Dr. Ala’ Khalifeh

Cloud Computing and BigData

Cloud for AI

- A cloud-based platform for running Jupyter notebooks and performing Machine Learning using Python:
- 🔗 [Google Colab](https://colab.research.google.com/)
- 🔑 Requires a Google account
- ✅ Comes with pre-installed libraries for Data Science and Machine Learning
- ⚡ Offers free access to GPU and TPU acceleration

Getting Started with Google Colab

- Google Colab is a cloud-based Jupyter notebook service.
- Free to use with Google account.
- Supports Python and many preinstalled libraries.
- Allows real-time collaboration.
- Provides access to GPU and TPU acceleration.
- Integrates with Google Drive for storage.
- No local setup required—runs in your browser.

What is Google Colab?

- Based on the Jupyter notebook interface.
- Runs entirely on the cloud—no installations needed.
- Ideal for Python programming, data science, and ML.
- Supports code execution, documentation, and visualization.
- Collaborators can comment or edit notebooks in real-time.
- Part of Google Research tools for open and reproducible research.
- Accessible from any device with internet access.
- Teaser! https://www.youtube.com/watch?v=inN8seMm7UI

Key Features Overview

- Interactive Python execution.
- Rich text, LaTeX, and HTML support via Markdown.
- Upload/download files directly from the notebook.
- GitHub integration for version control.
- Use of popular Python libraries pre-installed.
- Ability to visualize data using matplotlib/seaborn.
- Built-in tools for inspecting variables and managing files.

Requirements

- A valid Google account (Gmail).
- Modern web browser (Chrome/Firefox recommended).
- Basic knowledge of Python programming.
- Internet connection for cloud execution.
- Familiarity with Jupyter notebooks is helpful.
- Optional: Google Drive to store files and notebooks.
- Optional: GitHub account for loading notebooks from repos.

Accessing Google Colab

To learn it use the below link

- Educational materials and tutorials: https://colab.google/notebooks/

To try out use the below link

- Go to https://colab.research.google.com.

Sign in with your Google account.

Start from a blank notebook or open from Drive/GitHub.

Interface provides multiple templates to choose from.

Quick access to recent notebooks.

Choose 'Upload' tab to upload a local notebook.

Colab loads instantly in your browser.

Creating a New Notebook

Click on File > New Notebook.

Notebook automatically saved in Google Drive.

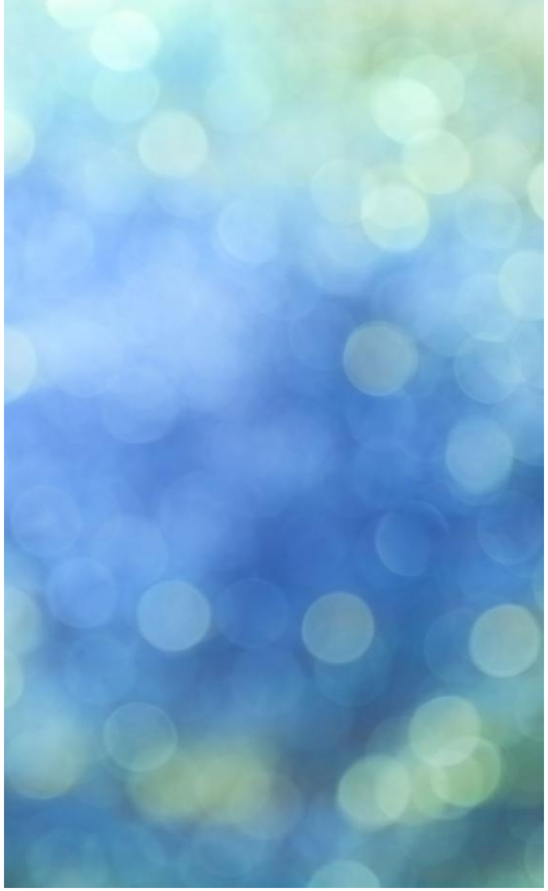
Default file name is Untitled0.ipynb (rename it).

