# Contexte du projet

Vous êtes un groupe de chercheurs en première année de sciences de la nature, travaillant sur un projet visant à analyser des données expérimentales dans le domaine de la biologie. Votre objectif est d’extraire, nettoyer, analyser et visualiser des données pour comprendre les relations entre différentes variables biologiques.

L’expérience consiste à étudier la croissance de plantes sous différentes conditions de lumière et d’arrosage. Les plantes ont été réparties en plusieurs groupes, chacun soumis à une combinaison spécifique de conditions de lumière (haute, moyenne, basse) et d’arrosage (haute, moyenne, basse). Les hauteurs initiales et finales des plantes ont été mesurées pour évaluer l’impact de ces conditions sur leur croissance.

# Jeu de données

Les données de l’expérience sont dans le fichier (***donnees\_plantes.csv***).

Ce fichier contient des données simulées sur la croissance des plantes sous différentes conditions de lumière et d’arrosage.

**Description des colonnes**

* **id** : Identifiant unique pour chaque plante.
* **hauteur\_initiale\_cm** : Hauteur initiale de la plante en centimètres.
* **hauteur\_finale\_cm** : Hauteur finale de la plante en centimètres.
* **condition\_lumiere** : Condition de lumière (haute, moyenne, basse).
* **condition\_arrosage** : Condition d’arrosage (haute, moyenne, basse).

# Livrables

* Un fichier Jupyter Notebook (.ipynb) contenant tout le code, les analyses et les visualisations.

# Évaluation

Lors de l’évaluation (**semaine 15**), vous aurez des questions se rattachant au projet, mais avec un autre jeu de données. Vous serez amenez à répondre à des questions dans Moodle **expliquer[[1]](#footnote-1)** votre démarche et vos résultats

Votre projet sera évalué sur les critères suivants :

* Qualité et propreté du code.
* Pertinence et précision des analyses statistiques.
* Clarté et esthétique des visualisations.
* Exhaustivité et clarté des commentaires et docstrings.

Voici la grille de correction qui sera utilisée :

## Préparation à l’évaluation :

Afin de vous préparer à cette évaluation, tout le long du projet, posez vous ces questions et surtout assurez-vous de pouvoir expliquer chaque étape de votre démarche.

1. Quels sont les principaux éléments à inclure dans l'algorithme ?
2. Quelles parties de votre code allez vous commenter ?
3. Comment allez-vous utiliser les bibliothèques scientifiques Pandas, NumPy et Matplotlib ?
4. Quels éléments de graphiques allez-vous d'utiliser pour afficher les résultats de votre traitement de données ?

# Objectifs du projet et Instructions

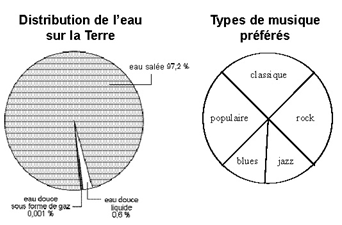
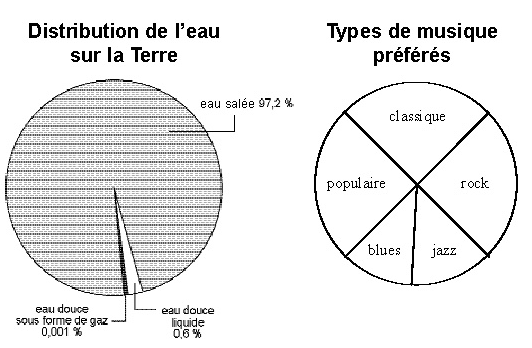
1. **Extraction, nettoyage et filtrage des données**
   * **Lecture du fichier CSV**
     + Lire et extraire les données du fichier CSV contenant les données.
     + Visionner les cinq premières lignes et les cinq dernières, afin d’identifier les valeurs manquantes (s’il y en a).
   * **Nettoyage des données**
     + Créer une fonction ***nettoyer\_donnees(dataframe)*** qui supprime les lignes contenant des valeurs manquantes.
       - Pour supprimer une ligne, il suffit de ne pas l’ajouter aux données nettoyées.
     + Cette fonction doit retourner le *dataframe* nettoyé.
2. **Analyse statistique descriptive, corrélation et régression linéaire**

