TP Pandas (≈ 1h) — Dataset Iris

# Objectifs

Télécharger un jeu de données existant (Iris via scikit-learn), le charger dans pandas puis réaliser des manipulations et des visualisations avec matplotlib.

# Prérequis

Python 3.x, pandas, scikit-learn, matplotlib.

# Consignes

* Importer le dataset Iris depuis scikit-learn (`from sklearn.datasets import load\_iris`).
* Construire un DataFrame pandas avec les colonnes : sepal\_length, sepal\_width, petal\_length, petal\_width, species.
* Effectuer des manipulations courantes (statistiques, filtrage, renommage, ajout/suppression de colonnes/lignes).
* Produire des graphiques avec matplotlib (pas de seaborn requis).

# Travail demandé

1. Charger Iris et afficher les 5 premières lignes + la taille du DataFrame.
2. Afficher df.info() et df.describe().
3. Renommer les colonnes en SepalLengthCm, SepalWidthCm, PetalLengthCm, PetalWidthCm, Species.
4. Ajouter la colonne PetalRatio = PetalLengthCm / PetalWidthCm et SepalRatio = SepalLengthCm \* SepalWidthCm.
5. Supprimer la colonne SepalRatio.
6. Supprimer les lignes où SepalLengthCm < 5.0.
7. Filtrer uniquement les lignes de l’espèce setosa.
8. Compter le nombre d’occurrences par espèce (table de fréquences).
9. Visualiser : (1) un histogramme d’une variable numérique, (2) un nuage de points entre deux variables, (3) un boxplot par espèce, (4) un diagramme en barres du compte par espèce.

# Conseils

* Utilisez load\_iris(as\_frame=True) pour obtenir un DataFrame facilement.
* Pour le scatter : plt.scatter(x, y) puis plt.xlabel, plt.ylabel, plt.title.
* Pour le boxplot : plt.boxplot avec une liste de séries (une par espèce).
* Pensez à plt.figure() avant chaque graphique pour éviter les chevauchements.

Durée estimée : 1h