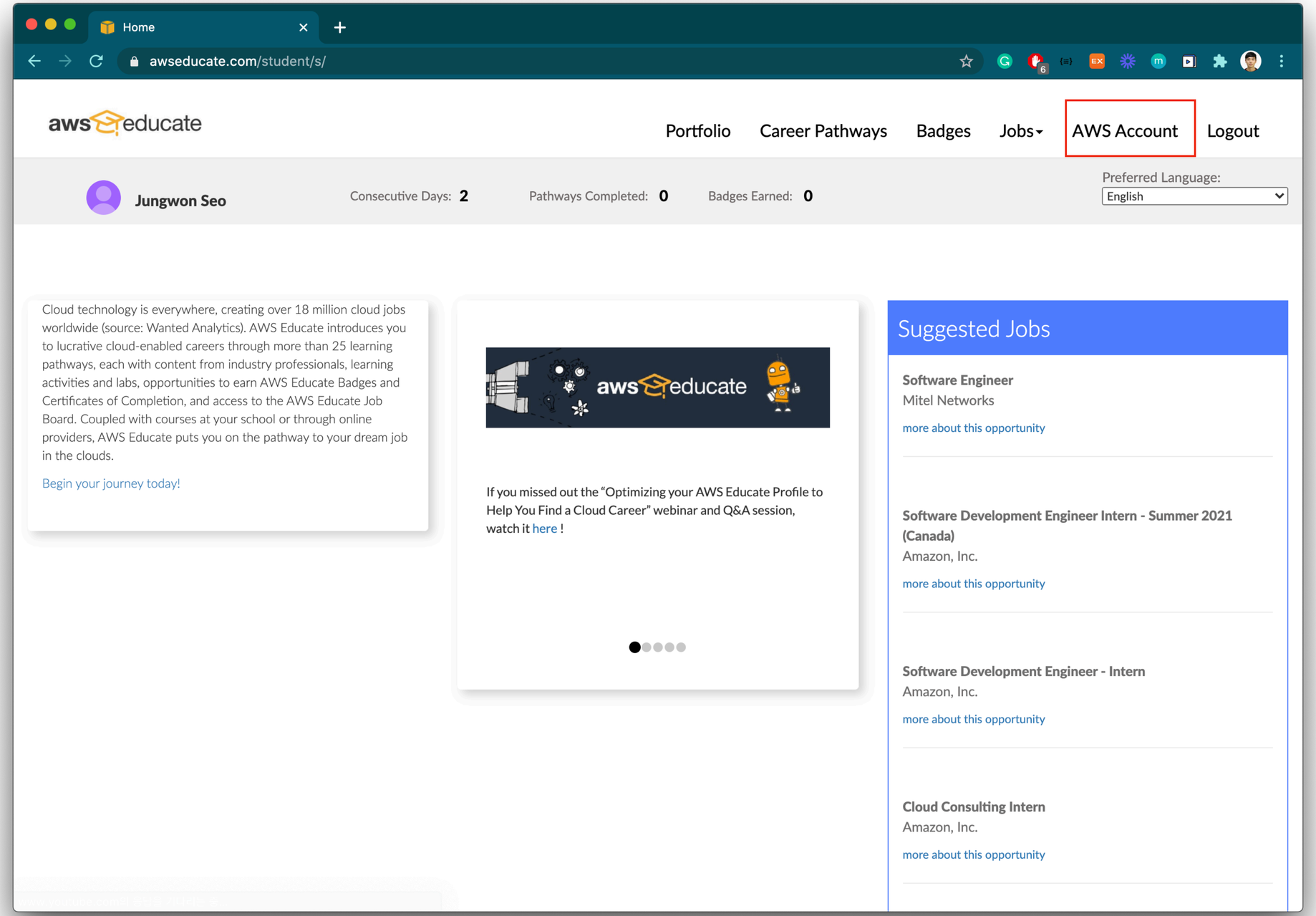
2. Educate 프로그램으로 AWS 인스턴스 생성

- EasyOCR이 요구하는 RAM이, 현재 저희가 사용하는 t2.micro 보다 높음
  - 더 좋은 성능의 인스턴스를 사용할 것입니다.
- 하지만 더 좋은 인스턴스는 무료 플랜에 들어있지 않음
  - AWS Educate 계정을 사용하여 무료로 사용하도록 하겠습니다!
- 주의!
  - 일반적으로 접속하는 AWS 계정과 Educate 계정은 다른 것입니다!
  - 혹시 기존에 사용하시던 계정에 추가적인 인스턴스나 서비스를 런칭하시면 비용이 나올 수도 있으니 꼭 우측 상단에 아이디가 어떻게 되어있는지 확인을 해주세요!