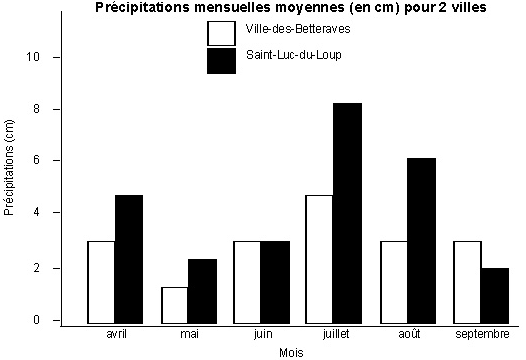
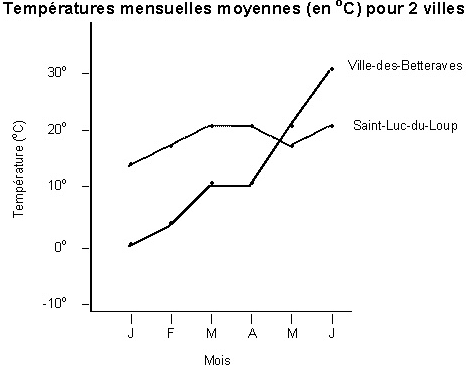
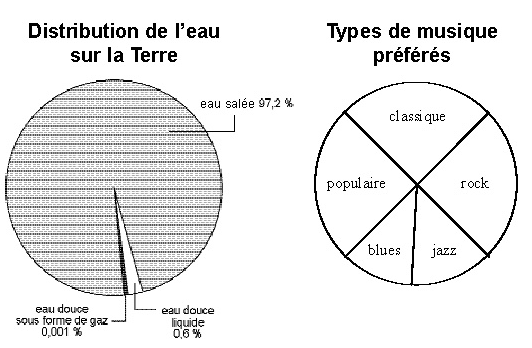
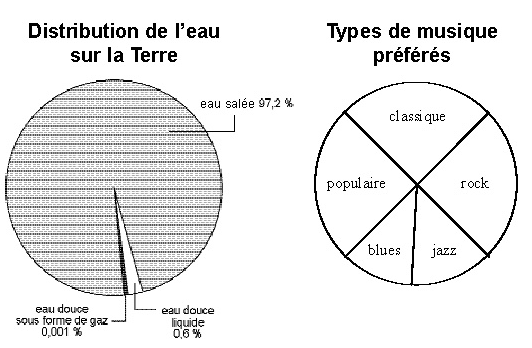
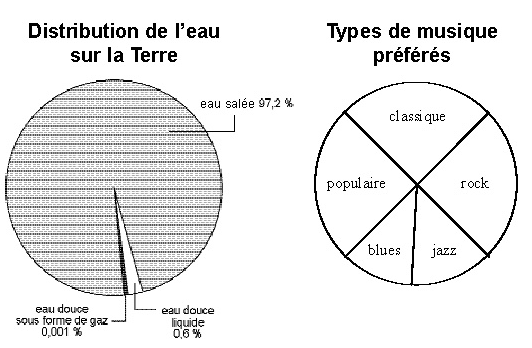
Le reste de l’analyse doit se faire sur le *dataframe* nettoyé.

* + **Calcul de la moyenne**
    - Créer une fonction ***calculer\_moyenne(dataframe, colonne)*** qui calcule la moyenne des valeurs d’une colonne en utilisant une boucle ***for***.
  + **Calcul de la corrélation**
    - Créer une fonction ***calculer\_correlation(dataframe, colonne1, colonne2)*** qui calcule la corrélation entre deux colonnes.
    - Écrire un paragraphe pour expliquer le résultat obtenu pour le c**oefficient de corrélation**.
  + **Régression linéaire**
    - Créer une fonction r***egression\_lineaire(dataframe, x\_col, y\_col)*** qui calcule la pente et l’ordonnée à l’origine de la droite de régression.
    - Cette fonction retourne la **pente** et **l’ordonnée** de la droite de regression.

1. **Création de graphiques**
   * **Graphique en nuage de points**
     + Créer un graphique en nuage de points montrant la relation entre la hauteur initiale et finale des plantes.
   * **Histogramme**
     + Créer un histogramme de la hauteur finale.
2. **Programmation lettrée (Commentaires et *Docstring*)**
   * **Ajout de commentaires**
     + Ajoutez des commentaires détaillés dans votre code pour expliquer les étapes et chaque fonction.
   * **Étape 2 : Ajout de docstrings**
     + Ajoutez des ***docstrings*** pour chaque fonction, décrivant leur rôle, leurs paramètres et leur valeur de retour.

Bonne chance et amusez-vous bien avec ce projet d’analyse de données scientifiques !





1. Dans le contexte d’un projet en programmation Python, “expliquer une démarche” signifie **détailler** les **étapes** et les **fonctions** et **méthodes utilisées** pour accomplir une tâche spécifique.  [↑](#footnote-ref-1)