

데이터처리프로그래밍

파이썬 개요



강원대학교 교육혁신원 송혜정

<hjsong@kangwon.ac.kr>



파이썬 개요

✓ 학습목표

- 파이썬 프로그래밍 언어를 이해한다
- 파이썬 개발환경을 이해한다

✓ 학습내용

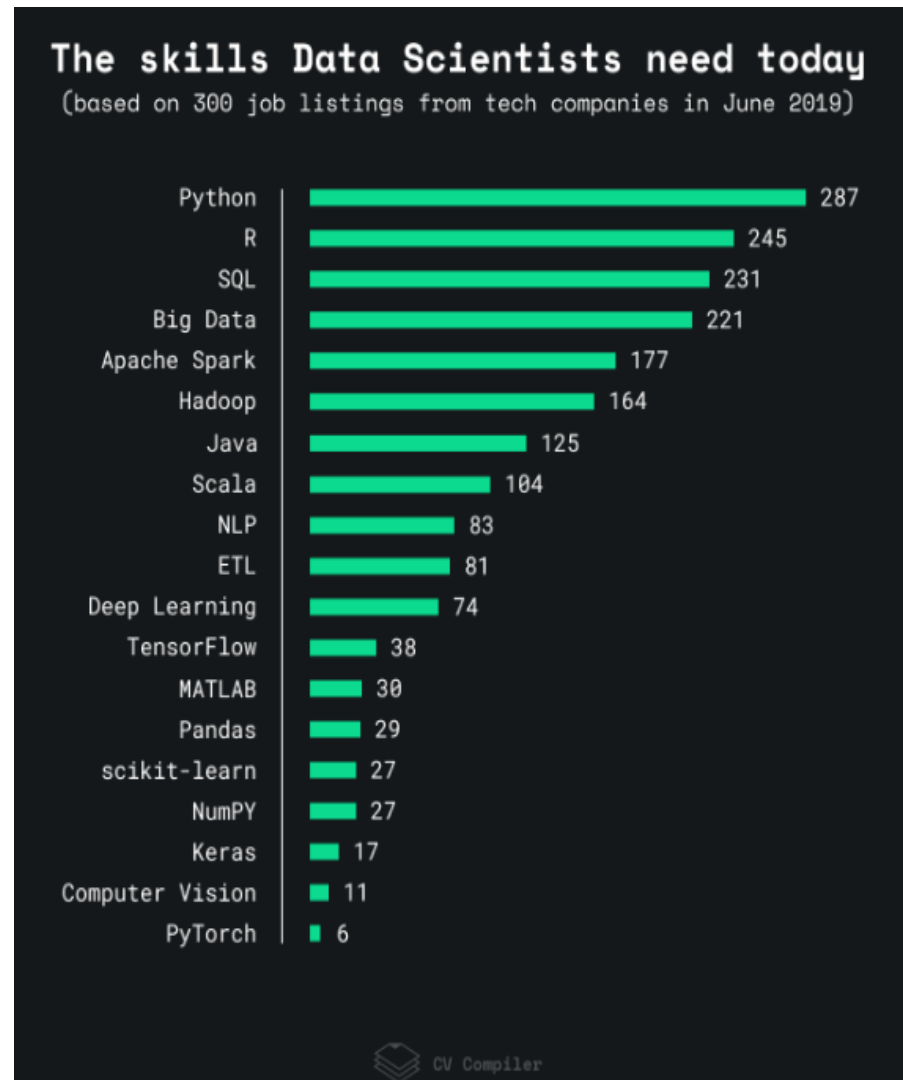
- 파이썬 프로그래밍 언어
- 파이썬 개발환경
- Jupyter Notebook
- Code cell
- Markdown cell (text cell)



강의에 앞서서..

- 본 강의자료는 아래의 자료들을 참고하여 만들어 졌음을 알립니다
1. 교재
 2. Python (<https://docs.python.org>)
 3. Jupyter Notebook (<https://jupyter.org>)
 4. KDnuggets (<https://www.kdnuggets.com>)

Why Python ?



<https://www.kdnuggets.com/2019/08/marketable-data-scientist.html>

Why Python ?

- Python은 데이터 사이언스 분야의 표준 프로그래밍 언어
- 데이터 사이언스를 위한 데이터 추출(Extract), 변환(Transform), 적재(Load), 시각화(Visualization), 통계(Statistics), 기계학습(Machine Learning)등 관련 다양한 라이브러리들을 제공
- 스크립트 프로그래밍 언어(scripting language)로 특정 분야의 문제를 해결하기 위해 빠르게 배워 활용 가능
- 범용 프로그래밍 언어로 접근성과 응용력이 우수하기 때문에 학계와 산업계의 많은 연구원, 프로그래머, 엔지니어 및 데이터 과학자에게 필수 도구임
- 스타일링이 있는 문서와 프로그램 코드를 하나로 묶어서 개발 가능한 주피터 노트북(Jupyter Notebook) 개발 도구를 이용

Python & Data Science

- 데이터 사이언스에 사용되는 파이썬 도구와 패키지

(1) IPython / Jupyter

- 파이썬으로 데이터 사이언스 작업을 하기 위한 개발 환경 제공
- 웹기반 대화형으로 실행 가능한 문서생성 제공
- <https://jupyter.org/>

(2) NumPy

- 다차원 데이터 배열을 위한 효율적인 저장 및 계산 기능을 제공
- <http://www.numpy.org/>

(3) SciPy

- 수학, 과학, 공학 분야의 과학계산용 알고리즘을 제공하는 수치도구로 고성능 선형대수, 함수최적화, 신호처리, 영상처리, 통계분포 등 포함
- <https://scipy.org/scipylib/index.html>

(4) Pandas

- DataFrame 개체를 이용한 데이터 조작, 필터링, 그룹화 및 변환관련 함수 제공
- <https://pandas.pydata.org/>

Python & Data Science

(5)Matplotlib

- 과학계산용 그래프 라이브러리로 다양한 플롯을 그릴 수 있는 인터페이스를 제공
- <https://matplotlib.org/>

(6)Scikit-Learn

- 파이썬의 기계학습 라이브러리, 데이터마이닝, 빅데이터분석을 위한 간단하고 효율적인 도구
- <https://scikit-learn.org/stable/index.html>

파이썬 특징

1. 플랫폼 독립적인 언어

어떤 운영체제든 상관없이 사용할 수 있는 언어를 말한다.

2. 인터프리터 언어

컴파일러 언어와 달리, 소스코드 자체가 바로 실행되는 특징이 있는 언어이다.

속도는 느리지만, 굉장히 간편하게 사용할 수 있다.

