

# **Business Analytics Workshop**

Yuan Tian

2025-10-13

# Table of contents

<b>Welcome</b>	<b>4</b>
<b>1 Command Lines and Terminal</b>	<b>5</b>
1.1 The Missing Piece . . . . .	5
1.2 Operating Systems Overview . . . . .	6
1.2.1 What Operating System does Google Colab use? . . . . .	7
1.3 What Is a Terminal? . . . . .	8
1.4 Why Learn Bash commands and Terminal? . . . . .	9
1.4.1 GUI, CLI, Terminal and Desktop . . . . .	10
1.5 Learning Bash commands in Colab . . . . .	11
1.5.1 File Directory in Google Colab . . . . .	11
1.5.2 Lab: Linux and bash . . . . .	12
1.5.3 Lab: Paths, Folders, Directories ( <code>pwd</code> ) . . . . .	12
1.5.4 Lab: List Directory ( <code>ls</code> ) . . . . .	13
1.5.5 Lab: Change Directory ( <code>cd</code> ) . . . . .	13
1.5.6 Lab: Make A Directory ( <code>mkdir</code> ) . . . . .	14
1.5.7 Lab: curl . . . . .	14
1.5.8 Lab: Clear the Screen ( <code>clear</code> ) . . . . .	14
1.5.9 Lab: Remove Directory ( <code>rmdir</code> ) . . . . .	15
1.5.10 Lab: Making Empty Files ( <code>touch</code> ) . . . . .	15
1.5.11 Lab: Copy a File ( <code>cp</code> ) . . . . .	15
1.5.12 Lab: Moving/Rename a File ( <code>mv</code> ) . . . . .	15
1.5.13 Lab: Stream a File ( <code>cat</code> ) . . . . .	16
1.5.14 Lab: Removing a File ( <code>rm</code> ) . . . . .	16
1.5.15 Lab: Exiting Your Terminal ( <code>exit</code> ) . . . . .	16
1.6 Summary Table – Common Bash Commands . . . . .	16
1.7 Directory Structure in GitHub Codespace Terminal . . . . .	17
1.7.1 Bash in VS codespaces . . . . .	18
1.7.2 Lab: create the <code>scripts</code> folder . . . . .	20
<b>2 Run .py scripts with Bash</b>	<b>22</b>
2.1 Why Run Python Scripts from Terminal? . . . . .	22
2.2 .ipynb Notebook vs .py Script Files . . . . .	22
2.2.1 Lab: Create and run a .py script file . . . . .	24

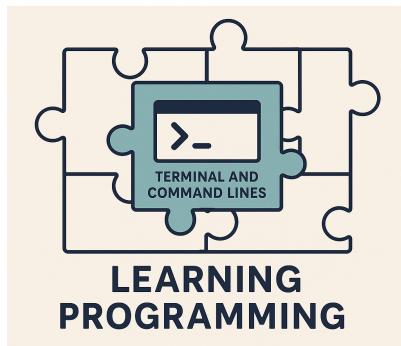
2.3	Other Python-related Commands . . . . .	24
2.3.1	Checking Your Python Version . . . . .	24
2.3.2	Different Python Commands . . . . .	25
2.3.3	Running pip command . . . . .	26
2.4	Working with Command-Line Arguments . . . . .	27
2.4.1	Creating a Script with Arguments . . . . .	27
2.4.2	Running with Arguments . . . . .	27
2.5	Running Scripts in Different Directories. . . . .	29
2.5.1	Absolute Paths . . . . .	29
2.5.2	Relative Paths . . . . .	30
2.5.3	Difference in absolute vs relative paths . . . . .	30
2.5.4	Changing Directories . . . . .	31
2.6	Make a .py file as a module . . . . .	31
2.6.1	Python module review . . . . .	31
2.6.2	Create a user-defined Module in .py file . . . . .	31
2.6.3	Download .ipynb file as a .py file . . . . .	35
<b>3</b>	<b>Develop an app with streamlit</b>	<b>36</b>
3.1	Start from a .ipynb file . . . . .	36
3.2	Make it a .py file with functions . . . . .	36
3.3	Learn a bit about Streamlit . . . . .	36
3.4	Vibe Coding with Github Copilot Agent . . . . .	36
<b>References</b>		<b>37</b>

# Welcome

Welcome to the Business Analytics Workshop 2025!

# 1 Command Lines and Terminal

## 1.1 The Missing Piece



Many students begin your programming journey by writing Python directly in platforms like **Google Colab** - no setup, no files, no terminals. These Jupyter notebook environments make coding easy; however, this convenience often means skipping an important part of programming: **working with real Python files (.py) and using the terminal to run and manage code.**

This “missing piece” is **essential** for understanding how programming works in real projects, **servers**, and **production systems**.

### ⚠ Importance of Command Line and Terminal

*If you skip learning command line skills or avoid the terminal, you'll struggle to work on real-world projects, collaborate effectively with teams, or operate in servers or cloud platforms — where graphical interfaces aren't available. The terminal isn't just a tool for experts; it's the foundation for professional workflows in data science and engineering.*

So, before diving deeper into advanced business data workflows, we'll start by filling this gap, and learn the command lines and terminal to navigate files, run Python scripts, and operate in professional computing environments.

### What You'll Learn Next

- Main operating systems: Windows, macOS, Linux/Unix

- GUI vs terminal and why terminals matter
- Bash shell basics for programming and data science

## 1.2 Operating Systems Overview

Most students in this course use **Windows**, which dominates *personal computers* with roughly **70–75%** of the global desktop market. **macOS** holds about **15–20%**, while **Linux** and others make up a small share of *personal use*.

In contrast, the **enterprise, cloud, AI/ML, and high-performance computing (HPC)** worlds are very different. **Linux** and other **Unix-like systems** are the backbone in web servers, cloud computing and supercomputers, making up nearly half of cloud workloads and being the OS for **all top 500 supercomputers**. Popular Unix and Linux systems include:

- **Ubuntu**
- **Debian**
- **Fedora**
- **Red Hat Enterprise Linux (RHEL)**

 macOS is Unix-based

Although **macOS** looks different, it is actually **Unix-based**, meaning the **terminal commands** and **Bash shell** you'll learn in this course work much the same on both macOS and Linux.

- **Servers** Linux holds a 62.7% market share for server operating systems.
  - Web servers: **77–88%** of public web servers run on **Linux or other Unix-like systems**. It is the most used operating system for web servers globally.
- **Cloud computing** Cloud workloads are heavily dependent on **Linux-based** operating systems. As of mid-2025, Linux powers 49.2% of all global cloud workloads.
- **Supercomputers** Linux has a complete monopoly in the supercomputing sector. 100% market share: Since 2017, **100% of the world's top 500 supercomputers have run on Linux**.
- **AI and ML workloads** Linux is the clear leader for AI and ML projects and infrastructure. In mid-2025, 87.8% of machine learning workloads ran on **Linux infrastructure**. Large ML and data science deployments predominantly run on Linux-based or Unix-based servers.
  - Cloud environments: Cloud providers like AWS, Google Cloud (GCP), **Colab**, and Microsoft Azure, which are leading providers for AI services, primarily offer Linux-based instances for running AI and ML tasks.

