**Devoir : *Construction d'un Modèle Prédictif pour la Réussite Académique***

(Une Approche Pratique de l’Apprentissage Automatique)

**Travail Obligatoire - À rendre avant le 13 janvier 2025 à 23 : 59**

**Contexte :**

En tant que Data Scientist débutant, vous allez analyser un ensemble de données réelles sur les performances académiques d’étudiants. Votre objectif est de comprendre les facteurs qui influencent les notes finales et de développer un modèle prédictif simple.

**Données :**

Le fichier 'student-mat.csv' contient des informations sur la performance d’étudiants en mathématiques, incluant :

* age : Âge de l’étudiant
* studytime : Temps d’étude hebdomadaire :
  + 1 = moins de 2 heures
  + 2 = 2 à 5 heures
  + 3 = 5 à 10 heures
  + 4 = plus de 10 heures
* G1 : Note du premier trimestre (0-20)
* G2 : Note du deuxième trimestre (0-20)
* G3 : Note finale (0-20)

**Questions à traiter :**

**Partie 1 : Exploration des Données**

1. Chargez les données et affichez les 5 premières lignes
2. Calculez et interprétez les statistiques descriptives des colonnes G1, G2 et G3

* *Que remarquez-vous concernant l'évolution des notes ?*

1. Identifiez la moyenne et l'écart-type des notes finales (G3)

* *Commentez la distribution des notes*

**Partie 2 : Visualisation**

1. Créez un histogramme de la distribution des notes finales (G3)

* *Décrivez la forme de la distribution*

1. Réalisez un Scatter Plot montrant la relation entre :

* *Les notes du premier trimestre (G1) et les notes finales (G3)*
* *Expliquez la relation observée*

1. Ajoutez des titres et des labels appropriés à vos graphiques

**Partie 3 : Modélisation**

1. Construisez un modèle de **régression linéaire** pour prédire G3 en utilisant G1 et G2
2. Évaluez votre modèle en calculant et interprétant le score R²

* *Que signifie concrètement ce score ?*

1. Faites une prédiction pour un nouvel étudiant ayant obtenu G1=15 et G2=16

* *La prédiction vous semble-t-elle cohérente ? Pourquoi ?*

**Important**

* Les résultats doivent être systématiquement interprétés. Un simple affichage des valeurs ou des graphiques sans explication ne sera pas accepté.
* Chaque visualisation doit être accompagnée d'un commentaire expliquant ce qu'elle révèle.
* Le code doit être commenté pour expliquer votre démarche.

**Livrables**

* Un notebook Jupyter (.ipynb) contenant votre code et vos interprétations
* Nom du fichier : « NOM\_Prenom\_PredictionNotes.ipynb »

**Critères d'Évaluation**

* Qualité et clarté du code (40%) :
* Pertinence des analyses et interprétations (40%)
* Qualité des visualisations (20%)

**Note Importante**

* Ce travail est obligatoire et individuel
* Les librairies nécessaires : **pandas, numpy, matplotlib, seaborn, scikit-learn**
* Utilisez le fichier 'student-mat.csv' fourni avec l’exercice