

Módulo 2:

Adquisición y Preparación de Datos

Google Colab –

MARIA DE LOS ANGELES CONSTANTINO GONZALEZ

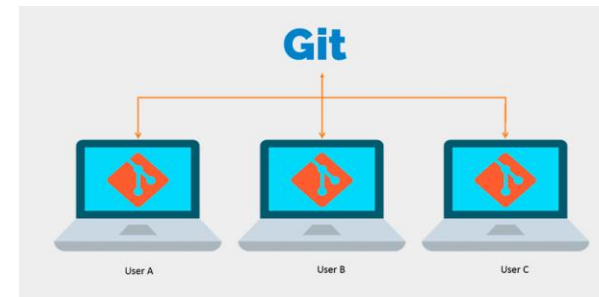


Tecnológico
de Monterrey



Herramientas para Desarrollo de proyectos en la nube

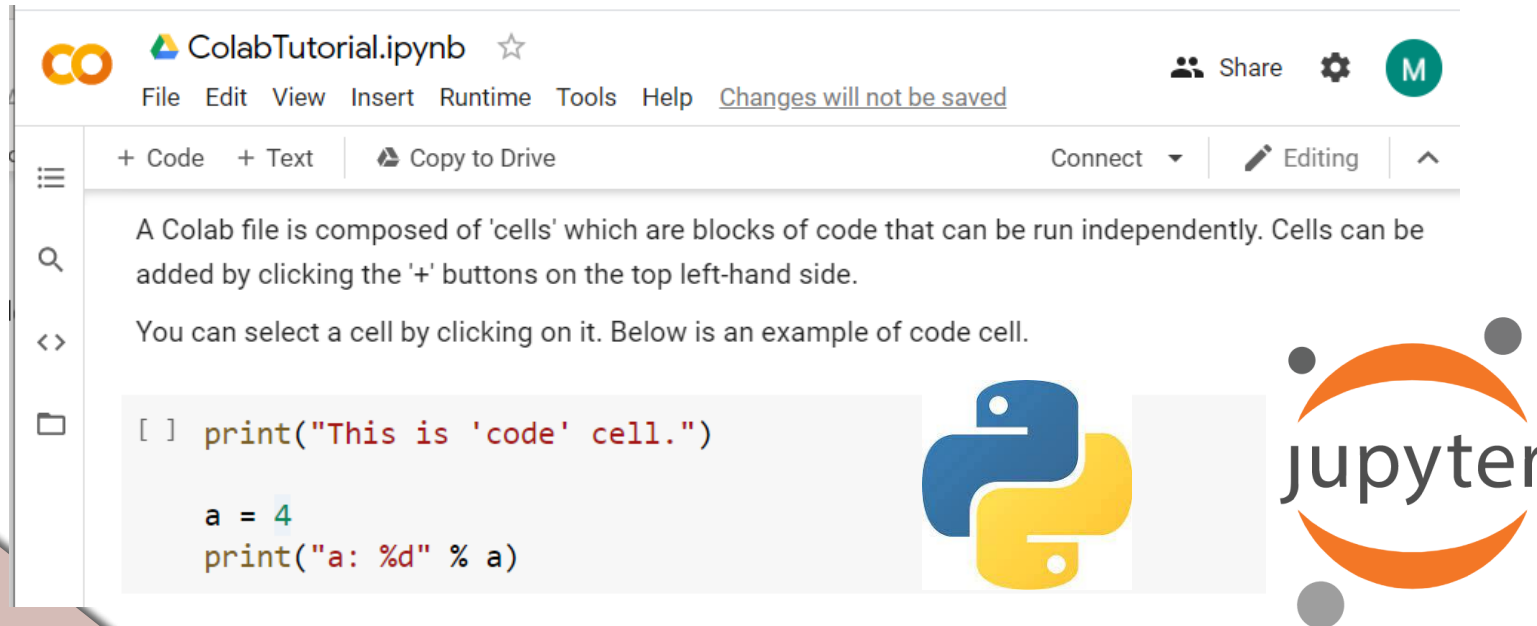
- Google Colab
 - Introducción
 - Notebooks
 - Google Drive
- GitHub
- Git



Google Colab (Colaboratory)

Introducción

- Herramienta **en la nube** creada por Google para codificar **código en Python** y facilitar el desarrollo de modelos de **Machine Learning**.
- Utiliza el formato de **Jupyter notebook**. Los cuadernos (*notebooks*) pueden guardarse en **Google Drive** y pueden **compartirse** entre varios desarrolladores.
- Permite **acceso gratuito** a procesamiento más rápido mediante el uso de GPUs y TPUs



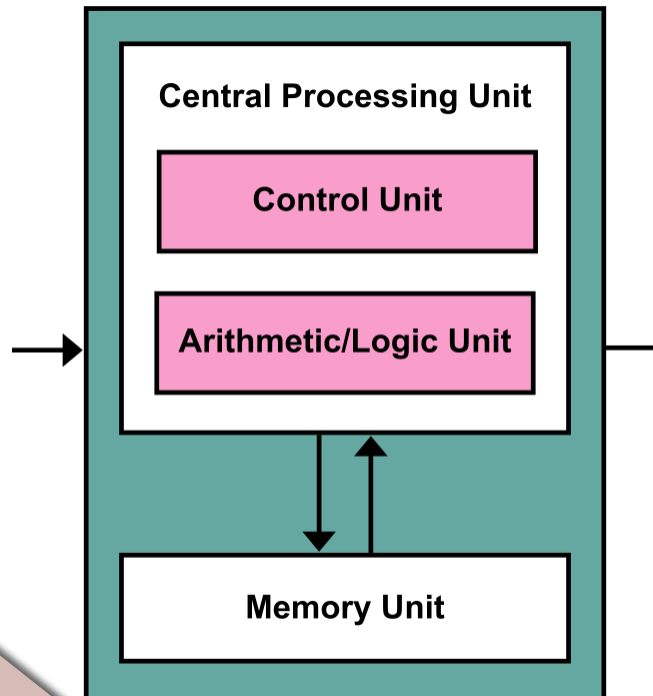
- Permite liberar a nuestra computadora de tener que llevar a cabo un trabajo demasiado costoso tanto en tiempo como en procesamiento

Google Colab (Colaboratory)