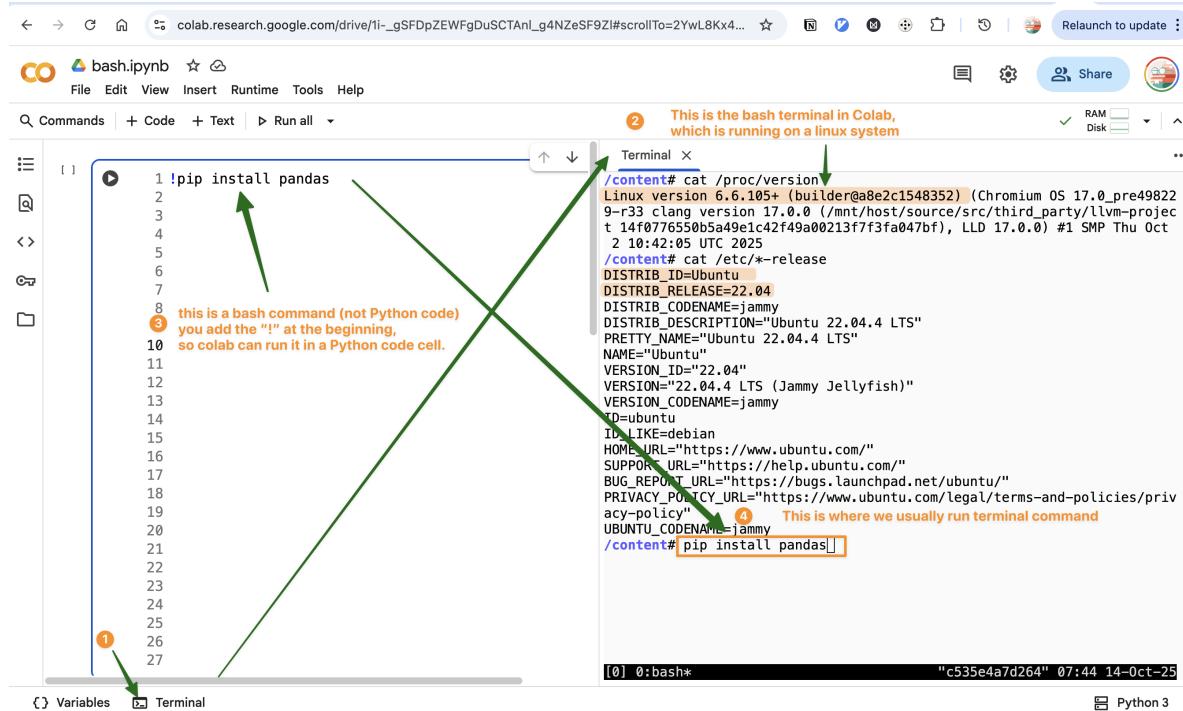
Source: [Wikipedia - Usage share of operating systems](#) [Azure Official Page](#), [Microsoft Tech Community Update \(Feb 2025\)](#)

### 1.2.1 What Operating System does Google Colab use?

Let's take a look at what is the operating system (OS) that running Google Colab. You can type the following command lines in the Terminal.

```
# Bash  
# Display info about the operating system  
cat /etc/*-release  
  
# Display the Linux kernel version and build info  
cat /proc/version
```

See image below for the outputs.



① This is the bash terminal in Colab, which is running on a linux system

② This is where we usually run terminal command

③ this is a bash command (not Python code) you add the "!" at the beginning, so colab can run it in a Python code cell.

④ RAM Disk

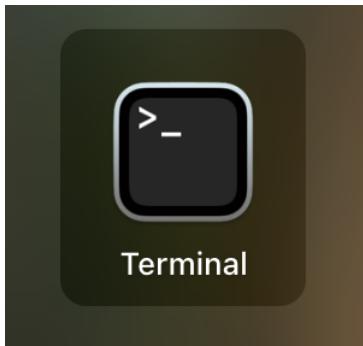
```
1 !pip install pandas  
2  
3  
4  
5  
6  
7  
8 this is a bash command (not Python code)  
9 you add the "!" at the beginning,  
10 so colab can run it in a Python code cell.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27
```

```
[content] cat /proc/version  
Linux version 6.6.105+ (builder@a8e2c1548352) (Chromium OS 17.0_pre49822  
9-r33 clang version 17.0.0 (/mmt/host/source/src/third_party/llvm-project  
14f076550b5a49e1c42f49a00213f7f3fa047bf), LLD 17.0.0) #1 SMP Thu Oct  
2 10:42:05 UTC 2025  
/content# cat /etc/*-release  
DISTRIB_ID=Ubuntu  
DISTRIB_RELEASE=22.04  
DISTRIB_CODENAME=jammy  
DISTRIB_DESCRIPTION="Ubuntu 22.04.4 LTS"  
PRETTY_NAME="Ubuntu 22.04.4 LTS"  
NAME="Ubuntu"  
VERSION_ID="22.04"  
VERSION="22.04.4 LTS (Jammy Jellyfish)"  
VERSION_CODENAME=jammy  
ID=ubuntu  
ID_LIKE=debian  
HOME_URL="https://www.ubuntu.com/"  
SUPPORT_URL="https://help.ubuntu.com/"  
BUG_REPORT_URL="https://bugs.launchpad.net/ubuntu/"  
PRIVACY_POLICY_URL="https://www.ubuntu.com/legal/terms-and-policies/priv  
acy-policy"  
UBUNTU_CODENAME=jammy
```

```
[0] 0:bash* "c535e4a7d264" 07:44 14-Oct-25
```

Python 3

## 1.3 What Is a Terminal?



A **terminal** (also called a **command line** or **shell**) is a text-based interface that lets you interact directly with your computer by typing commands.

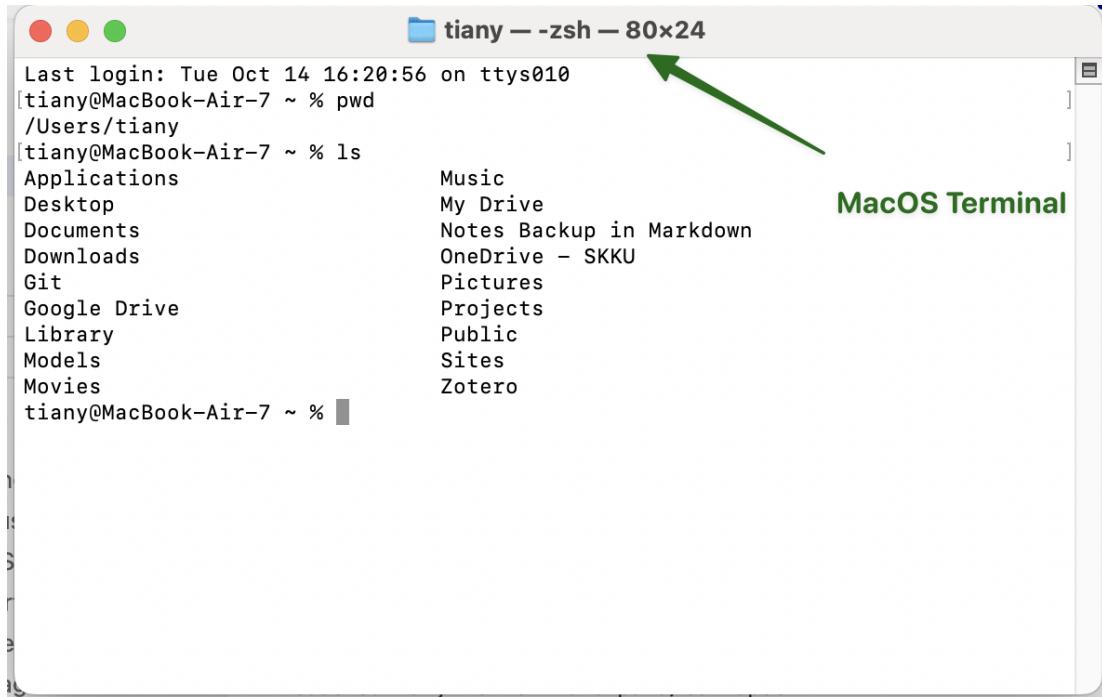
Before graphical interfaces (with windows, icons, and a mouse) were invented, the **terminal was the primary way users operated computers** — to run programs, manage files, and control hardware.

**EVERY** operating system includes a terminal app:

- **Windows:**
  - Command Prompt(`cmd`)
  - **PowerShell**
  - or Bash (through Windows Subsystem for Linux)
  - **Linux:** **Bash** is the default shell on most Linux systems (see [Colab terminal](#))
- **macOS:** **Zsh** in Terminal app (based on Unix) is the default terminal in MacOS, see image below.

**i** Bash vs Zsh

- Both **Bash** and **Zsh** are terminals that interpret your commands, and **they work almost the same**.



