

# Build an environment for Big data analysis

- ❖ Building a python development environment
- ❖ Database management system





01

# Python coding tools





# What is Python language..?

- ❖ Python is a programming language.
  - It is a high-level programming language published by programmer Guido van Rossum in 1991.
  - The name Python is derived from the comedy "Monty Python's Flying Circus", a favorite of its developer Guido.



# History of Python

- ❖ Python 2 was released on October 16, 2000, and Python 3 was released on December 3, 2008.
- ❖ The main features of Python 3 were also reflected in Python 2.6 and 2.7 versions, but there were still many problems in compatibility.
  - Support for Python 2 officially ended on January 1, 2020.
- ❖ The biggest advantage of Python compared to programming languages such as Java or C is that the syntax is very simple and close to a natural language.

# Grammar

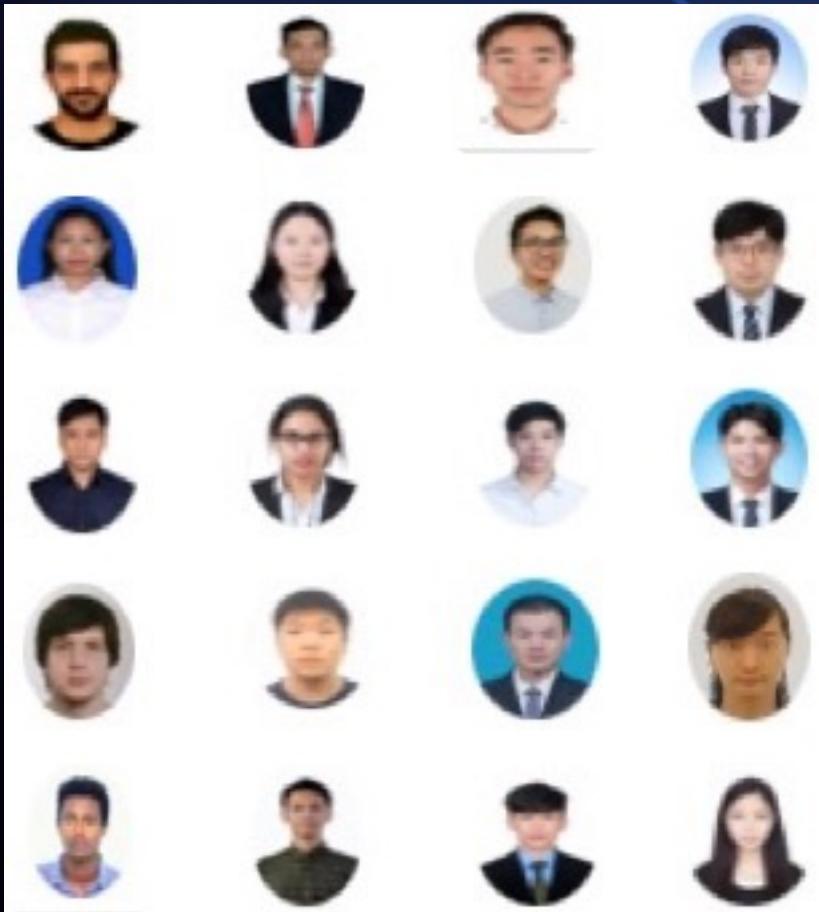


- ❖ The most well-known feature of Python's syntax is the block structure using indentation using 'tab' key.
- ❖ Python language (example)
- ❖ Java language (example)

```
def sayHello(name, gender):
    print('Hello, ')
    if gender == 'Female':
        print('Ms ' + name)
    else:
        print('Mr ' + name)
```

```
public String sayHello(String name, String gender) {
    String tmpStr = "Hello, ";
    if gender.equals("Female") {
        System.out.print(tmpStr + "Ms " + name);
    } else {
        System.out.print(tmpStr+"Mr " + name);
    }
}
```

# What is Language..?

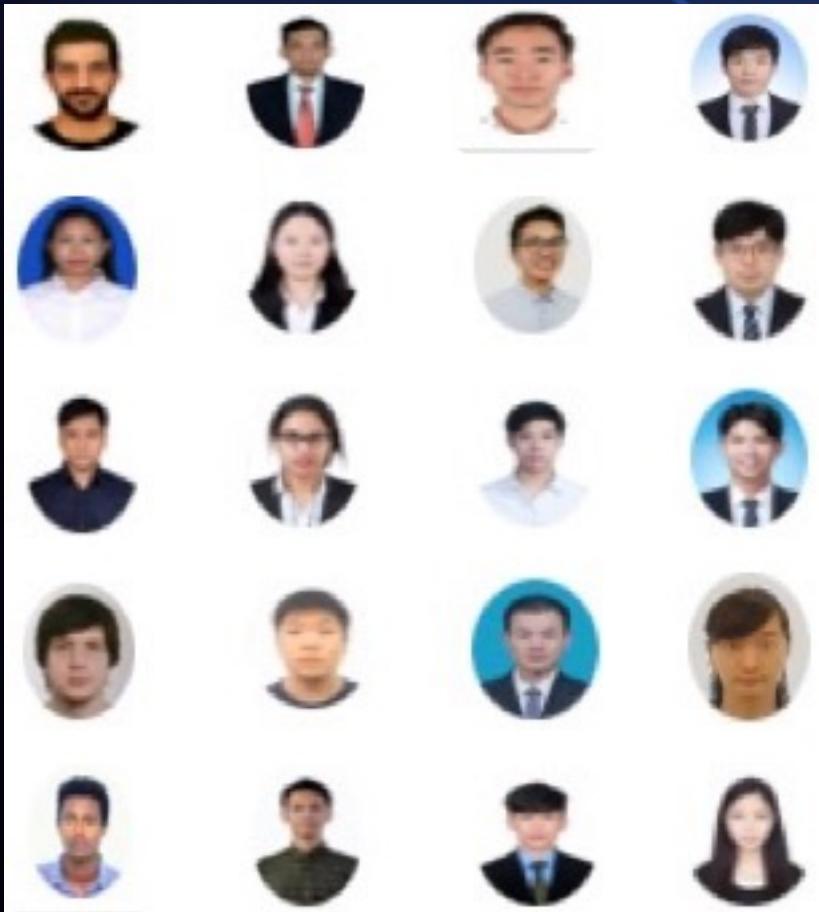


Hello~



안녕하세요!

# What is Computer language..?



**Python  
Java  
C  
Matlab  
R  
...**

01001000  
01100101  
01101100  
01101100  
01101111  
00100000



# Translator vs Compiler



## Translator

A screenshot of the Google Translate website. The interface shows a red border around the main input and output area. The source text "Hello" is in English, and the target text "안녕하십니까" is in Korean. Below the text, there is a Python code example: "def sayHello(name, gender): print('Hello, ')".

## Compiler

The diagram illustrates the compilation process. On the left, under the heading "Python language (example)", there is a block of Python code:

```
❖ Python language (example)  
  
def sayHello(name, gender):  
    print('Hello, ')\n    if gender == 'Female':\n        print('Ms ' + name)\n    else:\n        print('Mr ' + name)
```

On the right, there is a green computer monitor icon displaying the text "Hello, Ms Insung". Next to it is a green smartphone icon, representing the final compiled output.

# Compiler



- ❖ A compiler is a computer program that translates computer code written in one programming language (the source language) into another language (the target language).

