

Projet de statistique
exploratoire multivariée des
données

**Réalisé par Rebhi
Mohamed Amine**

IDS3

2023/2024



PLan de Projet

1. Comprehension macroscopique de Datasets
2. Présentation des variables
3. Presentation de travail a faire

Préserver la Dataset

 Ensemble de données sur l'attitude et le comportement des étudiants 🎓

Cet ensemble de données contient des informations collectées auprès d'étudiants universitaires via un formulaire Google. Il comprend des détails tels que les cours de certification, le genre, le département, la taille (en cm), le poids (en kg), les notes en 10e et 12e année, les notes universitaires, les passe-temps, le temps quotidien d'étude, l'environnement d'étude préféré, les attentes salariales, la satisfaction à l'égard de leur diplôme, la volonté de poursuivre une carrière liée à leur diplôme, l'utilisation des médias sociaux et des vidéos, le temps de déplacement, le niveau de stress et la situation financière. L'ensemble de données vise à fournir des informations sur le comportement des étudiants et peut être utilisé à des fins d'analyse et de recherche.

Cet Dataset est téléchargé à partir de site Kaggle

<https://www.kaggle.com/datasets/susanta21/student-attitude-and-behavior>

Plusieurs travaux ont été réalisés sur cette Dataset

- ☐ Des analyses EDA
- ☐ Développement des modèles de prédiction

🔍 See what others are saying about this dataset

What have you used this dataset for?

Learning 8

Research 7

Application 1

How would you describe this dataset?

Well-documented 4

Well-maintained 2

Clean data 3

















Original 3

High-quality notebooks 0

Other

Présentation des variables

Descriptions des colonnes :

- Cours de certification : Indique si l'étudiant a terminé un cours de certification ou non. 
- Genre : Le genre de l'étudiant. 
- Département : Le département ou le domaine d'études dans lequel l'étudiant est inscrit. 
- Taille (CM) : La taille de l'étudiant en centimètres. 
- Poids (KG) : Le poids de l'étudiant en kilogrammes. 
- Note de 10e : Les notes obtenues par l'étudiant en 10e année. 
- Note de 12e : Les notes obtenues par l'étudiant en 12e année. 
- Note universitaire : Les notes obtenues par l'étudiant dans son collège ou son université. 
- Passe-temps : Les passe-temps ou intérêts de l'étudiant. 
- Temps d'étude quotidien : Le temps que l'étudiant passe à étudier quotidiennement. 
- Préférez-vous étudier dans : L'environnement ou l'emplacement d'étude préféré de l'étudiant. 
- Attente salariale : L'attente salariale future de l'étudiant. 
- Aimez-vous votre diplôme ? : L'opinion de l'étudiant sur s'il aime son diplôme ou non. 
- Volonté de poursuivre une carrière basée sur leur diplôme : La volonté de l'étudiant de poursuivre une carrière liée à son diplôme. 
- Médias sociaux & Vidéo : L'engagement de l'étudiant avec les plateformes de médias sociaux et de vidéo. 
- Temps de déplacement : Le temps nécessaire à l'étudiant pour se rendre à son établissement d'enseignement. 

- Niveau de stress : Le niveau de stress perçu par l'étudiant. 😓
- Situation financière : La situation financière ou le contexte économique de l'étudiant. 💰
- Emploi à temps partiel : Si l'étudiant occupe un emploi à temps partiel ou non. 💼

On va utiliser cet ensemble de données à des fins d'analyse et de recherche ! 📊🔍

Le Travail à faire

D'accord, nous pouvons adapter les étapes de l'analyse en conséquence. Voici comment j'ai procédé :

1. Analyse univariée :
 - Pour chaque variable, effectuez une analyse univariée pour comprendre sa distribution et ses caractéristiques statistiques.
2. Analyse bivariée :
 - Examinez les relations entre paires de variables pertinentes. Vous pourriez, par exemple, étudier la relation entre le temps d'étude quotidien et les notes universitaires.
3. Matrice de corrélation :
 - Calculez une matrice de corrélation pour examiner les relations linéaires entre les variables quantitatives. Identifiez les paires de variables fortement corrélées.
4. Prédiction avec plusieurs modèles sur les variables stress et les notes universitaires :
 - Divisez les données en ensembles d'entraînement et de test.
 - Utilisez différents modèles de régression ou de classification pour prédire les notes universitaires en fonction du niveau de stress et d'autres variables pertinentes.

- Évaluez la performance des modèles à l'aide de mesures d'évaluation appropriées (par exemple, RMSE, R2, etc.).
5. Prédiction avec plusieurs modèles sur les notes universitaires :
 - Utilisez les mêmes modèles pour prédire les notes universitaires en fonction d'autres variables (comme les caractéristiques démographiques, les intérêts, etc.).
 - Comparez les performances des différents modèles.
 6. Analyse en composantes principales (PCA) :
 - Effectuez une PCA pour réduire la dimensionnalité des données.
 - Identifiez les composantes principales qui capturent le plus de variance dans les données.
 - Utilisez les composantes principales pour visualiser les relations entre les variables ou pour créer de nouvelles variables pour la modélisation.
 7. Discussion des résultats :
 - Résumez les principaux résultats de chaque analyse.
 - Discutez des implications des résultats pour comprendre comment le stress, les caractéristiques des étudiants et d'autres facteurs influencent les performances universitaires.
 - Identifiez les limitations de l'étude et suggérez des pistes pour de futures recherches.



Conclusion

En résumé, cette analyse met en évidence l'importance de comprendre les déterminants des performances académiques des étudiants, en mettant particulièrement l'accent sur le rôle du stress et des caractéristiques individuelles