3. 객체 지향 언어

해결해야 할 문제의 구성요소를 객체단위로 정의하여 프로그래밍하는 언어

각 객체는 기능(메서드)과 데이터(속성)로 정의하고 객체를 생성하고 메서드를 호출하는 방식으로 프로그램을 작성한다

4. 동적 타이핑 언어

프로그램의 실행 시점에서 각 프로그램 변수의 타입을 결정하는 언어이다



파이썬 코드 예

- 화면에 'Hello World!'라는 텍스트를 출력하는 프로그램
- 파이썬과 자바로 코드 비교

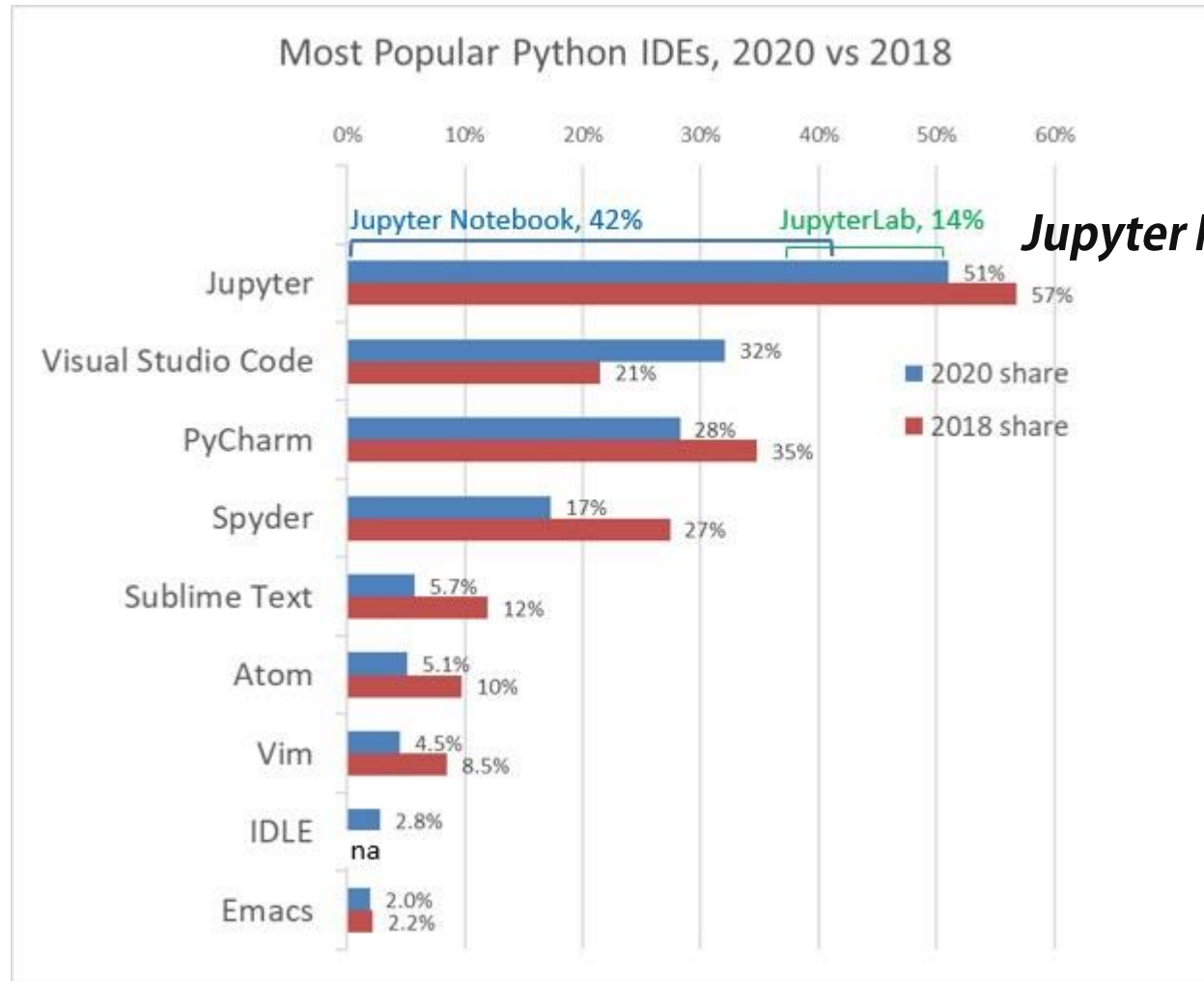
파이썬

```
print("Hello World!")
```

자바(Java)

```
class HelloWorldApp {  
    public static void main(String[] args) {  
        System.out.println("Hello World!");  
    }  
}
```

파이썬 통합개발환경



Jupyter Notebook

Anaconda

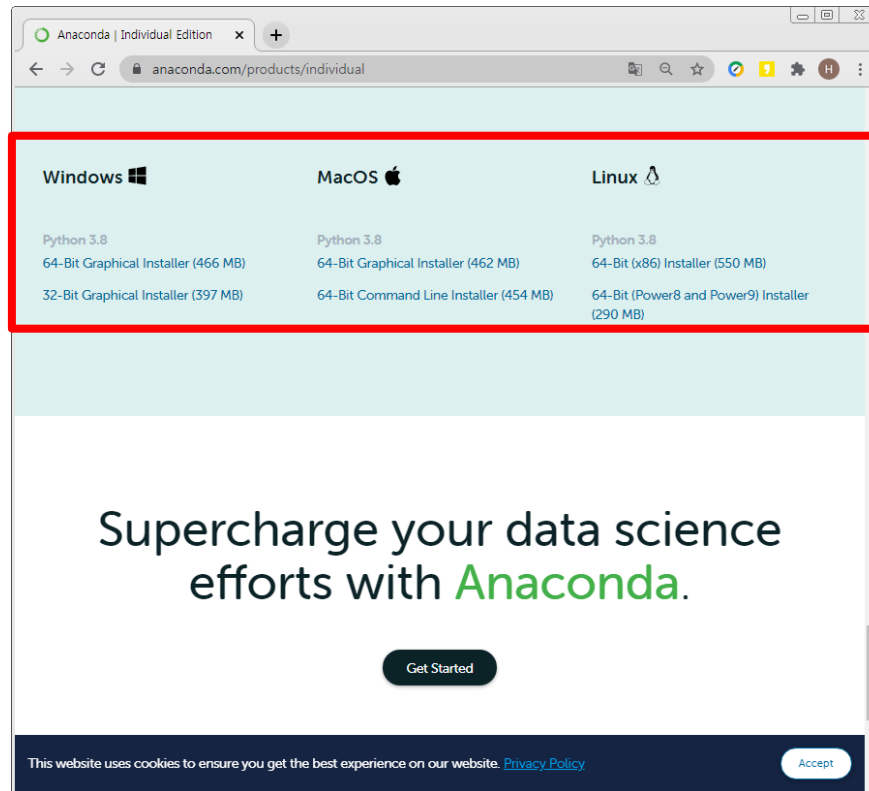
- 아나콘다(anaconda)는 파이썬 기반 데이터 사이언스를 위한 표준 플랫폼
- 파이썬 3.8 버전 지원
- 데이터 사이언스에 필요한 중요 패키지 기본 설치
 - IPython, Jupyter Notebook, numpy , scipy, matplotlib, pandas, scikitlearn 등
- 파이썬 패키지들을 가상환경에서 관리



<https://www.anaconda.com>

Anaconda

- Anaconda Distribution 설치
 - <https://www.anaconda.com/distribution/>



(1) 웹 기반 대화식 프로그래밍 환경 설치
파이썬 버전 3.8, 주피터노트북
(Jupyter notebook)

