Colab / Kaggle Tutorial

TA: 李冠儀、許景淯
2025.02.21
ntu-ml-2025-spring-ta@googlegroups.com

Outline

- Colab Tutorial (ref this Colab)
- Kaggle Tutorial (ref this Kaggle Notebook)

Outline

- Colab Tutorial
  - Introduction
  - Getting Started
  - Changing Runtime
  - Executing Code Block
  - Check GPU type
  - File Manipulation
  - Mounting Google Drive
  - Saving Notebook
  - Problems You May Encounter... (very important)
  - References
- Kaggle Tutorial

Introduction

What is Colab?

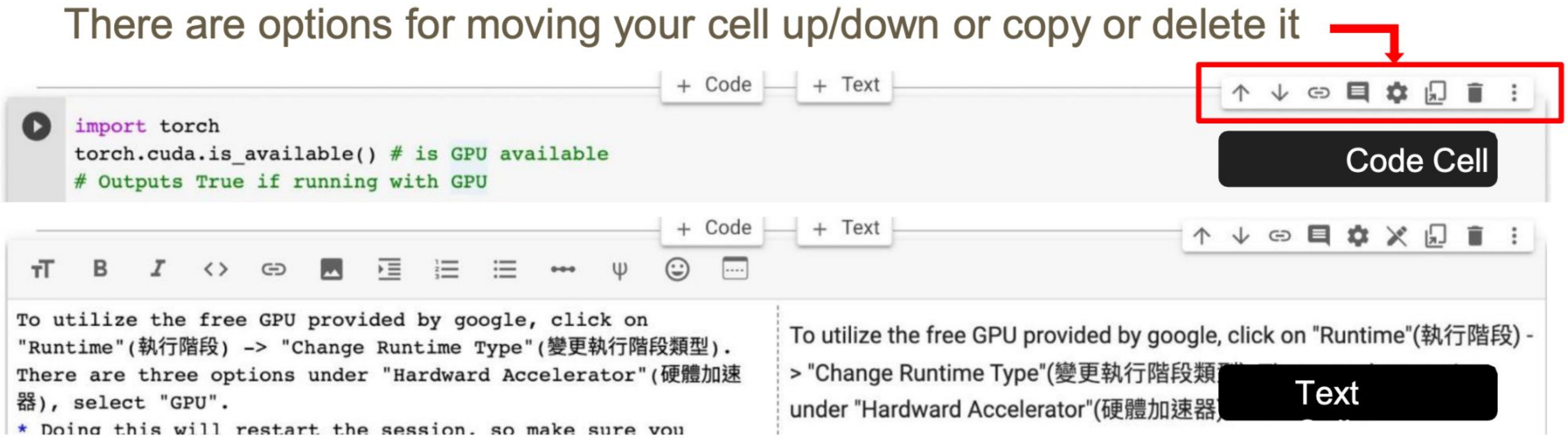
Colab, or "Colaboratory", allows you to write and execute Python in your browser, with

- Zero configuration required
- Free access to GPUs
- Easy sharing

Getting Started

Creating a new cell

You can create a new code cell by clicking on +Code, clicking on +Text generates a text cell.



There are options for moving your cell up/down or copy or delete it

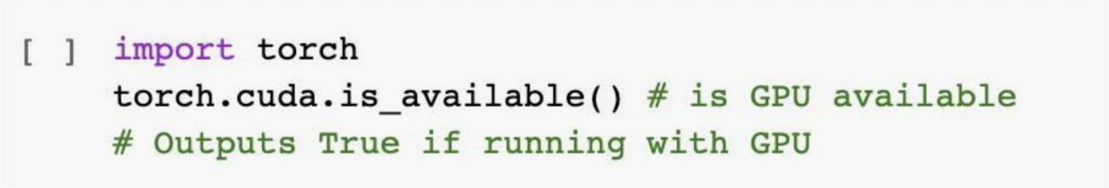
```
import torch
torch.cuda.is_available() # is GPU available
# Outputs True if running with GPU
```

```
To utilize the free GPU provided by google, click on
"Runtime"(執行階段) -> "Change Runtime Type"(變更執行階段類型).
There are three options under "Hardward Accelerator"(硬體加速
器), select "GPU".
* Doing this will restart the session, so make sure you
```

To utilize the free GPU provided by google, click on "Runtime"(執行階段) -> "Change Runtime Type"(變更執行階段類 under "Hardward Accelerator"(硬體加速器)

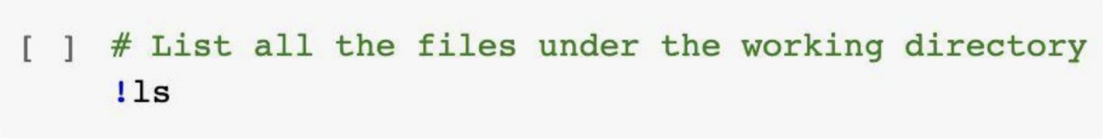
Getting Started

You can type python code in the code cell, or use a leading exclamation mark ! to change the code cell to treating the input as a shell script.



```
[ ]  import torch
     torch.cuda.is_available() # is GPU available
     # Outputs True if running with GPU
```

→ **python**



```
[ ]  # List all the files under the working directory
     !ls
```

→ **shell script**

Getting Started

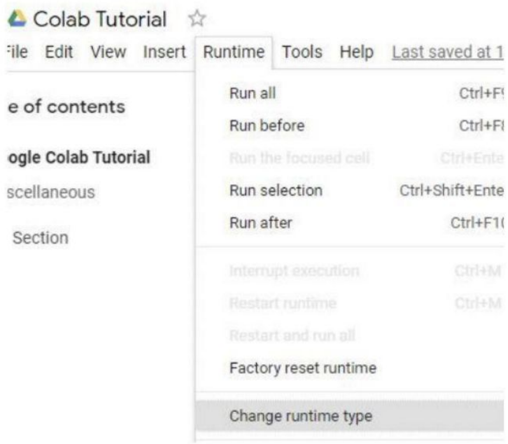
Using an exclamation mark (!) starts a new shell, does the operations, and then kills that shell, while percentage (%) affects the process associated with the notebook, and it is called a magic command.

Use % instead of ! for cd (change directory) command

other magic commands are listed here

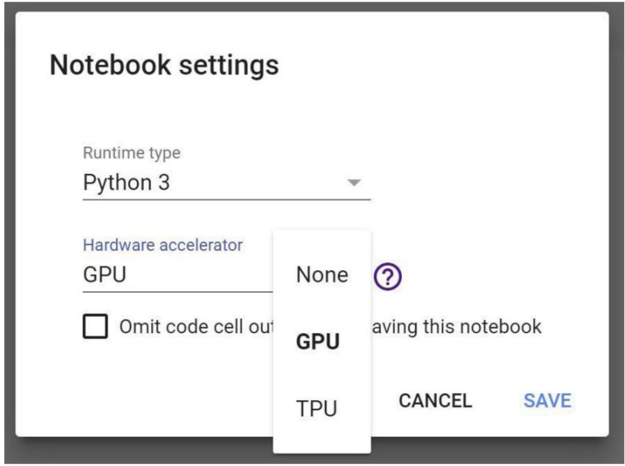
Changing Runtime

To utilize the free GPU provided by google, click on "Runtime"(執行階段) → "Change Runtime Type"(變更執行階段類型).



select "GPU" for "Hardware Accelerator"(硬體加速器)

Doing this will restart the session, so make sure you change to the desired runtime before executing any code.


