

Exercices – Traitements de données en tables

Les exercices précédés du symbole sont à faire sur machine.

Les exercices précédés du symbole sont à résoudre par écrit.

Notebook Capytale

Pour les exercices 1 et 2 : <https://capytale2.ac-paris.fr/web/c/0681-1609936>

Exercice 1

On utilise le fichier CSV `educ_cantal.csv` (encodage UTF-8).

Avertissement : certaines comparaisons portent sur des valeurs numériques, il faudra convertir les chaînes si nécessaire.

1. Importer ce fichier dans un programme Python et indexer les données sous la forme d'un tableau de dictionnaires nommé `table_educ`.
2. Combien d'enregistrements contient cette table ?
3. Opérations de **sélection** :
 - (a) En utilisant une boucle, définir le tableau `etab_AURILLAC` contenant les établissements situés à Aurillac.
 - (b) En compréhension, obtenir la table des établissements privés du Cantal.
4. Opérations de **projection** :
 - (a) Avec une boucle, obtenir le tableau des codes UAI.
 - (b) En compréhension, obtenir le tableau des noms d'établissements.
 - (c) Obtenir le tableau des communes sans répétition.
5. Opérations de **tri** :
 - (a) Trier par ordre croissant de codes postaux.
 - (b) Trier d'ouest en est puis du nord au sud.
 - (c) Trier par commune puis, dans chaque commune, statut (public avant privé).
6. Opération de **jointure** :
 - (a) Importer `population_Cantal.csv` dans un tableau de dictionnaires nommé `population`.
 - (b) Combien d'enregistrements contient cette table ?
 - (c) Effectuer la jointure entre `table_educ` et `population` dans `new_table` en ajoutant un champ `population`.
 - (d) Enregistrer `new_table` dans `exo1_jointure.csv`.
 - (e) Ouvrir le fichier dans un tableur, repérer les enregistrements sans population, expliquer la cause et proposer une correction.

À retenir

Lorsqu'on rapproche deux tables, il faut vérifier la cohérence des noms d'attributs, des formats de données (domaines de valeurs), des doublons, etc.

Un travail de **formatage** préalable des données est souvent nécessaire.

 **Exercice 2**

On reprend le fichier `population_Cantal.csv`. Pour chaque question, écrire des instructions Python.

1. Construire la table triée par population décroissante.
2. Construire le tableau des noms de communes dont le nom complet se termine par « AC ». Combien y en a-t-il ?
3. Construire la table des communes dont la population est comprise entre 1000 et 2000 habitants (bornes incluses).
4. Question ouverte : construire le tableau des noms de communes composés de plusieurs mots.