

# Google Colaboratory (autore: Vittorio Albertoni)

## Premessa

Con questo manualetto entriamo nel mondo del cloud computing.

Il cloud computing consiste nell'erogazione di servizi di stoccaggio di dati e/o di elaborazione di dati attraverso la rete Internet da parte di strutture attrezzate per questo.

Gli utenti di questi servizi sono aziende o persone che decidono di risolvere in tutto o in parte il loro fabbisogno di Information Technology senza investimento diretto in hardware e software ma ricorrendo a chi dà in uso questi strumenti, generalmente dietro corrispettivo.

Attualmente circa il 70% di servizi di questo tipo è offerto dai quattro principali fornitori, in ordine di importanza: Amazon (Amazon Web Services), Microsoft (Microsoft Azure), Google (Google Cloud Platform) e la cinese Alibaba (Alibaba Cloud). Ve ne sono tanti altri, tra i quali ricordo gli onnipresenti, quando si parla di IT, IBM e Oracle.

Tra questi, l'operatore tradizionalmente più aperto al mondo del software libero e più disponibile anche all'offerta di opportunità a dilettanti senza prosciugarne la borsa è Google, che ci offre la chicca che descrivo in questo manualetto: Google Colaboratory.

Google Colaboratory ci consente di scrivere codice in linguaggio Python su un browser web e di ottenerne l'elaborazione da parte di Google, con potenze di molto superiori a quelle normalmente disponibili sui nostri personal computer, tablet o addirittura smartphone, senza dover affrontare particolari configurazioni o predisposizioni.

Fino a certi livelli e pur con potenza variabile nel tempo in maniera imprevedibile possiamo fruire gratuitamente di questo servizio.

Per avere certezze sulla disponibilità dei livelli di servizio occorre ovviamente sottoscrivere versioni a pagamento.

Se le nostre esigenze non necessitano di particolari pianificazioni e, soprattutto, se da inesperti non abbiamo contezza di ciò che veramente ci serve abbiamo comunque la possibilità di provare ad «imbucarci» gratuitamente e vedere cosa succede.

Senza, peraltro, aver provato a risolvere i problemi con la nostra attrezzatura casalinga.

Ritengo, infatti, che l'aver a disposizione il linguaggio Python nel cloud non debba farci perdere l'interesse ad avere a disposizione lo stesso linguaggio a casa nostra, visto che non costa nulla in termini economici e costa poco in termini di occupazione del nostro disco fisso.

In altri termini ritengo che la vera utilità di Google Colaboratory sia semplicemente quella di fornirci una potenza di calcolo superiore a quella che ci può dare il nostro computer, se e quando serve.

# Indice

<b>1</b>	<b>Preliminare</b>	<b>3</b>
<b>2</b>	<b>Attrezzatura</b>	<b>3</b>
<b>3</b>	<b>Accesso al servizio</b>	<b>3</b>
<b>4</b>	<b>Come funziona</b>	<b>5</b>
<b>5</b>	<b>Cosa c'è dietro</b>	<b>6</b>
<b>6</b>	<b>Risorse hardware e loro disponibilità</b>	<b>7</b>
<b>7</b>	<b>Come caricare i dati in Google Colab</b>	<b>8</b>
7.1	Dati presenti nel filesystem locale . . . . .	8
7.2	Dati presenti in Google Drive . . . . .	9
7.3	Dati da Internet . . . . .	9

## 1 Preliminare

Visto che Google Colab (da qui in poi uso il termine abbreviato) è un servizio Google, per potervi accedere dobbiamo avere un account Google, quello che ci può servire anche per la posta elettronica e che, appunto, si identifica con un formato da indirizzo di posta elettronica del tipo

`nome@gmail.com`

Molto probabilmente già l'abbiamo e l'abbiamo sicuramente se possediamo uno smartphone equipaggiato Android.

