



Bravo ! vous êtes engagé comme **data scientist** auprès du ministre de l'Éducation nationale. Vous allez devoir épauler le ministre dans sa politique en lui fournissant les données nécessaires dont il a besoin pour effectuer des campagnes de communication, réaliser des budgets annuels ... etc Mais le ministre est très exigeant, vous devrez donc répondre vite à ses sollicitations !

Vos missions

Important : Vous réaliserez chaque mission dans un nouveau fichier vide Mission1.py, Mission2.py, etc...

Mission1.py : « Je vais mettre en place un plan de subvention aux collèges de France pour l'accès à la lecture, d'environ 5 000€ par établissement scolaire. Pouvez-vous chiffrer le cout total de cette réforme ? ».

Mission2.py : « Ouch, ça fait cher ! Heu.... On va limiter la réforme aux collèges qui disposent d'une SEGPA. Pouvez-vous chiffrer de nouveau le cout total de cette réforme ? »

Mission3.py : « Je vais faire une visite dans l'académie de Rouen, j'aurais besoin d'envoyer un email à l'ensemble des établissements scolaires présents dans le département 76. Il me faudra la liste des emails dans un fichier, un email par ligne »

Mission4.py : « Attends ne part pas ! J’aurais besoin aussi des adresses postales car on va envoyer un courrier... Est-ce que tu peux me faire un fichier CSV avec les mails dans la 1^{ère} colonne et les adresses postale (*format : 4 rue Maulévrier 76000 ROUEN*) dans la 2^{ème}. Va travailler maintenant, la pause-café est terminée. »

Mission5.py : « J’aimerais faire une visite dans l’une des plus petites structures du département 76. Peux-tu faire donner la liste des établissements avec leur nom, leur adresse postale et le nombre d’élève au format CSV dans un fichier **mission5.csv (disponible sur picassciences)**

Indice : *attention, parfois, la donnée nombre d’élève n’est pas disponible, ne pas inclure l’établissement dans la liste dans ce cas.*

Mission6.py : « ah, et il faudrait les trier par ordre croissant de nombre d’élève évidemment, sinon, ça ne sert à rien. »

Indice 1 : On charge ici le fichier **mission5.csv (disponible sur picassciences)** pour le trier (avec une méthode de tri intégrée à Python) et enregistrer le résultat dans **mission6.csv**

Indice 2 : *Si le tri n’est pas satisfaisant, utiliser les documents à votre disposition.*

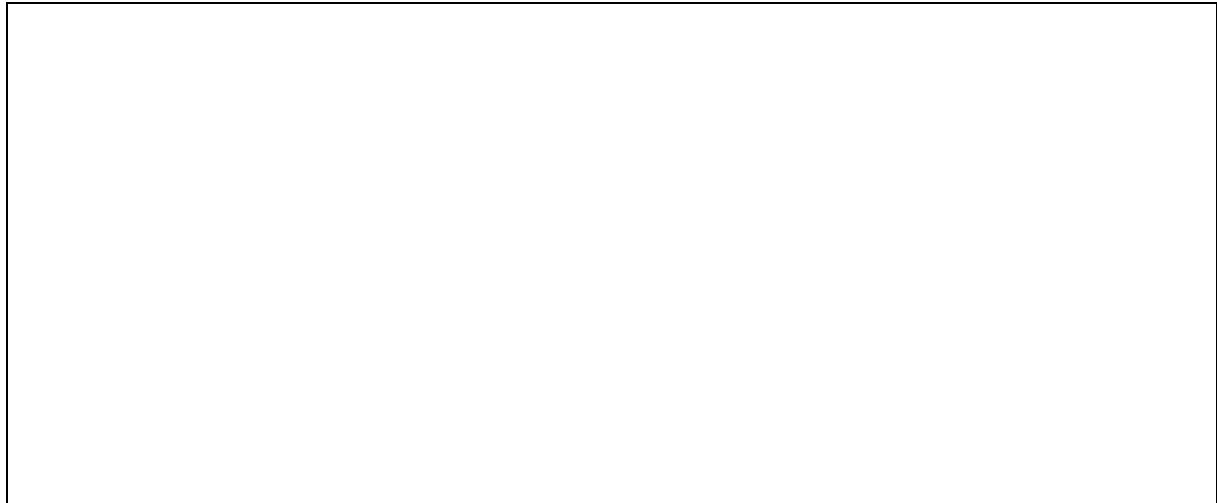
NSI – données en table

Mission7.py : J'ai utilisé ton travail de la Mission 5, mais j'ai l'impression qu'un de mes assistants à rajouter des doublons dans la liste, histoire de me faire une blague pour le 1^{er} avril... Tu peux vérifier les 1300 entrées de ce fichier, et me dire s'il y a des doublons ? et lesquels ?

Extrait de la base à analyser :

Ce fichier est disponible sur picassciences

```
Ecole maternelle la Ferme du Plan,49 rue des Hautes Haies 76240,181
Ecole maternelle Edouard Vaillant I,18 rue Charles Delescluzes 76610,145
Ecole élémentaire,8 rue de l'Eglise 76740,16
Ecole maternelle,Route de la Croix de Realcan 76740,28
Ecole élémentaire,Rue au Bout de la Rue aux Chiens 76740,22
Ecole primaire Maurice Decorde,9 rue de l'Epte LE FOSSE 76440,82
Ecole maternelle,3 rue des Moissonneurs 76440,45
Ecole élémentaire,5 rue de la Mairie 76440,48
Ecole élémentaire,54 route de Limpiville 76110,24
Ecole maternelle,23 rue de la Vallée 76110,19
```



Mission8.py : Horreur ! nous avons été la cible d'un ransomware ! Toutes nos données perdues ! J'ai eu la présence d'esprit de faire une sauvegarde de tes fichiers CSV. Peux-tu faire migrer les deux fichiers en un seul pour que ce soit cohérent.

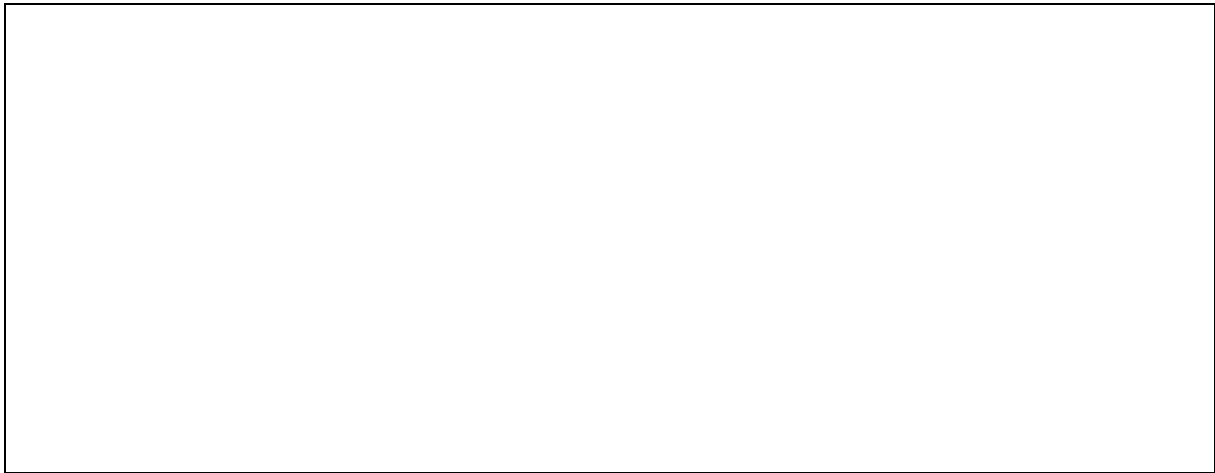
Extrait de la base Mission9a.csv : **Ce fichier est disponible** sur picassciences

```
0762119R,ECOLE SIMONE VEIL,Route de l'Abbaye,76840 ST MARTIN DE BOSCHERVILLE
0760881V,GROUPE SCOLAIRE THOMAS PESQUET,2 rue des Furets,76290 FONTENAY
0762118P,Ecole primaire,8 place de la Mairie,76700 ST LAURENT DE BREVEDENT
0762412J,Ecole élémentaire,67 route de la Mairie,76690 ST GERMAIN SOUS CAILLY
0762923P,Ecole primaire,1 rue de la République,76280 GONNEVILLE LA MALLET
0763252X,Ecole primaire Voyelles,1 rue de la Gare,76133 ROLLEVILLE
```