A screenshot of a Mac OS Terminal window titled "tiany — zsh — 80x24". The window shows a file listing in the current directory. A green arrow points from the text "MacOS Terminal" to the title bar of the terminal window.

```
Last login: Tue Oct 14 16:20:56 on ttys010
[tiany@MacBook-Air-7 ~ % pwd
/Users/tiany
[tiany@MacBook-Air-7 ~ % ls
Applications          Music
Desktop               My Drive
Documents             Notes Backup in Markdown
Downloads             OneDrive - SKKU
Git                   Pictures
Google Drive          Projects
Library               Public
Models                Sites
Movies                Zotero
tiany@MacBook-Air-7 ~ %
```

The terminal can do almost everything you normally do with a mouse:

- Navigate files and folders
- Run programs or scripts
- Install and manage software
- Connect to remote servers
- Automate repetitive tasks with shell scripts

Data scientists and developers rely on the **terminal** for its speed and automation, especially when working in **cloud environments** like Google Colab, GitHub codespaces, or on Linux servers.

## 1.4 Why Learn Bash commands and Terminal?

First, data science projects often run on **servers or cloud environments**, not personal laptops which lack the computational power for large-scale training, data processing, or deployment.

These servers — such as AWS EC2, Azure VMs, or Google Cloud Compute instances — usually run Linux or Unix systems and don't include a **graphical user interface (GUI)** by default. — they are managed entirely through the **command line interface (CLI)**. To interact with them efficiently, you use **Bash**, a powerful and widely used command-line shell.

### 💡 What is a GUI?

A **Graphical User Interface (UI)** is the visual part of your computer — windows, buttons, and menus you click with the mouse. However, **Linux servers** don't usually have this kind of visual interface.

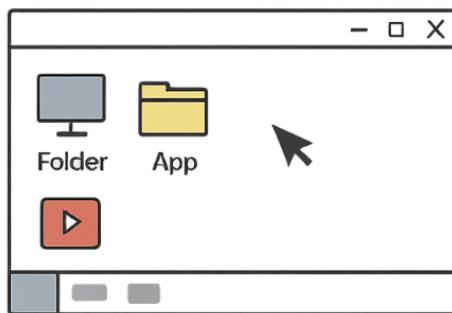
Instead, users interact with them through **script commands** typed into a **terminal** such as **bash**.

#### 1.4.1 GUI, CLI, Terminal and Desktop

## GUI vs Terminal

### GUI

(Graphical User Interface)



- Click to **open or run**
- Windows & icons
- Mouse-driven

### CLI

(Command Line Interface)

```
user@server: $  
> ls  
> cd data  
> python train.py
```

A screenshot of a terminal window. The window is dark with white text. It shows a command prompt "user@server: \$" followed by three command lines: "ls", "cd data", and "python train.py". The terminal has standard window controls at the top right.

- Type commands
- Text-based
- Keyboard-driven

- **GUI (Graphical User Interface)** – The visual interface you use with a **mouse, icons, and windows**, such as Windows desktop, macOS Finder. GUIs are user-friendly but less efficient for automation or remote access.

- **CLI (Command Line Interface)** – A **text-based interface** where you type commands instead of clicking.
- **Terminal** – The **program that provides access to the CLI**. It's like a window that lets you type commands and see text output, such as Windows PowerShell, macOS Terminal, Linux bash Terminal.
- **Desktop Environment** – The **collection of GUI components** that make up the user's graphical workspace — including the taskbar, file explorer, and app windows; such as Windows Desktop, macOS.

**i** Summary

- The **Terminal** gives you access to the **CLI**, while the **Desktop Environment** provides a **GUI**.
- Both let you control the same computer — one through text, the other through graphics.

## 1.5 Learning Bash commands in Colab

Mastering Bash is essential. It enables you to write scripts, manage jobs, and execute commands directly on compute servers — a critical skill when working with **large datasets** or **LLM pipelines**.

We are going to learn **basic** bash commands to:

- Navigate and manage files
- Run **Python (.py) scripts** and other programs (e.g., pip) directly from the command line
- Work efficiently within **server-based** or **local terminal** environments

We will use Google Colab to learn bash commands in Linux system.

Please create or open a new Colab notebook, and then open the Terminal panel.

### 1.5.1 File Directory in Google Colab

When you launch a new Colab notebook, the environment starts with a temporary Linux file system that looks like this:

```
/  
bin/  
boot/  
content/  
    drive/           ← your Google Drive (if mounted)  
    sample_data/     ← sample datasets provided by Colab  
    (your files)      ← any files you upload or create  
dev/  
etc/  
home/  
    root/  
lib/  
media/  
mnt/  
opt/  
proc/  
root/          ← default home directory if you type `cd ~`  
run/  
sbin/  
srv/  
sys/  
tmp/  
usr/
```

### 1.5.2 Lab: Linux and bash

- Display info about the operating system.

```
cat /etc/*-release
```

- Display the Linux kernel version and build info.

```
cat /proc/version
```

### 1.5.3 Lab: Paths, Folders, Directories (pwd)

- Print your current working directory (the folder you are “in”). A directory is a folder, directory and folder are the same thing.

```
pwd  
#/content
```

Please type `pwd` 5 times and each time say “print working directory”.

When to use `pwd?` if you lost in folders and don't know where you are in the directories or folders, `pwd` will tell you where you are.

```
/content# pwd  
#/content
```

#### 1.5.4 Lab: List Directory (`ls`)

The `ls` command is used to **list files and folders** in a directory.

Here are some of the most commonly used ones with options (such as `-a`, `-l`)

```
# List files and folders in the current directory  
ls  
  
# List **all** files, including hidden ones (those starting with .)  
ls -a  
  
# List files in a detailed (**long**) format - shows permissions, owner, size, and date  
ls -l  
  
# Combine options: show all files in detailed view  
ls -la  
  
# Sort files by modification **time** (newest first)  
ls -lt
```

#### 1.5.5 Lab: Change Directory (`cd`)

- `cd sample_data`: go the `sample_data` folder under the current directory.
- `cd ..`: go the parent folder.
- `cd ~`: go to the home folder. In Colab, the home folder is `\root`. If you are lost in a directory and want to start over from a **safe** directory – your home. You can type `cd ~`, and you will be taken to the home directory.

```
# go into the sample_data folder  
ls  
cd sample_data  
ls  
  
# Move up one folder (to the parent directory /content)
```

```
cd ..  
  
# Go back to your "home" folder (/root in Colab)  
cd ~  
pwd  
  
# To-do: find a way to go back to the /content folder.
```

### 1.5.6 Lab: Make A Directory (`mkdir`)

```
# Create a new folder named "data" under /content  
mkdir data  
  
# Make multiple folders at once  
mkdir project scripts results  
  
# Check that they were created  
ls
```

### 1.5.7 Lab: curl

Go to the "data" directory, and download a file from the internet. in `curl -O <URL>`, `-O` stands for saving file with the same name as on the web server.

```
# Go to the "data" Directory  
  
cd data  
  
# Download a small sample text file and save it with the same name.  
curl -O https://raw.githubusercontent.com/jbrownlee/Datasets/master/pima-indians-diabetes.dat  
  
# List files to confirm it's there  
ls
```

### 1.5.8 Lab: Clear the Screen (`clear`)

```
# Clear the terminal screen  
clear
```

### **1.5.9 Lab: Remove Directory (rmdir)**

```
# Create an empty folder named "temp_folder"  
rmdir temp_folder  
  
# Remove the empty folder  
rmdir temp_folder  
  
# Create multiple empty folders and remove them  
mkdir folder1 folder2  
rmdir folder1 folder2
```

### **1.5.10 Lab: Making Empty Files (touch)**

```
# Create an empty file named "notes.txt"  
touch notes.txt  
  
# Create multiple files at once  
touch a.txt b.txt c.txt  
  
# Verify files were created  
ls
```

### **1.5.11 Lab: Copy a File (cp)**

```
# Copy a file to a new file  
cp notes.txt notes_backup.txt  
  
# Create a folder to copy into  
mkdir backup  
  
# Copy a file into a different folder  
cp notes.txt backup/  
  
# Check the results  
ls backup
```

### **1.5.12 Lab: Moving/Rename a File (mv)**

```
# Move a file into a different folder  
mv notes_backup.txt backup/
```

```
# Rename a file  
mv notes.txt todo.txt
```

```
# Verify the changes  
ls
```

### 1.5.13 Lab: Stream a File (cat)

```
# Display the contents of a file  
cat todo.txt
```

```
# To-do: Display the README.md file in the "sample_data" folder:
```

```
# Display a system file (try this!)  
cat /etc/*-release
```