(2) 데이터분석용 라이브러리 설치
numpy, scipy, matplotlib,
pandas, scikitlearn

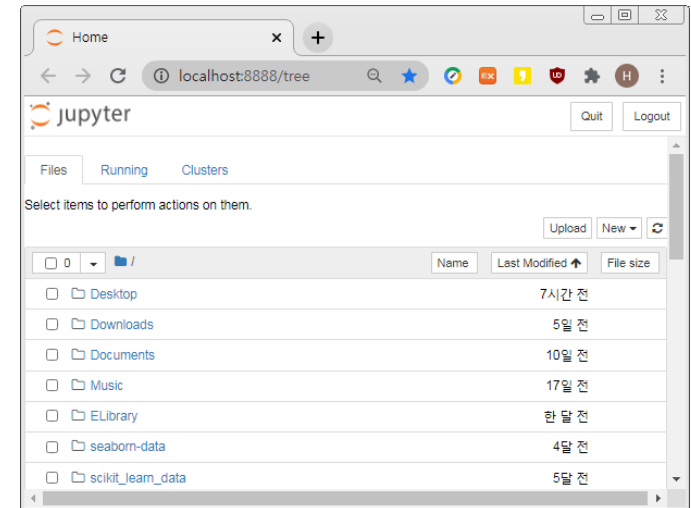
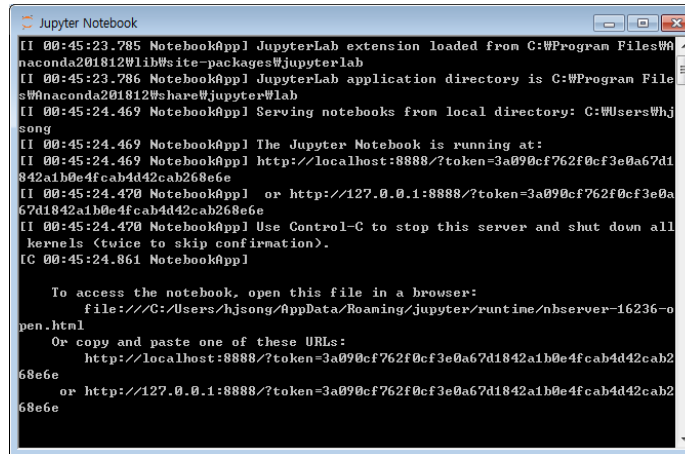
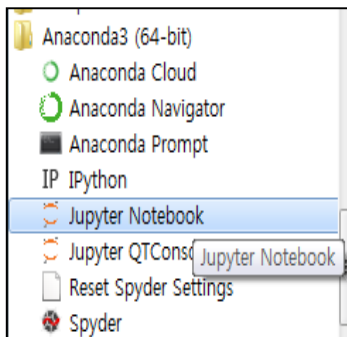
Jupyter Notebook

- 실행 가능한 코드를 삽입하여 문서를 작성할 수 있는 웹 기반 대화형 프로그래밍 환경
- [Running code](#)와 [markdown](#)을 이용하여 문서를 생성하고 공유하는 오픈 소스 웹 응용 프로그램
- 파이썬코드, 텍스트, 그래프, 수식, 이미지, 비디오 등을 포함한 노트북 파일을 작성
- 파이썬 이외의 다른 프로그래밍 언어(Ruby, R, JavaScript 등)도 지원
- Documentation: <https://jupyter-notebook.readthedocs.io/en/stable/>



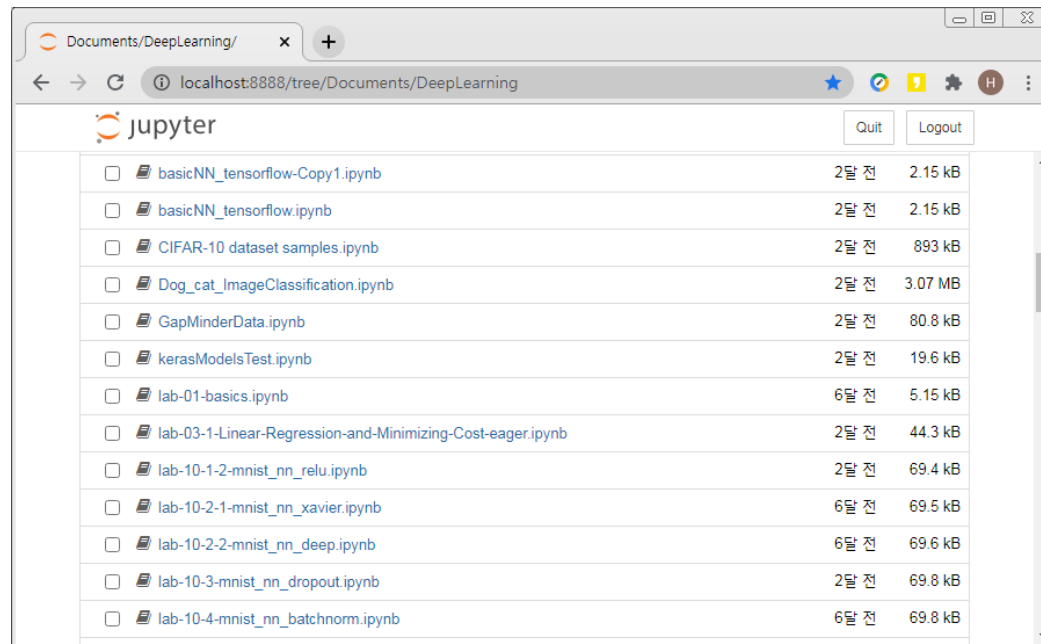
Jupyter Notebook

- Jupyter Notebook 시작하기
 - [시작]버튼 – [Ananconda3]–[Jupyter Notebook]
 - 콘솔에 노트북 서버에 대한 정보가 출력
 - 웹 브라우저가 웹 응용 프로그램의 URL (<http://localhost:8888/tree>) 로 open



