

TP Pandas (≈ 1h)

Objectifs :

Découvrir la bibliothèque pandas à travers la manipulation d'un jeu de données classique : le dataset Iris.

Consignes générales :

- Téléchargez le dataset Iris depuis `sklearn.datasets` ou un dépôt CSV en ligne.
- Chargez les données dans un DataFrame pandas.
- Effectuez différentes manipulations de données (statistiques descriptives, filtrage, ajout/suppression de colonnes ou lignes, renommage).
- Réalisez des visualisations simples avec matplotlib ou seaborn.

Travail demandé :

- 1 Importer le dataset Iris et afficher les 5 premières lignes.
- 2 Afficher les dimensions du jeu de données et la description statistique (`describe()`).
- 3 Renommer les colonnes avec des noms plus explicites (ex. `SepalLengthCm`).
- 4 Ajouter une colonne qui calcule le rapport longueur/largeur des pétales.
- 5 Supprimer une colonne inutile de votre choix.
- 6 Supprimer toutes les lignes où la longueur du sépale est inférieure à 5.0.
- 7 Filtrer uniquement les lignes correspondant à l'espèce 'setosa'.
- 8 Compter le nombre d'occurrences de chaque espèce.
- 9 Visualiser : (1) un histogramme d'une variable, (2) un nuage de points entre deux variables, (3) un boxplot comparant les espèces.

Durée estimée : 1h