Python  
Java  
C/C++  
Fortran  
Matlab

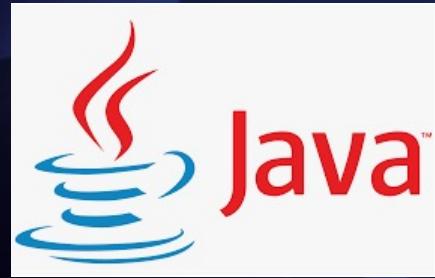
Front end

Middle end

Back end

```
001001 11101  
001000 00001  
001000 00001  
101011 11101  
001000 00010  
101011 11101  
001001 11101
```

# How can we use the exact compiler..?



```
C:\Users>javac
Usage: javac <options> <source files>
where possible options include:
```



```
[minigrace@miniisaiMacmini ~ % python3 -h
usage: /opt/homebrew/bin/python3 [option] ... [-c cmd
...]
```



```
[minigrace@miniisaiMacmini ~ % gcc -help
OVERVIEW: clang LLVM compiler
USAGE: clang [options] file...
```



# Why we need Editor for programming..?



## ❖ Without Editor

```
import matplotlib.pyplot as plt
import os
import numpy as np
get_ipython().run_line_magic('matplotlib', 'inline')

fileName = './corr_data/0816_corr_input.csv'
handle = open(fileName, 'w')

for (path, dir, files) in os.walk(INPUT_PATH):
    for filename in tqdm(files):
        time.sleep(0.1)
        ext = os.path.splitext(filename)[-1]

        if ext == '.csv':
            parsName = filename.split('_')

            country = ''
            country = parsName[1].strip()

            isTargetCoun = _isTargetCountry(country)

            if isTargetCoun == True:
                fileName2 = path + '/' + filename
                df = pd.DataFrame()
                df = pd.read_csv(fileName2, sep=',')
                df = df.replace(np.nan,0)

                df = df.drop(len(df)-1,axis=0)

                population = int(df.population[0])

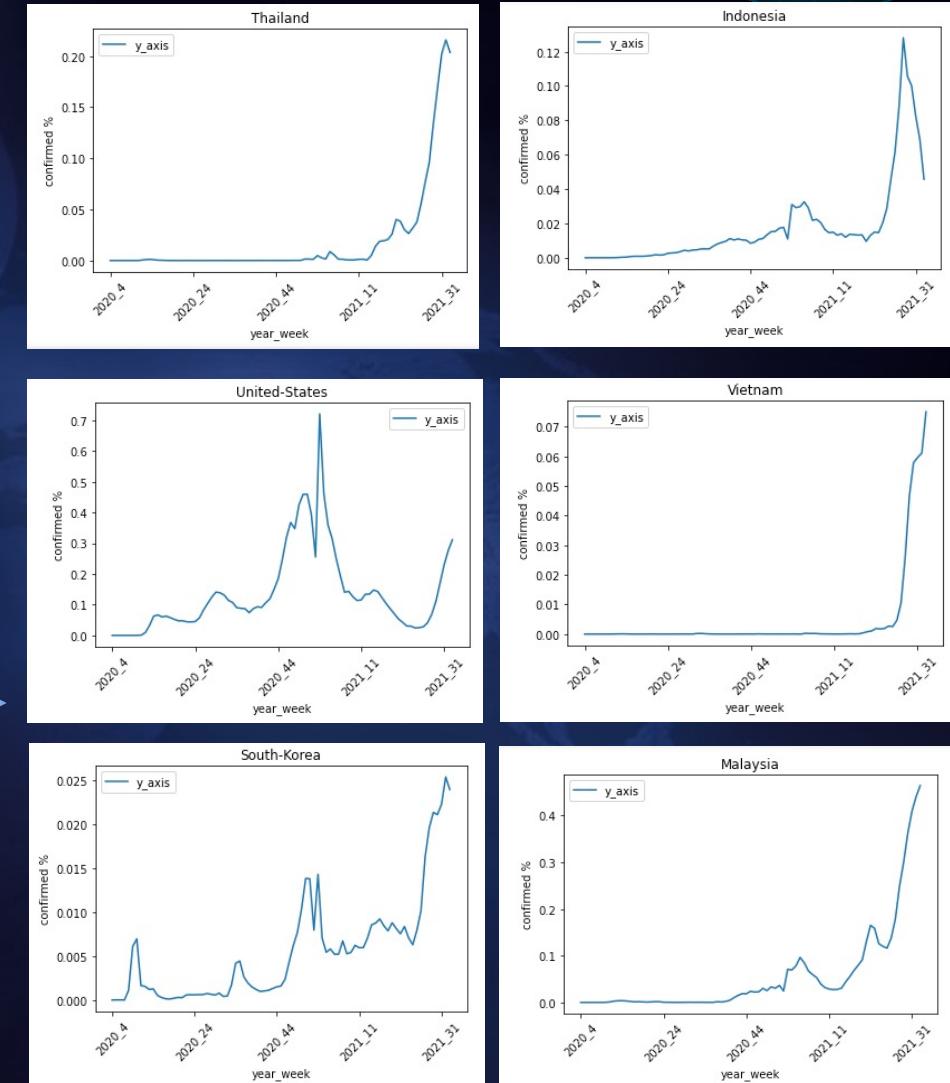
                cols = ['year', 'week']
```

analyzeCOVID19.py

Desktop %  
Desktop % python3 analyzeCOVID19.py

png

- Argentina\_confirmed.png
- Australia\_confirmed.png
- Austria\_confirmed.png
- Brazil\_confirmed.png
- Canada\_confirmed.png
- China\_confirmed.png
- France\_confirmed.png
- Germany\_confirmed.png
- India\_confirmed.png
- Indonesia\_confirmed.png



# Why we need Editor for compiling..?



## ❖ Using Editor

```
import matplotlib.pyplot as plt
import os
import numpy as np
get_ipython().run_line_magic('matplotlib', 'inline')

fileName = './corr_data/0816_corr_input.csv'
handle = open(fileName, 'w')

for (path, dir, files) in os.walk(INPUT_PATH):
    for filename in tqdm(files):
        time.sleep(0.1)
        ext = os.path.splitext(filename)[-1]

        if ext == '.csv':
            parsName = filename.split('_')

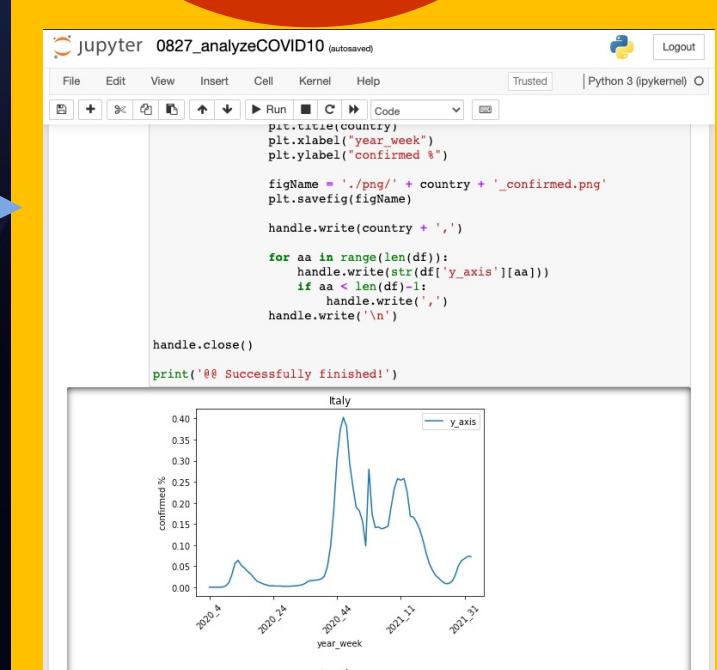
            country = ""
            country = parsName[1].strip()

            isTargetCoun = _isTargetCountry(country)

            if isTargetCoun == True:
                fileName2 = path + '/' + filename
                df = pd.DataFrame()
                df = pd.read_csv(fileName2, sep=',')
```