Notebook settings

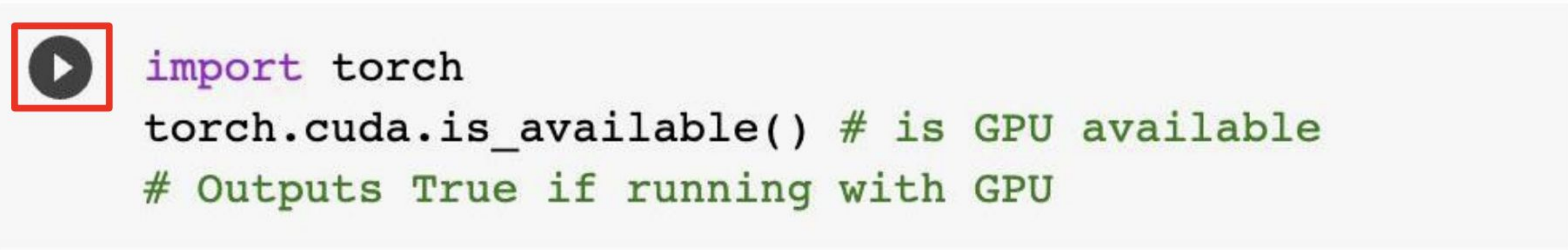
Runtime type
Python 3

Hardware accelerator
GPU

□ Omit code cell out aving this notebook

Executing Code Block

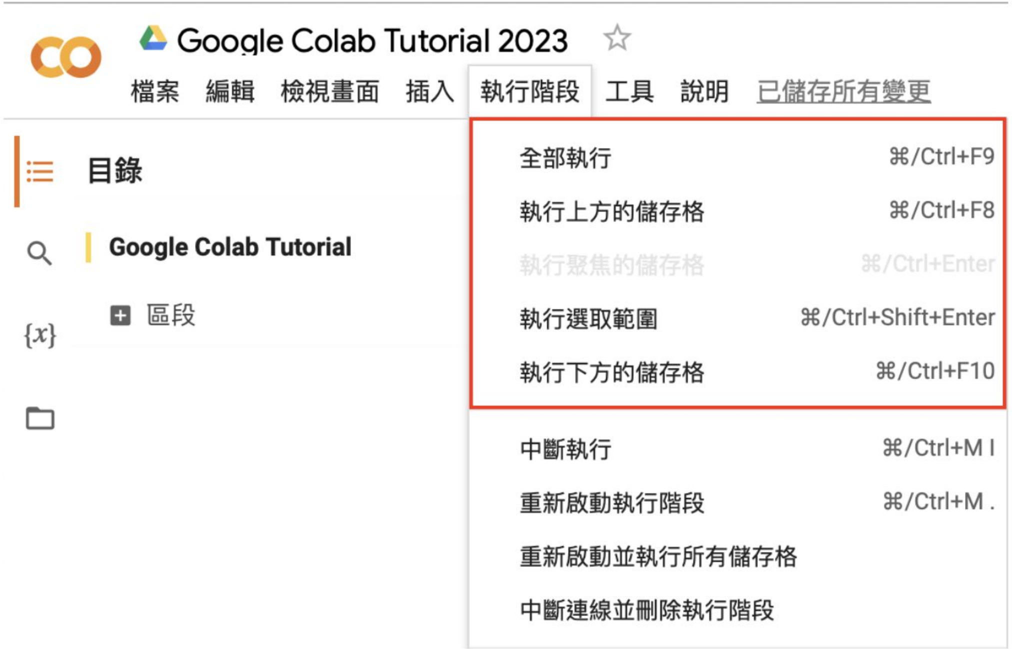
Click on the play button to execute a specific code cell



```python
import torch
torch.cuda.is_available() # is GPU available
# Outputs True if running with GPU
```

Executing Code Block

Other options to run your code



Google Colab Tutorial 2023
檔案 編輯 檢視畫面 插入 執行階段 工具 說明 已儲存所有變更
目錄
Google Colab Tutorial
區段
全部執行 ⌘/Ctrl+F9
執行上方的儲存格 ⌘/Ctrl+F8
執行聚焦的儲存格 ⌘/Ctrl+Enter
執行選取範圍 ⌘/Ctrl+Shift+Enter
執行下方的儲存格 ⌘/Ctrl+F10
中斷執行 ⌘/Ctrl+M I
重新啟動執行階段 ⌘/Ctrl+M .
重新啟動並執行所有儲存格
中斷連線並刪除執行階段

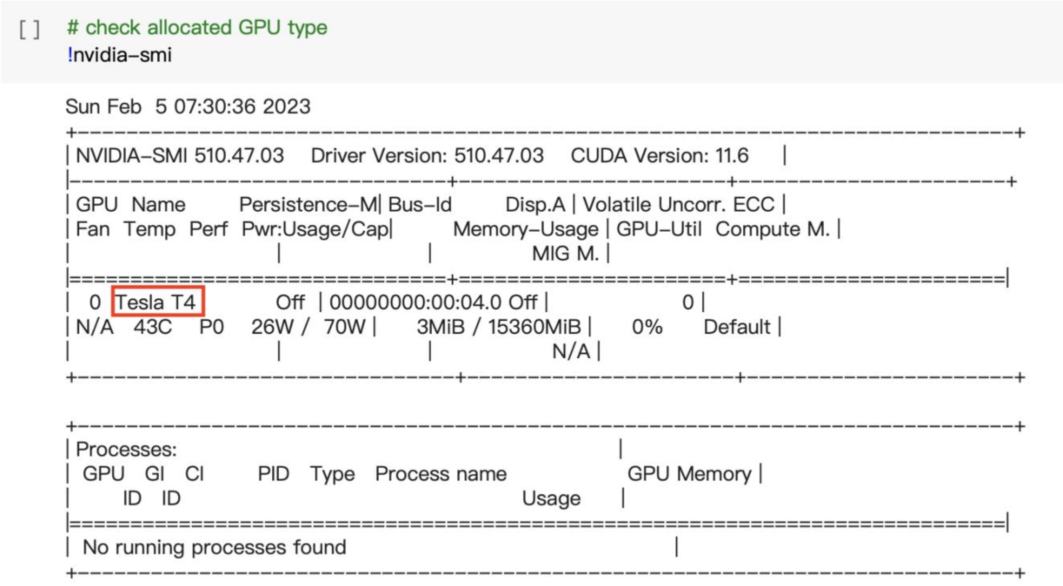
Check GPU Type

Use the command **nvidia-smi** to check the allocated GPU type

Available GPUs:

T4 > K80

(but most of the time you get K80 Using the free Colab)



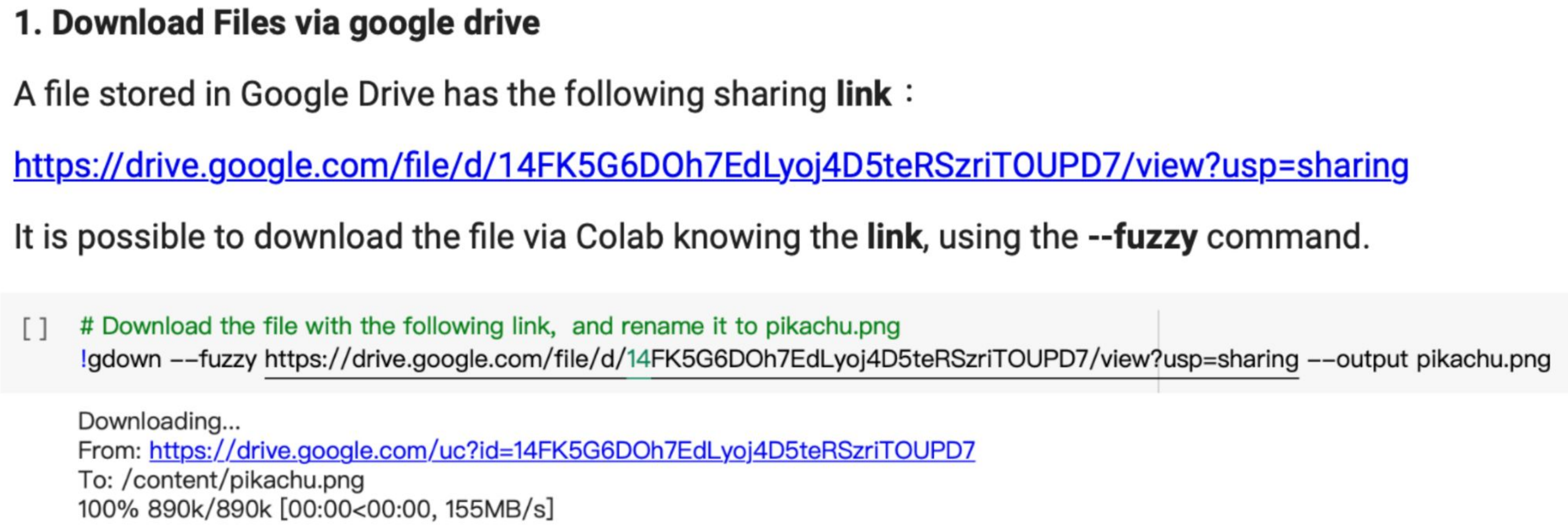
```
# check allocated GPU type
!nvidia-smi
```

```
Sun Feb  5 07:30:36 2023
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 510.47.03    Driver Version: 510.47.03    CUDA Version: 11.6     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla T4            Off  | 00000000:00:04.0 Off |                    0 |
| N/A   43C    P0    26W /  70W |      3MiB / 15360MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

File Manipulation

Download files via Google Drive



1. Download Files via google drive

A file stored in Google Drive has the following sharing **link** ：

https://drive.google.com/file/d/14FK5G6DOh7EdLyoj4D5teRSzriTOUPD7/view?usp=sharing

It is possible to download the file via Colab knowing the **link**, using the **--fuzzy** command.

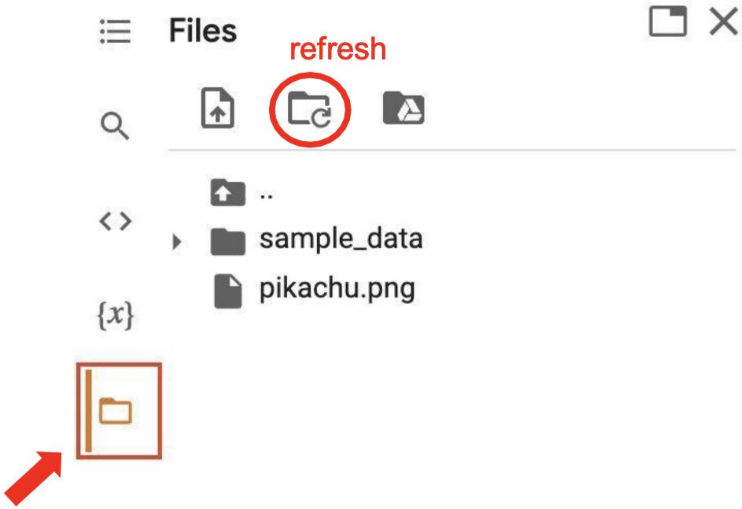
```
[ ]  # Download the file with the following link,  and rename it to pikachu.png
     !gdown --fuzzy https://drive.google.com/file/d/14FK5G6DOh7EdLyoj4D5teRSzriTOUPD7/view?usp=sharing --output pikachu.png
```