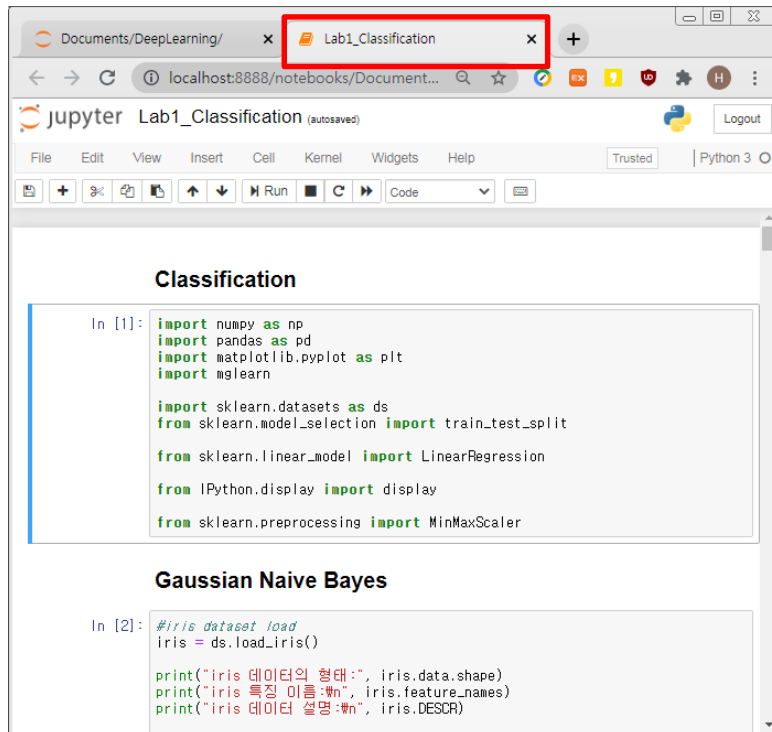
Jupyter Notebook

- Jupyter notebook dash board
 - 웹 애플리케이션의 방문 페이지 (<http://localhost:8888/tree>)
 - 현재 노트북의 디폴트 디렉토리(노트북 서버가 시작된 디렉토리)에서 노트북 파일(확장자 ipynb) 목록을 표시
 - Documents 폴더 아래에 작업폴더를 생성하여 이름을 변경하고 선택하여 새로운 노트북파일을 저장



Jupyter Notebook

- Jupyter notebook dash board
 - Files탭에서 노트북 파일을 선택하면 웹브라우저의 새로운 탭에서 활성화되어 코드확인, 수정, 실행 처리
 - 활성화된 모든 노트북은 대시보드의 Running탭에서 확인 및 Shutdown 버튼으로 제거



The screenshot shows a Jupyter Notebook interface with a code cell containing the following Python code:

```

In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import mglearn

import sklearn.datasets as ds
from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression

from IPython.display import display

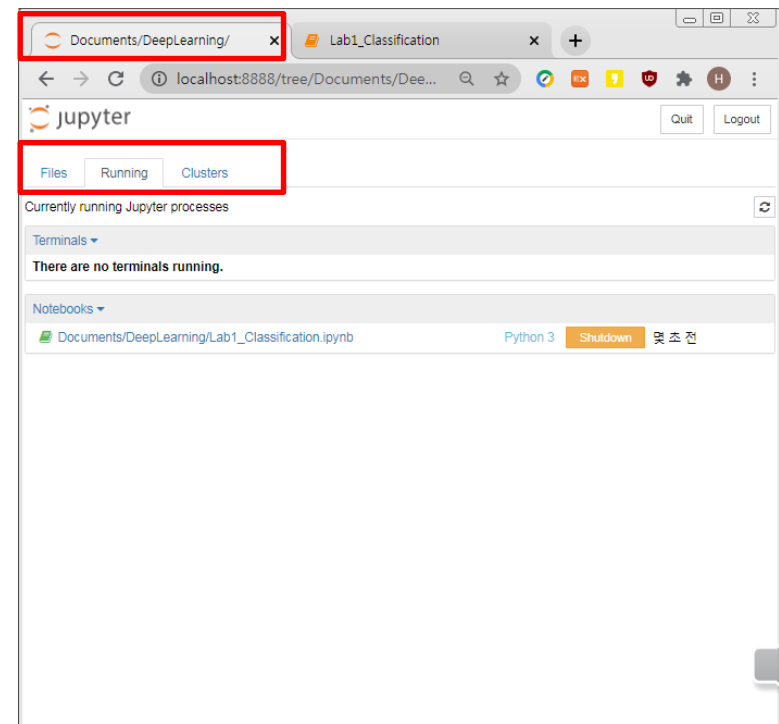
from sklearn.preprocessing import MinMaxScaler
  
```

Below the code cell, the title "Gaussian Naive Bayes" is visible, followed by another code cell:

```

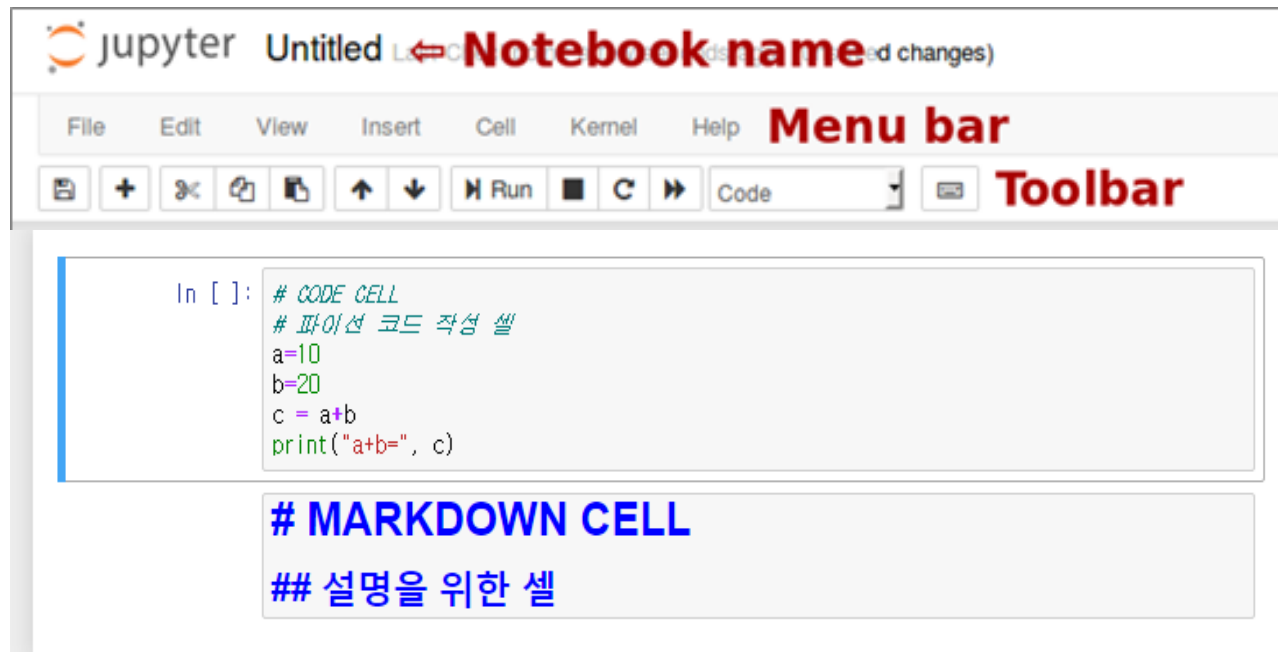
In [2]: #iris dataset load
iris = ds.load_iris()

print("iris 데이터의 형태:", iris.data.shape)
print("iris 특징 이름:\n", iris.feature_names)
print("iris 데이터 설명:\n", iris.DESCR)
  
