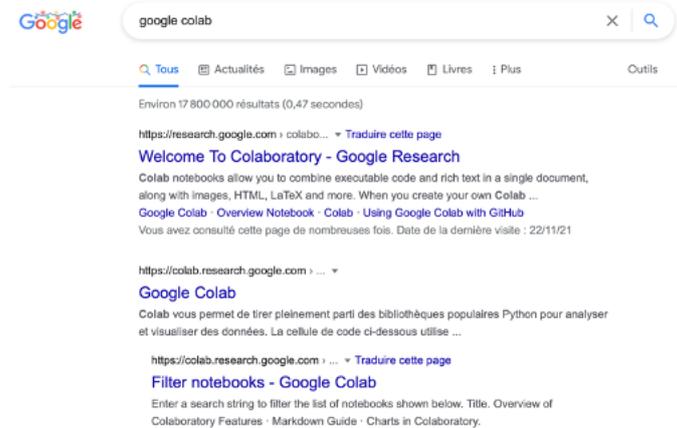
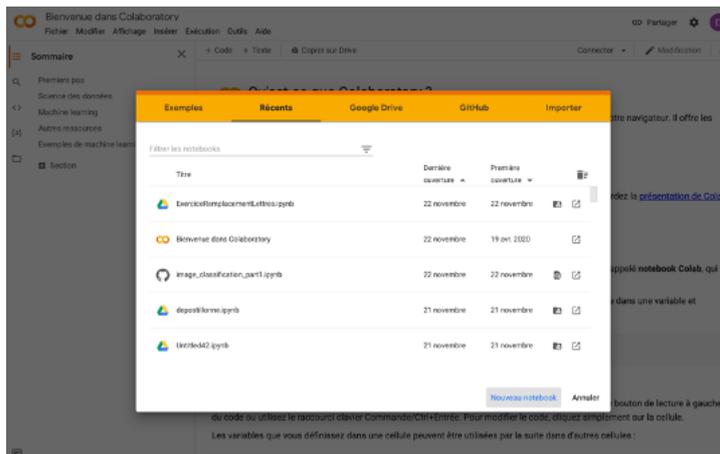


Google Colaboratory & Pandas

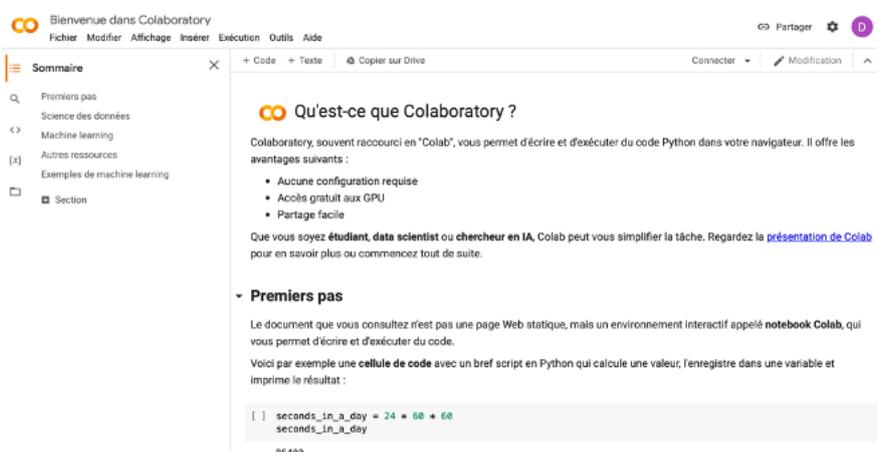
Recherchez « Google Colab » en utilisant votre moteur de recherche et cliquez sur « Welcome To Colaboratory » :



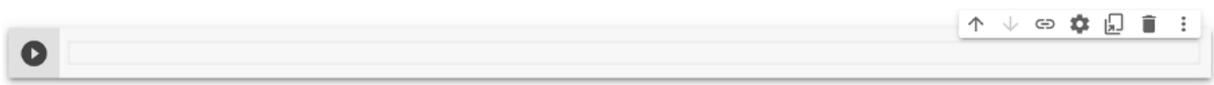
Cliquez sur « Nouveau notebook » pour ouvrir une page vierge. Dans le cas où vous souhaiteriez accéder à un ancien fichier Colaboratory, sélectionnez le fichier concerné dans la liste proposée.



