



Octubre 2023

# ACT1\_1: Gestión de ficheros en Google Colab



Carolina María Montesdeoca Álvarez  
SNS

## Actividad 1.1 – Gestión de ficheros en Google Colab

El objetivo de esta actividad es conocer con mayor detalle la gestión de ficheros en Google Colab. Para lo cual utilizaremos el cuaderno del siguiente enlace y completar el código para realizar cada uno de los ejercicios propuestos con el prefijo #Ejer:

### Ejercicio 01:

```
[9] #Ejer 01: Crear en el PC un fichero con un contenido de Lorem Ipsum y subirlo
#Importamos la librería
from google.colab import files

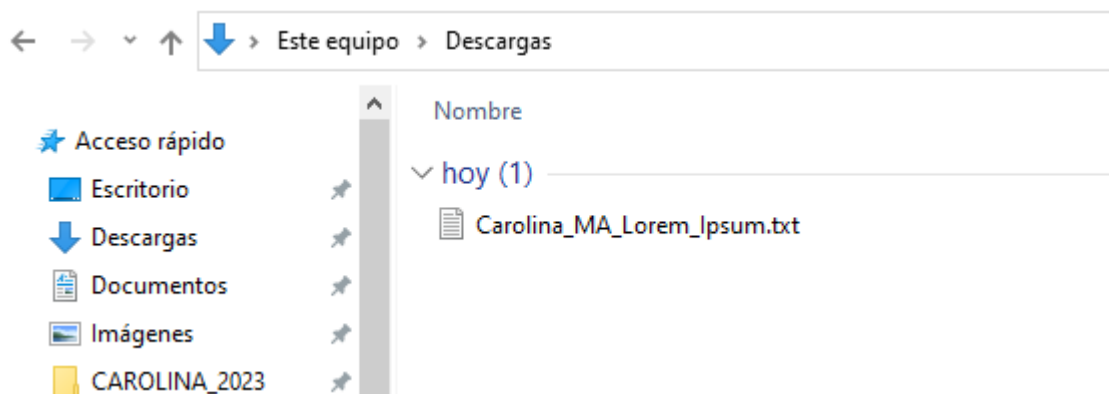
#Subimos el fichero y mostramos información de la subida
uploaded = files.upload()
for fn in uploaded.keys():
    print('User uploaded file "{name}" with length {length} bytes'.format(
        name=fn, length=len(uploaded[fn])))
```

Elegir archivos Ninguno archivo selec. Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.  
Saving Carolina\_MA\_Lorem\_Ipsum.txt to Carolina\_MA\_Lorem\_Ipsum.txt  
User uploaded file "Carolina\_MA\_Lorem\_Ipsum.txt" with length 1056 bytes

### Ejercicio 02:

```
#Ejer 02: Descargar el fichero subido en el apartado anterior

files.download('Carolina_MA_Lorem_Ipsum.txt')
```



## Ejercicio 03:

```

▶ #Importante: Se recomienda hacer todos los ejercicios tomando como punto de partida el directorio raíz de Drive,
#Ejer 03: Crear en drive un directorio llamado GoogleColab_test_Ejer_23_24 y dentro de ella un fichero con un contenido
# de Lorem Ipsum. Consultar/verificar su creación en Google Drive y en el navegador de Google Colab
#En las celdas anteriores ya se ha importado files y desmontamos la unidad de google drive

#Importamos las librerías
from google.colab import drive
from google.colab import files

#Montamos la unidad de google drive
drive.flush_and_unmount
drive.mount('/content/drive', force_remount=True)

Mounted at /content/drive

```

```

✓ [12] #A continuación se crea el directorio y el fichero
0s

folder_name = 'GoogleColab_test_Ejer_23_24'
folder_path = '/content/drive/My Drive/'+folder_name+'/'
file_name = 'Lorem Ipsum CMA.txt'
mkdir "{folder_path}"
with open(folder_path+file_name, 'w') as f:
    f.write('Lorem ipsum dolor sit amet consectetur adipiscing elit commodo nec, tempor dis montes interdum.')

#Si ponemos la ruta directamente se muestra el contenido de la carpeta
print("\nContenido de la carpeta:")
ls -l '/content/drive/My Drive/GoogleColab_test_Ejer_23_24/'

#Utilizando variables de Python se muestra el contenido del fichero
print("\nContenido del fichero:")
cat '{folder_path}{file_name}'

Contenido de la carpeta:
total 1
-rw----- 1 root root 95 Oct 23 08:04 'Lorem Ipsum CMA.txt'

Contenido del fichero:
Lorem ipsum dolor sit amet consectetur adipiscing elit commodo nec, tempor dis montes interdum.

```

Mi unidad ▾



Tipo ▾

Personas ▾

Modificado ▾

Nombre ↑	Propietario	Última modificación ▾	Tamaño de a	⋮
Colab Notebooks	yo	22 oct 2023 yo	—	⋮
GoogleColab_test_Ejer_23_24	yo	9:04 yo	—	⋮

Mi unidad &gt; GoogleColab\_test\_Ejer\_... ▾



Tipo ▾ Personas ▾ Modificado ▾

Nombre ↑	Propietario	Última modificación ▾	Tamaño de s	⋮
Carolina_Montesdeoca_Alvarez.txt	yo	9:16 yo	542 bytes	⋮
Lorem Ipsum CMA.txt	yo	9:04 yo	95 bytes	⋮

## Ejercicio 04:

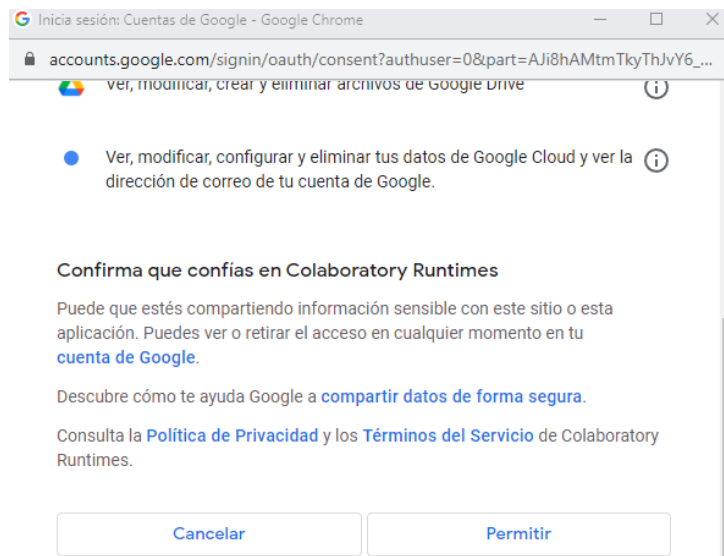
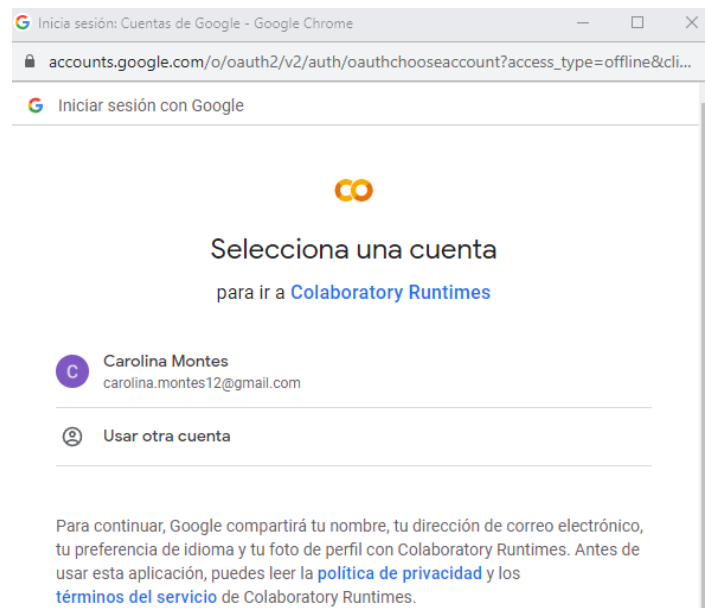
```
[13] #Ejer 04: En la carpeta creada en el ejercicio anterior (GoogleColab_test_Ejer_23_24), crear un fichero cuyo nombre sean tu nombre y apellidos,
# en su contenido incluir un texto LoremIpsum distinto al utilizado en el punto anterior (de 50 palabras y añadir tu nombre al principio)
# y al terminar mostrar su contenido.
# Similar al ejemplo. Validar su existencia consultando Gdrive.ListFile y el explorador de archivos de GoogleColab en este notebook

# Url's de referencia:
# https://pythonhosted.org/PyDrive/filemanagement.html
# https://pythonhosted.org/PyDrive/filelist.html#get-all-files-which-matches-the-query
# https://stackoverflow.com/questions/56434084/google-pydrive-uploading-a-file-to-specific-folder

#Importamos librerías
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials

#Primero hay que autenticarse para crear el cliente Drive
auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)
```