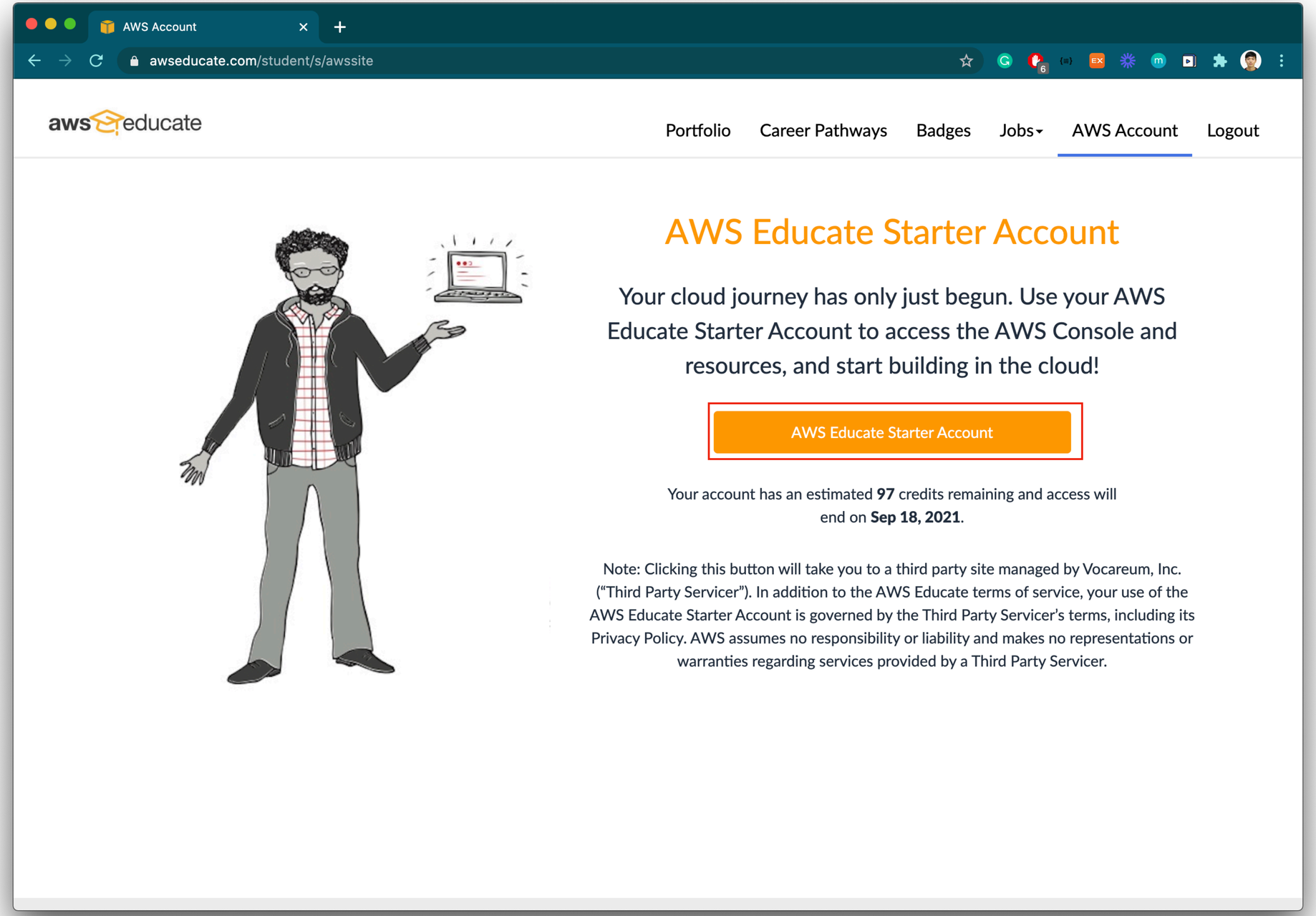
https://www.awseducate.com/signin 