```
Downloading...
From: https://drive.google.com/uc?id=14FK5G6DOh7EdLyoj4D5teRSzriTOUPD7
To: /content/pikachu.png
100% 890k/890k [00:00<00:00, 155MB/s]
```

File Manipulation

File Structure

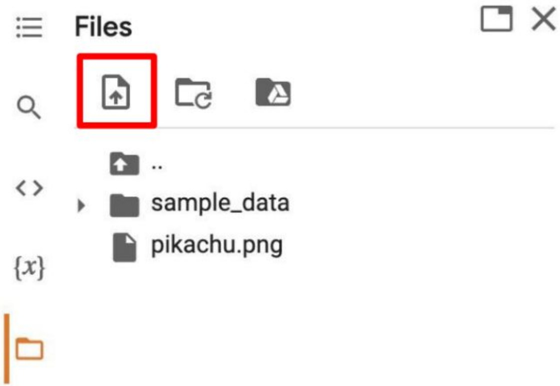
- You may click on the folder icon on the left to view your current files
- After downloading files, if the files are not immediately shown, click the refresh button
- Files are temporarily stored, and will be removed once you end your session.



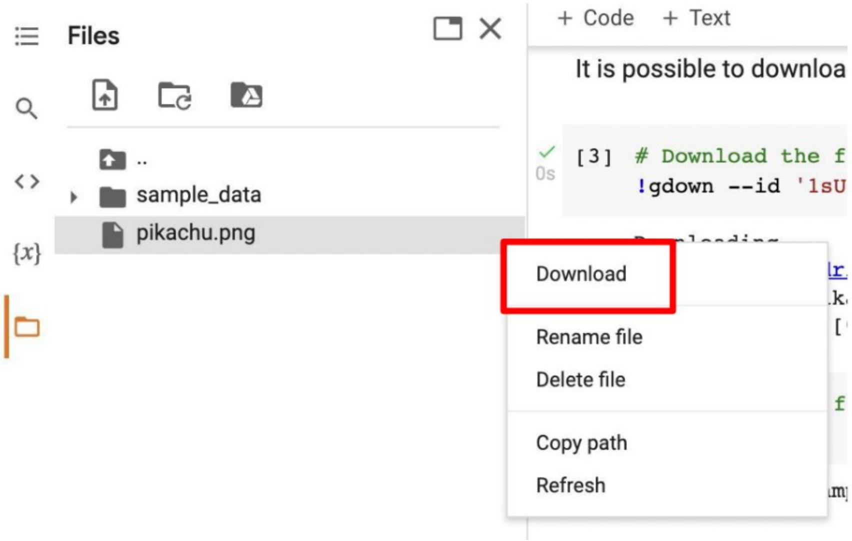
File Manipulation

Upload and Download Files

Click the upload icon to upload local files to your session



click ⋮ to download files to your local



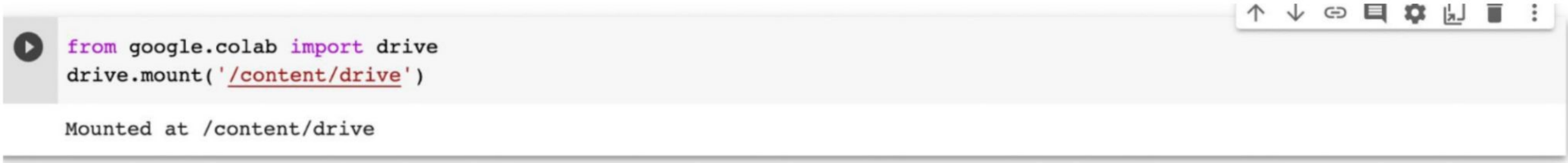
It is possible to downloa

```
0s [3] # Download the f
       !gdown --id '1sU
```

Mounting Google Drive

If you don't want to download the data every time you start a new session, or you want some files to be saved permantly,

you can mount your own google drive to colab and directly download/save the data to your google drive

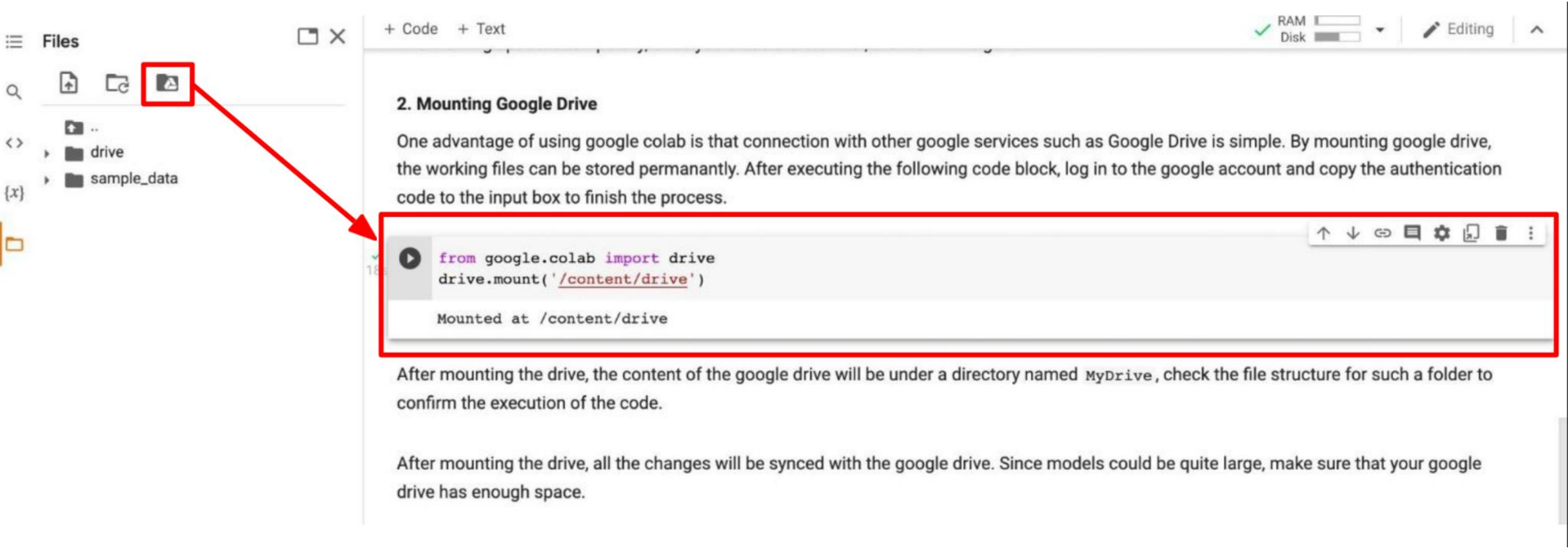


```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

Mounting Google Drive

Click on the Google Drive icon, the **Mount Drive** code block will be generated



2. Mounting Google Drive

One advantage of using google colab is that connection with other google services such as Google Drive is simple. By mounting google drive, the working files can be stored permanantly. After executing the following code block, log in to the google account and copy the authentication code to the input box to finish the process.