```



Jupyter Notebook

- User interface



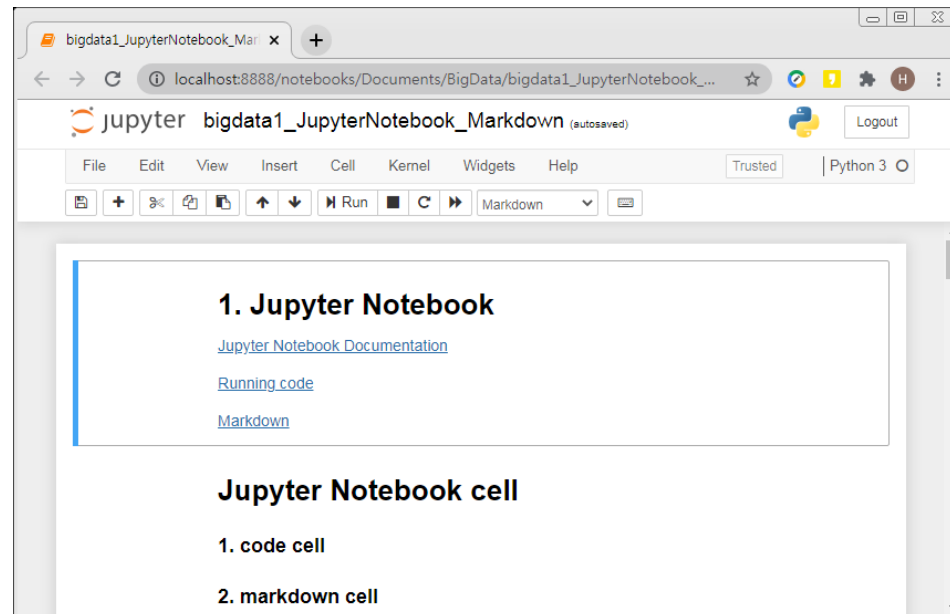
Jupyter Notebook

- Code cell

사용하는 프로그래밍 언어는 kernel에 따라 결정
기본 커널 (IPython)은 Python 코드를 실행.