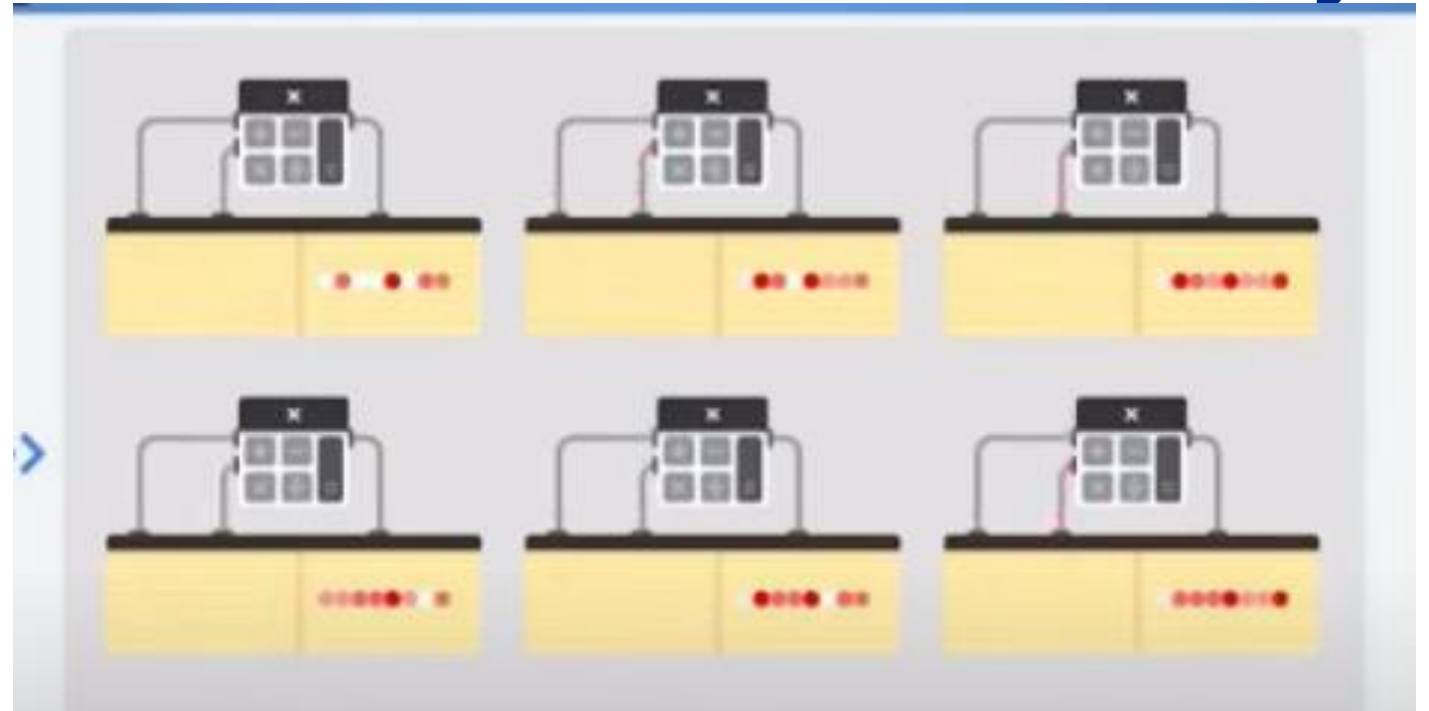
Introducción - GPUs

- Estrategia: Arquitectura que permite alojar entre 2500 y 5000 ALUs (Unidad Aritmética Lógica) para realizar operaciones más rápidamente que el CPU.
- Permite entrenar más rápidamente a los modelos de Deep Learning.

CPU



GPU



Google Colab (Colaboratory)

Introducción - TPUs

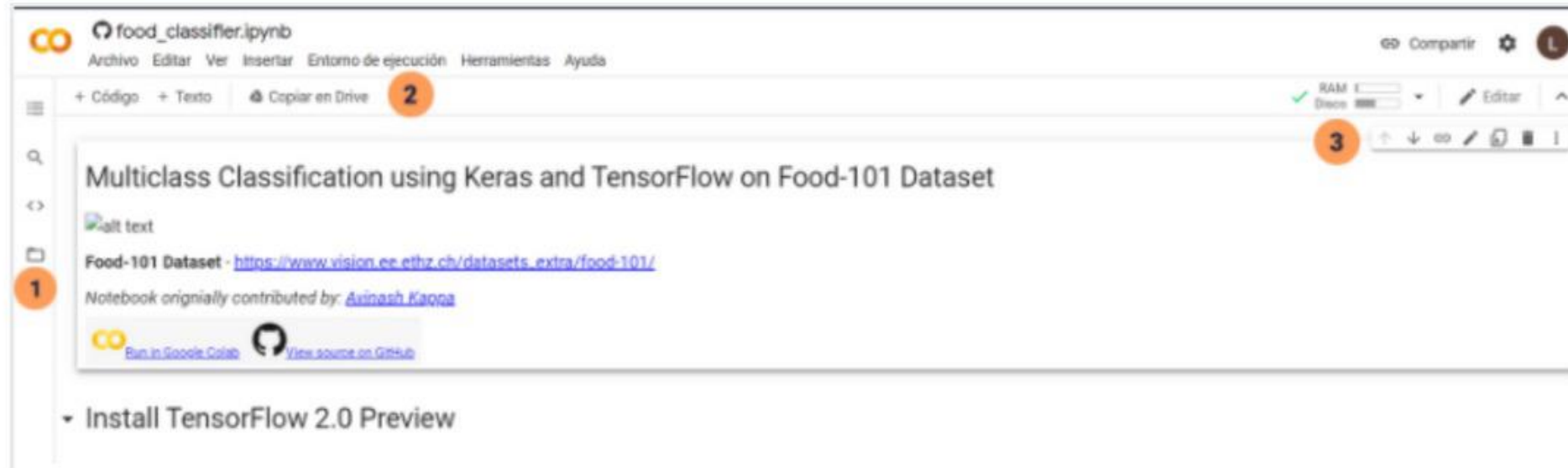
- Arquitectura desarrollada específicamente para Deep Learning
- Optimizada para la realización de miles de operaciones simultáneas con matrices, operaciones utilizadas frecuentemente en Redes Neuronales



Nota:

Para usar TPUs se requiere que los bloques de datos de entrenamiento y prueba sean múltiplos de 128

Google Colab



En la imagen se señala la pestaña (1) donde pueden subir archivos de datos que usarán para entrenar, la sección de entorno (2) donde están los ajustes para comprobar que el entorno dispone de GPU y la información del entorno (3) que indica si está conectado y de qué recursos dispone la máquina.

Si quieres acceder al sistema operativo que hay en la máquina asignada, puedes poner "!" al principio de la línea de código en la celda en la que estes y puedas escribir comandos

Google Colab

Introducción – Ventajas - Requisitos

- **Permite utilizar las bibliotecas más populares de Python para analizar y visualizar datos.**
 - Numpy
 - Matplotlib
 - Pandas
- **Tiene soporte para muchas bibliotecas de Machine Learning:**
 - TensorFlow y Keras
 - Pytorch
- **Se conecta con Google Drive y GitHub**
- **Puede importar datasets de Kaggle (comunidad de Ciencia de Datos)**