The screenshot shows a Google Colab notebook titled "Carolina\_Montesdeoca\_Alvarez\_Act\_1\_1\_Datos\_externos\_archivos\_locales\_Drive\_Hojas de cálculo\_y...". The code in the notebook is the same as shown in the previous block. A permission dialog box is displayed in the foreground, asking: "¿Permitir que este cuaderno acceda a tus credenciales de Google?". The dialog explains that this will allow the code to access Google Drive and Google Cloud data. There are two buttons: "No, gracias" and "Permitir".



```
[14] #Obtenemos el id de la carpeta en la raíz del directorio drive.
def get_folder_id_in_root(folder_name):
    file_list = drive.ListFile({'q': '"root" in parents and trashed=false'}).GetList()
    for f in file_list:
        if f['title'] == folder_name:
            return f['id']
    return 0

folder_name = 'GoogleColab_test_Ejer_23_24'
folder_id = get_folder_id_in_root(folder_name)

#Creamos el fichero dentro del directorio indicando su id por parámetro.
file_name_ejer04 = 'Carolina.Montesdeoca.Alvarez.txt'
file = drive.CreateFile({'title': file_name_ejer04, 'parents': [{'id': folder_id}]})
file.SetContentString('Sociis molestie eleifend sociosqu fames scelerisque libero rutrum, arcu risus tincidunt porta massa sem nam, class ultricies rhoncus inceptos semper facilisis. Dictumst odio auctor leo luctus posuere')
file.Upload()
```

```

✓ [15] #Comprobamos los ficheros del directorio y el contenido del fichero
file_list = drive.ListFile({'q': "'{}' in parents and trashed=false".format(folder_id)}).GetList()
for f in file_list:
    print('title: %s, id: %s' % (f['title'], f['id']))

check_file = drive.CreateFile({'id': file.get('id')})
print('\nContenido del fichero (ID: {}):\n {}".format(file.get('id'), check_file.GetContentString()))

title: Carolina_Montesdeoca_Alvarez.txt, id: 1zUtA-pzeeJEE6TyZFwYXhJPhdORGSh5
title: Lorem Ipsum CMA.txt, id: 1-D8oa2WhER9jfy7nO2I6waH1V5-vMZVp

Contenido del fichero (ID: 1zUtA-pzeeJEE6TyZFwYXhJPhdORGSh5):
"Sociis molestie eleifend sociosqu fames scelerisque libero rutrum, arcu risus tincidunt porta massa sem nam, class ultricies rhoncus inceptos semper facilisis. Dictumst odio auctor leo luctus posuere duis

```

Mi unidad &gt; GoogleColab\_test\_Ejer\_...

Tipo Personas Modificado


Nombre	Propietario	Última modificación	Tamaño de s
 Carolina_Montesdeoca_Alvarez.txt	 yo	9:16 yo	542 bytes
 Lorem Ipsum CMA.txt	 yo	9:04 yo	95 bytes


GoogleColab\_test\_Ejer\_23\_24



Detalles


Actividad


Hoy



 Has subido 2 elementos  
9:16 23 oct

 GoogleColab\_test\_Ej...
 

 Carolina\_Mont...
  Lorem Ipsum C...


 Has creado un elemento en  
9:04 23 oct

 Mi unidad
 

 GoogleColab\_t...