analyzeCOVID19.py

Desktop %  
Desktop % python3 analyzeCOVID19.py



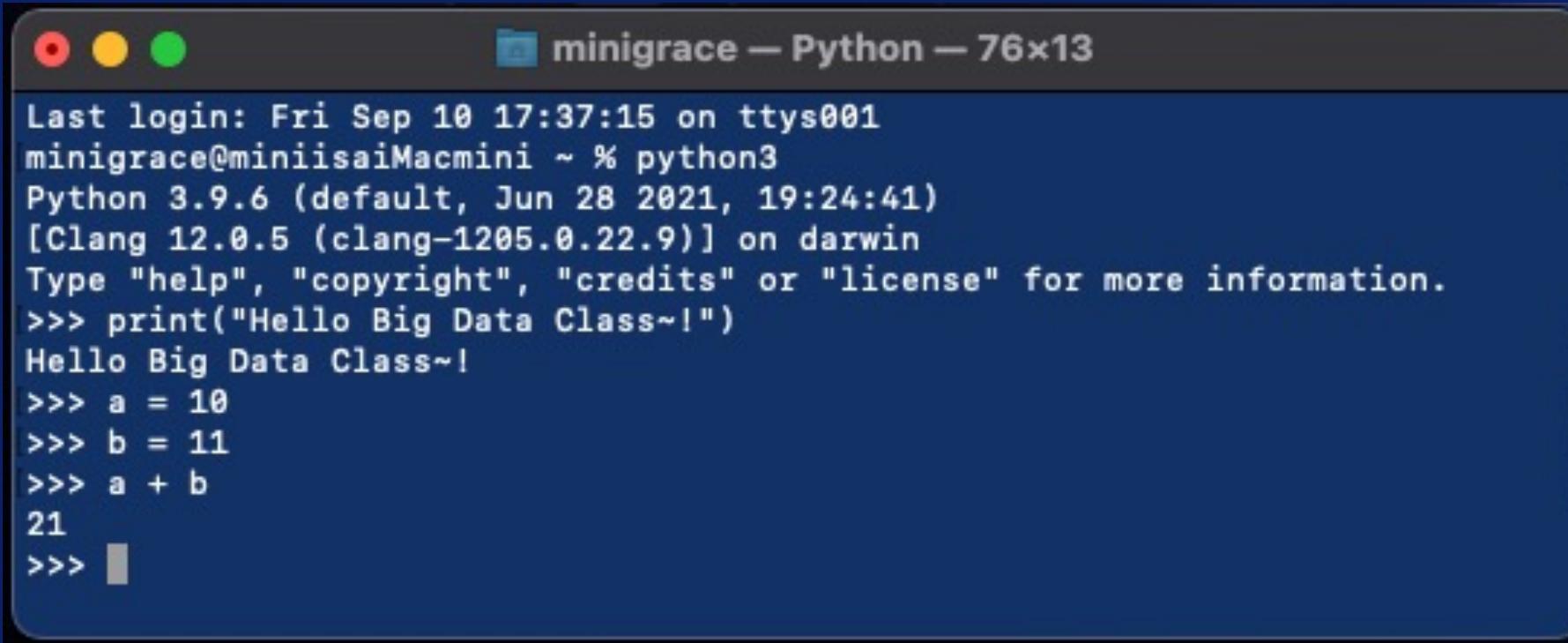
Python editor

png
Argentina_confirmed.png
Australia_confirmed.png
Austria_confirmed.png
Brazil_confirmed.png
Canada_confirmed.png
China_confirmed.png
France_confirmed.png
Germany_confirmed.png
India_confirmed.png
Indonesia_confirmed.png

# Python editors:

## 1. IDLE

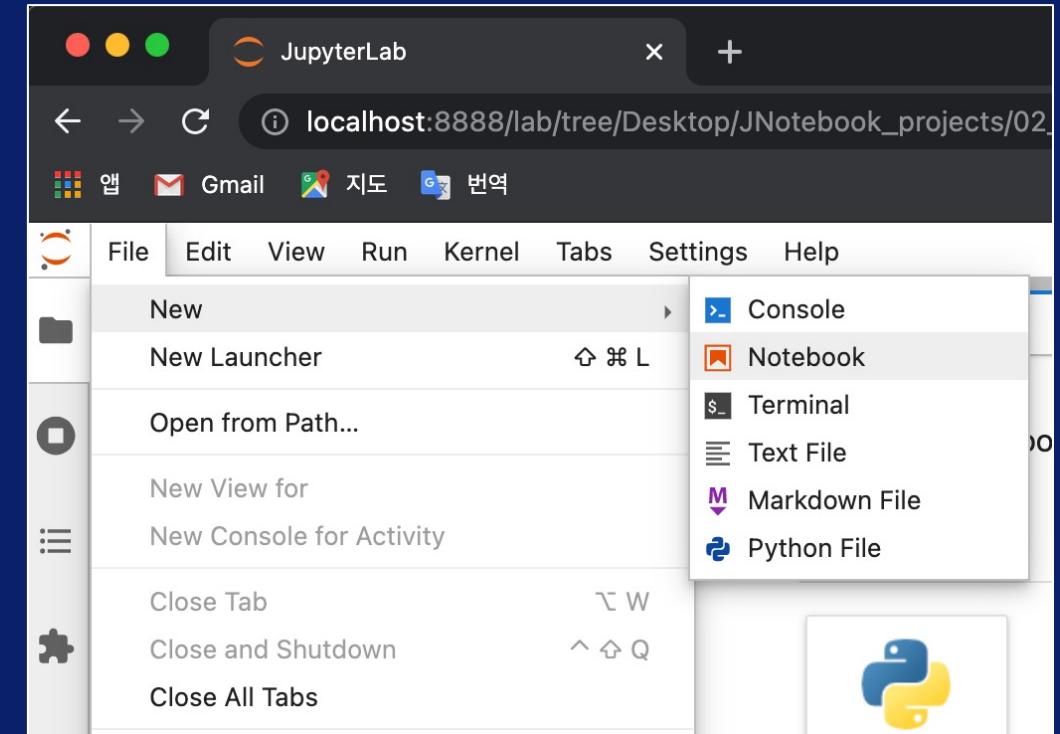
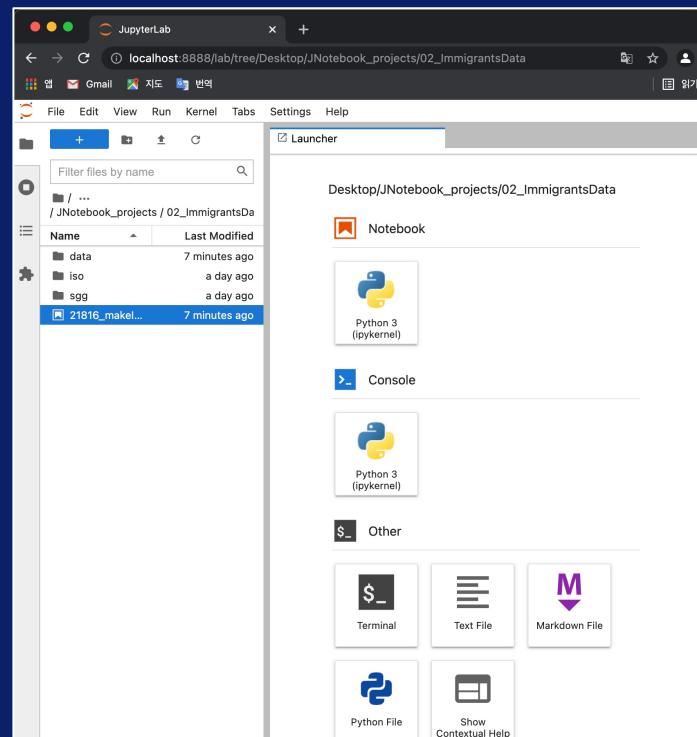
- ❖ IDLE (short for Integrated Development and Learning Environment) is an integrated development environment for Python.
- ❖ This program is automatically installed on your PC when you install the python.



```
Last login: Fri Sep 10 17:37:15 on ttys001
minigrace@miniisaiMacmini ~ % python3
Python 3.9.6 (default, Jun 28 2021, 19:24:41)
[Clang 12.0.5 (clang-1205.0.22.9)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Hello Big Data Class~!")
Hello Big Data Class~
>>> a = 10
>>> b = 11
>>> a + b
21
>>>
```

## Python editors: 2. Jupyter-notebook

- ❖ The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.
- ❖ JupyterLab is the next-generation web-based user interface of the Jupyter project.