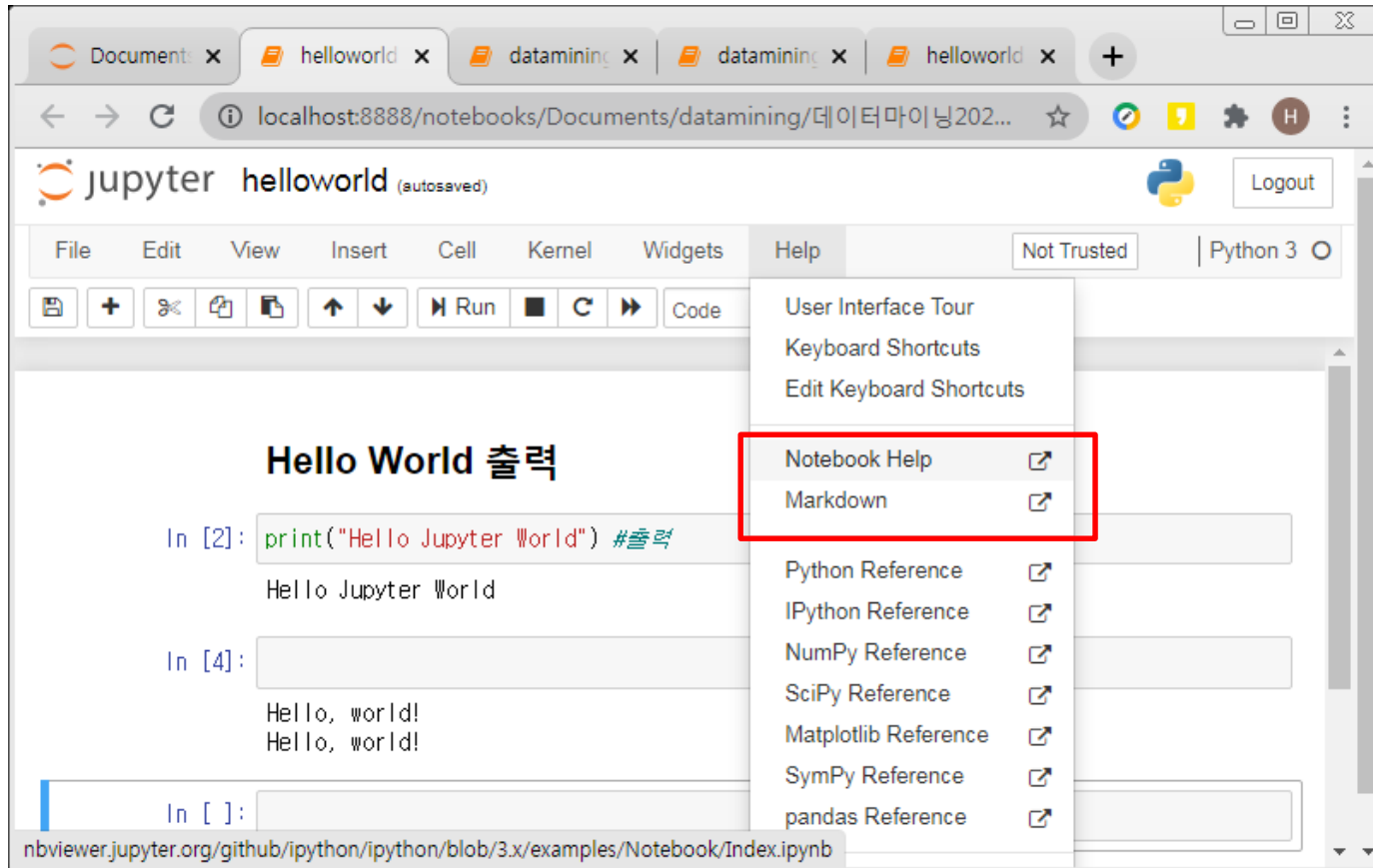
- Markdown cell

설명 텍스트를 표시하여 실행 코드를 문서화



Jupyter Notebook

- Help를 통하여 Notebook, Markdown 도움말 확인

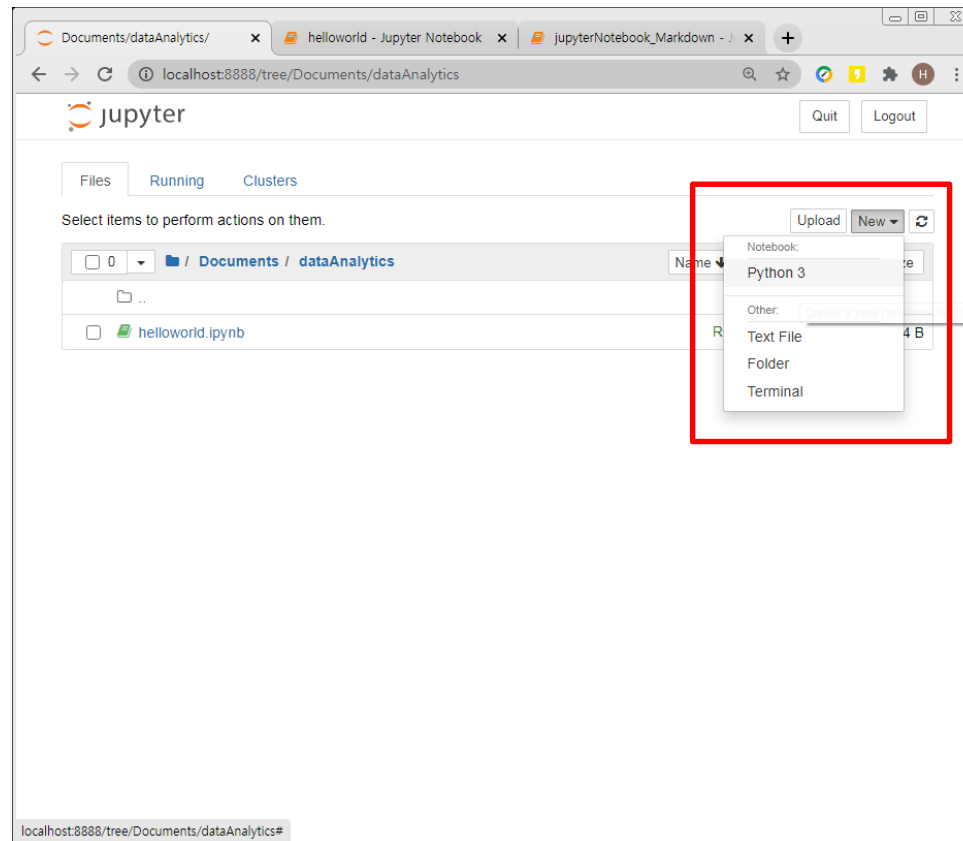


The screenshot shows a web browser window displaying a Jupyter Notebook. The browser tabs include 'Document', 'helloworld', 'datamining', and 'helloworld'. The address bar shows 'localhost:8888/notebooks/Documents/datamining/데이터마이닝202...'. The Jupyter interface has a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. The 'Help' menu is open, showing options: 'User Interface Tour', 'Keyboard Shortcuts', 'Edit Keyboard Shortcuts', 'Notebook Help', 'Markdown', 'Python Reference', 'IPython Reference', 'NumPy Reference', 'SciPy Reference', 'Matplotlib Reference', 'SymPy Reference', and 'pandas Reference'. The 'Notebook Help' and 'Markdown' options are highlighted with a red rectangle. The notebook content shows a code cell with 'print("Hello Jupyter World")' and its output 'Hello Jupyter World', followed by another code cell with 'Hello, world!' and its output 'Hello, world!'. The status bar at the bottom indicates 'nbviewer.jupyter.org/github/ipython/ipython/blob/3.x/examples/Notebook/Index.ipynb'.

