**Projet Scientifique : Analyse de Données Biologiques en Python**

**Contexte du Projet**

Ce projet vise à introduire les étudiants de première session en sciences de la nature aux bases de la programmation en Python à travers l’analyse de données biologiques. Le projet est divisé en quatre semaines, chacune se concentrant sur des compétences spécifiques en Python et en analyse de données.

**Objectifs du Projet**

* Apprendre à manipuler des données avec Pandas.
* Visualiser des données avec Matplotlib.
* Effectuer des analyses statistiques de base en utilisant des boucles et des listes.
* Comprendre et appliquer des concepts de régression linéaire et de dérivées.

import pandas as pd

# Exemple de données

data = {

'Ecosysteme': ['Forêt', 'Désert', 'Prairie', 'Marais', 'Océan'],

'Espece': ['Espèce A', 'Espèce B', 'Espèce C', 'Espèce D', 'Espèce E'],

'Valeur': [23, 45, 12, 67, 34] # Nombre d'individus

}

df = pd.DataFrame(data)

print(df)

**Semaine 1 : Introduction à Pandas et Extraction de Données**

1. **Installation et Configuration** :
   * Installer Python et les bibliothèques nécessaires : Pandas et Matplotlib.
   * Télécharger les fichiers .csv fournis.
2. **Exploration des Données** :
   * Charger les fichiers .csv avec Pandas.
   * Utiliser head(), info(), et describe() pour explorer les données.
   * Filtrer les données selon des critères spécifiques (ex : valeurs manquantes, plages de valeurs).
3. import pandas as pd
4. # Charger les données
5. df = pd.read\_csv('data1.csv')
6. # Exploration des données
7. print(df.head())
8. print(df.info())
9. print(df.describe())
10. # Filtrer les données
11. df\_filtered = df[df['valeur'] > 10]

**Semaine 2 : Visualisation des Données avec Matplotlib**

1. **Création de Graphiques Simples** :
   * Histogrammes, boîtes à moustaches, et graphiques en nuage de points.
   * Personnalisation des graphiques (titres, légendes, couleurs).
2. import matplotlib.pyplot as plt
3. # Histogramme
4. plt.hist(df['valeur'], bins=20, color='blue', edgecolor='black')
5. plt.title('Distribution des Valeurs')
6. plt.xlabel('Valeur')
7. plt.ylabel('Fréquence')
8. plt.show()
9. # Nuage de points
10. plt.scatter(df['variable1'], df['variable2'], color='red')
11. plt.title('Relation entre Variable1 et Variable2')
12. plt.xlabel('Variable1')
13. plt.ylabel('Variable2')
14. plt.show()

**Semaines 3 et 4 : Analyse Statistique et Régression Linéaire**

1. **Calculs Statistiques de Base** :
   * Moyennes, médianes, et écarts-types à l’aide de boucles et de listes.
2. # Calcul de la moyenne
3. valeurs = df['valeur'].tolist()
4. somme = 0
5. for valeur in valeurs:
6. somme += valeur
7. moyenne = somme / len(valeurs)
8. print(f'Moyenne: {moyenne}')
9. **Régression Linéaire et Dérivées** :
   * Implémentation de la régression linéaire à partir de zéro.
   * Calcul des dérivées pour analyser les tendances.
10. import numpy as np
11. # Régression linéaire
12. x = df['variable1'].values
13. y = df['variable2'].values
14. n = len(x)
15. xy = sum(x[i] \* y[i] for i in range(n))
16. x\_squared = sum(x[i] \*\* 2 for i in range(n))
17. a = (n \* xy - sum(x) \* sum(y)) / (n \* x\_squared - sum(x) \*\* 2)
18. b = (sum(y) - a \* sum(x)) / n
19. print(f'Equation de la droite: y = {a}x + {b}')
20. # Visualisation de la régression linéaire
21. plt.scatter(x, y, color='blue')
22. plt.plot(x, a \* x + b, color='red')
23. plt.title('Régression Linéaire')
24. plt.xlabel('Variable1')
25. plt.ylabel('Variable2')
26. plt.show()

**Fichiers .csv**

* **data1.csv** : Données sur la diversité des espèces dans différents écosystèmes.
* **data2.csv** : Données sur les niveaux de pollution et leur impact sur la biodiversité.
* **data3.csv** : Données sur la répartition des espèces en fonction des conditions environnementales.
* **data4.csv** : Données sur les changements dans la biodiversité au fil du temps.

Chaque fichier contient entre 50 et 100 lignes de données, avec des colonnes représentant différentes variables biologiques et environnementales.

**Conclusion**

Ce projet permettra aux étudiants de se familiariser avec les bases de la programmation en Python tout en appliquant ces compétences à des données réelles dans le domaine de la biologie cellulaire. Les étudiants apprendront à extraire, visualiser et analyser des données, tout en développant une compréhension des concepts statistiques et de régression linéaire.

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Une image contenant texte, capture d’écran, Police

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement