

## Tp - Iris

### Charger le jeu de données *Iris* dans un DataFrame Pandas.

1. Trouver la moyenne et la médiane de la colonne 'sepal\_length'.
2. Calculer le 75e percentile de la colonne 'petal\_width' pour chaque espèce du jeu de données *Iris*.
3. Créer une nouvelle colonne dans le DataFrame *Iris* appelée 'sepal\_area', qui est le produit de 'sepal\_length' et 'sepal\_width'.
4. Supprimer toutes les lignes du DataFrame *Iris* où 'petal\_length' est supérieure à deux fois l'écart-type de 'petal\_length' pour cette espèce.
5. Normaliser toutes les colonnes numériques du DataFrame *Iris* (sauf la colonne 'species') en utilisant la mise à l'échelle Min-Max (Min-Max scaling).
6. Trouver les trois combinaisons les plus fréquentes de 'sepal\_length', 'sepal\_width' et 'petal\_length' dans le jeu de données *Iris*.
7. Regrouper le DataFrame *Iris* par 'species' et trouver la ligne ayant la plus grande valeur de 'sepal\_width' pour chaque groupe.
8. Remplacer toutes les valeurs négatives de la colonne 'petal\_width' du DataFrame *Iris* par la moyenne des valeurs non négatives de cette colonne.
9. Calculer la matrice de corrélation des colonnes 'sepal\_length', 'sepal\_width', 'petal\_length' et 'petal\_width', puis trouver la variable ayant la plus forte corrélation absolue avec 'petal\_width'.