Jupyter Notebook

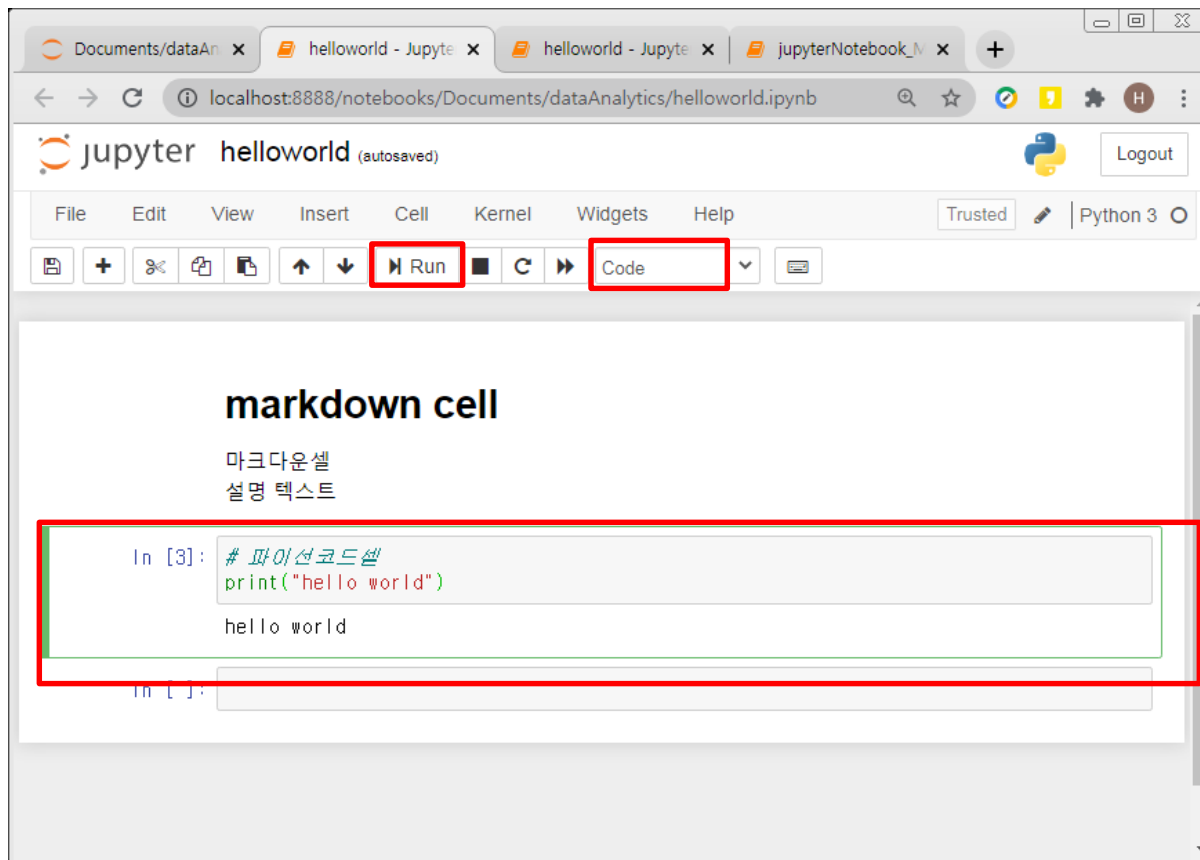
1) 대시보드에서 노트북 생성

- [New]-[Python3]
- 작업폴더에 Untitled.ipynb 파일로 저장



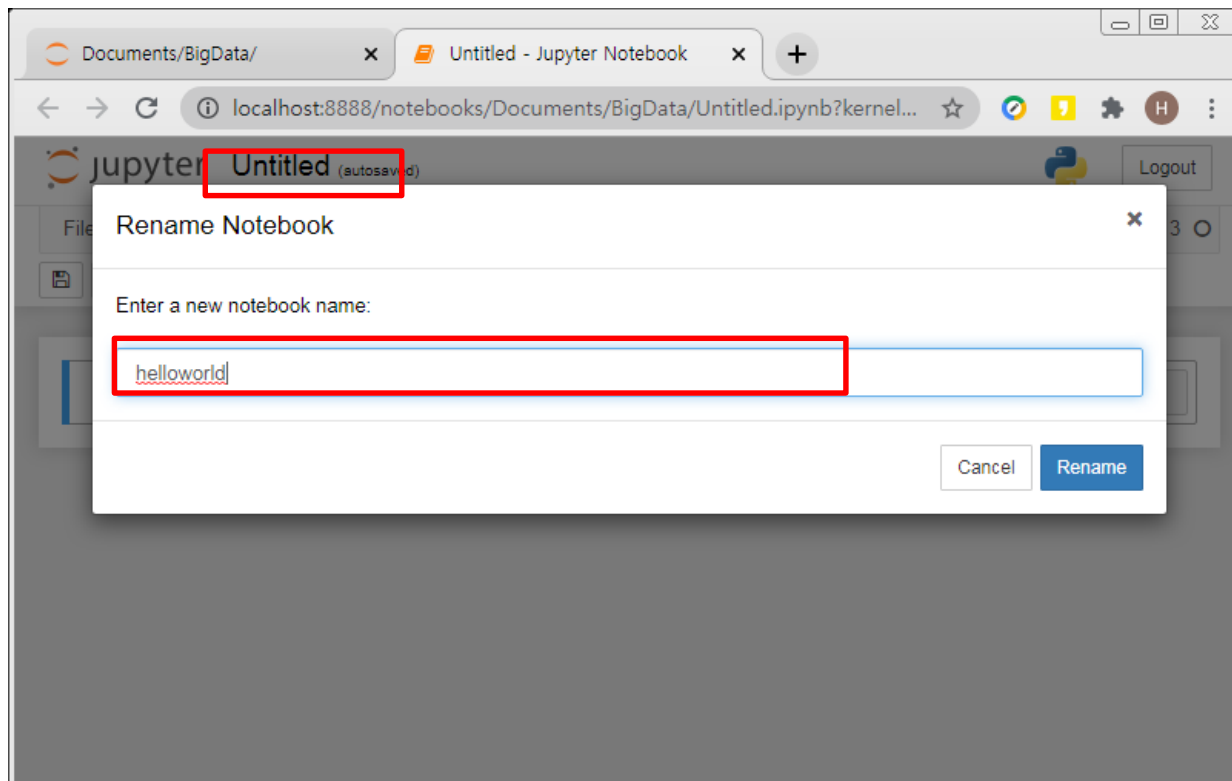
Jupyter Notebook

2) 코드 셀에 파이썬 코드 작성 및 실행(run cell)



Jupyter Notebook

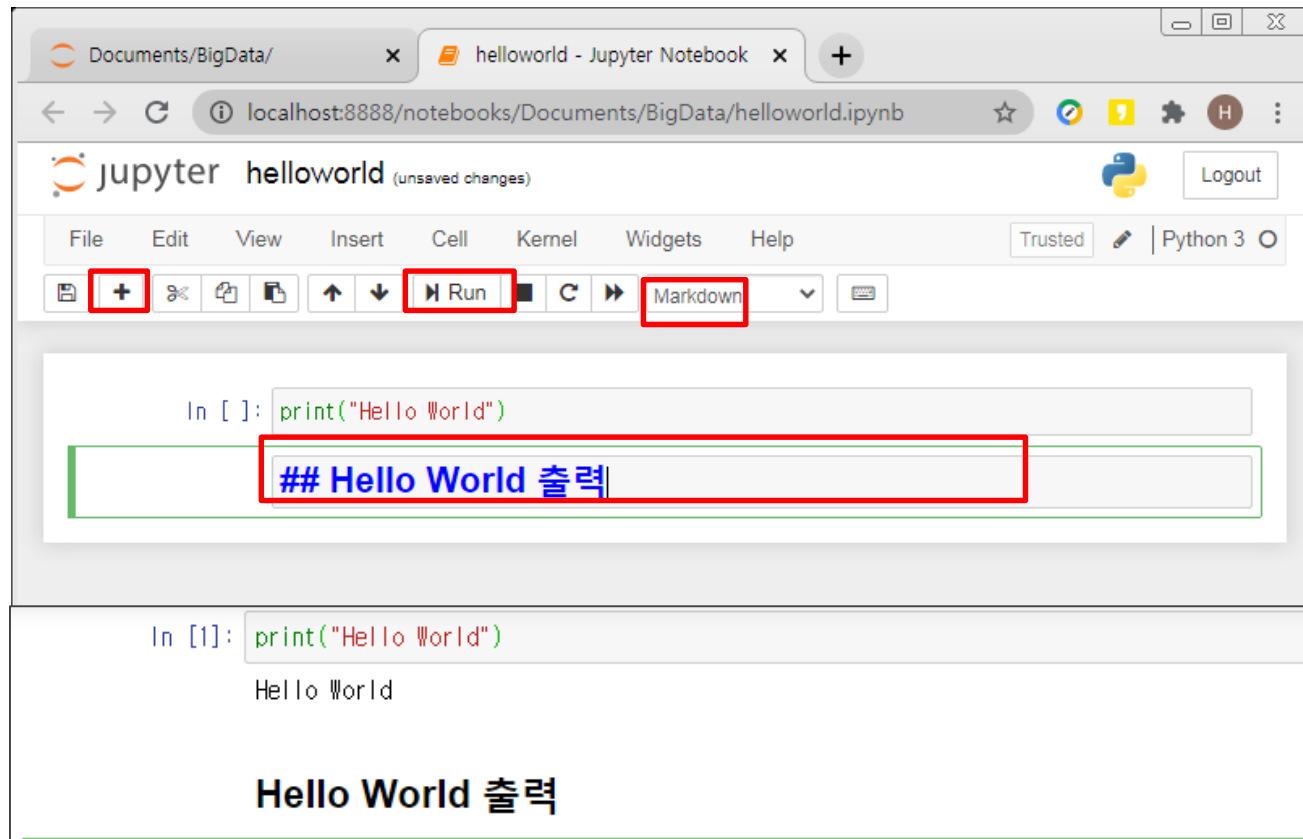
3) 파이썬 노트북 파일명 변경 [untitled] 클릭하여 새 이름 입력