Cliquez sur l'onglet « + Code » de la page d'accueil de Colaboratory :

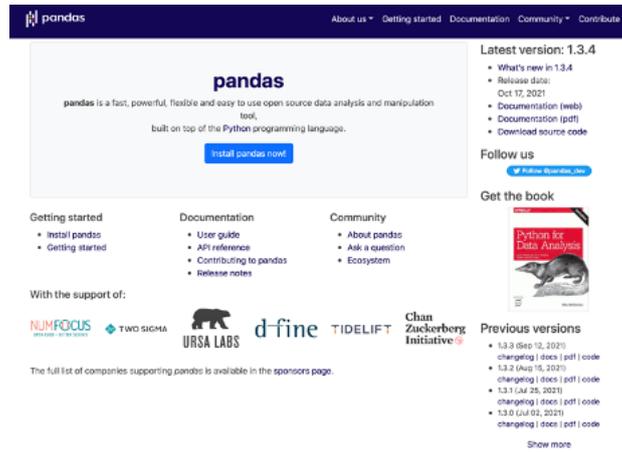


Un encadré permettant de saisir vos lignes de code s'affiche en bas de la page :



La bibliothèque pandas :

Pandas est une bibliothèque écrite pour le langage de programmation Python permettant la manipulation et l'analyse des données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numériques et de séries temporelles (<https://fr.wikipedia.org/wiki/Pandas>). Pour plus de détails : <https://pandas.pydata.org>



Cette bibliothèque sera notamment utilisée pour importer les jeux de données à traiter afin de les convertir en tableaux manipulables pour procéder aux analyses de données. Les principales structures de données gérées par la bibliothèque pandas sont :

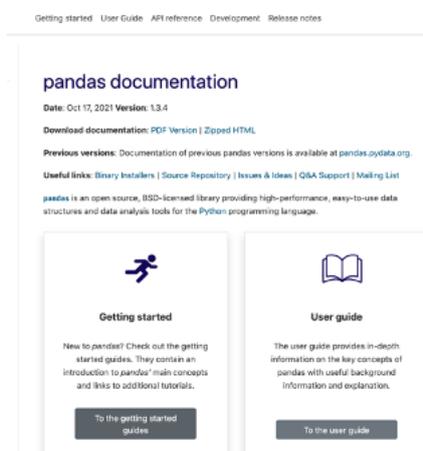
- Séries : stockage des données selon une dimension (grandeur en fonction d'un index)
- DataFrames : stockage des données selon 2 dimensions (lignes et colonnes)
- Panels : représentation des données selon 3 dimensions
- Panels4D ou DataFrames avec des index hiérarchiques aussi nommés MultiIndex : représentation des données selon plus de 3 dimensions - hypercube.

La bibliothèque pandas permet de :

- Manipuler des données aisément et efficacement avec des index pouvant être des chaînes de caractères via les DataFrames
- Lire et écrire des données structurées en mémoire depuis et vers différents formats : fichiers CSV, fichiers textuels, fichiers Microsoft Excel, base de données SQL...
- Gérer les données manquantes (NaN = not a number)
- Trier les données selon divers critères
- Redimensionner un tableau ;
- Fusionner et joindre de larges volumes de données
- Analyser des séries temporelles

Google Colaboratory & Pandas

Cliquez sur l'onglet « Documentation » pour accéder aux détails des commandes disponibles. Vous pouvez télécharger la documentation au format PDF ou accéder à la version en ligne.



Le format de données le plus utilisé est le DataFrame. Un DataFrame pandas est une structure de données étiquetée en 2 dimensions avec des colonnes de types potentiellement différents. Vous pouvez le voir comme une feuille de calcul ou une table SQL (ou un dict d'objets Series). Pour plus de détails, voir la documentation en page 180.

Importation des données sous Google Colaboratory

Pour importer vos données et les mettre sous la forme d'un tableau « pandas », saisissez les lignes de code suivantes sous Google Colaboratory :

```
# importe la bibliothèque pandas en spécifiant que l'abréviation pd sera utilisée par la suite pour faire
import pandas as pd

# importer les données
# importer les données
from google.colab import files
files.upload()
data=pd.read_csv('Shopping.txt', sep="\t")
```

Dans cet exemple, le fichier téléchargé est intitulé "Shopping.txt". Il s'agit d'un fichier ".txt" avec séparateur "Tabulation". Pour l'adapter à votre cas, vous devez changer le nom du fichier et, le cas échéant, le séparateur.