### 1.5.14 Lab: Removing a File (rm)

```
# Create some temporary files first  
touch old.txt temp.txt sample.txt
```

```
# Remove a single file  
rm old.txt
```

```
# Remove multiple files  
rm temp.txt sample.txt
```

```
# Remove an entire folder and its contents (be careful!)  
rm -r backup
```

### 1.5.15 Lab: Exiting Your Terminal (exit)

```
# Exit the current terminal session  
exit
```

## 1.6 Summary Table – Common Bash Commands

Command	Purpose	Example
<code>pwd</code>	Print working directory	<code>pwd</code>
<code>ls</code>	List files and folders	<code>ls -la</code>
<code>cd</code>	Change directory	<code>cd /content</code>
<code>mkdir</code>	Make a new directory	<code>mkdir data</code>
<code>rmdir</code>	Remove an empty directory	<code>rmdir temp_folder</code>
<code>curl</code>	Download a file from the internet	<code>curl -O https://example.com/file.txt</code>
<code>touch</code>	Create an empty file	<code>touch notes.txt</code>
<code>cp</code>	Copy a file	<code>cp notes.txt backup/</code>
<code>mv</code>	Move or rename a file	<code>mv old.txt new.txt</code>
<code>cat</code>	View contents of a file	<code>cat notes.txt</code>
<code>rm</code>	Remove a file or folder	<code>rm -r foldername</code>
<code>clear</code>	Clear the screen	<code>clear</code>
<code>exit</code>	Exit the terminal	<code>exit</code>

## 1.7 Directory Structure in GitHub Codespace Terminal

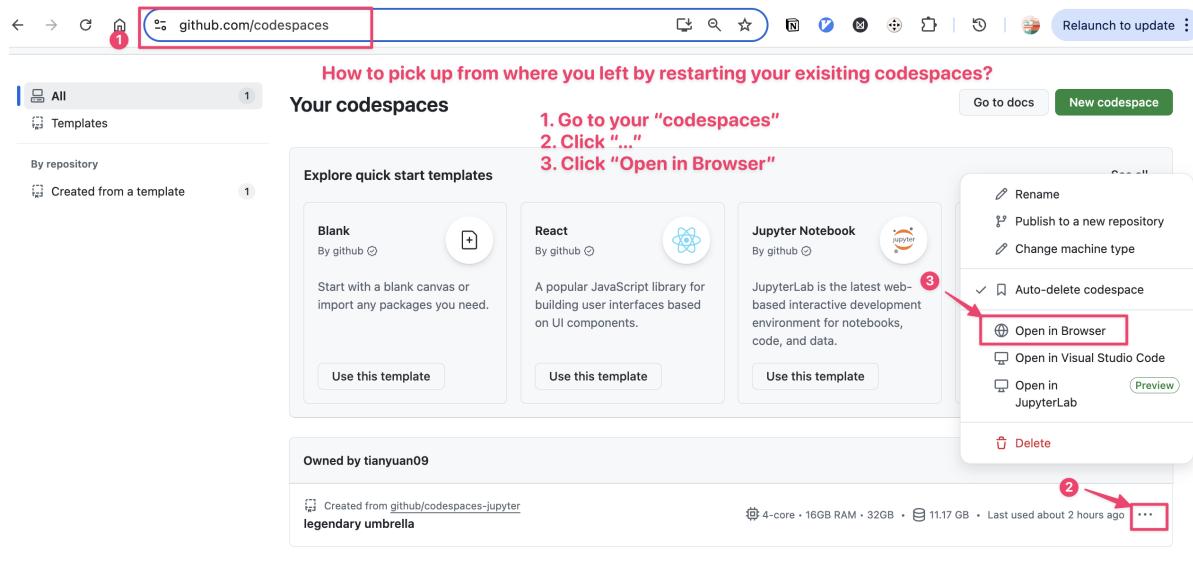
```

/
bin/
boot/
dev/
etc/
home/
    codespace/           ← your user home directory if you do `cd ~`
lib/
lib64/
media/
mnt/
opt/
proc/
root/
run/
sbin/
srv/
sys/
tmp/
usr/
workspaces/
    /codespaces-jupyter   ← your GitHub repo (default working dir)

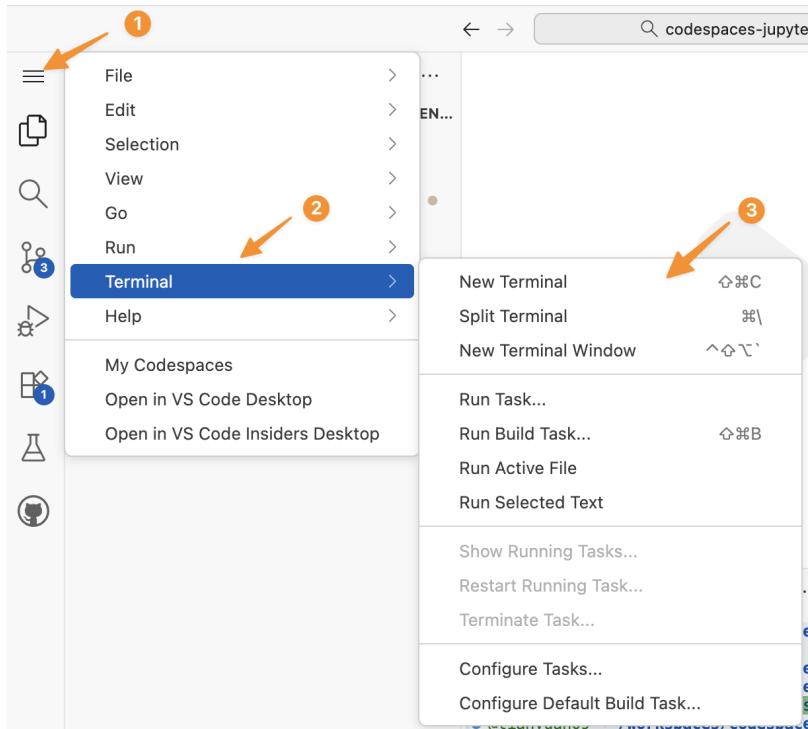
```

### 1.7.1 Bash in VS codespaces

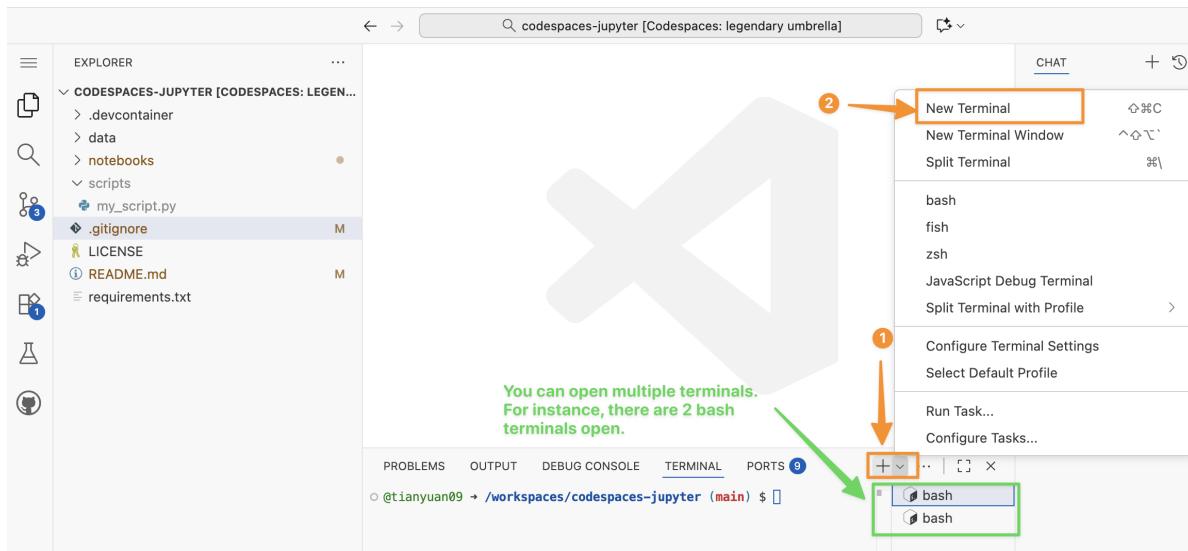
Restart your github codespaces.



Make sure you can see the Terminal panel. If you accidentally closed your terminal, you can always start one (or many) following the steps below.



You also can start a second terminal via the Terminal panel.



### 1.7.2 Lab: create the scripts folder

Using the bash terminal to create a `scripts` folder under GitHub default working dir `/workspaces/codespaces-jupyter`, and create an empty `my_script.py` file in the `scripts` folder.



Tip

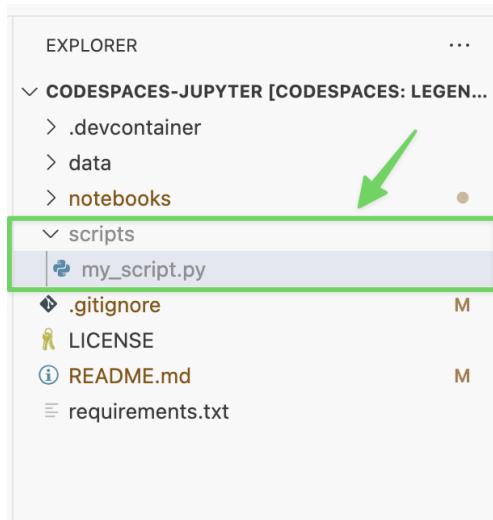
You may find the commands `pwd`, `mkdir`, `ls`, `cd`, and `touch` helpful for completing this exercise.

Your task is to:

- Verify your current working directory.
- Create a new folder called `scripts` inside the project root.
- List the directory contents to confirm that `scripts` was created successfully.
- Navigate into the `scripts` folder.
- Confirm it is empty.
- Create a new Python file named `my_script.py` inside the `scripts` folder.

Write the Bash commands needed to accomplish each step.

Once you complete the task, your explorer should look like below:



### 🔥 Solution

```
@tianyuan09 /workspaces/codespaces-jupyter (main) $ pwd
/workspaces/codespaces-jupyter
@tianyuan09 /workspaces/codespaces-jupyter (main) $ mkdir scripts
@tianyuan09 /workspaces/codespaces-jupyter (main) $ ls
LICENSE README.md data notebooks requirements.txt scripts
@tianyuan09 /workspaces/codespaces-jupyter (main) $ cd scripts
@tianyuan09 /workspaces/codespaces-jupyter/scripts (main) $ ls
@tianyuan09 /workspaces/codespaces-jupyter/scripts (main) $ touch my_script.py
```

# 2 Run .py scripts with Bash

## 2.1 Why Run Python Scripts from Terminal?

While interactive environments like **Jupyter notebooks** are great for exploration and prototyping, running **Python scripts (.py files)** from the terminal is essential for:

- **Production workflows:** Automated data processing, model training, and deployment
- **Server environments:** Most servers don't have graphical interfaces
- **Batch processing:** Running scripts on large datasets or multiple files
- **Scheduling:** Using cron jobs or task schedulers to run scripts automatically
- **Command-line arguments:** Passing parameters to scripts dynamically
- **Performance:** Scripts often run faster than notebooks for large tasks

### For Example

Data scientists often submit training jobs via Bash scripts like:

```
python train_model.py --epochs 10
```

It runs the Python script `train_model.py` to train the model for 10 epochs.

## 2.2 .ipynb Notebook vs .py Script Files

Figure 2.1 illustrates the difference between running Python code in `.ipynb` notebook versus in ‘`.py`’ script file.

### Understanding the Terminal

See [directory structure in codespaces](#) in the previous chapter.

In the terminal, it always starts with:

```
@tianyuan09 /workspaces/codespaces-jupyter (main) $.
```

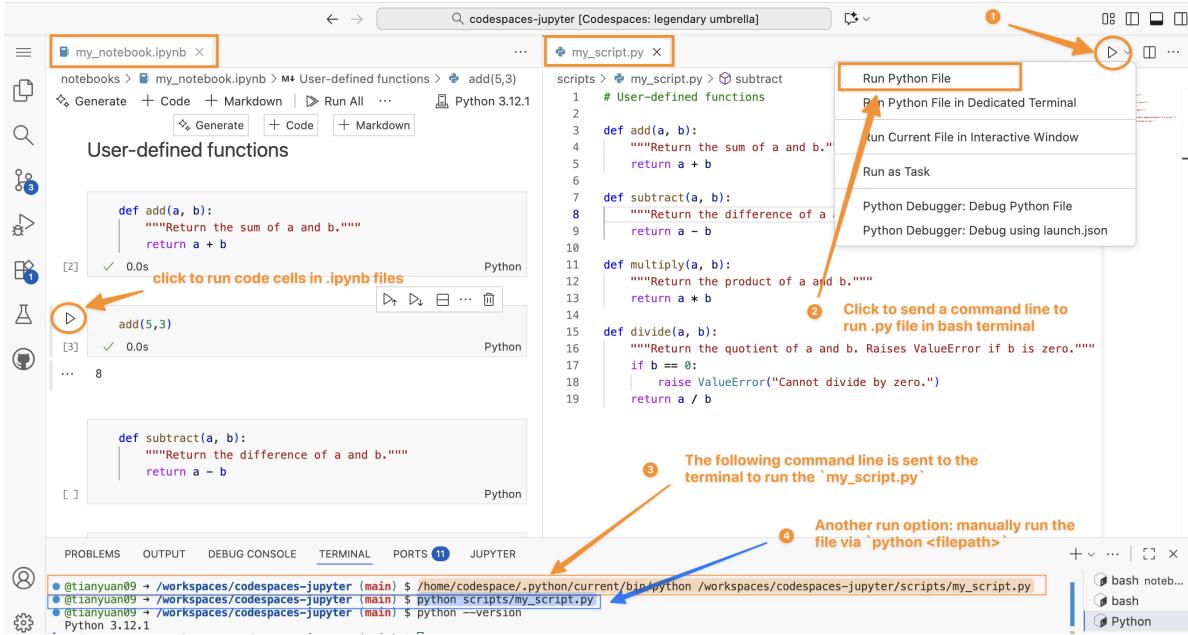


Figure 2.1: Comparison of Jupyter Notebook and Python Script formats

@tianyuan09 /workspaces/codespaces-jupyter (main) \$

Prompt symbol (\$):  
shows the terminal is ready for input

Current working directory:  
you're inside the folder "codespaces-jupyter"  
located under "/workspaces"

Arrow ( ):  
just a decorative separator in the prompt

Username (and sometimes host):  
"tianyuan09" - the current user logged into this environment

### Difference between .ipynb Notebook and .py Script File.

Feature	.ipynb (Jupyter Notebook)	.py (Python Script)
Structure	Structured JSON format combining code, text cells in Markdown, and outputs.	Plain text file containing only Python code and comments #.

Feature	.ipynb (Jupyter Notebook)	.py (Python Script)
<b>Execution</b>	Run one cell at a time, showing output immediately below each cell.	Executed all at once using a command like <code>python my_script.py</code> , seen in Figure 2.1.
<b>Use Case</b>	Ideal for data analysis, visualization, and teaching due to its interactive nature.	Better for automation, <b>deployment</b> of production-ready code.

### 2.2.1 Lab: Create and run a .py script file

First, let's create a `hello.py` under the `scripts` folder.

```
# file path: scripts/hello.py
print("Hello from Python!")
print("This script is running from the terminal.")

# Get current date and time
import datetime
now = datetime.datetime.now()
print(f"Current time: {now}")
```

You can directly use explorer or use bash command Figure 2.2.

**Expected output:**

```
Hello from Python!
This script is running from the terminal.
Current time: 2025-10-13 10:30:45.123456
```

## 2.3 Other Python-related Commands

### 2.3.1 Checking Your Python Version

```
# Check Python version
python --version
python3 --version
```

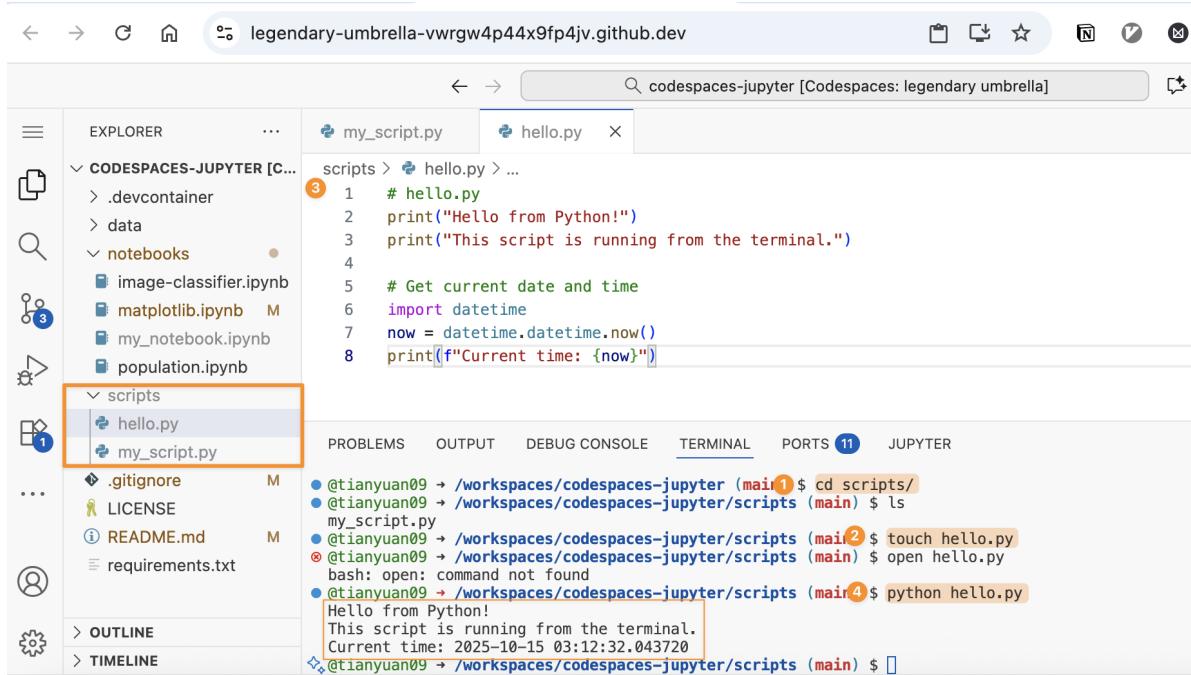


Figure 2.2: Create and run your 1st script file

```
# Check which Python executable you're using
which python
which python3
```

### 2.3.2 Different Python Commands

Depending on your system setup, you might need to use different commands:

```
# On systems with Python 3 as default
python hello.py

# OR
python3 hello.py

# OR Using specific Python version
python3.9 hello.py
python3.12 hello.py
```

### 2.3.3 Running pip command

#### ! Tip

pip is a **command-line tool**, not Python code. You actually should run `pip install ...` in your terminal.

Although you have run `pip` inside a code cell in Colab with `!`, such as `!pip install pandas`, the notebook actually sends it as a command line to the terminal to execute. It is equivalent to running `pip install pandas` in the terminal, see in Figure 2.3.

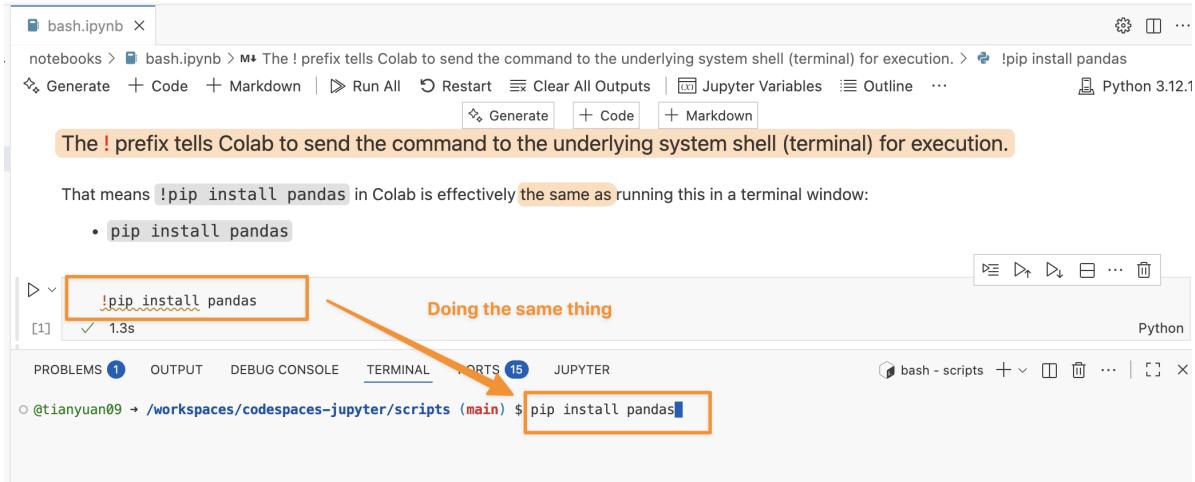


Figure 2.3: pip in terminal

#### ! How to install package for .py files?

In regular .py files (non-notebook files), pip commands **must be executed in the terminal**, not inside the script itself.

For example, to install the `streamlit` package, run this in your terminal:

```
pip install streamlit
```



List all the packages installed:

```
# Check installed packages
pip list
```

Please install streamlit in your codespaces.

## 2.4 Working with Command-Line Arguments

### 2.4.1 Creating a Script with Arguments

.py files with arguments allow the users to provide input directly when running the script. Please work on the following example.

```
# create and save in scripts/greet.py
import sys
print(sys.argv) # sys.argv is a list of string typed after python in the bash

# Check if arguments were provided
if len(sys.argv) < 2:
    print("Usage: python greet.py <name>")
    sys.exit(1)

name = sys.argv[1]
print(f"Hello, {name}!")

# Optional: Handle multiple arguments
if len(sys.argv) > 2:
    age = sys.argv[2]
    print(f"You are {age} years old.")
```

### 2.4.2 Running with Arguments

```
# Run with one argument
python greet.py Alice

# Run with multiple arguments
python greet.py Bob 25

# Without arguments (will show usage message)
python greet.py
```

The screenshot shows the VS Code interface with the following details:

- EXPLORER**: Shows a folder named "CODESPACES-JUPYTER [C...]" containing files like ".devcontainer", "data", "notebooks", "scripts" (with "greet.py" selected), "test\_logging.py", ".gitignore", "LICENSE", "README.md", and "requirements.txt".
- EDITOR**: Displays the content of "greet.py". The line `print(sys.argv)` is highlighted with an orange box.
- TERMINAL**: Shows the output of running the script. The first command is `python greet.py` followed by a usage message. The second command is `python greet.py alice 25`, which outputs "Hello, alice!" and "You are 25 years old.".
- PROBLEMS**: Shows one error: `@tianyuan09 > /workspaces/codespaces-jupyter/scripts (main) \$ touch greet.py`.

#### 2.4.2.1 `sys.argv` — what it really is

`sys.argv` is a **list of strings** containing everything typed after `python` in the terminal.

For example, if you run:

```
python greet.py Alice 25
```

Then in Python:

```
import sys
print(sys.argv)
# output: ['greet.py', 'Alice', '25'] # a list
```

#### 2.4.2.2 Index meaning

Index	Value	Meaning
<code>sys.argv[0]</code>	'greet.py'	the <b>name of the script file</b> being executed

Index	Value	Meaning
sys.argv[1]	'Alice'	the <b>first argument</b> typed after the script name
sys.argv[2]	'25'	the <b>second argument</b> (if given)

#### 2.4.2.3 why name = sys.argv[1]?

Because:

- sys.argv[0] is always the script name (`greet.py`)
- sys.argv[1] is the first real argument that the user provides. In this case, the person's name.

If you used `sys.argv[0]`, it would just say: Hello, `greet.py`!