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

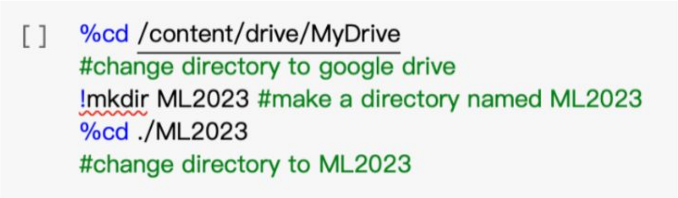
After mounting the drive, the content of the google drive will be under a directory named `MyDrive`, check the file structure for such a folder to confirm the execution of the code.

After mounting the drive, all the changes will be synced with the google drive. Since models could be quite large, make sure that your google drive has enough space.

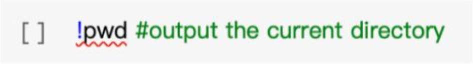
Mounting Google Drive

Execute the following three code blocks in order

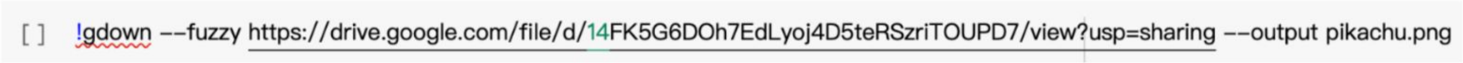
This will download the image to your google drive, and you can access it later



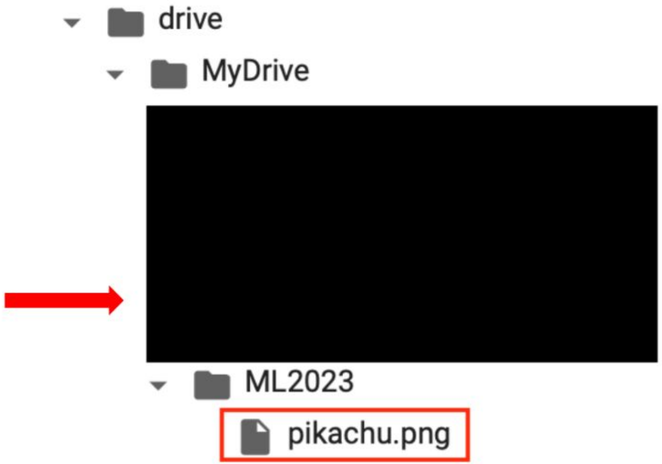
```
[ ]  %cd /content/drive/MyDrive
     #change directory to google drive
     !mkdir ML2023 #make a directory named ML2023
     %cd ./ML2023
     #change directory to ML2023
```



```
[ ]  !pwd #output the current directory
```



```
[ ]  !gdown --fuzzy https://drive.google.com/file/d/14FK5G6DOh7EdLyoj4D5teRSzriTOUPD7/view?usp=sharing --output pikachu.png
```



[redacted]

Saving Notebook

- Download the .ipynb file to your local device

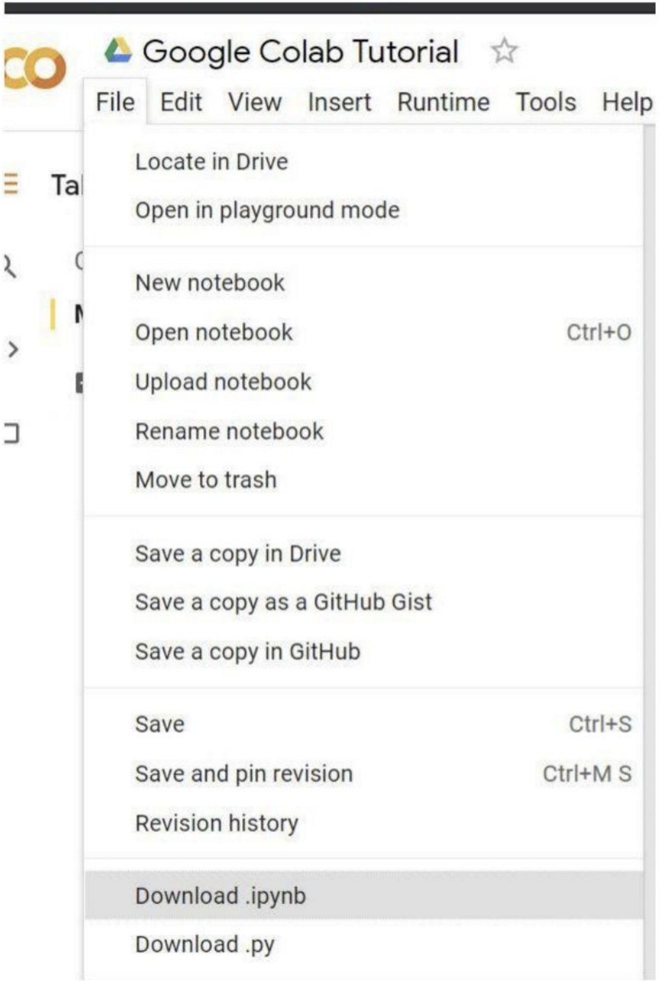
(File > Download .ipynb)

- Save the colab notebook to your google drive

(File > Save a copy in Drive).

- Convert .ipynb to .py and download

(File > Download .py)



Problems You May Encounter...

- Colab will **automatically disconnect** if idle timeout(90 min., sometimes varying) or when your screen goes black
  → solution: keep your screen on or try using javascript
- GPU usage is **not unlimited** ! (your account will be stopped for a period if you reached the max gpu usage 12 hrs)
  *** The cooldown period before you can connect to another GPU will extend from hours to days to weeks depending on your usage**
  → solution: open another account
- Best solution:
  - buy colab pro
  - use your own resource
  - Try kaggle

Reminder: TAs are not required to help you solve environment problems

Reference

- https://colab.research.google.com/drive/1tx06i2hlOwX3wK6ynSL_Bmjjtq2BGBAv
- https://research.google.com/colaboratory/faq.html

Outline

- Colab Tutorial
- Kaggle Tutorial
    - Introduction
    - Getting Started
    - Changing Runtime
    - Executing Code Block
    - Executing Whole Notebook
    - Check GPU type
    - Upload file / datasets
    - File Manipulation
    - Saving Notebook
    - Useful Linux Commands
    - Problems You May Encounter... (very important)

Introduction

What is Kaggle?

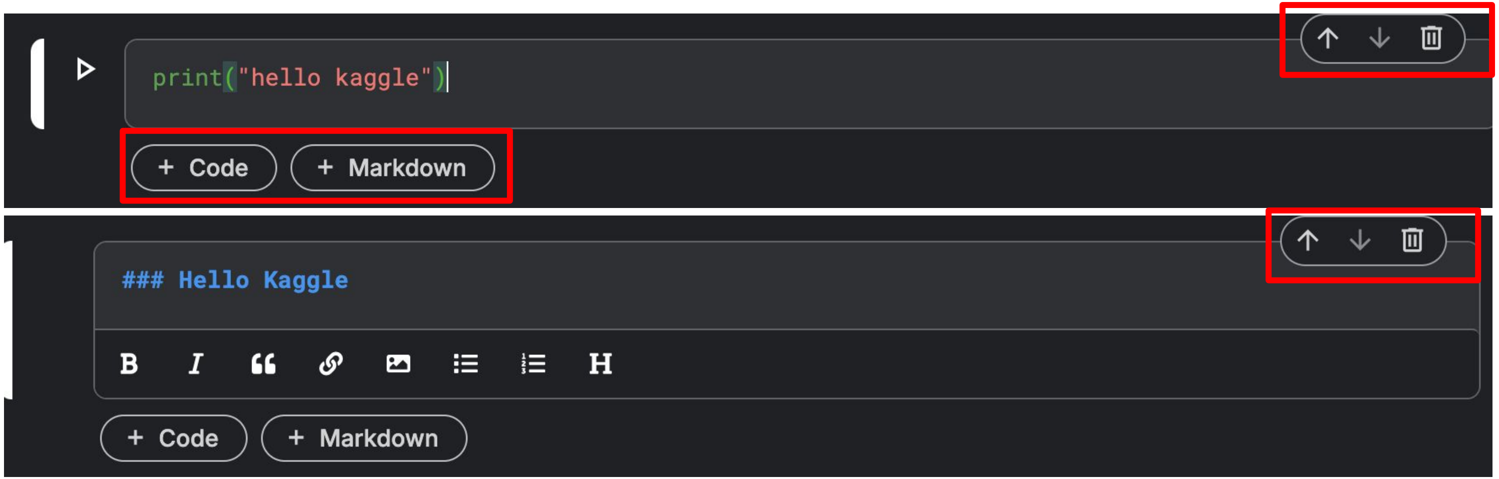
Kaggle also allows you to write and execute Python online, with

- Zero configuration required
- Free access to GPUs (30 hours / week)
- Easy sharing
- Stronger GPU resources
- Upload dataset, hold competitions……

Getting Started

Creating a new cell

You can create a new cell by clicking on +Code, clicking on +Markdown generates a markdown cell.