Se ancora non l'abbiamo possiamo facilmente ottenerlo andando all'indirizzo <https://support.google.com/accounts/answer/27441?hl=it#> dove scopriremo, se non ci interessa avere il servizio di posta elettronica gmail, di poter utilizzare anche il nostro di altra natura.

## 2 Attrezzatura

Per accedere a Google Colab ci serve semplicemente un browser web e un collegamento a Internet.

I browser su cui è testato il funzionamento ottimale del servizio sono Chrome, Firefox e Safari.

Probabilmente funziona bene anche con altri browser: posso personalmente garantire che ciò avviene per Chromium.

Non ha nessuna importanza il sistema operativo e la macchina su cui lavoriamo, basta che abbiamo la possibilità di collegarci a Internet e di utilizzare un browser web adatto al servizio.

L'unica altra cosa che serve è conoscere il linguaggio di programmazione Python.

Per imparare questo linguaggio, che considero abbastanza facile, esiste moltissima documentazione in libreria e in rete.

Nel mio blog all'indirizzo [www.vittal.it](http://www.vittal.it) si trovano alcuni articoli e manuali adatti allo scopo e il tutto è radunato nel testo «Tutto Python per principianti» che si trova su Amazon.

## 3 Accesso al servizio

Il modo normale per accedere a Google Colab è quello di andare all'indirizzo

<https://colab.research.google.com/?hl=it>

Se abbiamo il browser con l'accesso a Google già predisposto il collegamento al servizio avverrà immediatamente.

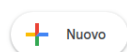
In caso contrario ci verrà richiesto di accedere a Google tramite l'account di cui abbiamo parlato nel Capitolo 1, inserendo nome dell'account e password.

Per questa via, ad accesso avvenuto, ci troviamo di fronte la finestra riprodotta all'inizio della prossima pagina.

Avendo il browser con l'accesso a Google già predisposto, possiamo accedere a Google Colab anche attraverso Google Drive.

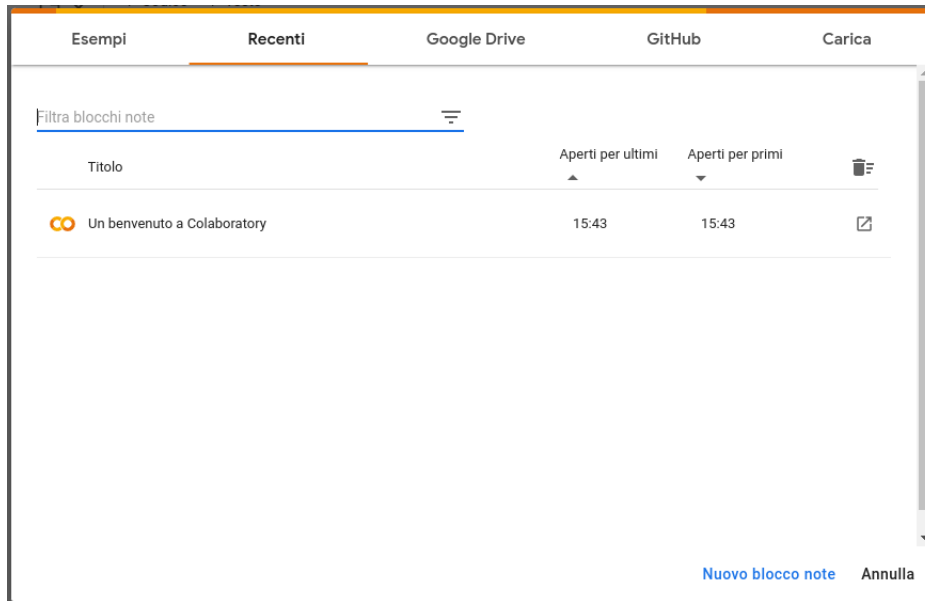
In questo caso, dal momento che le finestre di lavoro di Google Colab vengono salvate nel Drive in file con estensione `.ipynb`, possiamo aprire una finestra di lavoro già presente con doppio click su di essa.