# Python editors:

## 3. PyCharm

- ❖ PyCharm is an integrated development environment (IDE) used in computer programming, specifically for the Python language.
- ❖ It is developed by the Czech company JetBrains.

A screenshot of the PyCharm IDE. The main window shows a code editor with Python code for a COVID-19 analyzer. Below the editor is a 'Python Packages' tool window displaying the 'seaborn' package, version 0.11.2, which is described as a library for statistical graphics. The bottom of the screen shows the PyCharm interface with tabs like 'File', 'Edit', 'View', etc., and a status bar at the bottom.

```
03_covid19_Analyzer - 0829_covid19_Analyzer.py
03_covid19_Analyzer / 0829_covid19_Analyzer.py
handle = open(fileName, 'w')
for (path, dic, files) in os.walk(INPUT_PATH):
    for filename in tqdm(files):
        time.sleep(0.1)
        ext = os.path.splitext(filename)[-1]
        if ext == '.osv':
            parsName = filename.split('_')
            country = ''
            country = parsName[1].strip()
            isTargetCountry = _isTargetCountry(country)
            if isTargetCountry == True:
                fileName2 = path + '/' + filename
                df = pd.DataFrame()
                df = pd.read_csv(fileName2, sep=',')
                df = df.replace(np.nan, 0)
                df = df.drop(len(df) - 1, axis=0)
                population = int(df.population[0])
```





02

## Installation



# 1. Install Python

1. Visit the python's main homepage (<https://www.python.org>)
2. In Download tab, choose the file suitable for your PC's OS and download it. (Python 3.x)
3. Install the python program step-by-step.

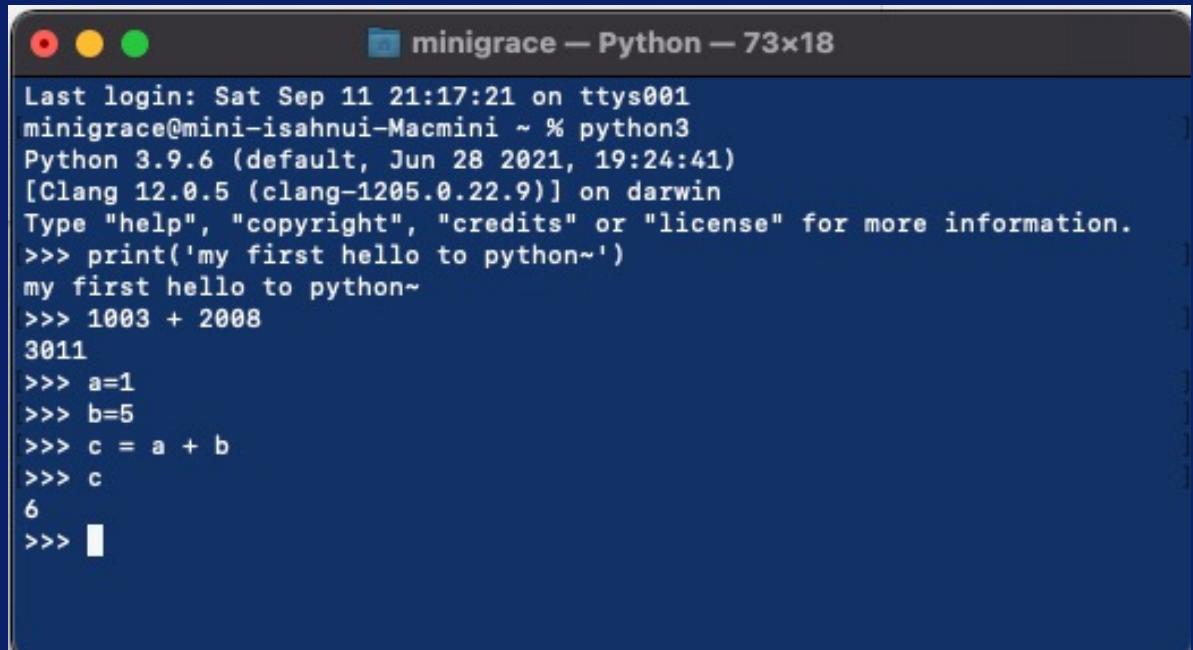
The screenshot shows the Python.org homepage with a dark blue header. The Python logo is on the left, followed by the word "python" in white. To the right are buttons for "Donate", "Search", "GO", and "Socialize". Below the header is a navigation bar with links: "About", "Downloads", "Documentation", "Community", "Success Stories", "News", and "Events". The main content area has a dark blue background with a graphic of two boxes descending from the sky on yellow and white parachutes. The text "Download the latest version for macOS" is prominently displayed, followed by a yellow button labeled "Download Python 3.9.7". Below this, there are links for other operating systems: "Windows", "Linux/UNIX", "macOS", and "Other". There are also links for "Prereleases" and "Docker images". A note mentions "Looking for Python 2.7? See below for specific releases". The bottom section is titled "Active Python Releases" and contains a table with the following data:

Python version	Maintenance status	First released	End of support	Release schedule
3.9	bugfix	2020-10-05	2025-10	PEP 596
3.8	security	2019-10-14	2024-10	PEP 569
3.7	security	2018-06-27	2023-06-27	PEP 537
3.6	security	2016-12-23	2021-12-23	PEP 494
2.7	end-of-life	2010-07-03	2020-01-01	PEP 373

## 2. Run python

---

1. Open the **console window**.
2. Type '**python3**' in the console window.
3. Run python
  - ✓ Print() function
  - ✓ Using variables



```
Last login: Sat Sep 11 21:17:21 on ttys001
minigrace@mini-isahnui-Macmini ~ % python3
Python 3.9.6 (default, Jun 28 2021, 19:24:41)
[Clang 12.0.5 (clang-1205.0.22.9)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> print('my first hello to python~')
my first hello to python~
>>> 1003 + 2008
3011
>>> a=1
>>> b=5
>>> c = a + b
>>> c
6
>>> █
```



### 3. Install JupyterLab

- ❖ What is pip
  - pip is a package management system that installs and manages package software written in Python.
    - pip install target-package-name
    - pip uninstall target-package-name
    - pip install –upgrade target-package-name
- ❖ pip3 install jupyterlab