Requisitos

- Navegador y Cuenta de Google
- Conocimiento básico de Python

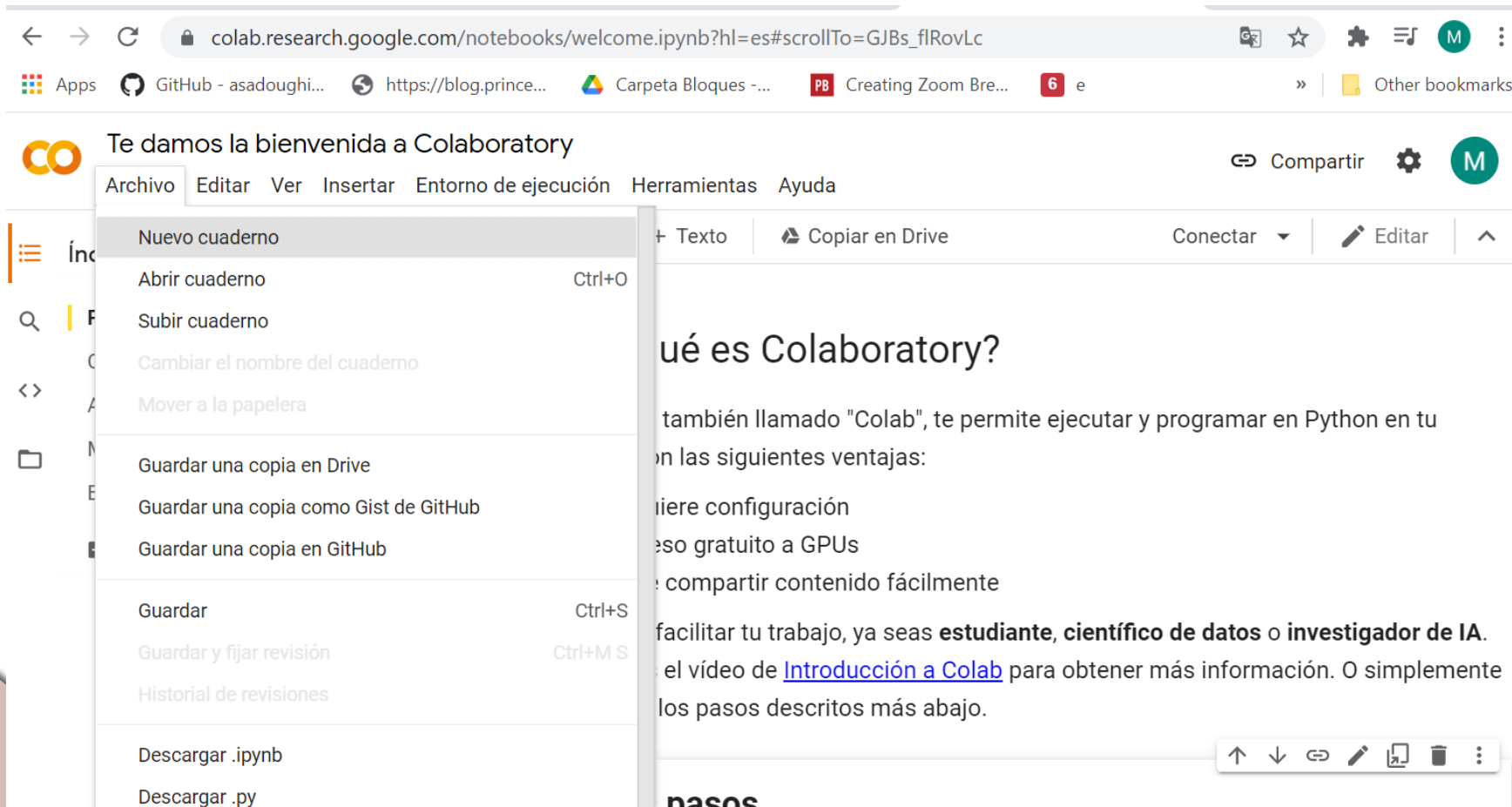
Google Colab Notebooks

¿Qué es un cuaderno?

- **Un cuaderno es un documento que contiene celdas con código ejecutable** (por ejemplo, Python) **y también celdas con elementos de texto enriquecido** (links, figuras, etc.)
- Las “**celdas**” que son bloques de código pueden ejecutarse de manera independiente.
- Este enfoque modular facilita cambiar partes del código sin tener que volver a ejecutar todo.

Google Colab Notebooks - Crear cuaderno

- Ingresar en el navegador a colab.research.google.com
- Abrir un nuevo cuaderno: Archivo – Nuevo cuaderno



Google Colab Notebook: Crear cuaderno

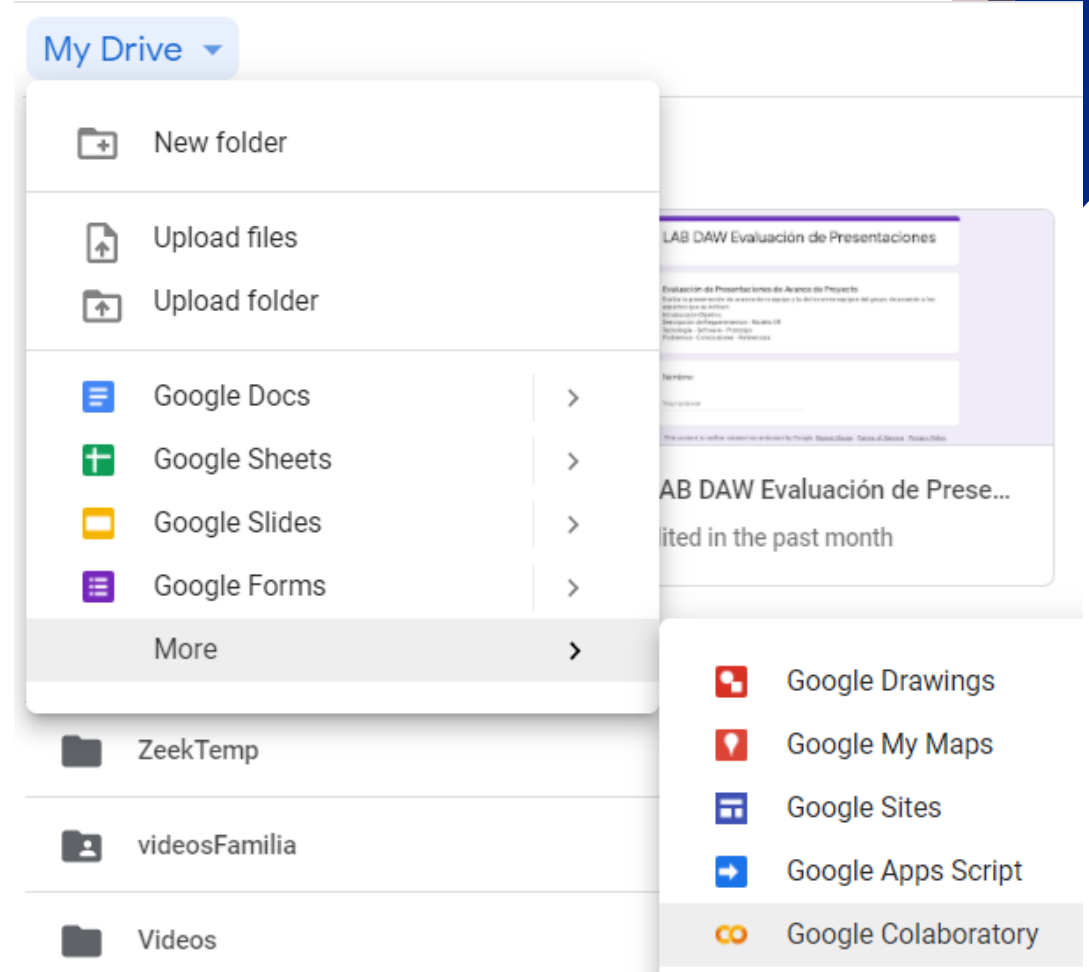


Renombrar: Ejemplo1
Guardar

Se guarda en drive.

Se crea automáticamente una carpeta llamada “**Colab Notebooks**” donde se pueden encontrar todos los cuadernos creados a partir de **Google Colab**.