Oppure possiamo aprire una nuova finestra in cui lavorare cliccando sul pulsante



nella finestra di Google Drive e scegliere dal menu che si presenta ALTRO ▷ GOOGLE COLABORATORY.

Con l'accesso normale, se è la prima volta che utilizziamo Google Colab, ci si presenta questa finestra

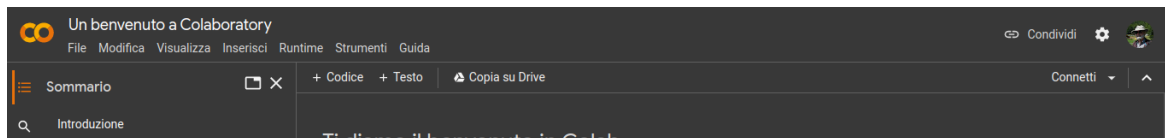


Essa si apre, per default, sulla scheda RECENTI, dove sono elencate le ultime cose che abbiamo fatto in modo da poterle richiamare.

Cliccando sulla voce visualizzata, UN BENVENUTO A COLABORATORY, oppure cliccando sul tasto ANNULLA apriamo una finestra di benvenuto in cui ci viene descritto il servizio:



Se clicchiamo sul pulsante  in alto a destra attiviamo la visualizzazione di una intestazione



utile in quanto contiene un menu che ci dà modo di avviare i nostri lavori.

Sempre in alto a destra abbiamo il pulsante  con cui apriamo una finestra di dialogo autoesplicativa per alcune impostazioni.

La finestra con sfondo scuro appena riprodotta deriva dalla scelta della modalità dark come tema del sito, in luogo della modalità adaptive di default che caratterizza la finestra di apertura precedentemente riprodotta.

Oltre alla scheda RECENTI, la finestra di apertura contiene le seguenti altre schede:

- . ESEMPI: che contiene file di esempio,
- . GOOGLE DRIVE: che visualizza e ci consente di aprire file salvati su Google Drive,
- . GITHUB: che consente di collegarsi ad un account GitHub,
- . CARICA: che consente di caricare un file da una directory locale.

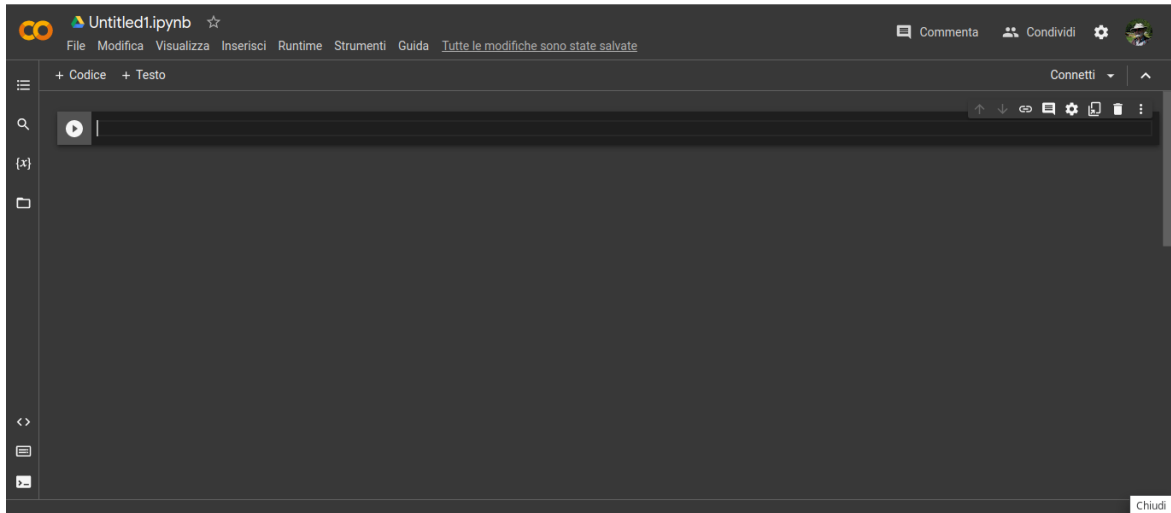
## 4 Come funziona

Dalla finestra di apertura possiamo avviare un nuovo lavoro cliccando sulla voce NUOVO BLOCCO NOTE in fondo a destra.