Ejercicio 05:

```

✓ [16] #Ejer 05
#Validamos la subida del fichero con Gdrive.ListFile
#Comprobamos los ficheros del directorio
file_list = drive.ListFile({'q': "'{}' in parents and trashed=false".format(folder_id)}).GetList()
for f in file_list:
    print('title: %s, id: %s' % (f['title'], f['id']))

title: Carolina_Montesdeoca_Alvarez.txt, id: 1zUtA-pzeeJEE6TyZFwYXhJPhdORGSh5
title: Lorem Ipsum CMA.txt, id: 1-D8oa2WhER9jfy7nO2I6waH1V5-vMZVp

```

Ejercicio 06:

```

# Ejer 06: Haciendo uso de la api (ejemplo anterior) descargar/mostrar el contenido del fichero que creaste con contenido LoremIpsum
# en la carpeta GoogleColab_test_Ejer y con el nombre tuyo.

b"111 Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s w passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum."

✓ [19] #Ejer 06
from google.colab import auth
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from oauth2client.client import GoogleCredentials

auth.authenticate_user()
gauth = GoogleAuth()
gauth.credentials = GoogleCredentials.get_application_default()
drive = GoogleDrive(gauth)

```

Carolina Montesdeoca Álvarez  
Página - 5 - de 7

```

#Ejer 06
#La solución consiste en obtener el id en el enlace que se genera al compartir el fichero en Google Drive
file_id = "1zUTA-pzeeJE6TyZFwYKH3Phd0RGSns"

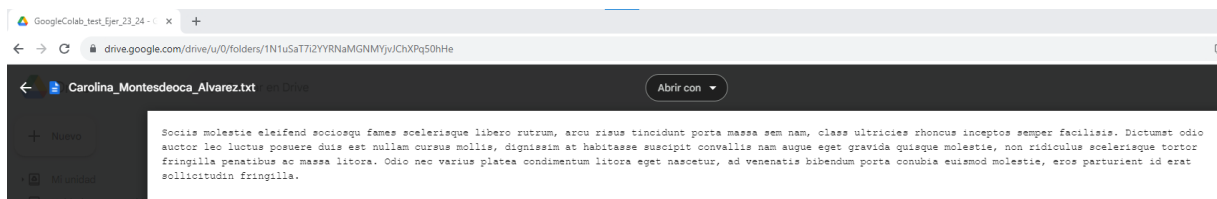
import io
from googleapiclient.http import MediaIoBaseDownload

request = drive_service.files().get_media(fileId=file_id)
downloaded = io.BytesIO()
downloader = MediaIoBaseDownload(downloaded, request)
done = False
while done is False:
    # _ is a placeholder for a progress object that we ignore.
    # (Our file is small, so we skip reporting progress.)
    _, done = downloader.next_chunk()

downloaded.seek(0)
print('Downloaded file contents are: {}'.format(downloaded.read()))

Downloaded file contents are: b'Sociis molestie eleifend sociosqu fames scelerisque libero rutrum, arcu risus tincidunt porta massa sem nam, class ultricies rhoncus inceptos semper facilisis. Dictumst odio auctor leo

```



## Ejercicio 07:

```

[00] # Ejer 07: Subir a un repositorio de GitHub una hoja de cálculo en formato csv y realizar la importación a un dataframe,
# utilizando como separador de campos un "punto y coma";

#Importamos la librería
import pandas as pd

#Importamos los datos
repositorio = 'https://raw.githubusercontent.com/carolProg/SNS_23_24/main/a05003423-producto-interior-bruto-pib-por-namas-de-actividad-cnae-09-islas-de-canarias-y-anos.csv'

df_train_pib = pd.read_csv(repositorio, sep=';')

df_train_pib

```

Fuente: [https://datos.gob.es/en/catalogo?theme\\_id=ciencia-tecnologia](https://datos.gob.es/en/catalogo?theme_id=ciencia-tecnologia)

[https://github.com/carolProg/SNS\\_23\\_24](https://github.com/carolProg/SNS_23_24)

<https://colab.research.google.com/drive/1b0XGVVyu6hdbp7G2N9tvx5NAEpOA1Yj#scrollTo=cJPamA5z49TD&uniqifier=1>