Notebook consists of cells—code and text.

You can immediately start typing Python code.

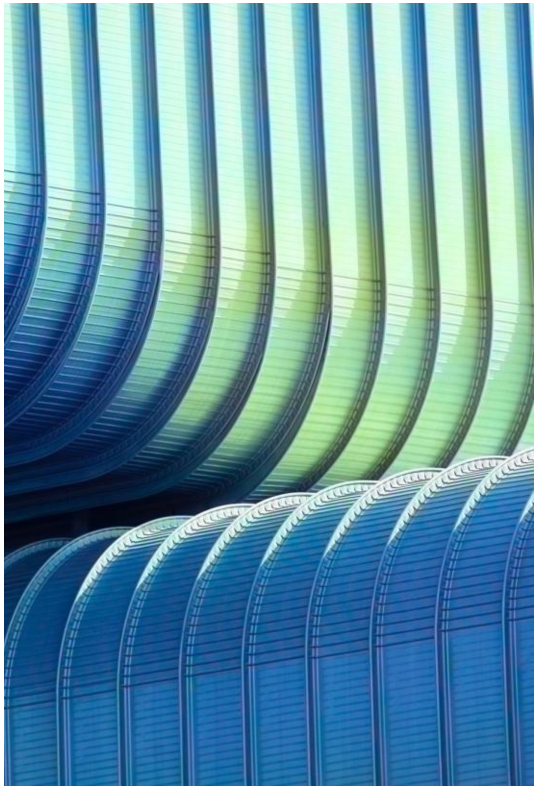
Add text cells using Markdown for documentation.

Name reflects in Google Drive instantly.



Interface Overview

- Menu bar with options (File, Edit, View, etc.).
- Toolbar with quick actions: run, restart, insert cell.
- Code cells for Python; text cells for instructions.
- Output appears directly under code cells.
- Left sidebar includes Table of Contents, Files, Variables.
- Top-right corner shows runtime status.
- Settings allow dark mode and preferences.



Working with Code Cells

Click on cell and type Python code.

Use Shift+Enter or the play icon to execute.

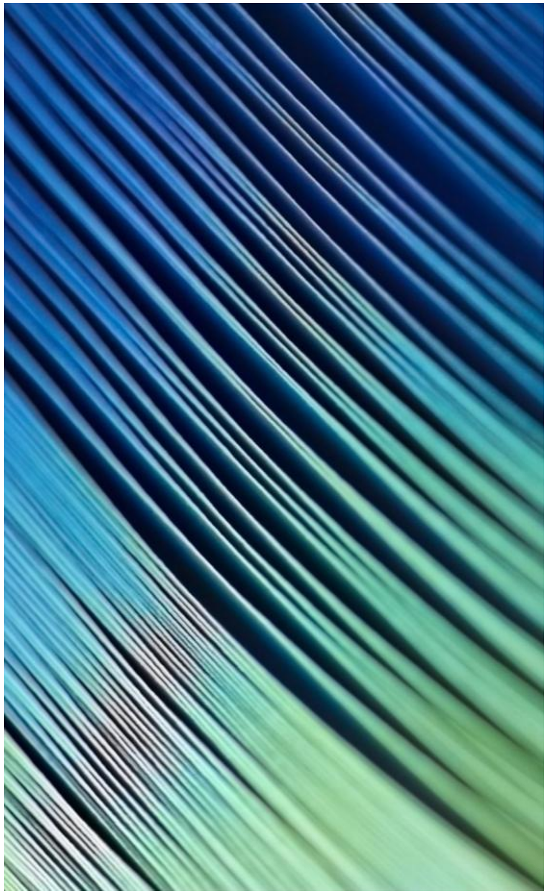
Outputs appear below the cell.

Syntax highlighting and auto-complete supported.

Use triple quotes for multiline strings.

Use semicolons to suppress output.

Add comments using '#' symbol.