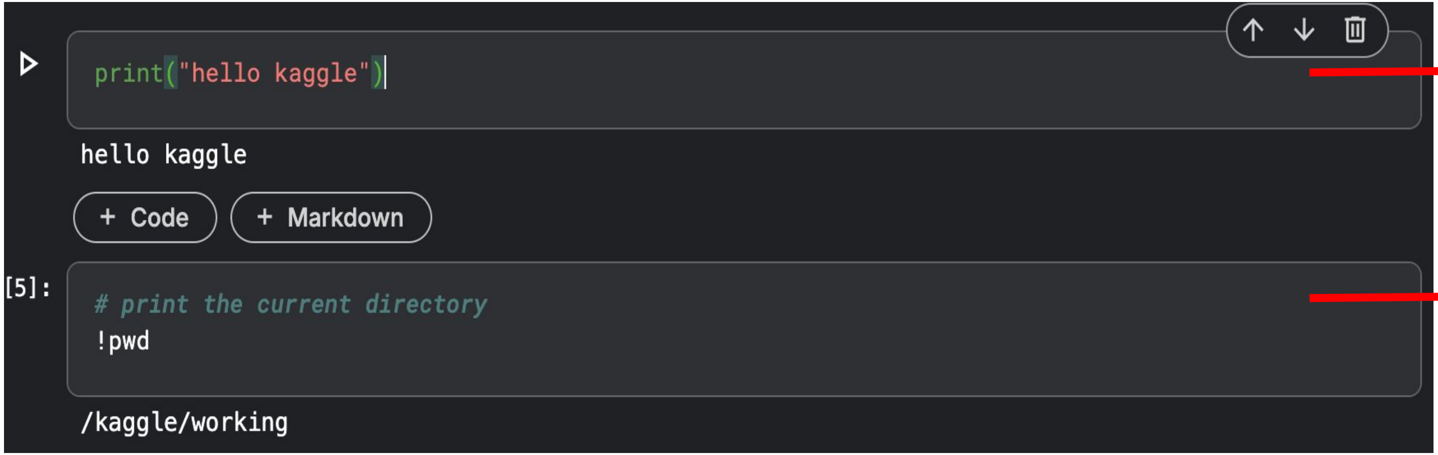
```
print("hello kaggle")
```

```
### Hello Kaggle
```

You can move the cell up / down, or remove the cell if needed.

Getting Started

You can type python code in the code cell, or use a leading exclamation mark ! to change the code cell to treating the input as a shell script.



```
print("hello kaggle")
```

```
hello kaggle
```

```
# print the current directory
!pwd
```

```
/kaggle/working
```

Python

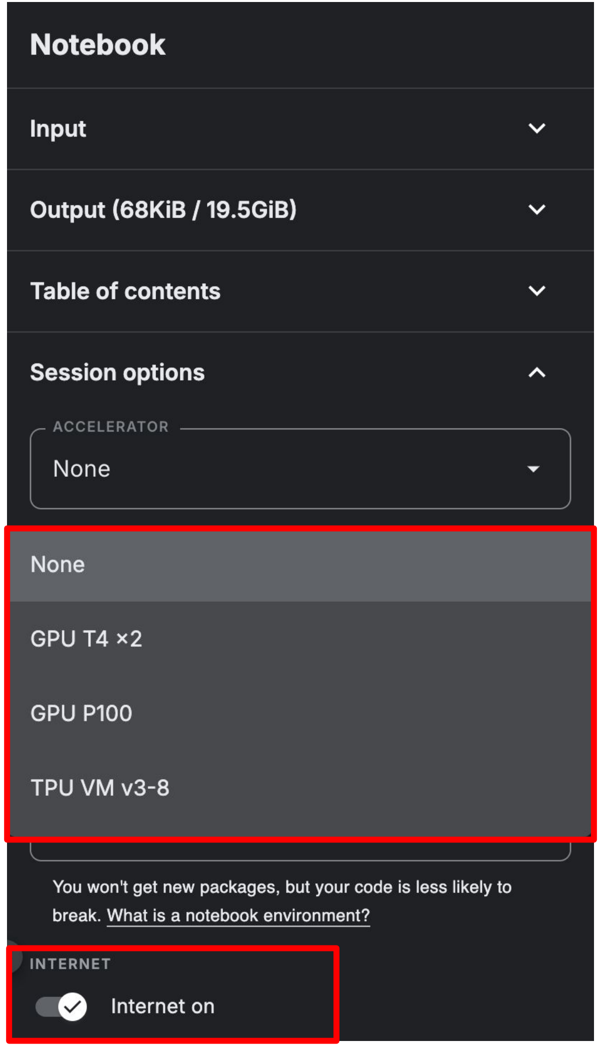
Shell script

Changing Runtime

To utilize the free GPU provided by kaggle, click on "Session options" → "Accelerator" → select the "GPU"

Doing this will restart the session, so make sure you change to the desired runtime before executing any code.

Make sure the Internet is toggled on in order to access to online resource



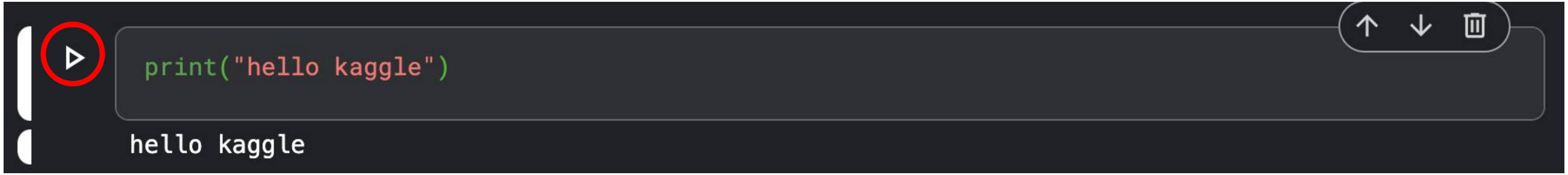
ACCELERATOR
None

You won't get new packages, but your code is less likely to break. What is a notebook environment?