로 접속



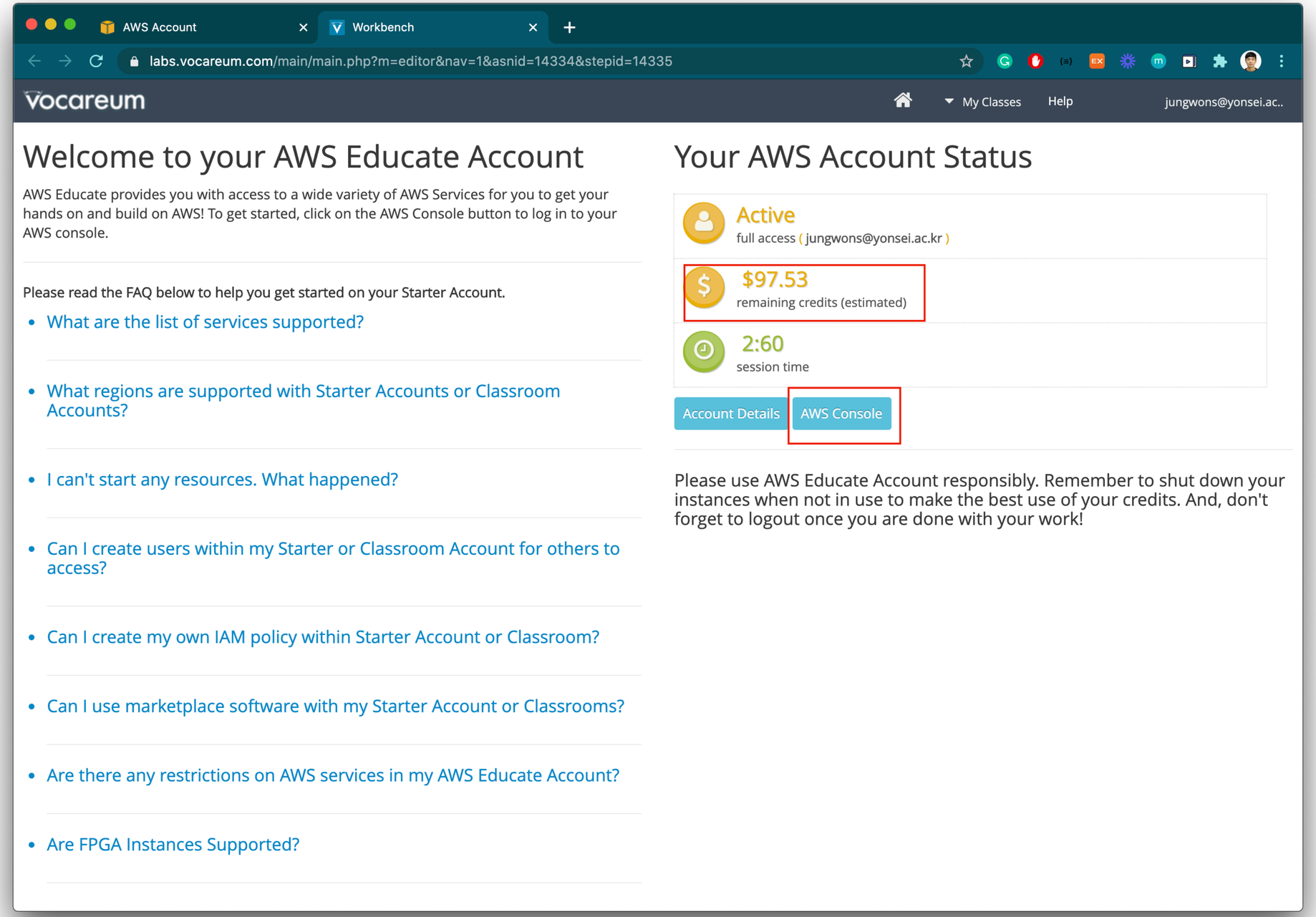
AWS Account 클릭!



