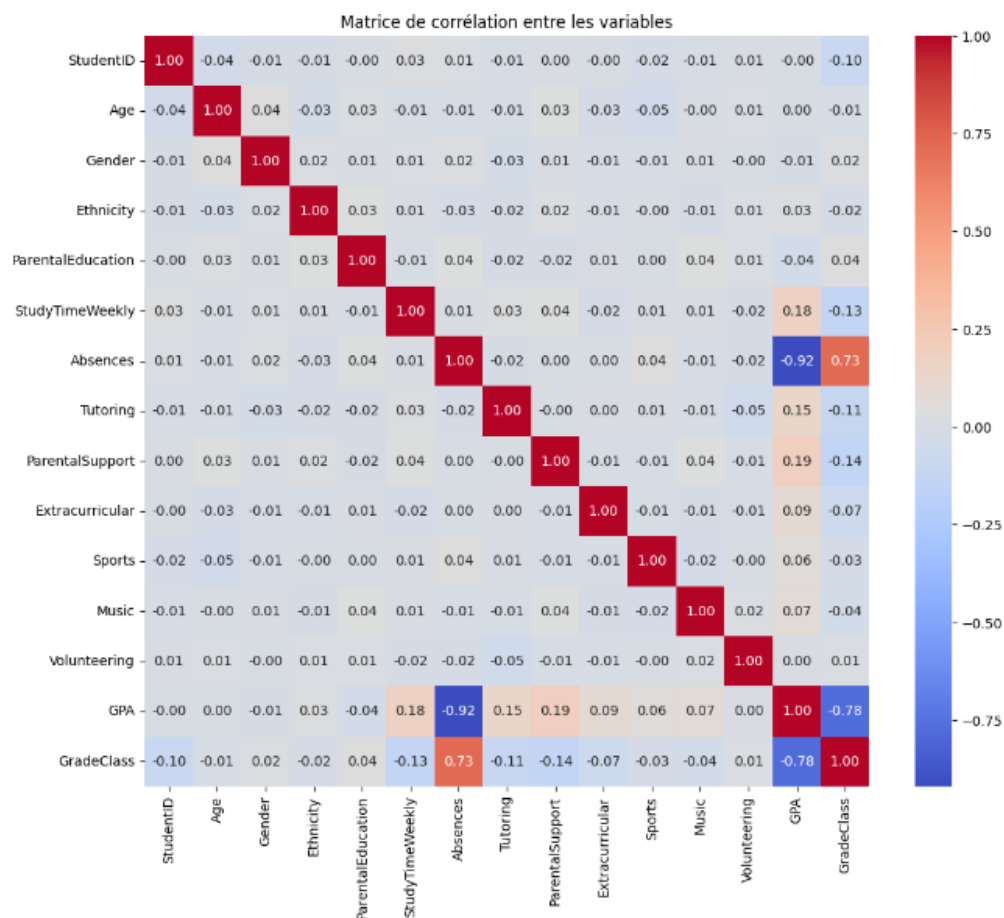
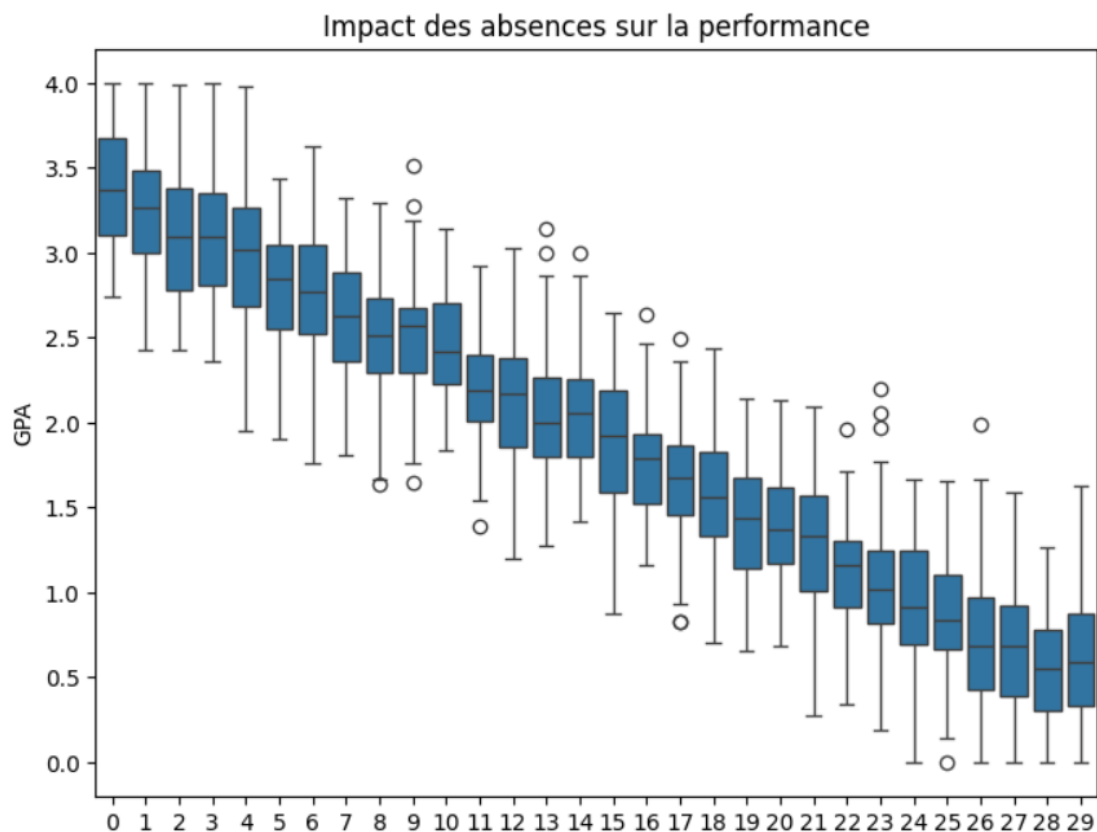