INTERNET
Internet on

Executing Code Block

Click on the play button to execute a specific code cell
(The log will be printed below)

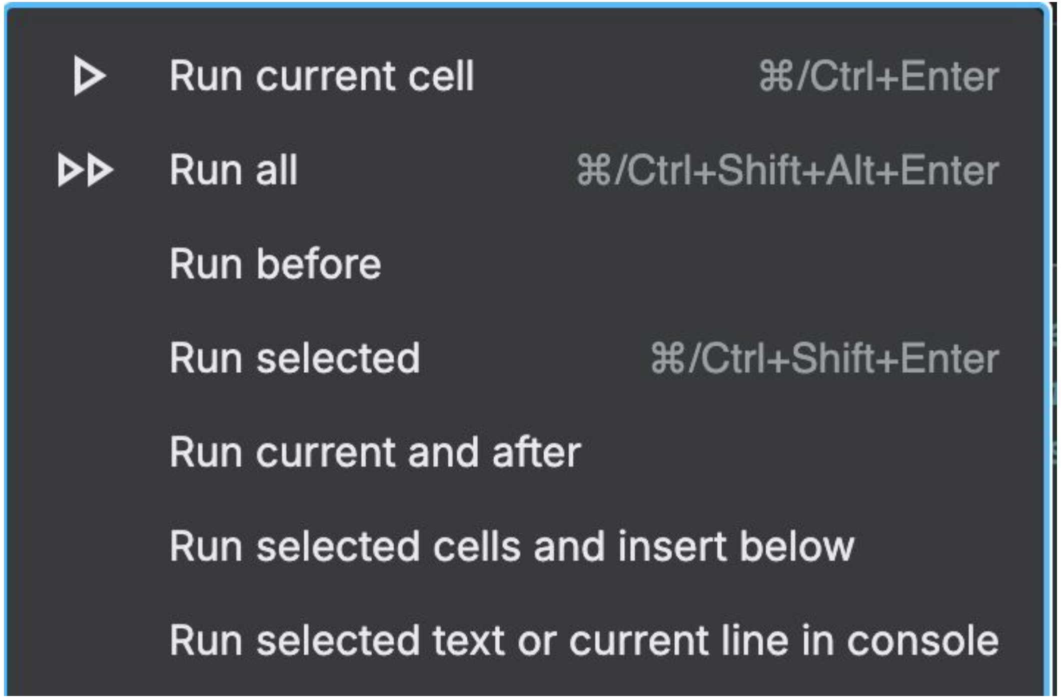


```
print("hello kaggle")
```

```
hello kaggle
```

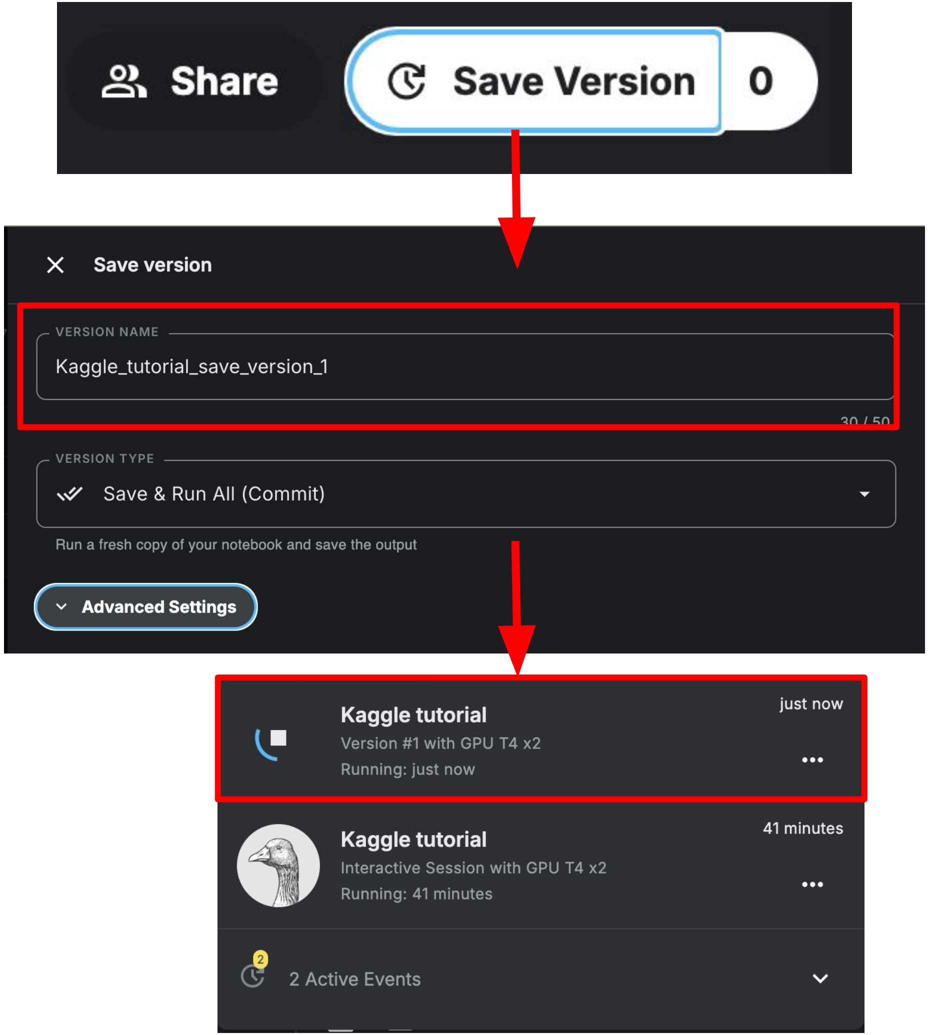
Executing Code Block

Other options to run your code


▷ Run current cell ⌘/Ctrl+Enter
▷▷ Run all ⌘/Ctrl+Shift+Alt+Enter
Run before
Run selected ⌘/Ctrl+Shift+Enter
Run current and after
Run selected cells and insert below
Run selected text or current line in console

Executing Whole Notebook

- save version
  - execute the whole notebook **without worrying** about connection loss
  - Run the notebook on online platform
  - At most 2 active events



Save version

VERSION NAME
Kaggle_tutorial_save_version_1
30 / 50

VERSION TYPE
Save & Run All (Commit)

Run a fresh copy of your notebook and save the output

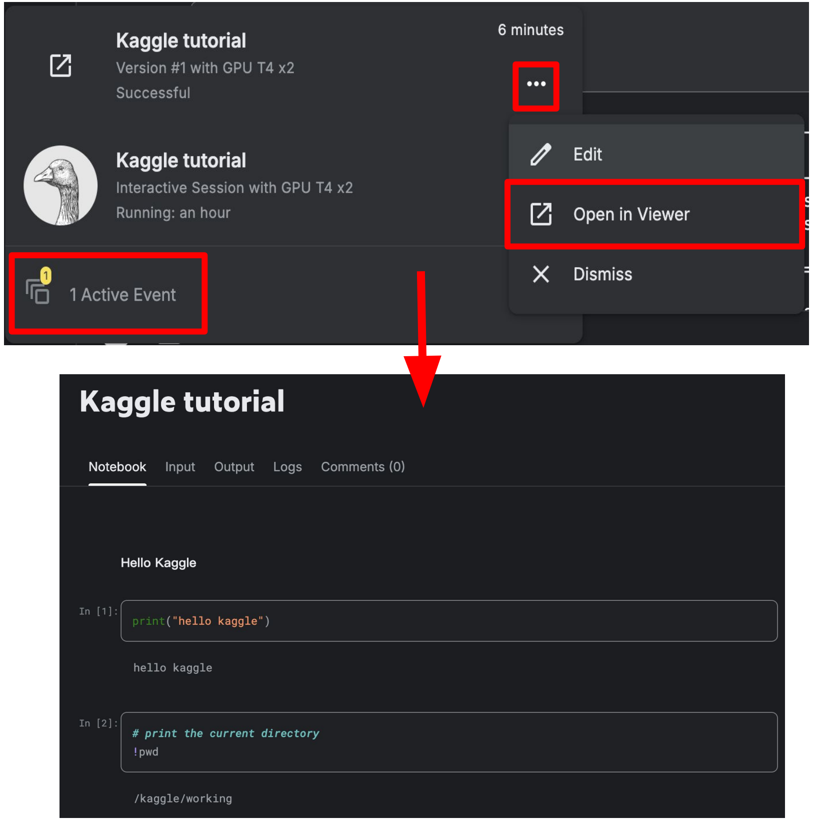
Kaggle tutorial
Version #1 with GPU T4 x2
Running: just now
just now

Kaggle tutorial
Interactive Session with GPU T4 x2
Running: 41 minutes
41 minutes

2 Active Events

Executing Whole Notebook

- See the execution result
  - Press the icon in left-bottom
    → Open in Viewer