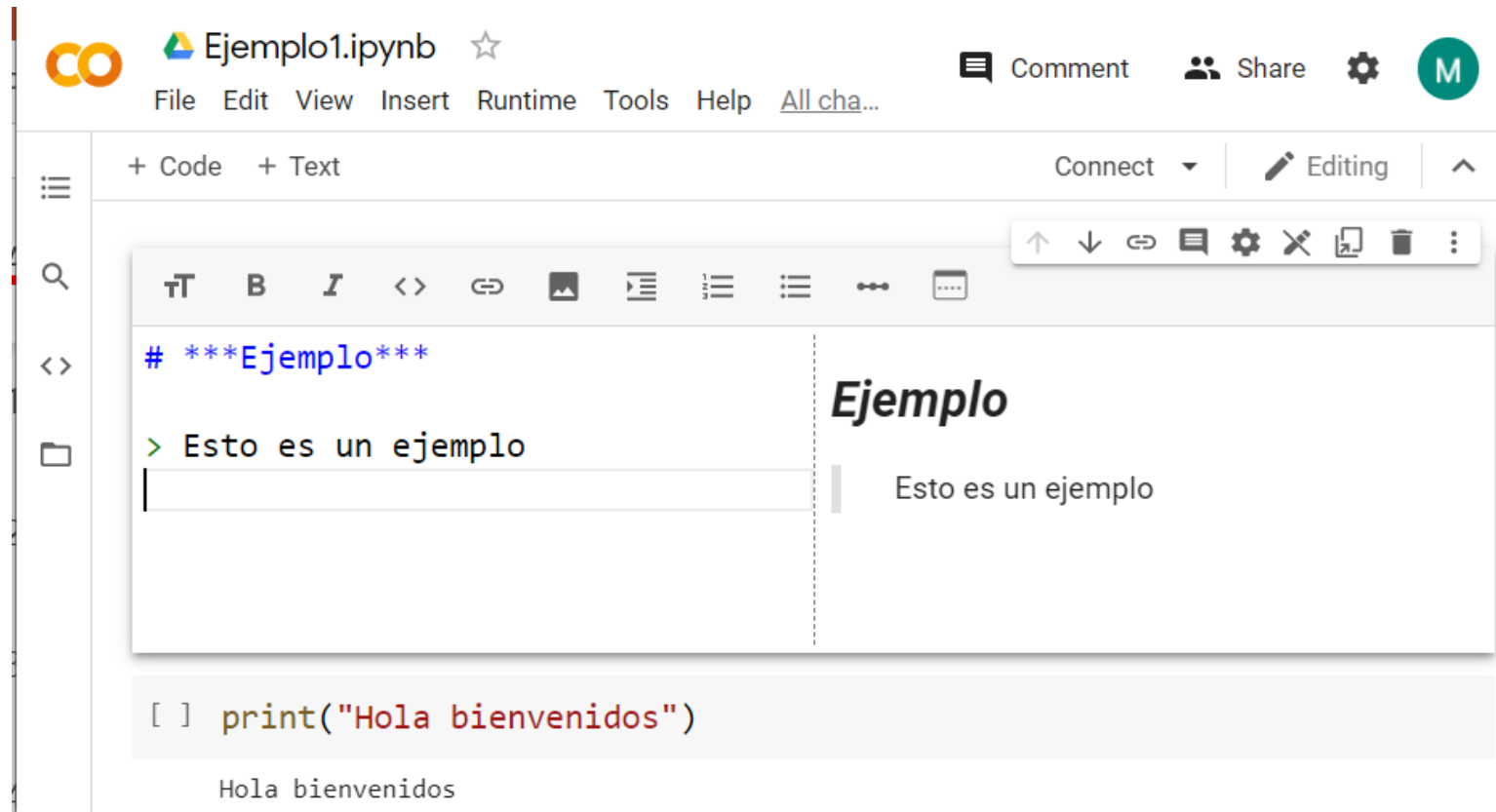
Desde Google Drive



Google Colab

Cuaderno - Texto enriquecido

- Las celdas de texto se formatean usando un lenguaje de etiquetas llamado Markdown.



Google Colab

Notebook - Definir Entorno de Ejecución

- Se puede definir el entorno de ejecución (runtime) a utilizar:
 - CPU (None),
 - GPU
 - TPU



Google Colab Notebooks - Ejecución

▼ Ejemplo

Esto es un ejemplo

```
print("Hola, bienvenidos")
```

Hola, bienvenidos

```
a=5
```

```
[2] b=10
```

```
[3] a+b
```

```
-----  
NameError                                Traceback (most recent call last)  
<ipython-input-3-ca730b97bf8a> in <module>()  
----> 1 a+b
```

NameError: name 'a' is not defined

SEARCH STACK OVERFLOW



Ejemplo1.ipynb ☆

File Edit View Insert Runtime Tools Help [Saving...](#)