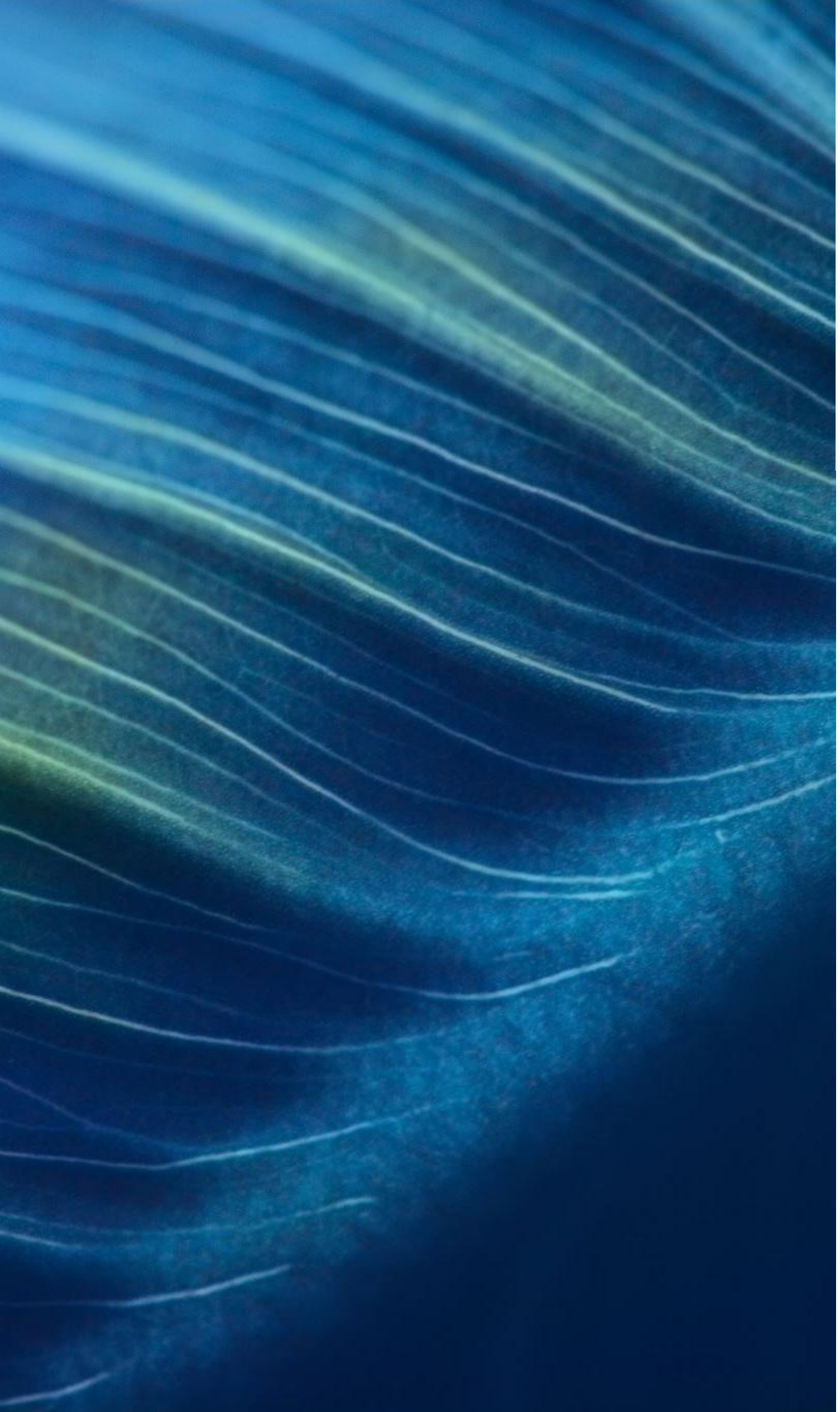


Markdown for Documentation

- Markdown cells allow text formatting.
- # for headings, **bold**, *italic*, `code`.
- Include lists using -, + or *.
- Insert links using [text](url).
- Add images with ![alt](image_url).
- Render LaTeX equations using $$ or \(\).
- Great for explaining steps in analysis.