Per proseguire o modificare un lavoro già avviato possiamo avvalerci della scheda RECENTI o GOOGLE DRIVE oppure, se il file è stato salvato in una directory locale, CARICA.

Possiamo fare le stesse cose, se siamo nella finestra di benvenuto, agendo da menu FILE ▷ NUOVO BLOCCO NOTE, oppure FILE ▷ APRI BLOCCO NOTE e FILE ▷ CARICA BLOCCO NOTE.

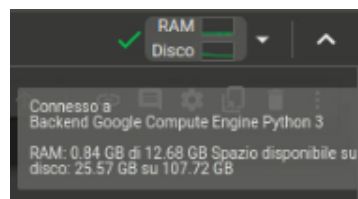
Una nuova finestra di lavoro si presenta così



e scopriamo che si tratta di un notebook Jupyter: da qui il ricorrente appellativo di blocco note o, in inglese, notebook con cui si identificano le finestre di lavoro di Google Colab.

Possiamo rinfrescare la nostra conoscenza di Jupyter ricorrendo al manualetto in formato PDF «Jupyter» allegato all'articolo «Un editor speciale» postato nel luglio 2022 sul mio blog all'indirizzo [www.vittal.it](http://www.vittal.it).

Per cominciare a lavorare dobbiamo connetterci e lo facciamo cliccando sulla voce CONNETTI, in alto, sulla destra, sotto la barra di intestazione. Dopo qualche istante saremo connessi e, in luogo della voce CONNETTI, avremo una doppia finestrella su cui possiamo posizionare il mouse ottenendo



conferma di essere connessi ad un motore Python3, con quanto ci è inizialmente assegnato di una RAM di complessivi 12,68 GB e quanto spazio ci è inizialmente assegnato su un disco di 107,72 GB.

La prima cosa che facciamo è assegnare un nome al nostro notebook, selezionandone con il mouse, nella barra dell'intestazione, quello assegnato per default (Untitled1) e modificandolo a nostro piacimento, conservando l'estensione .ipynb.

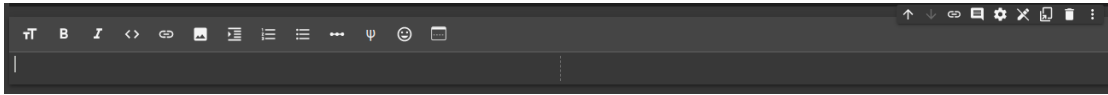
All'apertura il notebook presenta la finestra per l'immissione di codice



A cella selezionata abbiamo, in alto a destra, una piccola barra di strumenti: passando il mouse sulle icone troviamo le spiegazioni.

Il codice inserito si esegue cliccando sul triangolino sulla sinistra oppure premendo insieme i tasti MAIUSCOLO e INVIO della tastiera.

Per aprire una finestra che serva per immettere testo clicchiamo sulla voce + TESTO nella barra degli strumenti ed otteniamo l'inserimento nel nostro note book della riga



Con la finestra aperta per inserire testo o modifiche, oltre alla solita piccola barra di strumenti in alto a destra, abbiamo una barra di strumenti con icone che facilitano l'immissione di tag nel linguaggio Markdown, utilizzato in Jupyter Notebook per scrivere testi formattati.

Per fissare nel notebook il testo chiudiamo la finestra di inserimento premendo insieme i tasti MAIUSCOLO e INVIO della tastiera.

Per tornare all'inserimento di codice, posizionati nel punto del notebook dove vogliamo fare l'inserimento, clicchiamo sulla voce + CODICE e così via.

