

ANALYSEUR D'INSCRIPTIONS, DE NOTES ET DE COHERENCE PEDAGOGIQUE

Vous disposez d'un jeu de données ***en mémoire*** (liste de tuples) représentant des notes d'étudiants. Certaines lignes sont incohérentes (notes invalides, champs vides, matière inconnue, etc.). Votre objectif est de construire un programme qui ***nettoie, structure, analyse et détecte des anomalies*** en utilisant uniquement ***les structures de données*** Python vues en cours.

Le projet peut être fait dans un seul fichier (tp_numero_groupe.py). Vous pouvez aussi découper en modules si préférable mais ce n'est pas obligatoire.

Données fournies

Chaque enregistrement est un tuple : (nom, matière, note, groupe)

```
donnees = [
    ("Sara", "Math", 12, "G1"),
    ("Sara", "Info", 14, "G1"),
    ("Ahmed", "Math", 9, "G2"),
    ("Adam", "Chimie", 18, "G1"),
    ("Sara", "Math", 11, "G1"),
    ("Bouchra", "Info", "abc", "G2"),
    ("", "Math", 10, "G1"),
    ("Yassine", "Info", 22, "G2"),
    ("Ahmed", "Info", 13, "G2"),
    ("Adam", "Math", None, "G1"),
    ("Sara", "Chimie", 16, "G1"),
    ("Adam", "Info", 7, "G1"),
    ("Ahmed", "Math", 9, "G2"),
    ("Hana", "Physique", 15, "G3"),
    ("Hana", "Math", 8, "G3"),
]
```

Travail demandé

Partie 1 : Nettoyage et validation

- Écrire une fonction *valider(enregistrement)* qui renvoie (*True, ""*) si la ligne est valide, sinon (*False, "raison : note/nom/matière erronés"*).

Validité : nom non vide, matière non vide, groupe non vide, note numérique, note dans [0, 20].

- Diviser les données dans 3 structures différentes :

valides : liste des tuples nettoyés (note convertie en float, etc.)

erreurs : liste de dictionnaires {"ligne": enregistrement, "raison": "...”}

doublons_exact : set des enregistrements répétés exactement (même tuple)

Partie 2 : Structuration

- À partir de l'ensemble des données valides, identifier toutes les matières distinctes présentes dans le jeu de données et regrouper les dans une structure ne contenant aucun doublon. Cette structure devra permettre de connaître rapidement l'ensemble des matières enseignées.
- Organiser les données de manière hiérarchique afin de pouvoir accéder facilement aux notes d'un étudiant par matière. Pour chaque étudiant, il devra être possible de retrouver toutes ses matières, et pour chaque matière, l'ensemble des notes associées.
- Regrouper les étudiants par groupe pédagogique. Pour chaque groupe, on devra pouvoir connaître l'ensemble des étudiants qui y sont inscrits, sans répétition inutile.

Partie 3 : Calculs et statistiques

- Mettre en place un calcul de somme basé sur la récursivité (une fonction qui s'appelle elle-même). À partir de cette somme récursive, définir un calcul de moyenne.
- À partir de ces fonctions, calculer pour chaque étudiant sa moyenne générale ainsi que ses moyennes par matière.

Partie 4 : Analyse avancée et détection d'anomalies

- Mettre en place une structure dédiée au regroupement des alertes détectées lors de l'analyse. Cette structure devra permettre d'identifier clairement le type de problème et les éléments concernés.
- Les situations suivantes doivent être détectées :
 1. les étudiants ayant plus qu'une seule note pour une même matière ce qui peut indiquer une saisie erronée ou un problème de duplication ;
 2. les étudiants dont le profil est incomplet : n'ayant pas des notes dans toutes les matières ;
 3. les groupes pédagogiques présentant une moyenne générale faible (inférieure à un seuil donné) pouvant signaler un problème global ;
 4. les étudiants présentant des écarts très importants entre leurs notes minimale et maximale, suggérant des incohérences ou des performances très instables.