Kaggle tutorial
Version #1 with GPU T4 x2
Successful

6 minutes

Kaggle tutorial
Interactive Session with GPU T4 x2
Running: an hour

Kaggle tutorial

Hello Kaggle

In [1]:

```
print("hello kaggle")
```

```
hello kaggle
```

In [2]:

```
# print the current directory
!pwd
```

```
/kaggle/working
```

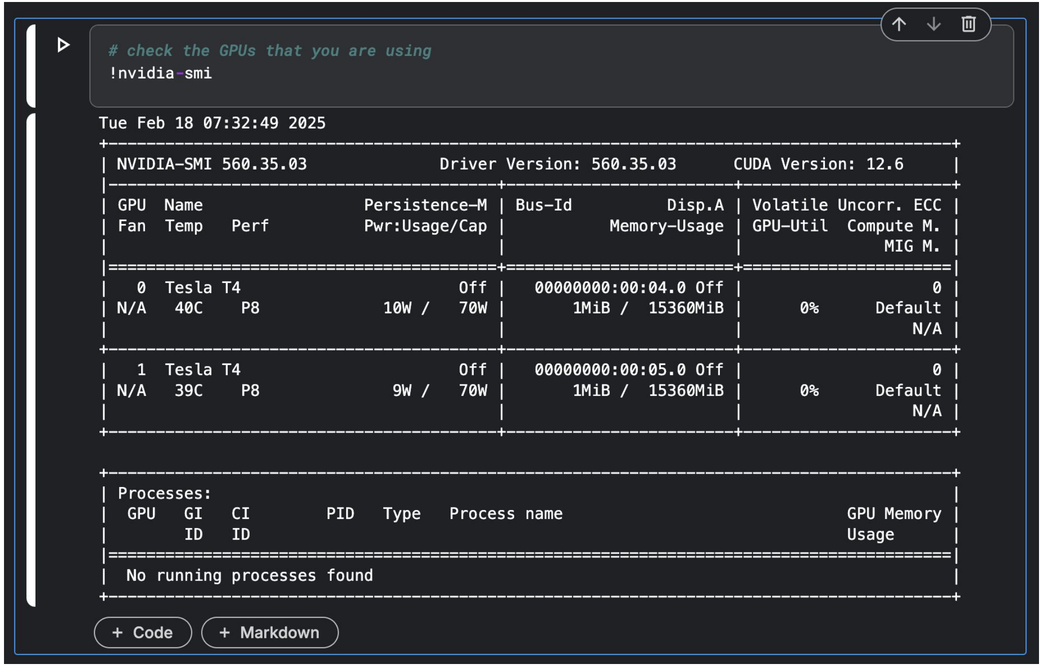
Check GPU Type

Use the command **nvidia-smi** to check the allocated GPU type

Available GPUs:

P100 - better speed

T4 x 2 - 2 GPUs at a time



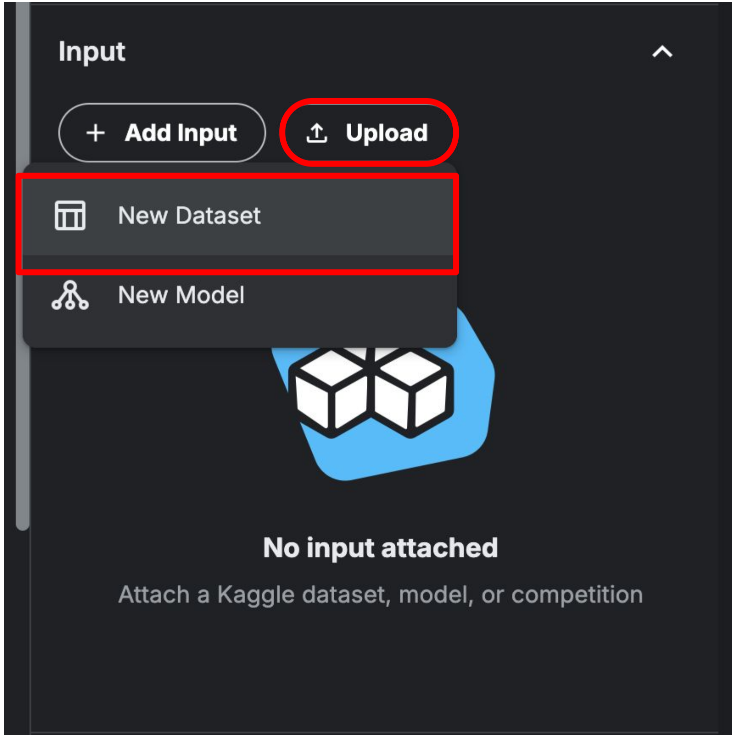
```
# check the GPUs that you are using
!nvidia-smi
```

```
Tue Feb 18 07:32:49 2025
+-----------------------------------------------------------------------------------------+
| NVIDIA-SMI 560.35.03              Driver Version: 560.35.03      CUDA Version: 12.6     |
|-----------------------------------------+------------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage | GPU-Util  Compute M. |
|                                         |                        |               MIG M. |
|=========================================+========================+======================|
|   0  Tesla T4                       Off |   00000000:00:04.0 Off |                    0 |
| N/A   40C    P8             10W /   70W |       1MiB /  15360MiB |      0%      Default |
|                                         |                        |                  N/A |
+-----------------------------------------+------------------------+----------------------+
|   1  Tesla T4                       Off |   00000000:00:05.0 Off |                    0 |
| N/A   39C    P8              9W /   70W |       1MiB /  15360MiB |      0%      Default |
|                                         |                        |                  N/A |
+-----------------------------------------+------------------------+----------------------+

+-----------------------------------------------------------------------------------------+
| Processes:                                                                              |
|  GPU   GI   CI        PID   Type   Process name                              GPU Memory |
|        ID   ID                                                               Usage      |
|=========================================================================================|
|  No running processes found                                                             |
+-----------------------------------------------------------------------------------------+
```

Upload Files / dataset

- Upload→New dataset
- set the dataset name, visibility…
- show the file structure in the Input area

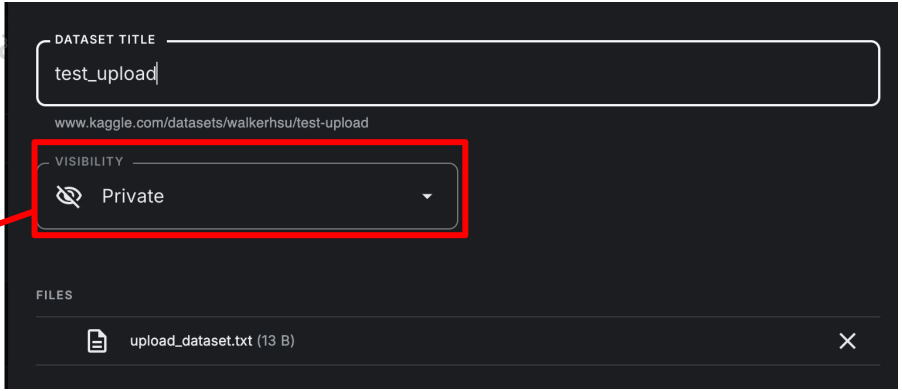


No input attached

Attach a Kaggle dataset, model, or competition

Upload Files / dataset

- Upload→New dataset
- set the dataset name, visibility…
- show the file structure in the Input are



DATASET TITLE

test_upload

www.kaggle.com/datasets/walkerhsu/test-upload

VISIBILITY

Private

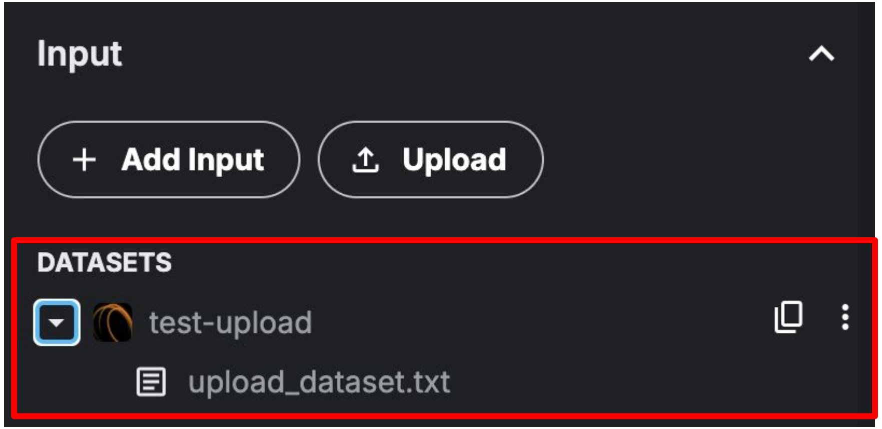
FILES

upload_dataset.txt (13 B)

You can set the visibility depending on who can see the dataset uploaded

Upload Files / dataset

- Upload→New dataset
- set the dataset name, visibility...
- show the file structure in the Input area