Tutto ciò che facciamo viene automaticamente salvato nel nostro Google Drive, in una directory Colab Notebooks. A fine lavoro abbiamo comunque la possibilità di salvare utilizzando le voci del menu, una delle quali ci consente di salvare il nostro notebook su GitHub.

Possiamo stampare il nostro lavoro da menu FILE > STAMPA, avendo modo di scegliere tra stampante collegata oppure salvataggio in formato PDF.

Con menu FILE > SCARICA possiamo salvare il file del notebook in locale sul nostro computer, dove possiamo lavorarlo avvalendoci dello Jupyter Notebook che abbiamo sul nostro computer ed eventualmente ricaricarlo poi su Google Colab da menu FILE > CARICA BLOCCO NOTE.

## 5 Cosa c'è dietro

Già abbiamo visto, esplorando la modalità di connessione, di essere connessi ad un motore Python3 (Python2 non è più supportato).

Potremmo volerne conoscere la versione e possiamo farlo inserendo in una finestrella del codice il seguente codice

```
import platform
print(platform.python_version())
```

che, eseguito, fornisce il risultato



Altra curiosità quella di conoscere il sistema operativo su cui lavoriamo nel cloud e possiamo appagarla inserendo in una finestrella del codice il seguente codice

```
import platform
print(platform.system())
```

che, eseguito, fornisce il risultato



Possiamo anche conoscere la versione di singoli moduli che importiamo per l'uso. Per esempio, per conoscere la versione di numpy che ci viene fornita, inseriamo il codice

```
import numpy as np
print (np.__version__)
```

ottenendo



## 6 Risorse hardware e loro disponibilità

Da quanto visto finora sappiamo che la risorsa software che ci offre Google Colab è un interprete Python 3 utilizzabile attraverso Jupyter Notebook su sistema operativo Linux.

Sul piano dell'hardware, nella versione gratuita e per default, ci vengono messi a disposizione una RAM fino a 12,68 GB, un disco di memoria fino a 107,72 GB e una CPU senza acceleratore.

Penso che il problema minore sia quello del disco di memoria, che ha una capacità che mi sembra veramente elevata, almeno per contenere dati numerici.

Quanto alla RAM possiamo avere un'idea di cosa significa avere a disposizione 12,68 GB valutando che in questa dimensione può essere contenuta una lista di poco oltre 1.500.000 elementi.

Lo possiamo sperimentare eseguendo in un notebook Google Colab il seguente codice

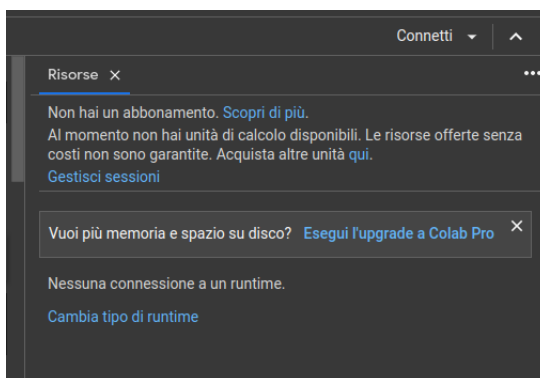
```
l = []
while(True):
    l.append('ciao')
```

che va in crash per esaurimento della RAM quando nella lista abbiamo la parola «ciao» ripetuta per circa 1.520.000 volte.

La CPU senza acceleratore è praticamente una cosa simile a quella che probabilmente abbiamo sul nostro computer casalingo.

Sempre nella versione gratuita possiamo utilizzare il nostro notebook continuamente per 12 ore e dopo 30 minuti di inattività veniamo disconnessi.

Se clicchiamo, in alto a destra, sul triangolino di fianco alla voce CONNETTI e, dal menu che si apre, scegliamo la voce RISORSE apriamo questa finestra



nella quale ci viene ricordato che siamo degli «imbucati» senza abbonamento, che non possediamo unità di calcolo e che nulla ci è garantito circa la disponibilità di risorse.

In realtà questa mancata garanzia si riferisce unicamente alle risorse di CPU.