Jupyter Notebook

4) 새로운 셀에 설명 추가

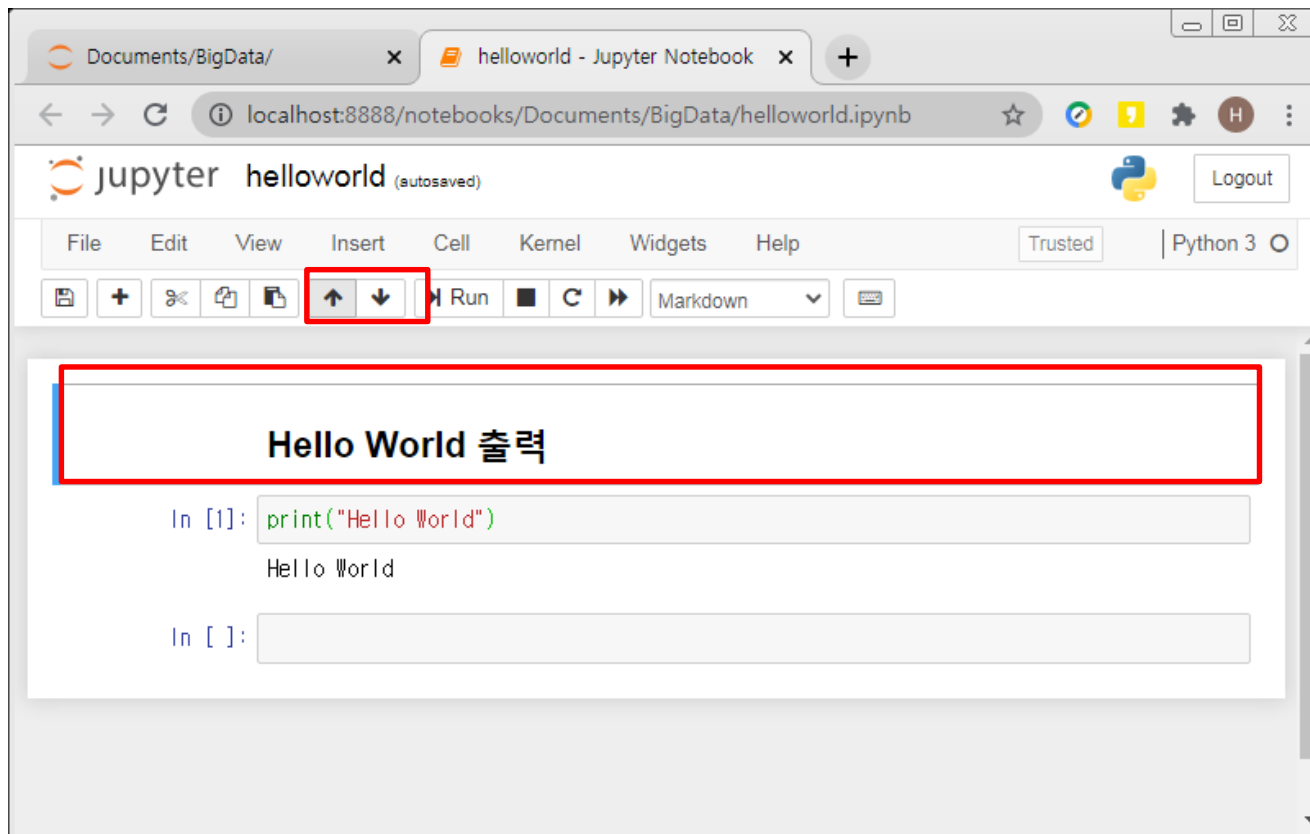
[+]로 새로운 셀 생성 - [markdown] 선택 후 마크다운 셀에 설명입력
[Run cell] 로 확인



Jupyter Notebook

5) 셀 이동

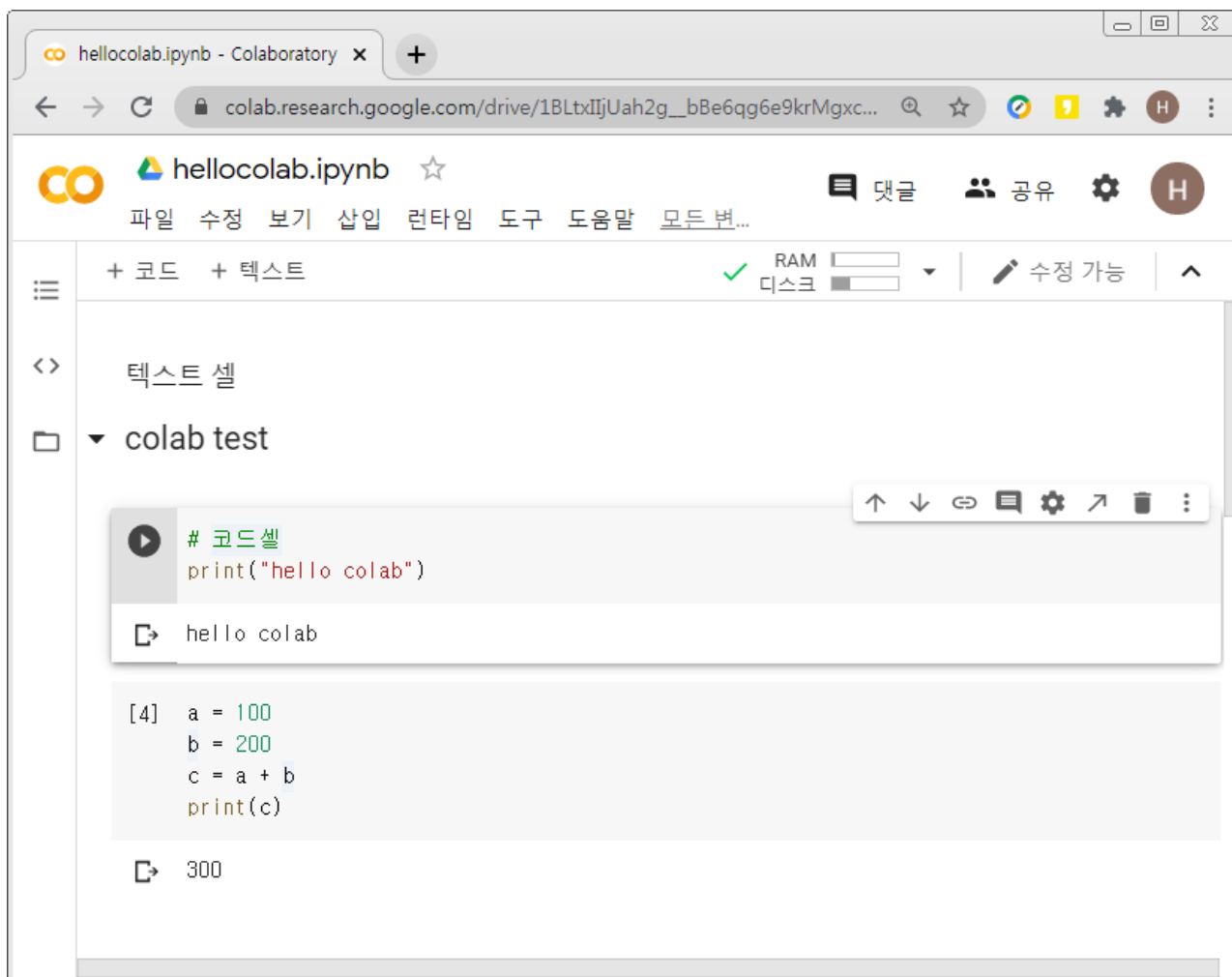
이동할 셀을 선택하여 위로/아래로 이동 가능



Colab

- 주피터 노트북 개발을 위한 구글의 웹서비스
(<https://colab.research.google.com/>)
- Colaboratory(Colab)을 사용하면 브라우저에서 Python을 작성하고 실행 가능함
- 구글 지메일(gmail) 계정으로 로그인하여 사용
- 코랩 노트북의 장점.
 - 파이썬 설치 없이 크롬 웹 브라우저에서 주피터 노트북 개발 가능
 - Google drive와 연동하여 가능
 - 다른 사용자들과 공유가 용이
 - 파이썬 기반 데이터분석 라이브러리들이 미리 설치됨(numpy, pandas, matplotlib, scikit-learn, Tensorflow, kerase등)
 - GPU기반 딥러닝 실행 가능

Colab



QUIZ!

- 과제1. Jupyter Notebook의 파이썬 코드로 다음을 작성하여 제출하세요.

- (1) 과제에 대한 설명은 마크다운셀로 작성
- (2) 간단한 자기소개를 코드셀에 파이썬 프로그램으로 작성

<출력예시>

성명 : ...

좌우명 : ...

장래희망: ...

- 학습활동 결과는 e-루리에 제출바랍니다.
- 제목 : Report1. 파이썬 문제
- 제출내용 : Report1_성명.ipynb 노트북파일을 제출
- 제출기한 : 2021년 3월 8일 월요일 오후 11:55