## 2.5 Running Scripts in Different Directories.

### 2.5.1 Absolute Paths

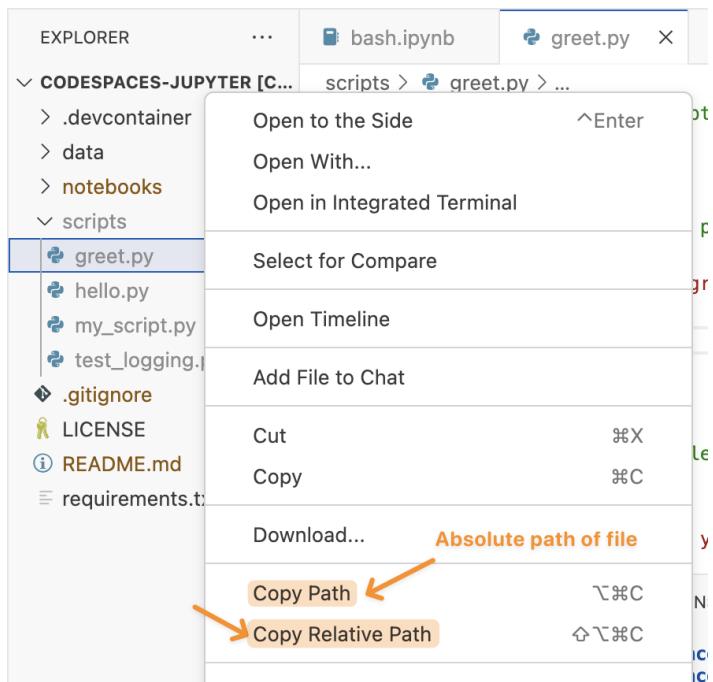
```
# Run script from anywhere using absolute path
python /Users/username/projects/my_script.py
```

/Users/username/projects/my\_script.py is a absolute file path.

Absolute paths:

- On macOS/Linux, starts with /.
- On Windows, starts with a drive letter like C:/.
- In short, it always starts from the root of the filesystem (the top level).

You can find either the relative or absolute file paths in codespaces:



### 2.5.2 Relative Paths

```
# Run script in current directory
python ./script.py

# Run script in subdirectory
python scripts/data_analysis.py

# Run script in parent directory
python ../utilities/helper.py
```

The paths above (e.g. `../utilities/helper.py`) are relative paths.

- They **don't start with /** or a drive letter (e.g. C:/).
- They may include `.` (current folder) or `..` (parent folder).
- They starts from your current working directory (`pwd`).

### 2.5.3 Difference in absolute vs relative paths

Starts With	Type	Meaning
/ (Linux/Mac)	Absolute	Starts at root of file system

Starts With	Type	Meaning
C:\ (Windows)	Absolute	Starts at root of drive
. or ..	Relative	Based on current working directory
No / or C:\	Relative	Implied to start from current folder

### 2.5.4 Changing Directories

```
# Navigate to script directory, then run
cd /path/to/scripts
python analysis.py

# Or combine in one line
cd /path/to/scripts && python analysis.py
```

## 2.6 Make a .py file as a module

### 2.6.1 Python module review

**import**

Python modules and packages generally fall into four categories: built-in, standard library, third-party, and user-defined (see in Figure 2.4). So far, you've worked with the first three — modules that come with Python (e.g., `datetime`) or are installed from external sources (e.g., `pandas`).

### 2.6.2 Create a user-defined Module in .py file

In this section, you'll learn how to create your own .py module with reusable functions – one that can be both imported into a Colab notebook and run independently for testing.

Create `my_util.py` in the `scripts` folder.

```
# scripts/my_utils.py

def greet(name):
    """Return a personalized greeting."""
    return f"Hello, {name}!"
```

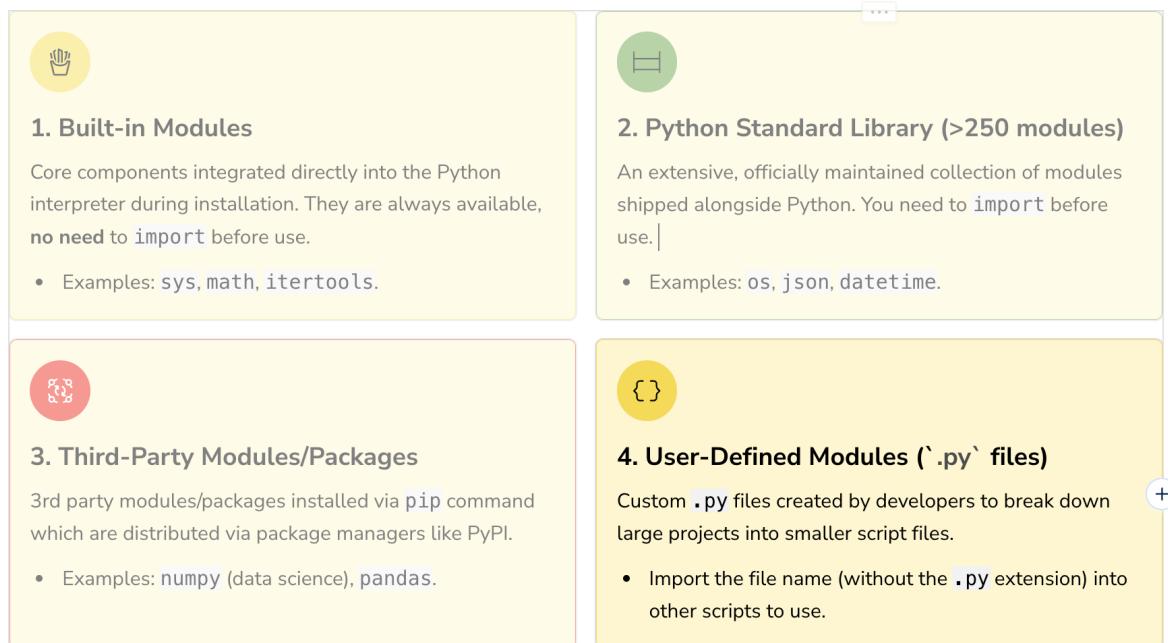


Figure 2.4: Categories of Python Modules/Packages

```

def add_numbers(a, b):
    """Return the sum of two numbers."""
    return a + b

# many functions that you may want to re-use.

# This block runs only if you execute `python my_utils.py`
# This will make the .py file both run alone or import as a module
if __name__ == "__main__":
    print("Running my_utils.py as a standalone script...")
    print(greet("Tester"))
    print("5 + 7 =", add_numbers(5, 7))

```

#### Option 1: Run as a standalone .py file

```
python scripts/my_utils.py
```

```

EXPLORER      ...
CODESPACES-JUPYTER [C...]
> .devcontainer
> data
> notebooks
scripts > my_utils.py > ...
1 # scripts/my_utils.py
2
3 def greet(name):
4     """Return a personalized greeting."""
5     return f"Hello, {name}!"
6
7 def add_numbers(a, b):
8     """Return the sum of two numbers."""
9     return a + b
10
11 # many functions that you may want to re-use.
12
13 # This block runs only if you execute python my_utils.py
14 if __name__ == "__main__":
15     print("Running my_utils.py as a standalone script...")
16     print(greet("Tester"))
17     print("5 + 7 =", add_numbers(5, 7))

PROBLEMS    OUTPUT    DEBUG CONSOLE    TERMINAL    PORTS 16
@tianyuan09 → /workspaces/codespaces-jupyter (main) $ cd scripts
@tianyuan09 → /workspaces/codespaces-jupyter/scripts (main) $ touch my_utils.py
@tianyuan09 → /workspaces/codespaces-jupyter/scripts (main) $ python my_utils.py
Running my_utils.py as a standalone script...
Hello, Tester!
5 + 7 = 12
@tianyuan09 → /workspaces/codespaces-jupyter/scripts (main) $ 

```

## Option 2: Run as a user-defined modules with reusable functions

Create a `test_util.ipynb` file in the **same** folder as your `my_utils.py` file. Then, you can do `import my_utils` to use it like any module/package that you have used before.

```

EXPLORER      ...
CODESPACES-JUPYTER [C...]
> .devcontainer
> data
> notebooks
scripts > my_utils.py > ...
1 # scripts/my_utils.py
2
3 def greet(name):
4     """Return a personalized greeting."""
5     return f"Hello, {name}!"
6
7 def add_numbers(a, b):
8     """Return the sum of two numbers."""
9     return a + b
10
11 # many functions that you may want to re-use.
12
13 # This block runs only if you execute python my_utils.py
14 if __name__ == "__main__":
15     print("Running my_utils.py as a standalone script...")
16     print(greet("Tester"))
17     print("5 + 7 =", add_numbers(5, 7))

test_utils.ipynb > ...
scripts > test_utils.ipynb > import my_utils
Generate + Code + Markdown | Run All ... Python 3.12.1
1 import my_utils
[1] ✓ 0.0s Python
2 my_utils.greet("BA MBA Workshop stdtuents")
[2] ✓ 0.0s Python
... 'Hello, BA MBA Workshop stdtuents!'
3 my_utils.add_numbers(5,10)
[4] ✓ 0.0s Python
... 15

Make sure they are in the same folder

