

## Exercice : *Analyse des Performance des Étudiants*

### Introduction :

Dans cet exercice, vous allez analyser les performances académiques d'un groupe d'étudiants en fonction de leurs habitudes d'étude et de leur mode de vie. L'objectif est de découvrir les facteurs qui influencent le plus les résultats scolaires et de créer un modèle prédictif simple.

### Contexte :

Les établissements scolaires souhaitent souvent comprendre les éléments qui contribuent au succès des étudiants. En analysant des données liées aux habitudes d'étude et aux performances académiques, il est possible d'identifier des tendances permettant d'améliorer les méthodes d'apprentissage.

### Données :

Le fichier `performances_etudiants.csv` contient les données suivantes :

- `heures_etude` : Nombre d'heures d'étude hebdomadaires (hors temps de cours).
- `aux_presence` : Taux de présence en cours (en pourcentage, entre 60 et 100).
- `exercices_completes` : Nombre d'exercices pratiques complétés.
- `heures_sommeil` : Moyenne d'heures de sommeil par jour.
- `note_finale` : Note finale sur 20

### Note :

Ce fichier a été généré automatiquement avec le code suivant :

```
import pandas as pd
import numpy as np

np.random.seed(42)
n_etudiants = 50

data = {
    'heures_etude': np.random.uniform(2, 15, n_etudiants).round(1),
    'taux_presence': np.random.uniform(60, 100, n_etudiants).round(1),
    'exercices_completes': np.random.randint(5, 30, n_etudiants),
    'heures_sommeil': np.random.uniform(5, 9, n_etudiants).round(1),
    'note_finale': np.random.uniform(8, 18, n_etudiants).round(1)
}

df = pd.DataFrame(data)
df.to_csv('performances_etudiants.csv', index=False)
```

## Tâches :

Votre travail est organisé en trois parties principales :

### 1. Exploration et analyse des données (6 points) :

- Affichez les 5 premières lignes du Dataset
- Calculez la moyenne des notes finales
- Déterminez le nombre moyen d'heures d'étude

### 2. Visualisation des données (7 points) :

Créez deux visualisations :

- Un graphique montrant la relation entre les heures d'étude et les notes finales
- Un graphique comparant les notes finales selon le taux de présence

### 3. Modélisation prédictive (7 points) :

Construisez un modèle simple :

- Utilisez une régression linéaire pour prédire les notes finales à partir des heures d'étude
- Faites une prédiction pour un étudiant qui étudie 10 heures par semaine

## Questions pour guider l'analyse :

1. Quelles habitudes d'étude semblent avoir le plus d'impact sur les notes finales ?
1. Existe-t-il des relations surprenantes dans les données ?
2. Quelle est la fiabilité des prédictions dans des scénarios réels ?
3. Quelles données supplémentaires pourraient améliorer l'analyse ?

## Livrables :

- Un notebook Jupyter nommé : `NOM_Prenom_AnalysePerformance.ipynb`.
- Le notebook doit inclure :
  - Le code pour chaque tâche.
  - Des commentaires explicatifs.
  - Des visualisations avec des interprétations concises.

### **Critères Évaluation :**

- Qualité du code (40%) : Clarté, commentaires et organisation.
- Pertinence des analyses (40%) : Exactitude et pertinence des réponses.
- Qualité des Visualisations (20%) : Graphiques clairs, bien étiquetés et pertinents.

### **Notes Importantes :**

- Cet exercice est à réaliser **individuellement**.
- Utilisez les librairies suivantes : `pandas`, `numpy`, `matplotlib`, `seaborn`, `scikit-learn`.
- La date limite de soumission est le **12 janvier 2025, à 23:59**.