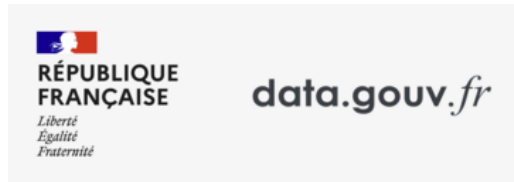


## TP - Traitement de données ouvertes

Dans ce TP, tu vas récupérer des données dites « ouvertes », c'est à dire mise à disposition publiquement dans le nuage.



Plus précisément, tu vas récupérer des données concernant les *lycées d'Occitanie* à partir de la plateforme ouverte des données publiques française [data.gouv.fr](https://www.data.gouv.fr). Après les avoir analysé, tu les traiteras avec un tableau puis par de la programmation.

### A. Récupération de données ouverte sous différents formats

1. Effectue une recherche sur le site <https://www.data.gouv.fr> (<https://www.data.gouv.fr>) pour obtenir les données ouvertes disponibles sur les **lycées d'Occitanie**.
2. Selon les informations disponibles sur la page Web dédiée au jeu de données sur les lycées d'occitanie :
  - Quelles **données** sont susceptibles d'être récupérées ?
  - En observant les « critères de Qualité » des métadonnées, essaie de définir ce qu'est « une métadonnée ».
3. Sous quels **formats de fichier** ces données sont-elles distribuées ?

#### Notepad++



L'application *Notepad++* est un *éditeur de texte* qui permet d'éditer du texte brut. Tu la trouvera sur *IoRdi* dans le portail applicatif MCNL, rubrique SNT/NSI.

4. Télécharge puis ouvre le fichier au format CSV avec Notepad++.

- A. Quelles données sont présentes dans ce fichier ? Quel est leur **type** ?
- B. Est-il facile de trouver l'adresse du *lycée Pierre d'Aragon* en parcourant manuellement le contenu du fichier ? Que peux-tu en conclure ?

#### VALIDATION

Appelle le professeur pour faire valider tes réponses rédigées sur ton **carnet de bord**.

### B. Structuration des données : Analyse du format de fichier CSV

1. En examinant le contenu du fichier au format CSV avec *Notepad++* puis réponds aux questions suivantes :
  - A. Explique le nom du format de fichier « comma-separated value ».
  - B. Quelles différences observes-tu entre les deux premières lignes ?
2. Ouvre maintenant ce même fichier CSV avec le tableur *LibreOffice Calc* et dis ce que représente par rapport aux données :
  - A. une cellule du tableau ?
  - B. une ligne du tableau ?
  - C. une colonne du tableau ?
3. Sélectionne la colonne correspondant au nom du département puis trie-la par ordre alphabétique avec l'option « sélection courante ».
  - A. Décris le résultat obtenu.
  - B. Les données sont-elles encore exploitables ? Pourquoi ?

#### LibreOffice Calc



LibreOffice Calc est un tableur qui permet de visualiser et de traiter facilement des quantités de données relativement importantes à condition qu'elles soient *structurées*.

## VALIDATION

Appelle le professeur pour faire valider tes réponses rédigées sur ton **carnet de bord**.

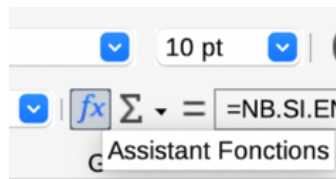
## C. Traitement des données par tableur et par programmation

Utilise le tableur pour répondre aux questions suivantes :

1. Quelle est l'adresse du lycée Bellevue d'Albi?
2. Détermine le nombre de lycée du Tarn.
3. Détermine le pourcentage de lycées privés parmi l'ensemble des lycées d'Occitanie.
4. Représente cette information sous le forme d'un graphique.

### Fonction de comptage de LibreOffice Calc

- Pour compter un nombre de cellules selon une certaine *condition*, tu peux utiliser les fonctions `NB.SI()` et `NB.SI.ENS()` du tableur.
- Au besoin, utilise l'*assistant des fonctions* :



- Par exemple, on utilise l'expression suivante pour déterminer le \_nombre de lycée qui possèdent un *internat* :

`=NB.SI(P100:P183 ; "oui")`

## VALIDATION

Appelle le professeur pour faire valider tes réponses rédigées sur ton **carnet de bord**.

5. Analyse puis exécute le script Python ci-dessous.
6. Modifie le script pour compter le nombre de lycée du tarn (voir bulles Python ci-dessous).
7. Que permet le traitement des données par programmation par rapport au traitement par tableur?

Entrée[ ]:

```
1 import csv
2
3 # Chargement du fichier csv en mémoire
4 with open("lycees-occitanie.csv", mode='r', encoding='utf-8') as
5
6     # Lecture du fichier csv,
7     # le contenu est stocké ensuite dans la variable nommée lecteur_csv
8     lecteur_csv = csv.reader(fichier, delimiter=';')
9
10    # Boucle permettant de parcourir chaque ligne du fichier csv
11    for ligne in lecteur_csv:
12
13        # Affichage de la deuxième valeur de la ligne sans retour
14        print(ligne[2], end=" ")
15
16        # Affichage de la quatrième valeur de la ligne
17        print(ligne[4])
18
```

### Stocker une valeur dans une variable



Pour créer une variable et y stocker une valeur, il suffit de lui donner un nom et d'utiliser le symbole `=` pour indiquer la valeur que l'on veut y stocker.

Par exemple :

```
x = 1
pi = 3.14159269
ma_variable = "super"
```

On peut y stocker toutes sortes de *type* de données, comme des nombres entiers ou décimaux, mais aussi des mots ou *chaîne de caractère* que l'on met entre guillemets `"..."` ou `'...'`.

### Compter le nombre de répétition d'une boucle en Python



1. Il faut tout d'abord définir un variable `compteur` avant la boucle qui stocke la valeur 0 (en effet, on n'a pas encore commencé les répétitions).

```
compteur = 0
```

2. Dans la boucle, on récupère la valeur stockée dans la variable, à laquelle on ajoute 1 avec l'expression :

```
compteur + 1
```

3. Enfin on doit stocker cette nouvelle valeur dans la même variable `compteur` :

```
compteur = compteur + 1
```

### Instruction « Si ... sinon ... » avec Python



Pour exécuter une instruction uniquement **si** une variable contient une certaine valeur, on utilise une *instruction conditionnelle* où la condition est un *test d'égalité* `==` :

```
if nom_personne == 'Albert':  
    print("Tu dois être un génie!")  
else:  
    print("Dommage pour toi, tu ne t'appelle pas Albert")
```

### VALIDATION

Appelle le professeur pour faire valider tes réponses rédigées sur ton **carnet de bord** et montre-lui ton script.