# 4. jupyterLab

## ❖ How to start?

- <http://localhost:8888/lab>
- Browser: Chrome, Safari, Firefox

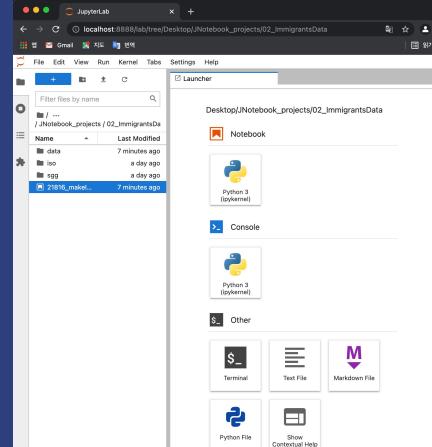
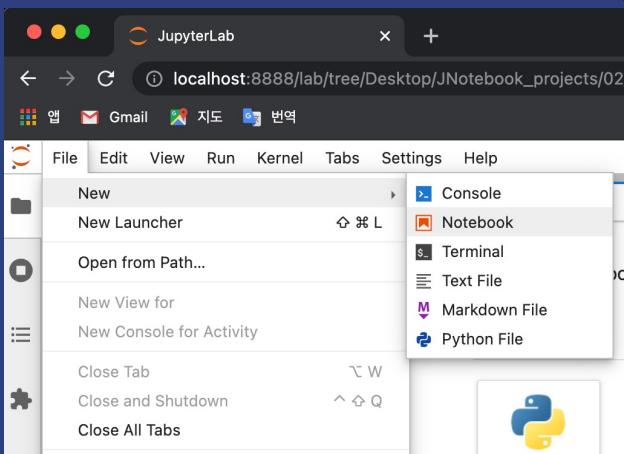
```
gracenotebook@aninseocBookPro ~ % jupyter-lab
[2021-09-11 20:09:35.945 ServerApp] jupyterlab | extension was successfully loaded
[2021-09-11 20:09:35.958 ServerApp] Writing Jupyter server cookie secret to /var/folders/...
[2021-09-11 20:09:36.358 ServerApp] nbclassic | extension was successfully loaded
[2021-09-11 20:09:36.401 ServerApp] nbclassic | extension was successfully loaded
[2021-09-11 20:09:36.403 LabApp] JupyterLab extension loaded from /usr/local/lib/python3.8/dist-packages/jupyterlab
[2021-09-11 20:09:36.403 LabApp] JupyterLab application directory is /usr/local/share/jupyter/lab
[2021-09-11 20:09:36.406 ServerApp] jupyterlab | extension was successfully loaded
[2021-09-11 20:09:36.407 ServerApp] Serving notebooks from local directory:
[2021-09-11 20:09:36.407 ServerApp] Jupyter Server 1.11.0 is running at:
[2021-09-11 20:09:36.407 ServerApp] http://localhost:8888/lab?token=a289dec8...
[2021-09-11 20:09:36.407 ServerApp] or http://127.0.0.1:8888/lab?token=a289...
[2021-09-11 20:09:36.407 ServerApp] Use Control-C to stop this server and sh...
[2021-09-11 20:09:36.415 ServerApp]
```

## ❖ How to stop?

- Ctrl + c

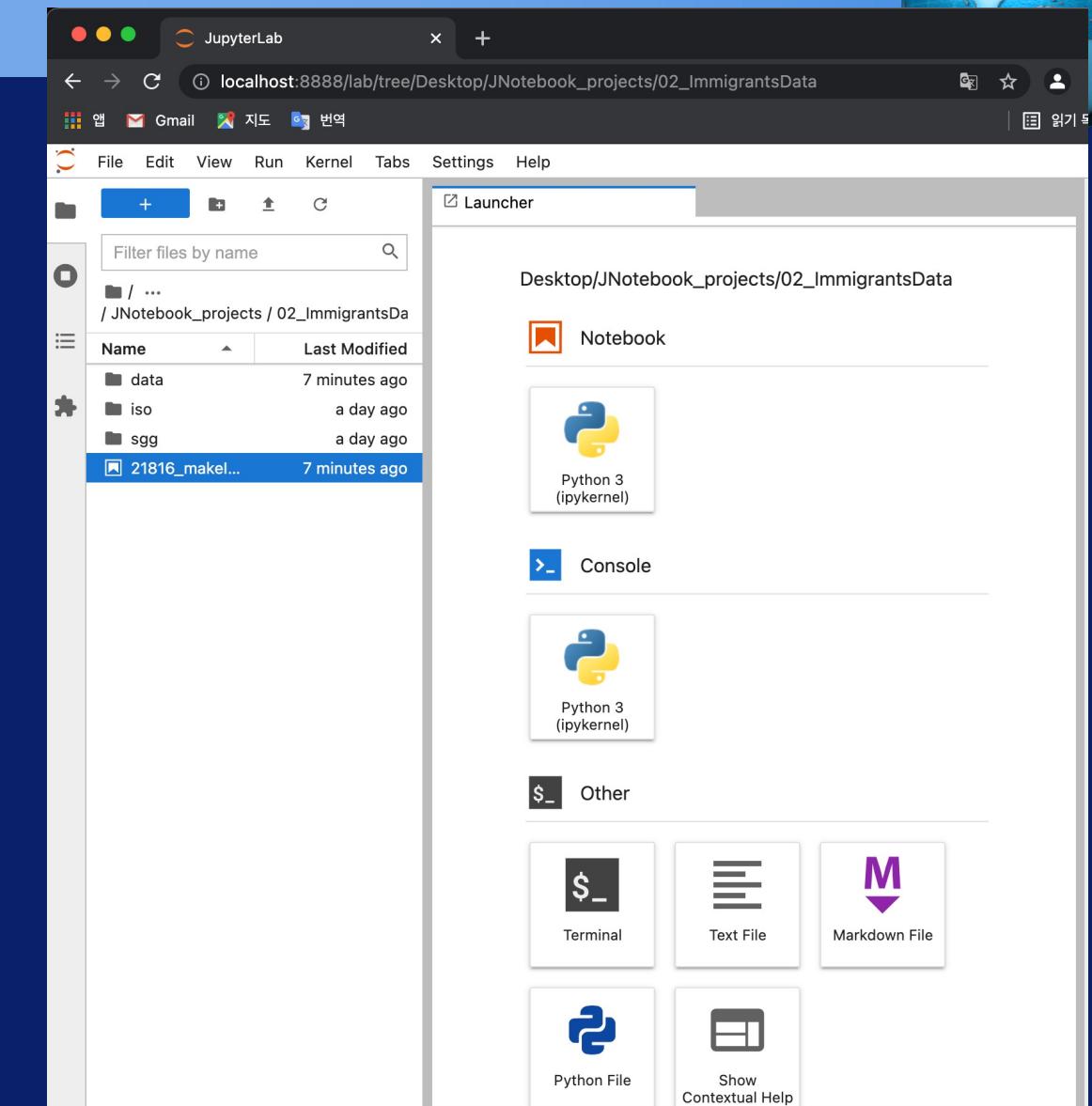
```
Shutdown this Jupyter server (y/[n])? y
[E 2021-09-12 00:19:08.381 ServerApp] Shutdown confirmed
[I 2021-09-12 00:19:08.389 ServerApp] Shutting down 2 extensions
[I 2021-09-12 00:19:08.389 ServerApp] Shutting down 3 kernels
[I 2021-09-12 00:19:08.390 ServerApp] Kernel shutdown: 27e71abf-ba86-410...
[I 2021-09-12 00:19:08.391 ServerApp] Kernel shutdown: df6f9199-4370-41e...
[I 2021-09-12 00:19:08.391 ServerApp] Kernel shutdown: e39517da-92a1-4f1...
[I 2021-09-12 00:19:08.610 ServerApp] Shutting down 2 terminals
[I 2021-09-12 00:19:08.681 ServerApp] EOF on FD 96; stopping reading
[I 2021-09-12 00:19:08.786 ServerApp] Terminal 1 closed
[Websocket closed]
[I 2021-09-12 00:19:08.813 ServerApp] EOF on FD 98; stopping reading
[I 2021-09-12 00:19:08.914 ServerApp] Terminal 2 closed
[Websocket closed]
gracenotebook@aninseocBookPro ~ %
```