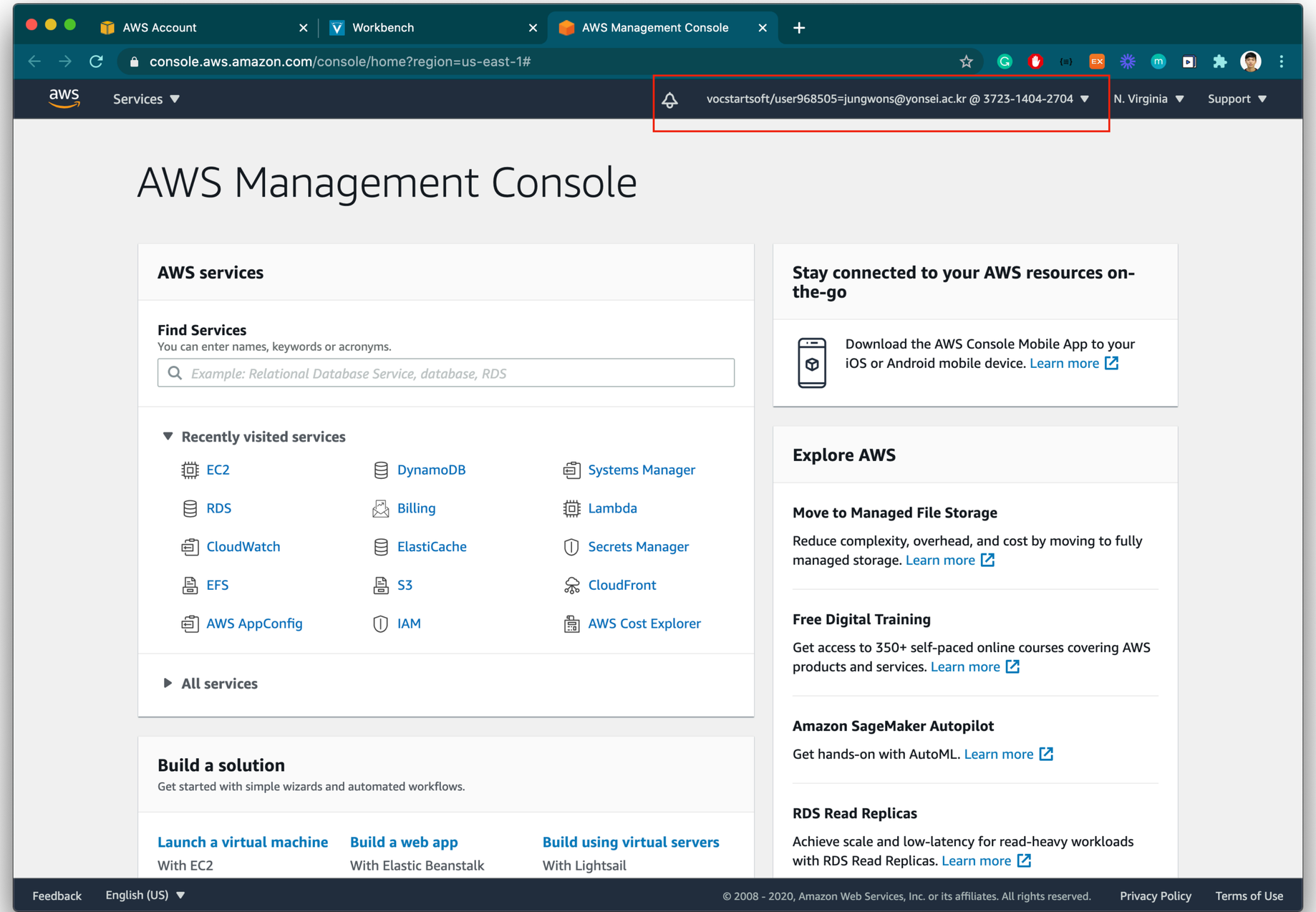
AWS Educate Starter Account 클릭!



잔액 확인 후, AWS Console 클릭!



우측 상단에 유저 이름과 Region이 이런식으로 나오면 정상적으로 Educate계정으로 접속이 된것입니다!



E.O.D