```

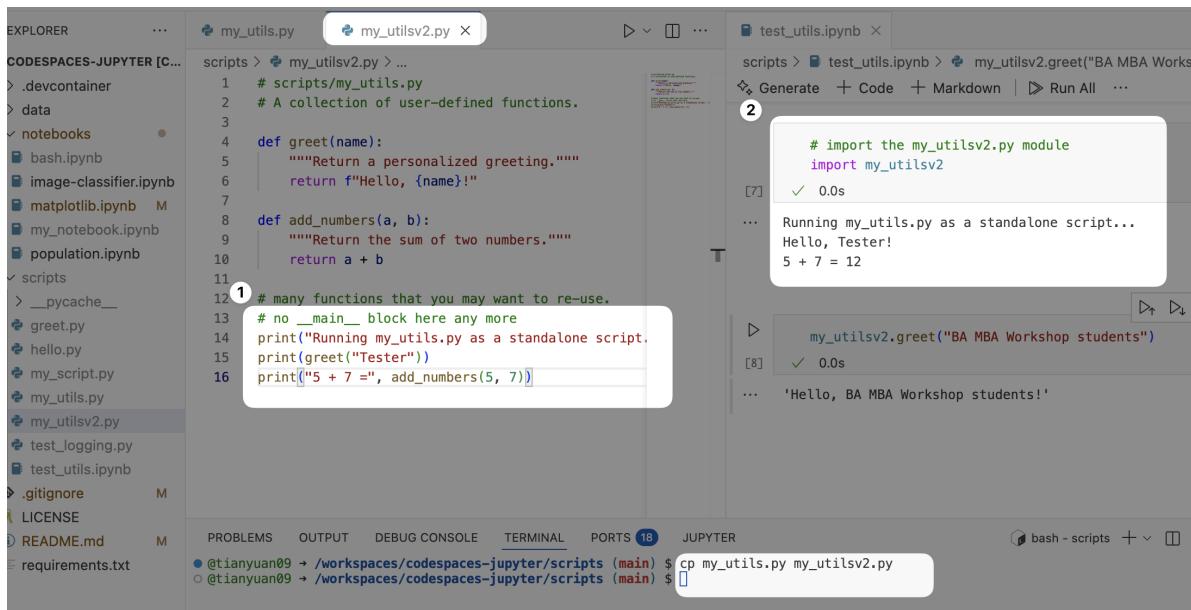
### What does `if __name__ == "__main__":` do?

`__name__` is a special variable that Python automatically sets when a file is run or imported.

- If the file is run directly (e.g., `python my_utils.py`), `__name__` becomes "`__main__`".
- If the file is imported (e.g., `import my_utils`), `__name__` becomes the module's name ("`my_utils`").
- `if __name__ == "__main__":` ensures that code inside it runs only when the file is executed directly, not when imported. This allows a `.py` file to act as both:
  - a reusable module (when imported)
  - a standalone script for testing or demos (when run directly).

### What if you remove the `if __name__ == "__main__":` line

If you do that, and then import the module, any code at the bottom of the file will **run automatically during import**, which can **cause unwanted behavior (like printing, running tests, or altering data)**, see Figure 2.5.



The screenshot shows the VS Code interface with the following details:

- EXPLORER:** Shows files in the workspace, including `my_utils.py` and `my_utilsv2.py`.
- CODESPACES-JUPYTER [C...]:** Shows the file structure and code for `my_utilsv2.py`. The code includes functions `greet` and `add_numbers`. A comment at the bottom indicates no `__main__` block.
- TERMINAL:** Shows a terminal session with the command `cp my_utils.py my_utilsv2.py` being run.
- BASH - SCRIPTS:** Shows a terminal session with the command `my_utilsv2.greet("BA MBA Workshop students")` being run.

The code in `my_utilsv2.py` is as follows:

```
1 # scripts/my_utils.py
2 # A collection of user-defined functions.
3
4 def greet(name):
5     """Return a personalized greeting."""
6     return f"Hello, {name}!"
7
8 def add_numbers(a, b):
9     """Return the sum of two numbers."""
10    return a + b
11
12 # many functions that you may want to re-use.
13 # no __main__ block here any more
14 print("Running my_utils.py as a standalone script...")
15 print(greet("Tester"))
16 print(f"{5 + 7} =", add_numbers(5, 7))
```

Figure 2.5: Removing `__main__`

The `if __name__ == "__main__":` line **isn't required** for a module to work, but it helps

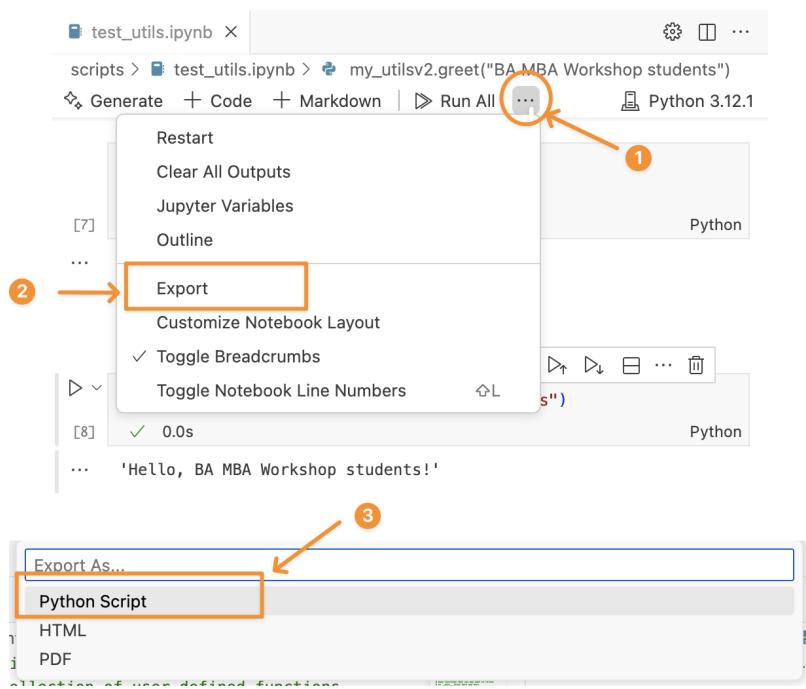
keep your code tidy and organized, allowing the file to be used both as a reusable module and as a standalone script with testing.

You also can import a single or multiple functions in a .py file.

```
# import the greet(), add_numbers() functions
from my_utils import greet, add_numbers

print(greet("Bob"))          # Output: Hello, Bob!
print(add_numbers(10, 5))    # Output: 10 + 5 = 15
```

### 2.6.3 Download .ipynb file as a .py file



# 3 Develop an app with streamlit

## 3.1 Start from a .ipynb file

Files:

- `sp500_data.csv` – Financial data.
- `sp500_tickers.csv` – List of tickers.

## 3.2 Make it a .py file with functions

## 3.3 Learn a bit about Streamlit

### 💡 Questions that I should ask myself

- What is Streamlit, and where can I find its API reference and official documentation?
- What can Streamlit do — what kinds of apps or problems is it best at solving?
- What small examples or experiments can I build to quickly discover and learn its core features?

- **Start with the official docs and examples** — skim the Streamlit docs homepage and gallery to get a mental map of what's possible before diving into code.
- **Learn by doing small experiments** — build micro-apps (e.g., one with a button, one with a chart) to turn reading into muscle memory.
- **Read code, not just tutorials** — explore community demos or open-source Streamlit apps to see how others structure layouts, manage state, and handle inputs.
- **Iterate with feedback loops** — run `streamlit run app.py` often and tweak one thing at a time; immediate visual feedback accelerates understanding.
- **Reflect and generalize** — after each mini-project, note what patterns repeat (e.g., sidebar widgets, caching, layout control) to build knowledge for future tools.

## 3.4 Vibe Coding with Github Copilot Agent

## **References**