File Manipulation

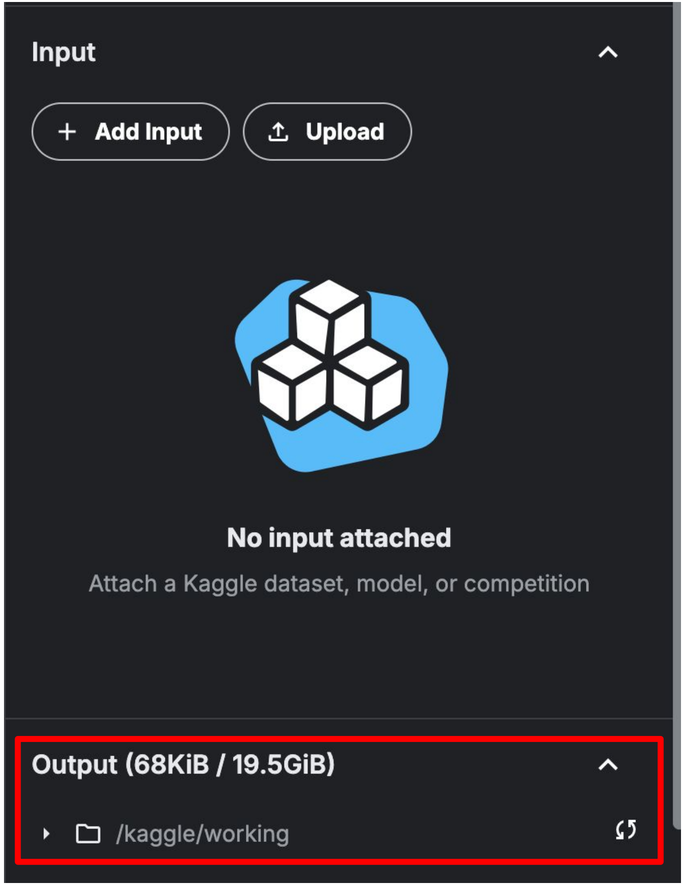
File Structure

- You may click on the output icon on the right to view your current files

- After any files are generated / downloaded, if the files are not immediately shown, click the refresh button

- Files are temporarily stored, and will be

removed once you end your session.



No input attached

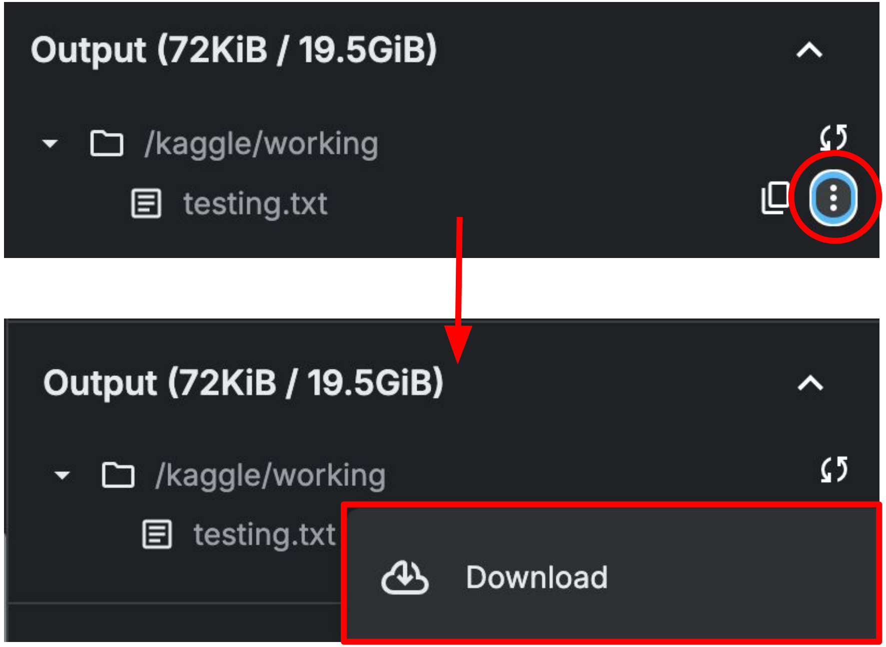
Attach a Kaggle dataset, model, or competition

Output (68KiB / 19.5GiB)

/kaggle/working

File Manipulation

Download the files by clicking



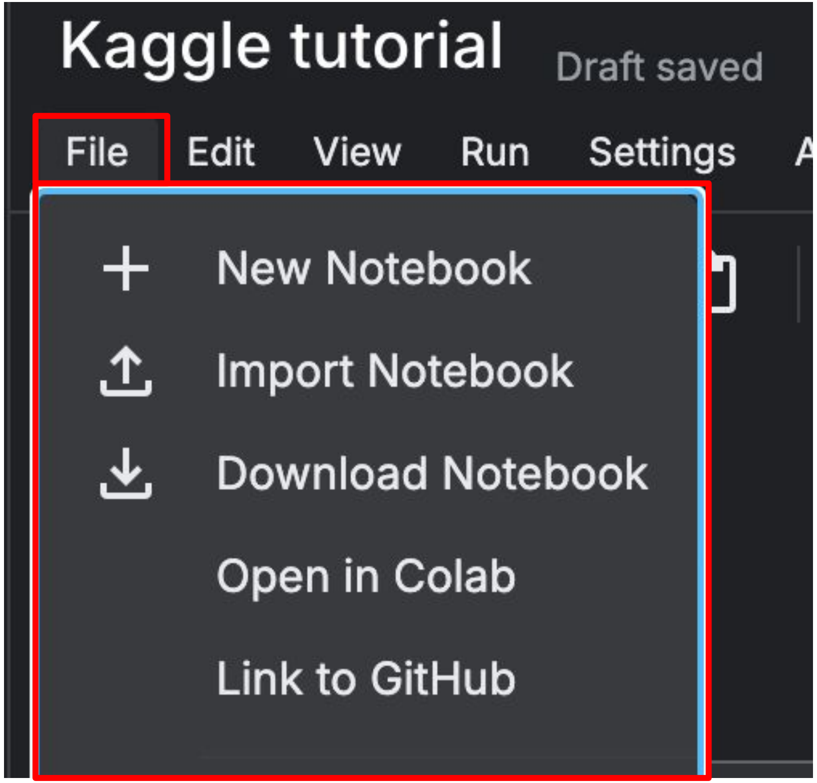
Saving Notebook

- Download the .ipynb file to your local device

(File > Download Notebook)

- Open your notebook in Colab

(File > Open in Colab).



Useful Linux Commands

ls : List all files in the current directory

ls -l : List all files in the current directory with more detail

pwd : Output the working directory

mkdir : Create a directory

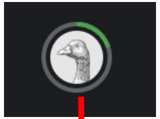
cd : Move to directory gdown : Download files from google drive

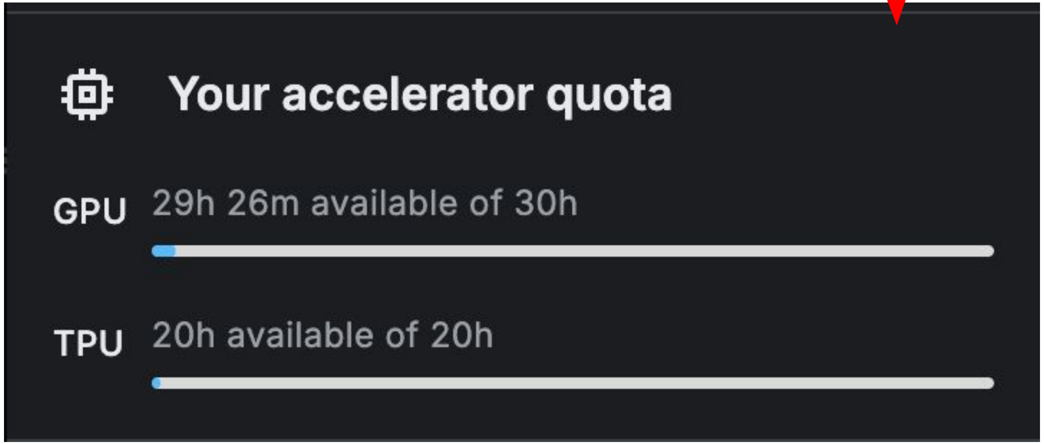
wget : Download files from the internet

python : Executes a python file

Problems You May Encounter...

- GPU on Kaggle usage is **not unlimited** ! (your account will be stopped for a period if you reached the max gpu usage 30 hrs / week)
  - You can see the quota from homepage -> icon head on right-top





Your accelerator quota

GPU 29h 26m available of 30h

TPU 20h available of 20h

Reminder: TAs are not required to help you solve environment problems