Code Execution Basics

- Each code cell runs independently.
- State is preserved across cells in the same session.
- Variables and imports stay available.
- Use Runtime > Restart to clear memory.
- Run all cells using Runtime > Run all.
- Use Ctrl+M+B to add new cells below.
- Use Ctrl+Enter to run without moving.



Using GPUs and TPUs

Runtime > Change runtime type.

Select GPU or TPU from hardware accelerator menu.

Free GPU/TPU resources are time-limited.

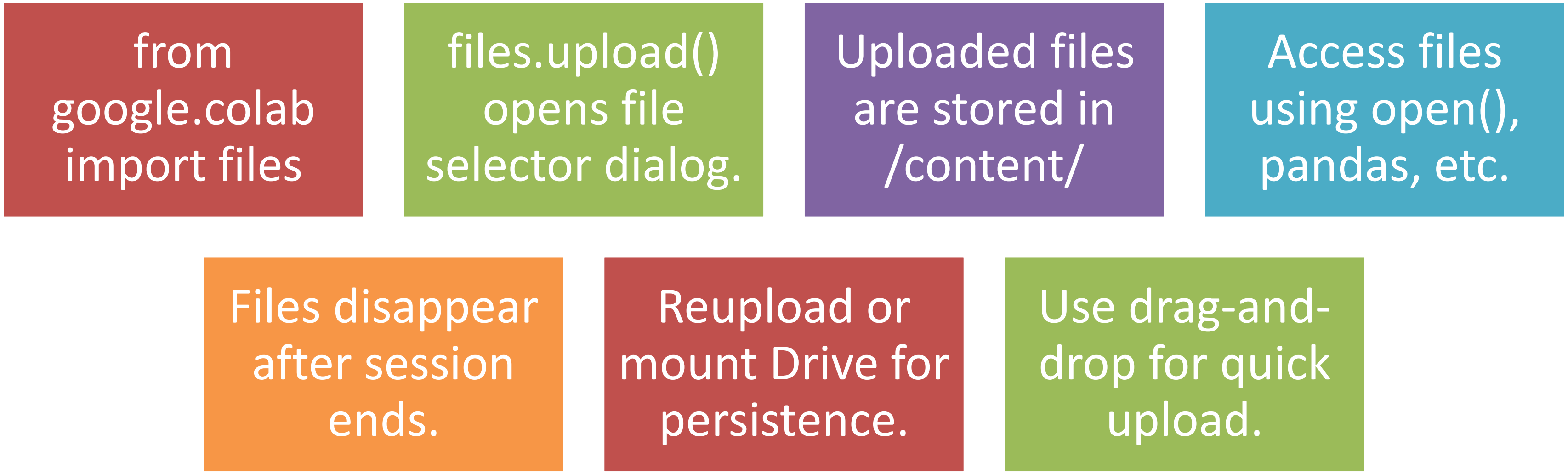
Check connection in top-right status bar.

Use 'nvidia-smi' to check GPU status.

Use TensorFlow or PyTorch with hardware acceleration.