# Basic manual for jupyterLab



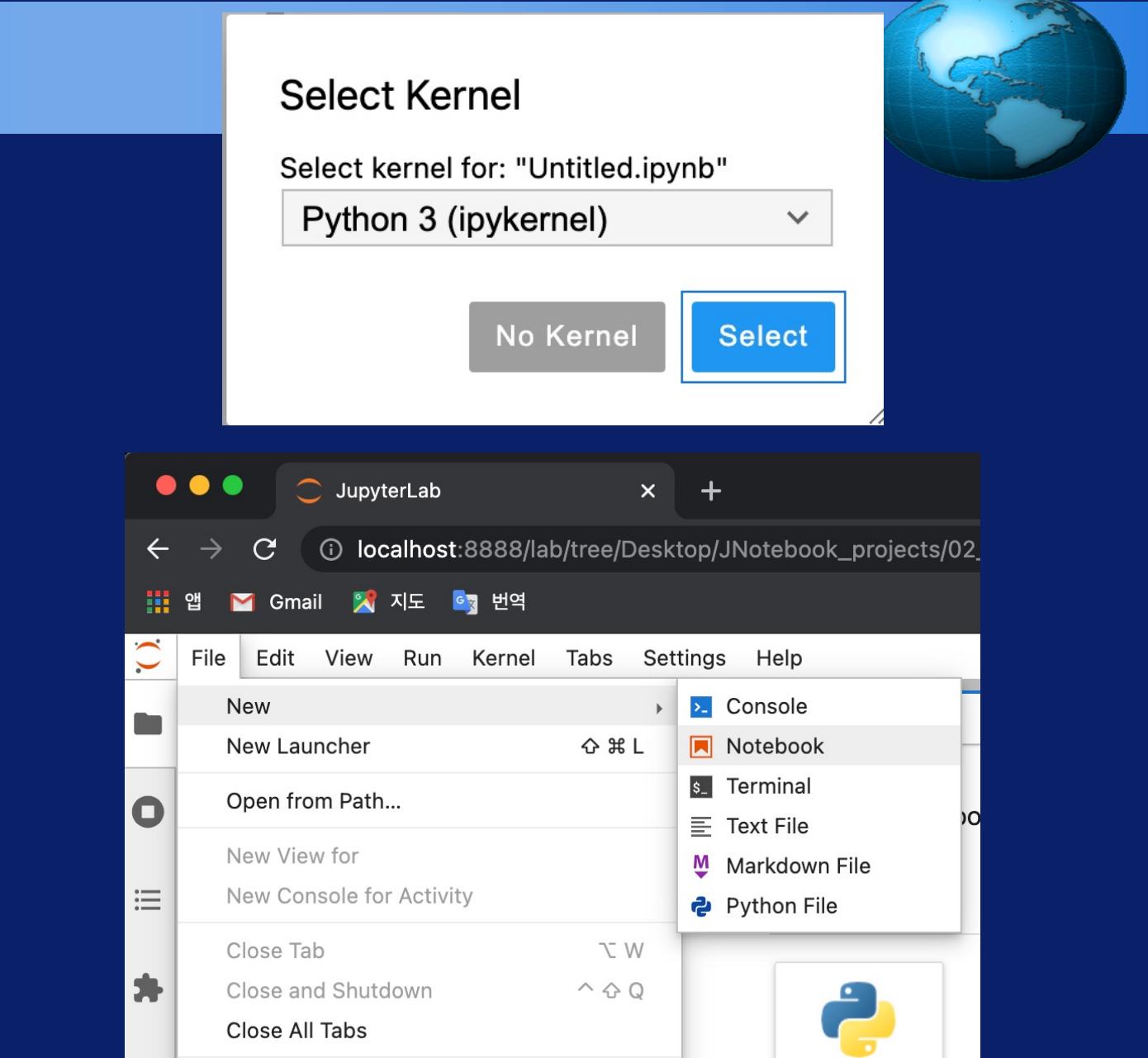
# 4.1 Main page

- This is the first screen that appears when jupyterLab is started.
- On the Launcher page, various types of coding environments provided by jupyterLab are presented.



## 4.2 Create New Notebook

- To Start, you must select an environment to work with.
- To take full advantage of JupyterLab, select a notebook environment.



## 4.3 Mode & Style of jupyter

- Command Mode used to edit cells, and the shortcut key is "esc".
- Edit mode is used to edit the contents of a cell, and the shortcut key is "enter".

```
[ ]: This is a command mode
[ ]: print('This is a edit mode~!')
## This is a Markdown style
This is a Markdown style
Esc + m
[ ]: Esc + y
```

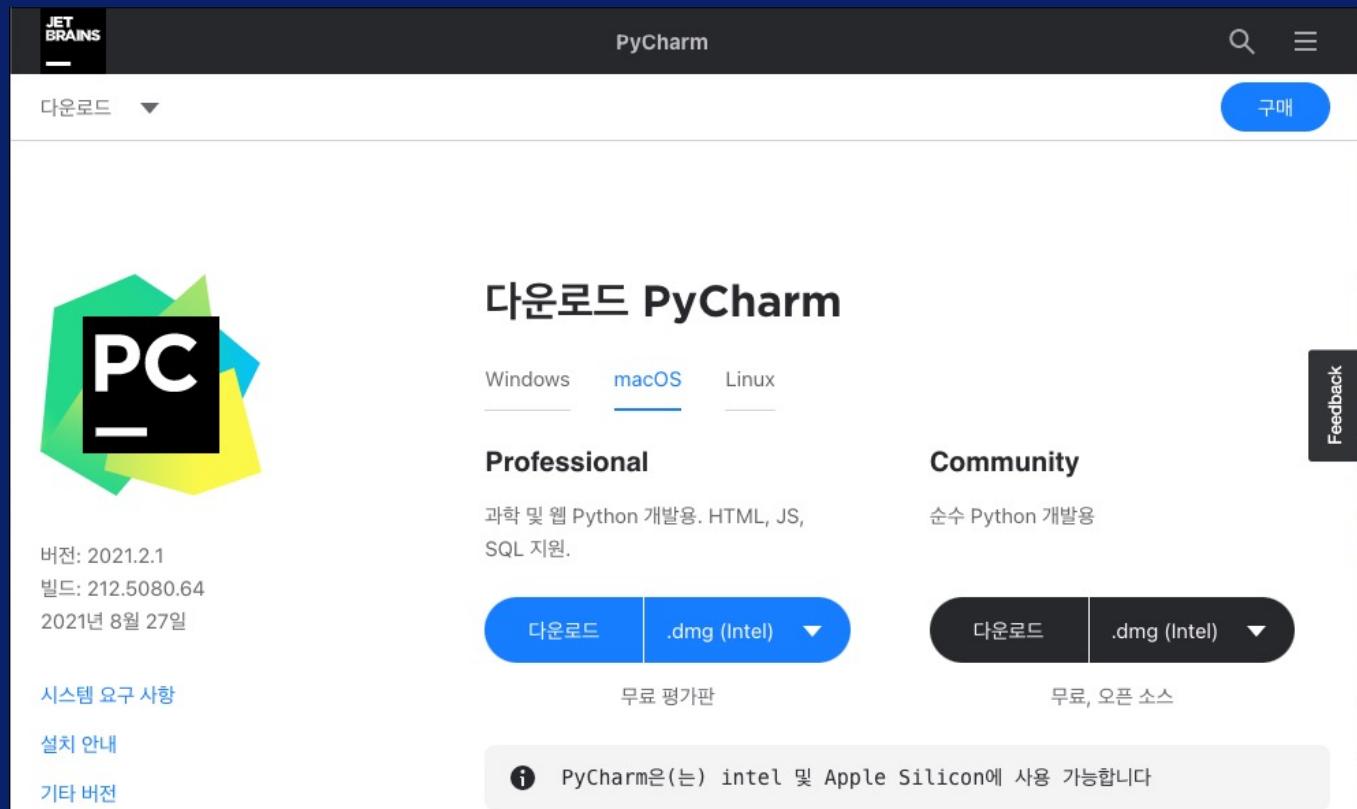
## 4.4 Useful keyboard shortcuts to know



- ❖ shift + enter : run the cell
- ❖ a in command mode: create a new cell above the current cell
- ❖ b in command mode: create a new cell below the current cell
- ❖ x : delete cell in command mode
- ❖ z : revert cell in command mode

# 5. Install PyCharm

1. Visit PyCharm's homepage and download the program suitable for your PC's OS.
2. Professional product is a paid version, and Community product is for free, so you should choose this.