+ Code + Text

▼ Ejemplo

Esto es un ejemplo

```
[ ] print("Hola, bienvenidos")
```

Hola, bienvenidos

```
[5] a=5
```

```
[2] b=10
```

```
▶ a+b
```

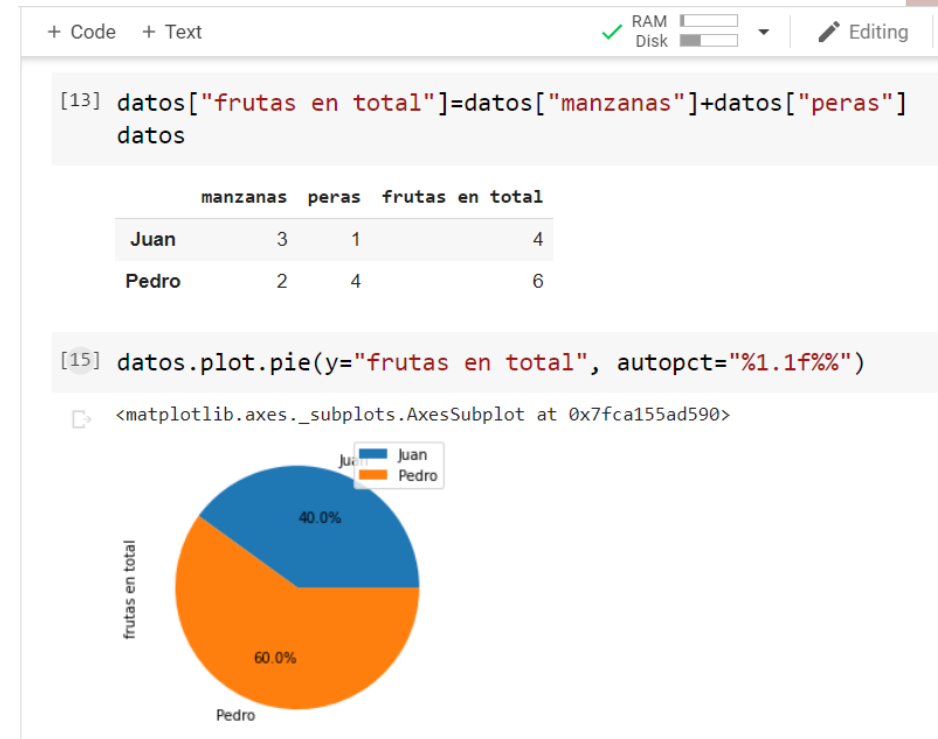
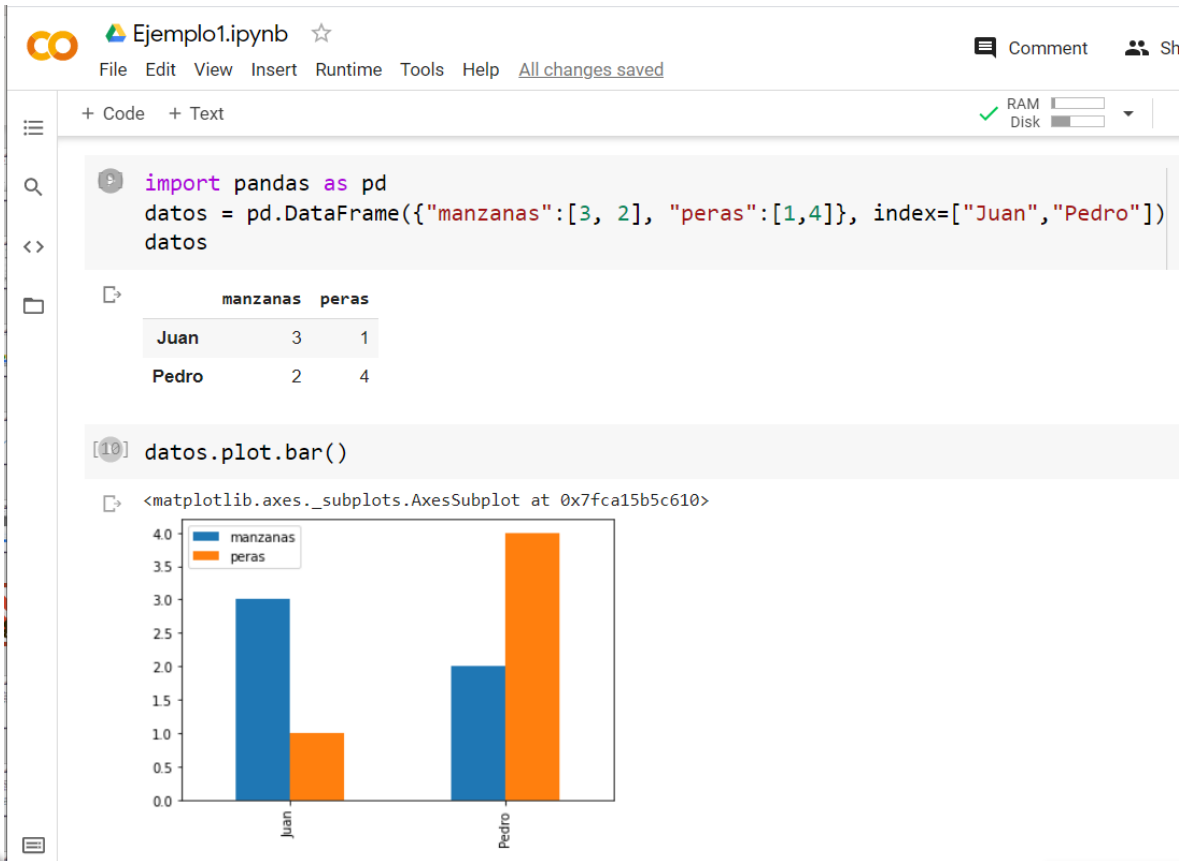
15

Ejecutar código

Ctrl+Enter (ejecutar celda) o
Shift+Enter (ejecutar celda y saltar a la siguiente)

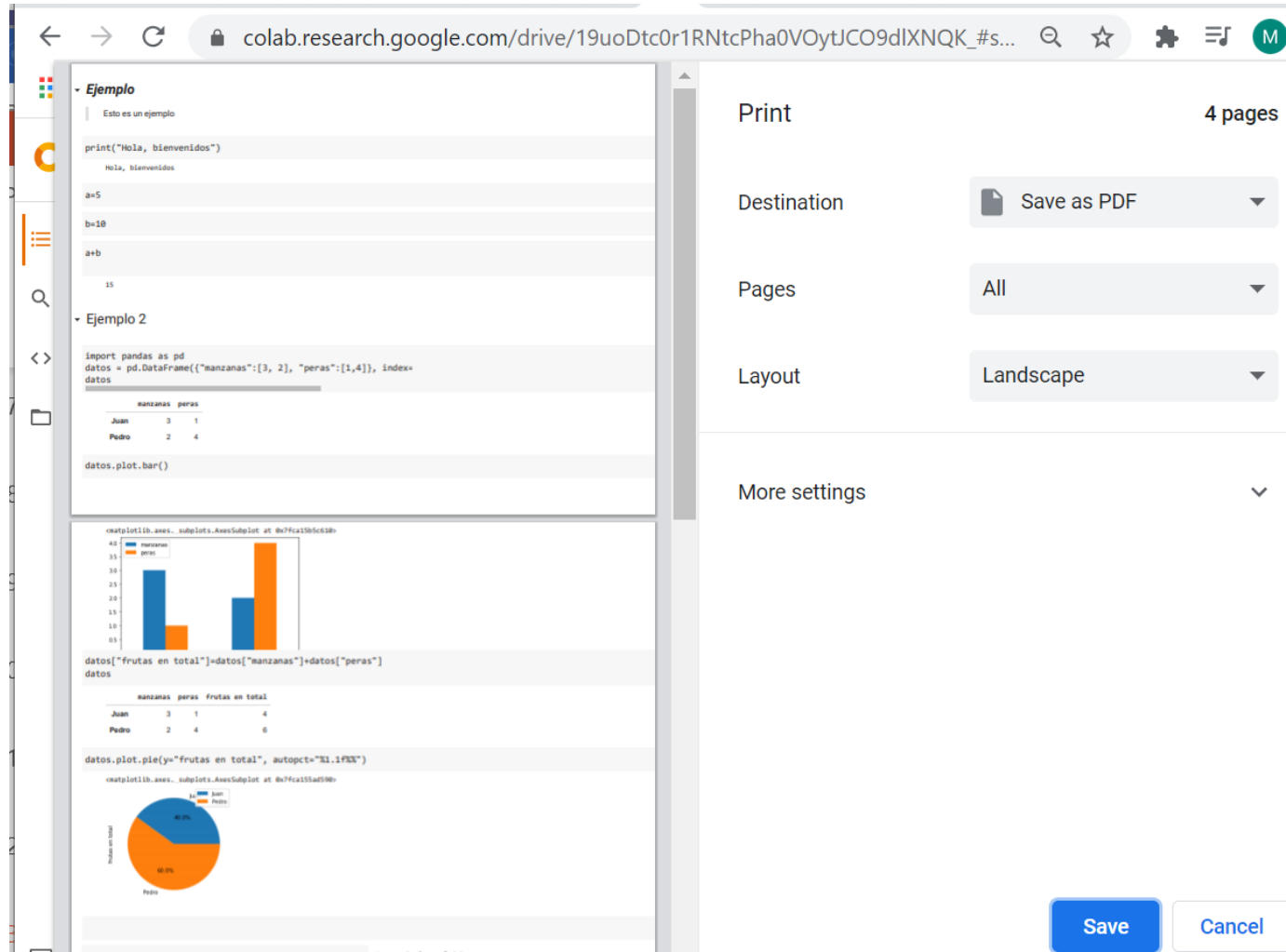
Google Colab

Notebooks - Uso de librerías



Google Colab Notebooks - Guardar como PDF

Print – Save as PDF



The screenshot shows the Google Colab interface. The notebook content includes:

- Ejemplo**
Esto es un ejemplo

```
print("Hola, bienvenidos")
```

Hola, bienvenidos

```
a=5  
b=10  
a+b
```

15
- Ejemplo 2**

```
import pandas as pd  
datos = pd.DataFrame({"manzanas": [3, 2], "peras": [1, 4]}, index=[  
datos
```

	manzanas	peras
Juan	3	1
Pedro	2	4

```
datos.plot.bar()
```

A bar chart is displayed with 'manzanas' in blue and 'peras' in orange. The y-axis ranges from 0 to 40.

```
datos["frutas en total"] = datos["manzanas"] + datos["peras"]  
datos
```

	manzanas	peras	frutas en total
Juan	3	1	4
Pedro	2	4	6

```
datos.plot.pie(y="frutas en total", autopct="%1.1f%%")
```

A pie chart is displayed showing the distribution of 'frutas en total' for Juan and Pedro. Juan's share is 40.0% (blue) and Pedro's share is 60.0% (orange).

The print sidebar on the right shows:

- Print** 4 pages
- Destination** Save as PDF
- Pages** All
- Layout** Landscape
- More settings** (dropdown arrow)

At the bottom right, there are **Save** and **Cancel** buttons.

Google Colab

Notebooks: Acceso a Datos en Google Drive

Desde Google Colab se pueden cargar o guardar datos en Google Drive

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



Go to this URL in a browser
Enter your authorization code:
.....
Mounted at /content/drive

Por seguridad pide código de autorización.
Copiar código de autorización y pegarlo en el lugar indicado dentro del notebook.

Google

Sign in

Please copy this code, switch to your application and paste it there:

4/1AY0e-g5dPzhLgUC-zg-ijz9NvbPiivs3nG-
zwjDXHFS1_MYbg08ywXPcbkE



Google Colab

Notebooks – Leer Datos de Google Drive

Table of contents

- Ejemplo
- Ejemplo 2
- Leer Datos desde Google Drive
- Section

+ Code + Text

RAM Disk Editing

Leer Datos desde Google Drive

```
[16] from google.colab import drive
drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

```
[18] cd "/content/gdrive/My Drive/TC2004B"
```

/content/gdrive/My Drive/TC2004B

Leer datos dentro del directorio