Extrait de la base Mission9b.csv : **Ce fichier est disponible** sur picassciences

```
0761701L,0761701l@ac-rouen.fr,629
0761703N,ce.0761703n@ac-rouen.fr,304
0761704P,ce.0761704p@ac-rouen.fr,312
0761730T,ce.0761730t@ac-rouen.fr,205
0761734X,accueil@sthild.fr,244
0761711X,lesormeaux.lehavre@wanadoo.fr,724
```

Vous remarquerez que les données ne sont pas dans le même ordre.



Éléments à votre disposition

Document 1 : Comment lire des données au format CSV ?

Voir document 1 sur picassciences (jusqu'à 8:19) :

Le chemin vers un fichier s'écrit comme suit :

```
import csv

CHEMINVERSFIChIER = "C:\\unrepertoire\\autre repertoire\\"

f = open( CHEMINVERSFIChIER + "annuairecomplet.csv",
encoding="utf8")

csv_reader = csv.reader(f, delimiter=';')
for row in csv_reader:
    print(row)
```

Utilisez **utf8** dans la plupart des cas. Utiliser **latin1** si les caractères accentués ne s'affichent pas correctement ou si vous obtenez une erreur d'encodage.

Note : si vous désirez mettre plusieurs boucles dans un même fichier, vous devez remettre à zéro l'itérateur de ligne avec `f.seek(0)`

Document 2 : La structure du fichier CSV étudié

Voici les informations présentes dans chaque numéro de colonne. Pour y accéder, il suffit de reprendre le document 1 et changer `print(row)` par `print(row[0])` pour obtenir la 1^{ère} information, le code établissement.

Information	Information
0 Code établissement	26 Segpa;
1 Nom_etablissement;	27 Section_arts;
2 Type_etablissement;	28 Section_cinema;
3 Statut_public_privé;	29 Section_theatre;
4 Adresse_1;	30 Section_sport;
5 Adresse_2;	31 Section_internationale;
6 Adresse_3;	32 Section_europeenne;
7 Code postal;	33 Lycee_Agricole;
8 Code_commune;	34 Lycee_militaire;
9 Nom_commune;	35 Lycee_des_metiers;
10 Code_departement;	36 Post_BAC;
11 Code_academie;	37 Appartenance_Education_Prioritaire;
12 Code_region;	38 GRETA;
13 Ecole_maternelle;	39 SIREN_SIRET;
14 Ecole_elementaire;	40 Nombre_d_eleves;
15 Voie_generale;	41 Fiche_onisep;
16 Voie_technologique;	42 position;
17 Voie_professionnelle;	43 Type_contrat_privé;
18 Telephone;	44 Libelle_departement;
19 Fax;	45 Libelle_academie;
20 Web;	46 Libelle_region;
21 Mail;	47 coordonnee_X;
22 Restauration;	48 coordonnee_Y;
23 Hebergement;	49 epsg;
24 ULIS;	50 nom_circonscription;
25 Apprentissage;	

Document 3 : Comment écrire un fichier CSV ?

```
CHEMINVERSFICHIER = "C:\\unrepertoire\\autre repertoire\\"
```

```
hs = open(CHEMINVERSFICHIER + "out.csv", "w")
hs.write(str(chaine_a_ecrire_dans_le_fichier))
hs.close()
```

Ce code doit être placé en fin de script, il ne peut être utilisé dans une boucle. Vous devez donc au préalable préparer une chaîne de caractère `chaine_a_ecrire_dans_le_fichier` par concaténation, puis écrire cette chaîne de caractère dans un fichier.

Pour réaliser des sauts de lignes dans la chaîne de caractère, on utilise + `"\n"`

Document 4 : Découper des chaînes de caractères

```
Chaine = "bonjour"
print(Chaine[0:2])
```