Per le risorse in termini di memoria e spazio su disco abbiamo quelle viste sopra e per averne di più dobbiamo sottoscrivere l'abbonamento a Colab Pro (11,28 € al mese) o Colab Pro+ (51,54 € al mese).

Quanto, invece, alla CPU, il non avere unità di calcolo disponibili significa semplicemente che non ci è garantito di poter rafforzare la CPU, non che non lo possiamo fare.

Ci possiamo pertanto provare.

Se nella finestra appena sopra riprodotta clicchiamo sulla dizione CAMBIA TIPO DI RUNTIME (oppure scegliamo da menu RUNTIME > CAMBIA TIPO DI RUNTIME) possiamo accedere ad una finestra nella quale aprire un menu a tendina con cui possiamo scegliere un acceleratore hardware tra Nessuno, GPU e TPU.

A scelta effettuata, se non compaiono messaggi che lo negano, possiamo dedurre di essere riusciti a rafforzare la CPU con l'acceleratore scelto.

Con GPU (Graphics Process Unit) possiamo arrivare ad una velocità di elaborazione dalle 30 alle 80 volte superiore a quella della CPU senza acceleratore.

Con TPU (Tensor Process Unit) possiamo arrivare ad una velocità dalle 15 alle 30 volte superiore a quella della GPU.

Per avere la certezza di ottenere queste accelerazioni e di averle a disposizione per elaborazioni di una certa durata dobbiamo acquistare unità di calcolo.

Possiamo farlo senza avere abbonamenti, pagando per quello che usiamo, acquistando 100 unità per € 11,28 oppure 500 unità per € 51,54.

Sottoscrivendo l'abbonamento a Colab Pro abbiamo 100 unità al mese incluse nell'abbonamento, che sono 500 nel caso di abbonamento a Colab Pro+.

## 7 Come caricare i dati in Google Colab

Google Colab ha una directory radice che si chiama `content` nella quale deve essere presente il file dei dati da elaborare.

Il formato di questo file deve essere della specie riconosciuta dal modulo Pandas di Python. Tra i tanti cito i più noti: `.csv`, `.excel`, `.json`, `.html`, `.sas`, ecc.

Per il caricamento dei dati in questa directory abbiamo procedure diverse, a seconda di dove è collocato il file di dati da caricare.

### 7.1 Dati presenti nel filesystem locale

Se il file dei dati è archiviato sul computer su cui stiamo lavorando lo carichiamo eseguendo le seguenti istruzioni da scrivere come codice:

```
from google.colab import files
uploaded = files.upload()
```

che, all'esecuzione, mostra quanto segue




Cliccando su SCEGLI FILE apriamo una finestra di navigazione attraverso la quale raggiungiamo il file da caricare e il file, una volta caricato, è raggiungibile nella directory `/content` di Google Colab.


Questo è un esempio di elaborazione su un file `dati.csv` presente sul computer su cui lavoriamo e caricato in Google Colab secondo la procedura appena vista:



Possiamo caricare il file di dati presente sul computer su cui stiamo lavorando anche in quest'altro modo.