## PyCharm 2021.1.3 (Community Edition)

Build #PC-211.7628.24, built on June 30, 2021

Runtime version: 11.0.11+9-b1341.60 aarch64

VM: OpenJDK 64-Bit Server VM by JetBrains s.r.o

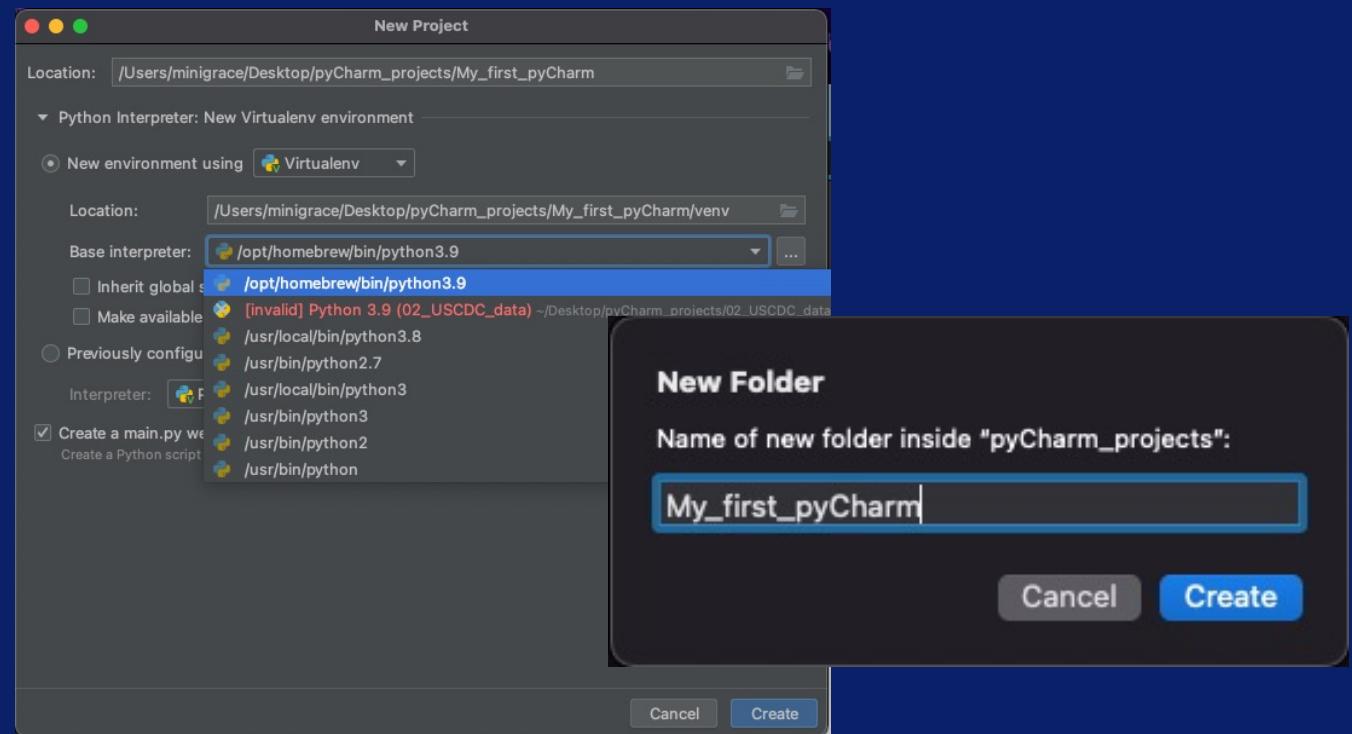
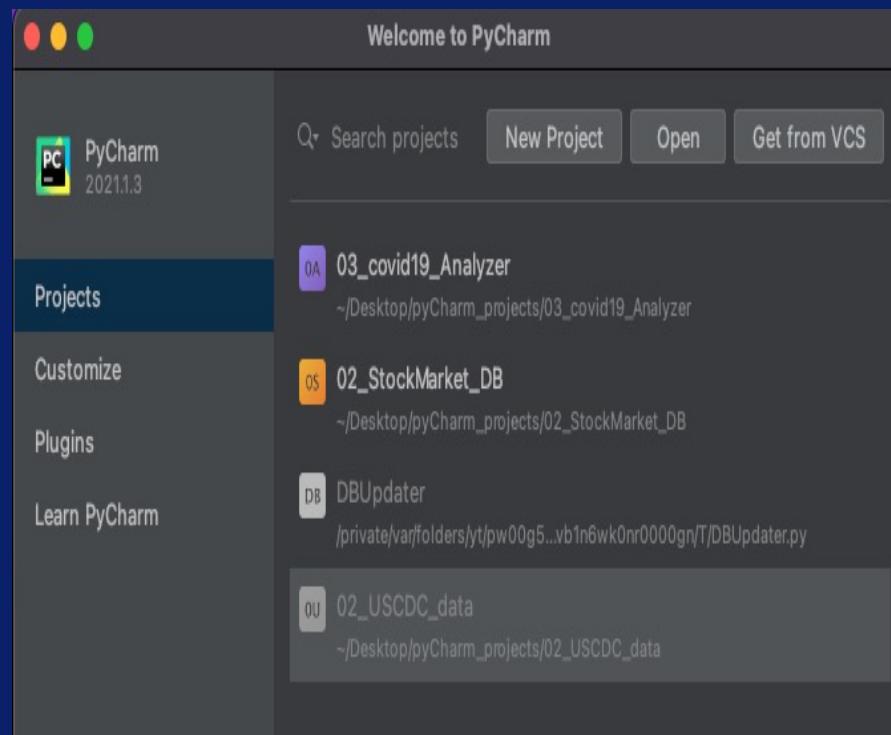
Powered by [open-source software](#)

A screenshot of the PyCharm IDE. The top navigation bar shows the title '0829\_covid19\_Analyzer.py'. The left sidebar shows the 'Project' structure with files like '03\_covid19\_Analyzer', 'corr\_data', 'result', '0829\_covid19\_Analyzer.py', 'External Libraries', and 'Scratches and Consoles'. The main code editor window contains Python code for analyzing COVID-19 data. Below the code editor is the 'Python Packages' tool window, which lists packages from the PyPI repository. The 'seaborn' package is selected, showing its version (0.11.2) and a brief description: 'Seaborn is a library for making statistical graphics in Python. It is built on top of matplotlib and closely integrated with pandas data structures.' A scrollable list of seaborn's functionality is also visible.

# Basic manual for PyCharm

# 5.1 Main page

- ❖ PyCharm is basically project-based.
- ❖ After creating a project folder in the desired location, create a new virtual environment.



# 5.2 Project page

- ❖ In the project, you just need to remember the 4 windows.
- ❖ Project main window is the default screen when starting a project.
- ❖ New → Python File

## Project main window

The screenshot shows the PyDev IDE's Project main window. The code editor displays the following Python script:

```
03_covid19_Analyzer > 0829_covid19_Analyzer.py
def _get_filepath_list_pair(_target_dir):
    target_dir = os.path.normpath(_target_dir)
    lstName = []

    for (path, dir, files) in os.walk(target_dir):
        for fname in files:
            fullname = path + "/" + fname
            lstName.append(str(fullname))
    return lstName

def search(dirname):
    LDir_raw = []
    filenames = os.listdir(dirname)
    for filename in filenames:
        full_filename = os.path.join(dirname, filename)
        LDir_raw.append(str(fullname))
    return LDir_raw

def _isTargetCountry(_country):
    isTarget = False
    for coun in lstOfCountry:
        if country == coun:
            isTarget = True
    return isTarget
```

## Terminal window

The screenshot shows the PyDev IDE's Terminal window with the following command history:

```
Terminal: Local x +
(venv) minigrace@mini-isahnui-Macmini 03_covid19_Analyzer % ls
0829_covid19_Analyzer.py      corr_data      result
(venv) minigrace@mini-isahnui-Macmini 03_covid19_Analyzer %
```

## Python packages window



## Python console window

The screenshot shows the PyDev IDE's Python Console window with the following session:

```
Python Console x
Type 'copyright', 'credits' or 'license' for more information
IPython 7.25.0 -- An enhanced Interactive Python. Type ? for help.
Pydev console: using IPython 7.25.0

Python 3.9.6 (default, Jun 28 2021, 19:24:41)
[Clang 12.0.5 (clang-1205.0.22.9)] on darwin
In[2]: print('This is IDLE console!')
This is IDLE console!

In[$]:
```



## 6. WHAT IS DBMS..?

- ❖ DBMS stands for Database Management System.
- ❖ MySQL is the most popular DBMS program.
- ❖ MySQL is open source and provides various APIs for several programming languages.