```
[19] mis_datos = pd.read_csv("titanic.csv")
```

```
[20] mis_datos.head()
```

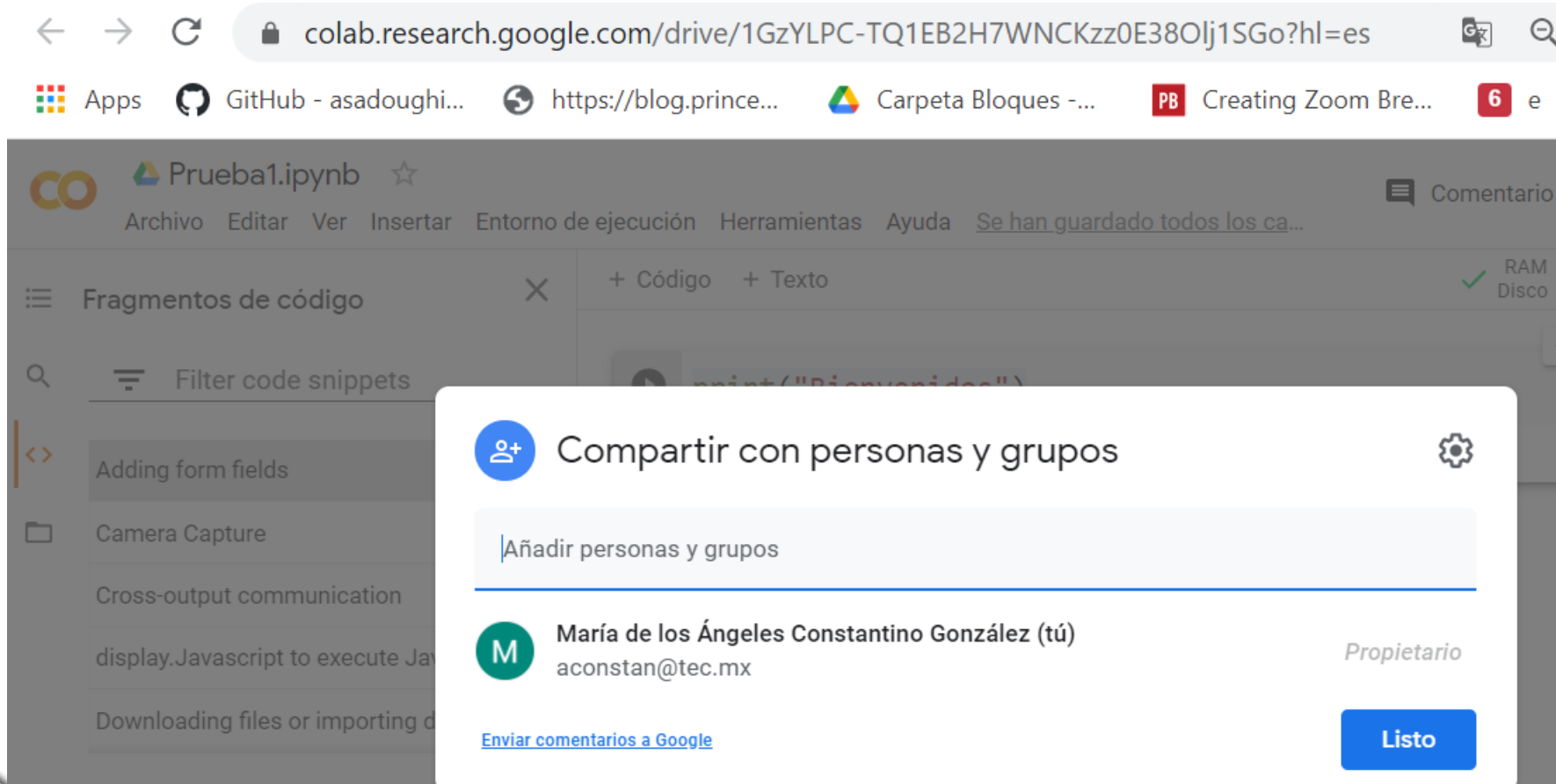
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171

Cumings.

Moverse al directorio en donde se encuentra los archivos

Todos los archivos en Google serán accesibles desde la carpeta llamada "My Drive".

Google Colab – Notebooks - Compartir Cuaderno



•Puedes compartir los resultados con los demás. .

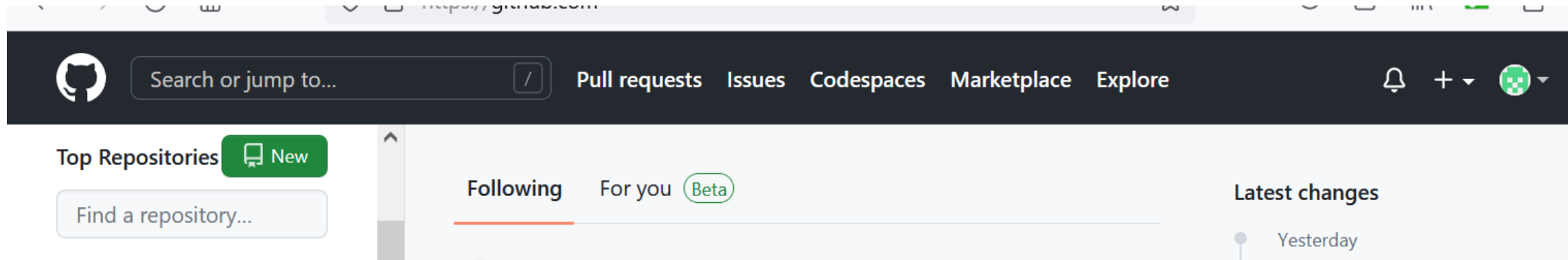
•Nota: No puede haber más de una persona editando un cuaderno al mismo tiempo.

Compartir link:

<https://colab.research.google.com/drive/19uDtc0r1RNtcPha0VOytJCO9dIXNQK?usp=sharing>

GitHub

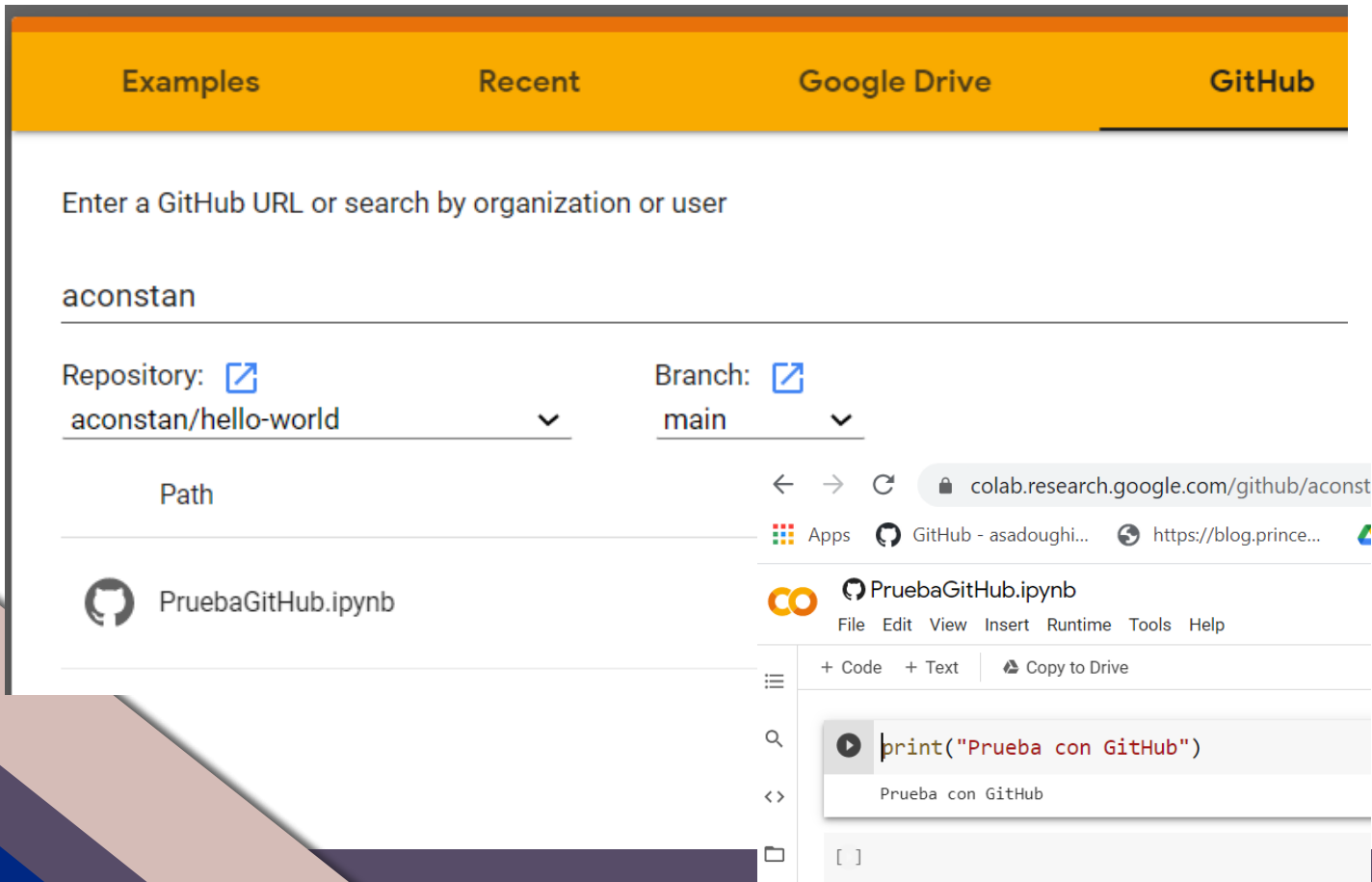
- GitHub es una plataforma que permite crear repositorios web para alojar proyectos y que otras personas puedan acceder a ellos
- Está basada en el sistema de control de versiones Git, el cual permite llevar un estricto control de los cambios que varias personas realizan sobre un archivo.



Google Colab / GitHub

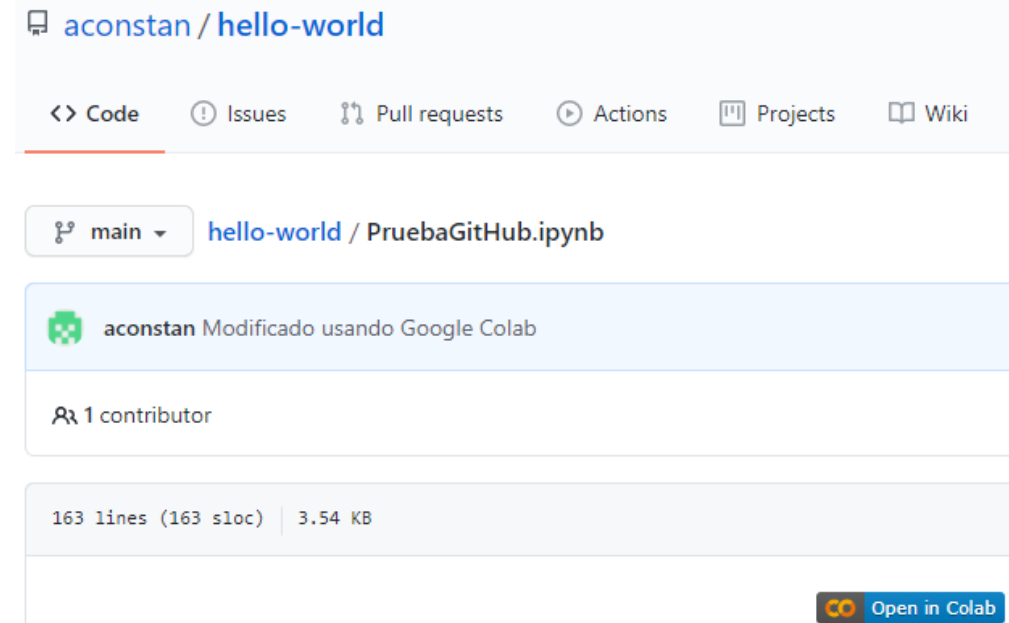
- Google Colab proporciona integración con Github.
- Es posible cargar cuadernos directamente de los repositorios, ya sea desde Google Colab o desde Git-Hub

Google Colab



The screenshot shows the Google Colab web interface. At the top, there is a navigation bar with tabs: Examples, Recent, Google Drive, and GitHub. Below this, a search bar prompts the user to "Enter a GitHub URL or search by organization or user". The search results show the repository "aconstan/hello-world" on the "main" branch. The file "PruebaGitHub.ipynb" is selected. The interface includes a path bar, a file explorer, and a code editor. The code editor shows a single line of Python code: `print("Prueba con GitHub")`. The bottom of the interface displays the file name "PruebaGitHub.ipynb" and a toolbar with options like "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help".

GitHub



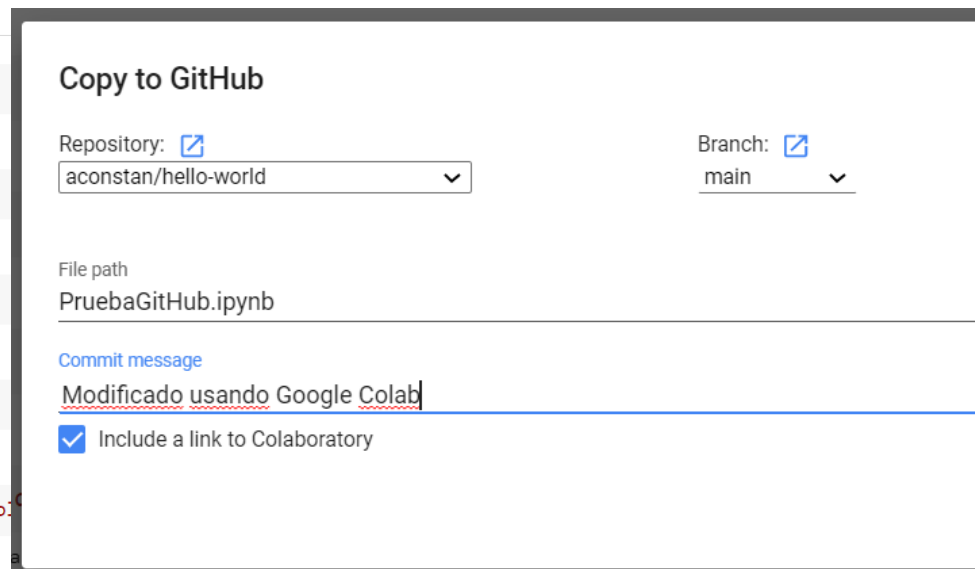
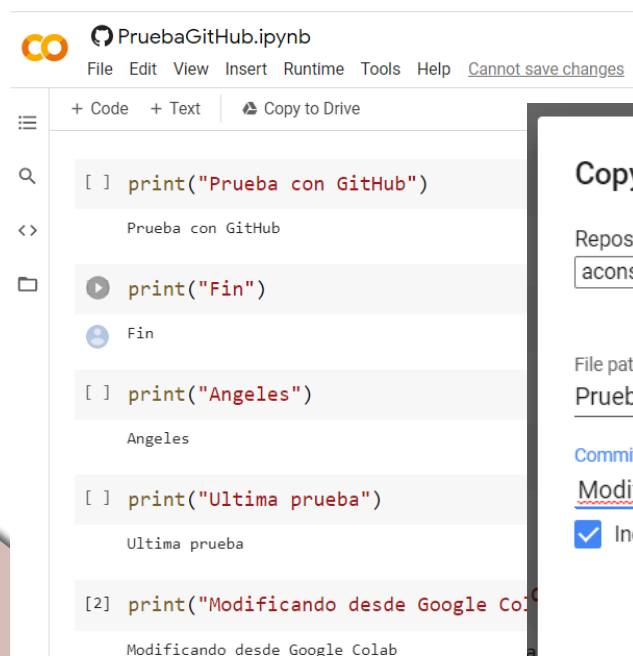
The screenshot shows the GitHub repository page for "aconstan/hello-world". The page includes a navigation bar with links to Code, Issues, Pull requests, Actions, Projects, and Wiki. The repository is on the "main" branch. The file "PruebaGitHub.ipynb" is selected. The file details show it was modified using Google Colab, has 1 contributor, and contains 163 lines of code (163 sloc) with a size of 3.54 KB. At the bottom right, there is a button labeled "Open in Colab" with the Google Colab logo.



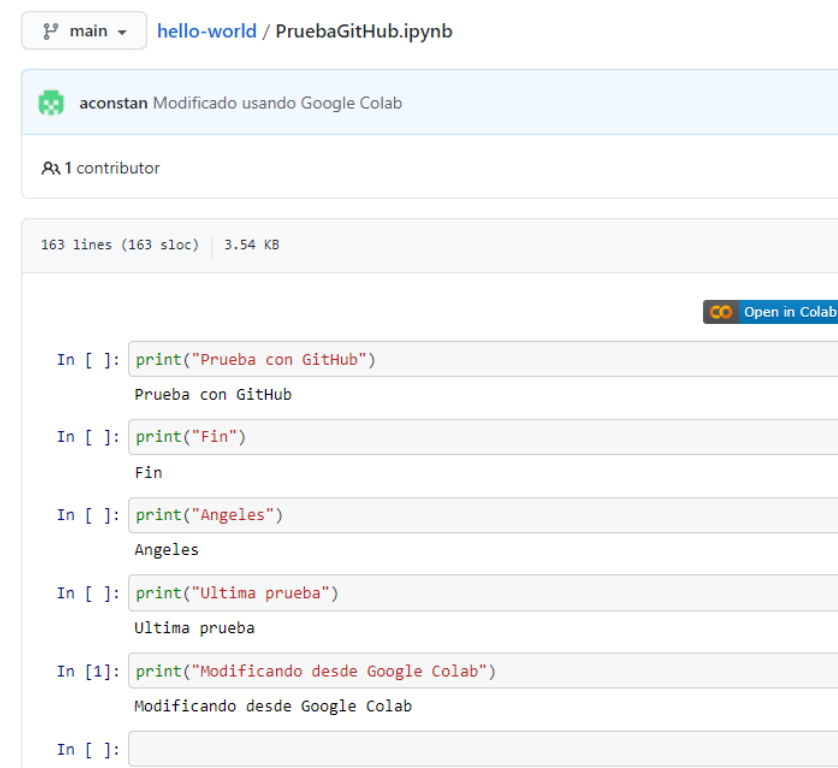
Google Colab / Git-Hub

- Es posible guardar cuadernos directamente en Github, indicando URL del repositorio
- Es equivalente a hacer commit. En **“Commit Mesagge”** incluir detalle de los cambios realizados

Google Colab



GitHub



References

- Introducción a Google Colab: Una noble y completa guía:
https://www.youtube.com/watch?v=8VFYs3Ot_aA
- Google Colab Tutorial 2021 en Español:
<https://www.youtube.com/watch?v=0mOV4pIF2Xo>
- https://en.wikipedia.org/wiki/Von_Neumann_architecture