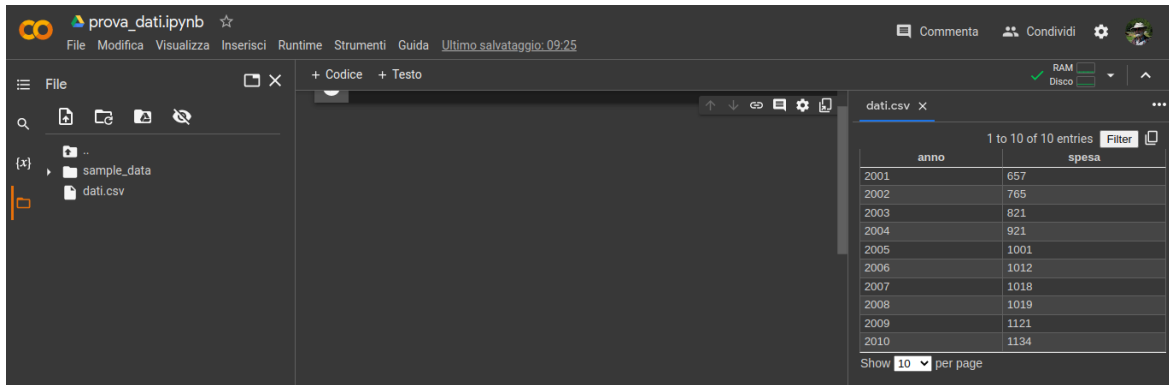


Se clicchiamo sull'icona  presente nella barra verticale sulla sinistra della finestra del notebook abbiamo accesso alla directory `/content`, nella quale, se non abbiamo ancora caricato dati, è presente una sottodirectory `sample_data`, contenente alcuni file utili per esercitazione.

Per caricare il nostro file locale possiamo cliccare sull'icona  e andare a scegliere il file attraverso la finestra di navigazione che si presenta e caricarlo.

Il file caricato è ora visibile nella directory `/content`.

Con doppio click sul suo nome possiamo anche vederne il contenuto.





## 7.2 Dati presenti in Google Drive

Se il file dei dati su cui lavorare si trova in Google Drive, dobbiamo montare il Drive in Google Colab.

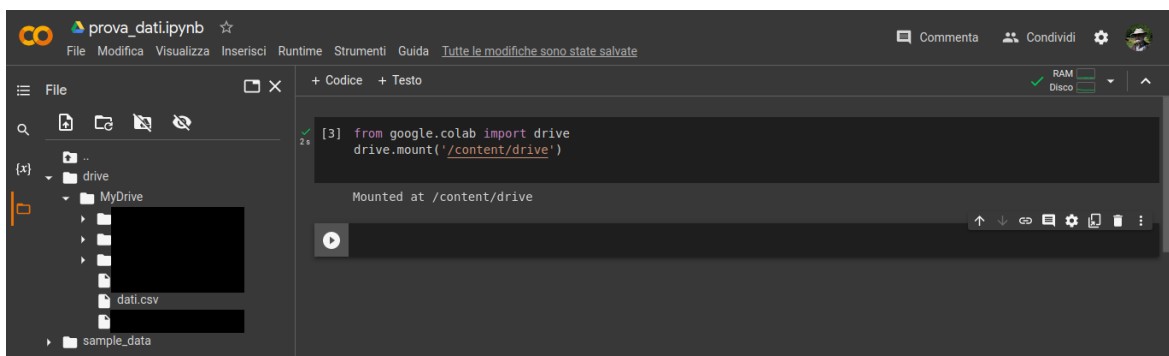
Lo possiamo fare eseguendo il seguente codice

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

e seguendo i passaggi che ci vengono proposti in una serie di finestre di dialogo.

In maniera più rapida otteniamo lo stesso risultato cliccando sull'icona  presente nella barra verticale sulla sinistra della finestra del notebook e successivamente sull'icona .

A drive montato la directory `/content` appare così



e il path per raggiungere il file di dati con Pandas diventa `'/content/drive/MyDrive/dati.csv'`.

## 7.3 Dati da Internet

La rete è una ricchissima fonte di dati di ogni tipo.

I siti che rendono disponibili dati molte volte offrono all'utente la possibilità di scaricare i dati selezionati semplicemente premendo determinati pulsanti o addirittura cliccando sul nome del dataset. E' il caso, per esempio, del nostro Istituto Nazionale di Statistica, all'indirizzo <https://www4.istat.it/it/istituto-nazionale-di-statistica>, dove basta cliccare sul nome del dataset e viene scaricato il relativo file in formato Excel.

Altre volte dobbiamo provvedere noi a scaricare il dataset e lo possiamo fare su Google Colab eseguendo la seguente istruzione

```
!wget <indirizzo_url> -P '/content'
```

per avere il dataset disponibile nella directory radice di Google Colab, da dove lo possiamo utilizzare, potendolo anche scaricare se vogliamo salvarlo nel file system del computer.

