

Introduction aux données structurées — Le format CSV

Durée : 1h

1 Préparation

Faire chaque activité dans un nouveau dossier est une bonne pratique dans toutes les matières et en dehors du lycée. Prenez l'habitude de bien organiser vos dossiers pour retrouver vos documents plus tard.

EXERCICE 1

Créez un dossier nommé « tp-csv » dans votre zone personnelle.
Téléchargez-y le fichier à l'adresse : https://git.jean-cloud.net/adrian/cours-snt/raw/branch/master/donnees_structurees/tp_decouverte_csv/programme.py.
Ouvrez ce programme avec pyzo ou edupython.

Depuis quelques années, l'état et les collectivités locales sont tenues de mettre des données relatives à leurs activités en libre accès. Cette données sont dites « ouvertes » car libre d'accès pour tous et toutes.

EXERCICE 2

Rendez-vous à cette adresse : <https://www.data.gouv.fr/fr/datasets/lieux-de-vaccination-contre-la-covid-19/>
Téléchargez le fichier CSV « centres-vaccination.csv » et copiez son contenu dans le fichier CSV « centres-vaccination.csv ».
Téléchargez la documentation de ce fichier, disponible sur la même page web, et essayez de comprendre la signification des trois premiers descripteurs.

Ce fichier CSV n'utilise pas la virgule comme séparateur, car c'est un caractère trop commun en français, il utilise le point virgule. Avoir un séparateur dans un champs empêcherait un programme de le lire correctement, il aurait l'impression qu'il y a plus de valeurs que prévu.

EXERCICE 3

En bas du fichier python que vous avez téléchargé, appelez la fonction « `afficher_descripteurs` » comme suit :

```
afficher_descripteurs()
```

Vérifiez que vous avez bien une liste de champs qui s'affiche comme sur l'exemple suivant :

```
gid
nom
xy_precis
id_adr
adr_num
adr_voie
com_cp
com_insee
com_nom
lat_coor1
long_coor1
```

2 Identifier un problème social

EXERCICE 4

La ligne suivante permet de compter le nombre de valeurs renseignées pour chaque champs :

```
afficher_compte_colonnes()
```

Quels sont les trois champs les moins renseignés? Cela vous semble t'il normal?

3 Identifier les incohérences techniques

EXERCICE 5

Remplacez votre affichage par le code suivant :

```
afficher_contenu_colonne('lieu_accessibilite')
```

Il permet d'afficher les différentes valeurs renseignées dans le champs « `lieu_accessibilite` » et le nombre de fois qu'elles apparaissent dans le fichier CSV.

Selon vous, pourquoi les valeurs sont aussi hétérogènes? Quelles valeurs devrait-on pouvoir trouver dans ce champs selon la documentation disponible avec le fichier CSV?

EXERCICE 6

L'URL suivante est extraite du fichier CSV et permet de réserver dans un centre de vaccination :

```
https://partners.doctolib.fr/centre-depistage-covid/chateauroux/  
centre-de-vaccination-covid-ssr-chateauroux-36?speciality\_id=5494&enable\_cookies\_  
consent=1
```

Les paramètres qui sont passés à la page via cette url sont les suivants :

```
speciality_id=5494
```

```
enable_cookies_consent=1
```

Que peut signifier le dernier paramètre? Quelles implications peut-il avoir?