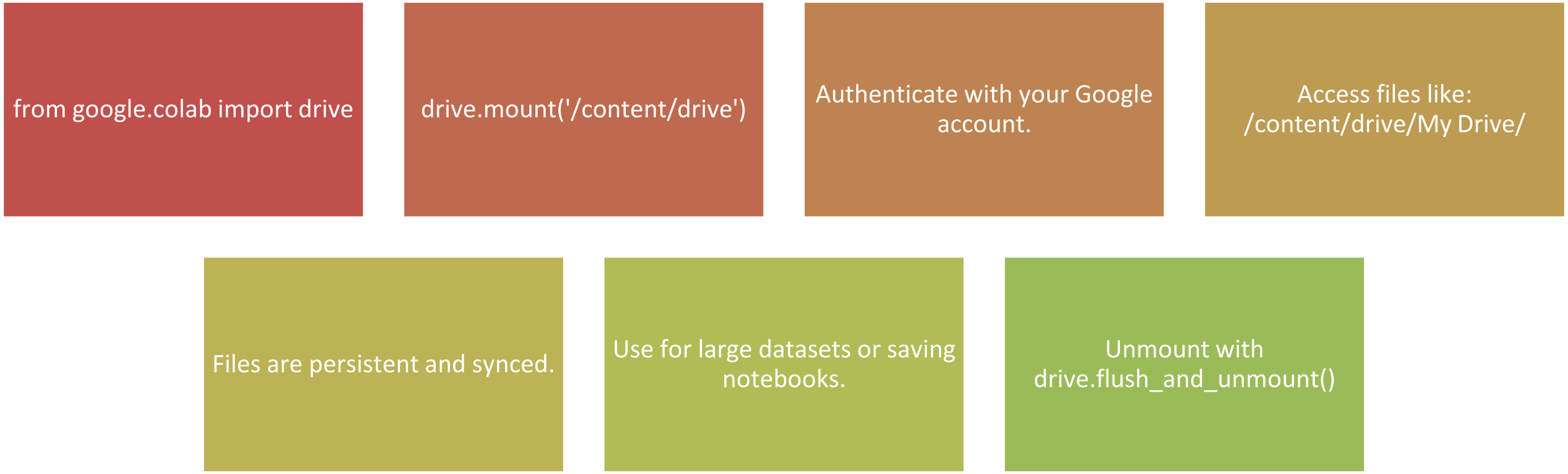
Reconnect if idle for too long.

Uploading Files



- from google.colab import files
- files.upload() opens file selector dialog.
- Uploaded files are stored in /content/
- Access files using open(), pandas, etc.
- Files disappear after session ends.
- Reupload or mount Drive for persistence.
- Use drag-and-drop for quick upload.

Mounting Google Drive



from google.colab import drive

drive.mount('/content/drive')

Authenticate with your Google account.

Access files like: /content/drive/My Drive/

Files are persistent and synced.

Use for large datasets or saving notebooks.

Unmount with drive.flush_and_unmount()

Installing Packages

USE !PIP INSTALL PACKAGE-NAME

EXAMPLE: !PIP INSTALL SEABORN

WORKS IN NOTEBOOK CELL LIKE A TERMINAL COMMAND.

PACKAGES ARE INSTALLED INTO COLAB VM.

SESSION-SPECIFIC INSTALLS (LOST ON DISCONNECT).

CAN USE CONDA VIA %SHELL MAGIC IF NEEDED (TO BE EXPLAINED LATER)

COMMON LIBRARIES: NUMPY, PANDAS, MATPLOTLIB, ETC.

Working with DataFrames

Use pandas for tabular data manipulation.

```
import pandas as pd
```

```
df = pd.read_csv('file.csv')
```

df.head() shows first few rows.

df.describe() for statistics.

Use df.plot() for quick visualization.

df.to_csv() to export data.

Visualizing Data

- Use matplotlib and seaborn for plots.
- import matplotlib.pyplot as plt
- plt.plot(x, y), plt.scatter(), etc.
- sns.heatmap(df.corr()) for correlation matrix.
- Plots appear inline in the notebook.
- Interactive charts possible with Plotly.
- Customize with titles, labels, legends.

Practical tutorials

- Overview of Collaboratory Features

- Colab Primier
- https://colab.research.google.com/github/google/picatrix/blob/main/notebooks/Quick_Primer_on_Colab_Jupyter.ipynb

- https://colab.research.google.com/notebooks/basic_features_overview.ipynb

- Basic Python tutorial
- https://colab.research.google.com/github/cs231n/cs231n.github.io/blob/master/python-colab.ipynb

- Detailed Python tutorial
  https://colab.research.google.com/drive/1E8_Hxr4AECBJNAwoyah0m0ZwQCrekIRQ?usp=sharing

- Other resources
- https://colab.google/resources/