Teniamo comunque presente che il modulo Pandas, una volta che conosciamo l'indirizzo dove si trova il file del dataset e il formato del file che lo contiene, può collegarsi direttamente ad esso senza bisogno di scaricare nulla.

Esiste, per esempio, su GitHub un dataset che contiene tutto ciò che è avvenuto nella famigerata esperienza della recente epidemia Covid.

Il file del dataset è disponibile all'indirizzo

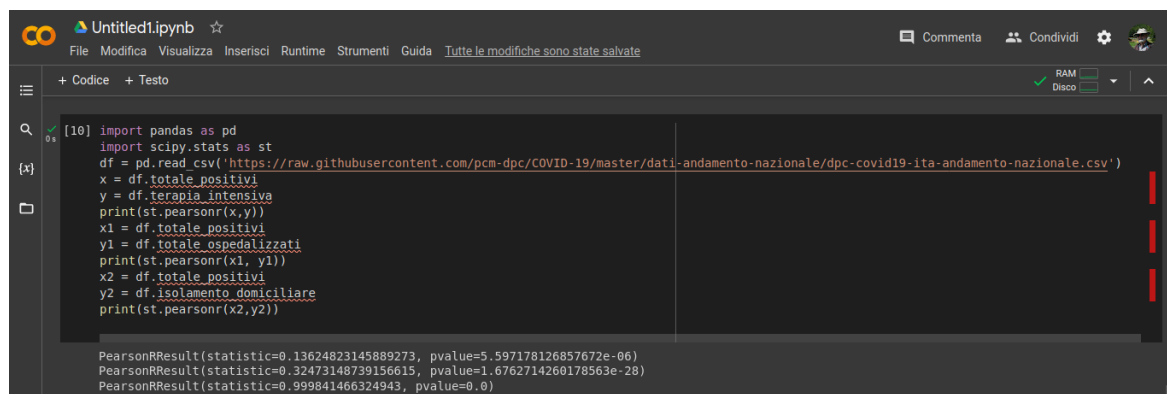
<https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv>

Per avere ciò che ci serve dobbiamo optare per il formato Raw, cliccando sul relativo pulsante nella barra degli strumenti che si trova nell'intestazione della tabella dei dati.

In tal modo l'indirizzo che ci interessa diventa

<https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv>

Con l'esercizio qui riportato, collegando direttamente Pandas all'URL, ho verificato quale sia stata la correlazione tra la serie giornaliera dei positivi e i ricoverati in terapia intensiva, i normali ospedalizzati e gli isolati domiciliari:



```
import pandas as pd
import scipy.stats as st
df = pd.read_csv('https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv')
x = df.totale_positivi
y = df.terapia_intensiva
print(st.pearsonr(x,y))
x1 = df.totale_positivi
y1 = df.totale_ospedalizzati
print(st.pearsonr(x1, y1))
x2 = df.totale_positivi
y2 = df.isolamento_domiciliare
print(st.pearsonr(x2,y2))
```

PearsonResult(statistic=0.13624823145889273, pvalue=5.597178126857672e-06)  
PearsonResult(statistic=0.32473148739156615, pvalue=1.6762714260178563e-28)  
PearsonResult(statistic=0.999841466324943, pvalue=0.0)

I risultati evidenziano come tra positivi e ricoverati in terapia intensiva non vi sia stata alcuna correlazione (indice 0,136), come la correlazione tra positivi e ospedalizzati normali sia stata molto bassa (indice 0,325) e come invece vi sia stata correlazione perfetta tra positivi e isolati domiciliari (indice 0,999).

Come dire che il ricovero in ospedale raramente si è rivelato necessario per combattere il morbo e meno ancora si è rivelata necessaria la terapia intensiva.

Pertanto è stato brutto ma poteva anche andare peggio.