## 6.1 Advantages of MySQL

- ❖ It's **free to use** because it's under an open source license.
- ❖ It can be used on **various operating systems** and supports **several programming languages**.
- ❖ Even large data sets can be processed very **quickly and effectively**.

## 6.2 Install MySQL

- ❖ Visit the MySQL Homepage → Find the MySQL Community (GPL) version → Download the program suitable for your PC environment and install it in order.

**MySQL Enterprise Edition**  
MySQL Enterprise Edition includes the most comprehensive set of advanced features, management tools and technical support for MySQL.

[Learn More »](#)  
[Customer Download »](#)  
[Trial Download »](#)

---

**MySQL Cluster CGE**  
MySQL Cluster is a real-time open source transactional database designed for fast, always-on access to data under high throughput conditions.

[MySQL Cluster](#)  
[MySQL Cluster Manager](#)  
[Plus, everything in MySQL Enterprise Edition](#)

[Learn More »](#)  
[Customer Download » \(Select Patches & Updates Tab, Product Search\)](#)  
[Trial Download »](#)

---

[MySQL Community \(GPL\) Downloads »](#)

## MySQL Community Downloads

- [MySQL Yum Repository](#)
- [MySQL APT Repository](#)
- [MySQL SUSE Repository](#)
- [MySQL Community Server](#)
- [MySQL Cluster](#)
- [MySQL Router](#)
- [MySQL Shell](#)
- [MySQL Workbench](#)

**General Availability (GA) Releases** [Archives](#) [?](#)

### MySQL Community Server 8.0.26

Select Operating System: [macOS](#) [Looking for previous GA versions?](#)

Select OS Version: [All](#)

**!** Packages for Big Sur (11) on x86 are compatible with Catalina (10.15)

macOS 11 (ARM, 64-bit), DMG Archive	8.0.26	413.8M	<a href="#">Download</a>
(mysql-8.0.26-macos11-arm64.dmg)	MD5:	f052981e6c26e124d85f784d55fcc0ca	<a href="#">Signature</a>

macOS 11 (x86, 64-bit), DMG Archive	8.0.26	418.3M	<a href="#">Download</a>
(mysql-8.0.26-macos11-x86_64.dmg)	MD5:	aae4bb6572c1646ce92c32a8f326de5c	<a href="#">Signature</a>



## 6.3 create user in mysql

- ❖ Create password for root
  - mysql -uroot -p
  - alter user 'root'@'localhost' identified by 'your\_password';
- ❖ Create user
  - mysql -uroot -p
  - use mysql;
  - create user your\_ID;
  - create user your\_ID@localhost identified by 'your\_password';
  - create database your\_DB;
  - grant all privileges on your\_DB.\* to your\_ID@localhost identified by 'your\_password';



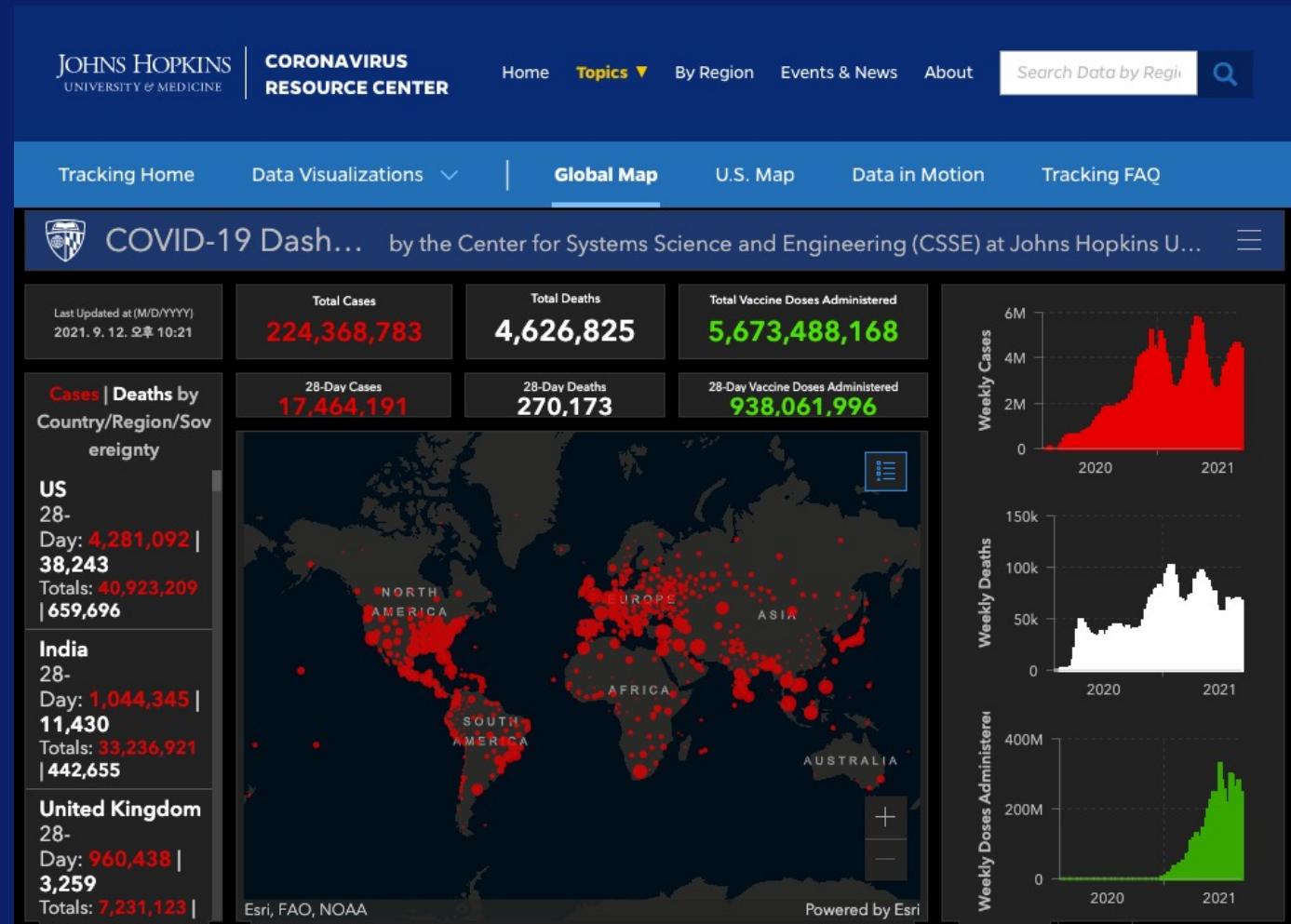
03

# COVID-19 Dashboard



# Preview: COVID-19 map from Johns Hopkins University

- ❖ <https://coronavirus.jhu.edu/map.html>
- ❖ This is a daily map showing the status of COVID-19 provided by Johns Hopkins University.
- ❖ They collect and publish the number of COVID-19 confirmed, deaths, and recovered patients by country by day.





# Assignment (practice)

1. Install the Python3 program on your PC.
2. Select and install either JupyterLab or PyCharm on your PC.
3. Install the MySQL program suitable for your OS.
4. Take a look at the COVID-19 map page and think about how you can get the information you want from this page in the form of data.
5. Short presentations: Mr Hizkia Gultom from KIST, Mr Tergel Munkhbat from KISTI , Ms 신평화 from KITECH .

# THANK YOU