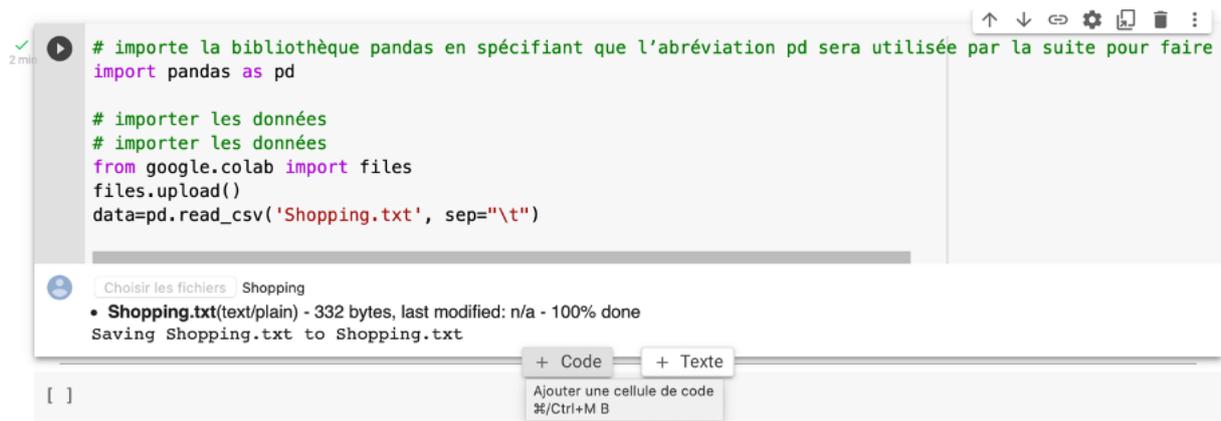
Cliquez sur  pour exécuter le code.

Dans le cas où la bibliothèque pandas n'est pas installée, vous pouvez l'installer en utilisant pip install :

- !pip install pandas

Cliquez sur « Choisir les fichiers » pour rechercher le fichier « Shopping.txt » sur votre ordinateur afin qu'il soit importé sur Google Colaboratory.

Google Colaboratory & Pandas



```
# importe la bibliothèque pandas en spécifiant que l'abréviation pd sera utilisée par la suite pour faire
import pandas as pd

# importer les données
# importer les données
from google.colab import files
files.upload()
data=pd.read_csv('Shopping.txt', sep="\t")
```

Choisir les fichiers Shopping

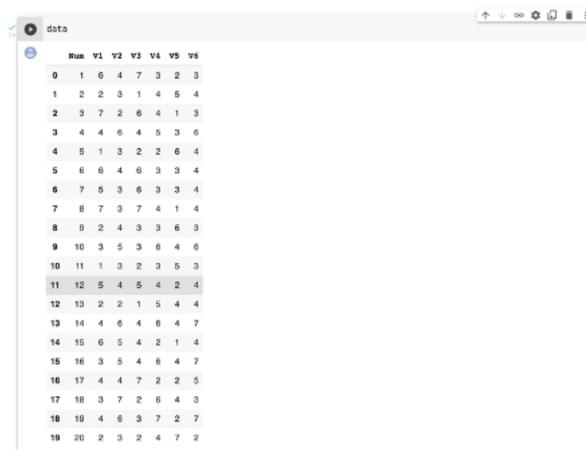
- **Shopping.txt**(text/plain) - 332 bytes, last modified: n/a - 100% done
Saving Shopping.txt to Shopping.txt

+ Code + Texte

Ajouter une cellule de code
⌘/Ctrl+M B

Une fois l'importation effectuée, survolez le centre bas de la fenêtre et cliquez sur « + Code » pour qu'une nouvelle ligne de commande vierge s'affiche. Vous pourrez utiliser cette ligne afin de saisir les codes décrivant les instructions à réaliser.

Ecrire « data » et lancer l'exécution. Le contenu du DataFrame « data » s'affiche à l'écran.



	num	v1	v2	v3	v4	v5	v6
0	1	6	4	7	3	2	3
1	2	2	3	1	4	5	4
2	3	7	2	6	4	1	3
3	4	4	6	4	5	3	6
4	5	1	3	2	2	6	4
5	6	6	4	6	3	3	4
6	7	5	3	6	3	3	4
7	8	7	3	7	4	1	4
8	9	2	4	3	3	6	3
9	10	3	5	3	6	4	6
10	11	1	3	2	3	5	3
11	12	5	4	5	4	2	4
12	13	2	2	1	5	4	4
13	14	4	6	4	6	4	7
14	15	6	5	4	2	1	4
15	16	3	5	4	6	4	7
16	17	4	4	7	2	2	5
17	18	3	7	2	6	4	3
18	19	4	6	3	7	2	7
19	20	2	3	2	4	7	2

Un DataFrame est un objet à 2 axes : "axis 0" et "axis 1". "axis 0" représente les lignes et "axis 1" représente les colonnes. Nous demandons à afficher les colonnes.

Reportez-vous à la documentation de la bibliothèque pandas pour plus de détails sur les fonctionnalités proposées.