COURS

Traitement de données en tables : CSV, sélection, tri, fusion

1re

Programme de première générale, enseignement de spécialité — BO spécial n°1 du 22 janvier 2019

Relue le 14 novembre 2026

Résultats sportifs, horaires de trains, données ouvertes de l'État : une grande partie des données du monde est rangée en tables. Savoir les lire, filtrer, trier et croiser en Python, c'est la base de l'analyse de données.

1 Une table de données

DÉFINITION

Ensemble d'enregistrements ayant les mêmes descripteurs (colonnes). En Python, on la représente par une **liste de dictionnaires** : chaque dictionnaire est une ligne.

DÉFINITION

Format texte où chaque ligne est un enregistrement et les valeurs sont séparées par un délimiteur (virgule ou point-virgule). La première ligne contient les descripteurs.

RÈGLE

Lecture : `import csv` puis `list(csv.DictReader(fichier, delimiter=";"))`. Les valeurs lues sont des **chaînes** : il faut les convertir (`int`, `float`).

2 Sélectionner et trier

RÈGLE

Sélection par compréhension : `[l for l in table if l["pays"] == "France"]`. Projection : `[l["nom"] for l in table]`. Tri : `sorted(table, key=lambda l: int(l["age"]), reverse=True)`.

3 Fusionner des tables

DÉFINITION

Croisement de deux tables qui partagent un descripteur commun (une **clé**) pour produire des lignes enrichies.

RÈGLE

Il faut vérifier la **cohérence** : doublons, valeurs manquantes, formats différents (« 03 » et « 3 »).

EXEMPLE

Trier les élèves par moyenne décroissante.

1. Convertir la moyenne en nombre : `float(l["moyenne"])`.
2. Utiliser `sorted(eleves, key=lambda l: float(l["moyenne"]), reverse=True)`.
3. Vérifier sur trois lignes que l'ordre est correct.

On trie avec `sorted` et une fonction `key`, après avoir converti les chaînes en nombres.

LE PIÈGE QUI COÛTE LE PLUS DE POINTS

Trier des nombres lus dans un CSV sans les convertir. — Les chaînes se trient dans l'ordre alphabétique : « 10 » passe avant « 9 ».

Le réflexe : Convertir avec int ou float dans la clé de tri.

EXPLOITER UNE TABLE CSV

1. Ouvrir le fichier et lire les lignes avec DictReader.
2. Convertir les valeurs numériques.
3. Sélectionner, projeter ou trier par compréhension et sorted.
4. Contrôler le résultat sur quelques lignes.

À VÉRIFIER

- Ai-je défini le vocabulaire ?
- Mon code est-il testé sur un exemple ?
- Ai-je traité les cas limites ?
- Ai-je justifié la complexité ou le résultat ?

À RETENIR

- Table = **liste de dictionnaires** ; **CSV** : texte, délimiteur, ligne d'en-tête.
- Sélection et projection par compréhension ; tri avec `sorted(..., key=...)` ; conversion des chaînes.
- **Fusion** sur une clé commune ; cohérence des données.