RAPPORT DE L'ANALYSE DE NOTE



Corrélation entre les heures d'étude et la moyenne des notes : 0.1792751268560882
 Corrélation entre l'âge et la moyenne des notes : 0.00027538819833585246
 Corrélation entre le genre et la moyenne des notes : -0.013359974928969455
 Corrélation entre le niveau d'éducation des parents et la moyenne des notes : -0.035853639627925205
 Corrélation entre le support scolaire et la moyenne des notes : 0.14511891996780377
 Corrélation entre les absences et la moyenne des notes : -0.9193135764208646
 Corrélation entre le groupe ethnique et la moyenne des notes : 0.027759583777886312
 Corrélation entre le support parental et la moyenne des notes : 0.19077372781932933
 Corrélation entre les heures d'étude et la moyenne des notes est de : -0.9193135764208646



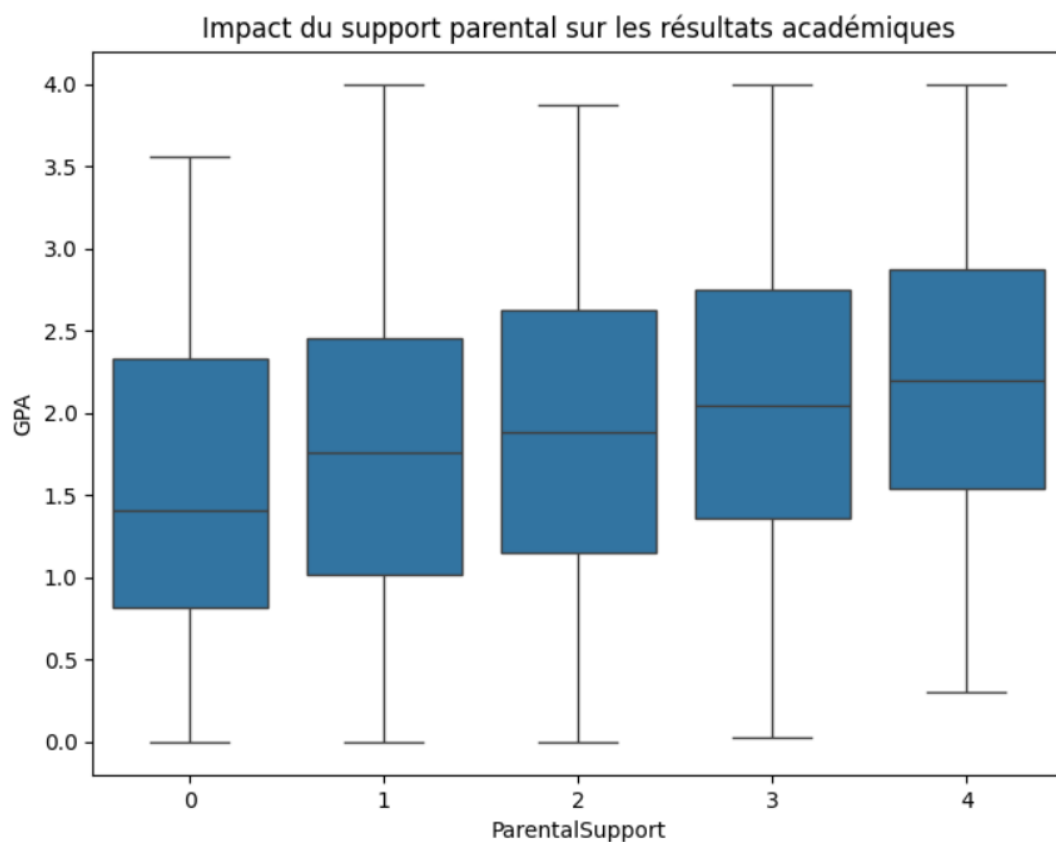
Il y a une faible corrélation entre les heures d'étude, c'est-à-dire que le plus le nombre d'heures d'étude par semaine augmente le moyen des notes à tendance à augmenter.

Au vu de la corrélation d'entre les absences, on peut dire que l'absence affecte négativement la moyenne des notes.

Les supports parentaux exercent une influence positive sur la note générale de l'étudiant

L'influence des activités extra-scolaires sur la moyenne des notes est presque nulle donc les activités extra-scolaires n'influencent quasiment pas la moyenne des notes.

L'âge et/ou le niveau d'éducation des parents n'influe pas du tout sur la performance des étudiants



Les facteurs qui semblent être les meilleurs pour prédire les notes finales sont :

- Les absences
- Le tutorat
- Le support parental
- Les heures d'études

NB : les activités extra-scolaires les influencent mais pas remarquablement

En utilisant un modèle de régression, vous pouvez modéliser la performance des étudiants à partir des autres variables. Voici une approche possible:

Importance des variables dans la forêt aléatoire:

	Variable	Importance
1	Absences	0.854596
0	StudyTimeWeekly	0.059451
2	ParentalSupport	0.032520
10	Tutoring	0.014651
3	Extracurricular	0.008124
4	Sports	0.007510
9	ParentalEducation	0.006940
8	Age	0.006806
5	Music	0.004475
7	Gender	0.002752
6	Volunteering	0.002175

Régression linéaire:

MSE: 0.03857657618962719

R2: 0.95334979472927

Au vu des résultats du MSE et R2, le modèle qui décrit le plus les données et le modèle de régression linéaire.

Car il décrit plus de 95% des données.

Arbre de décision:

MSE: 0.11136195560518695

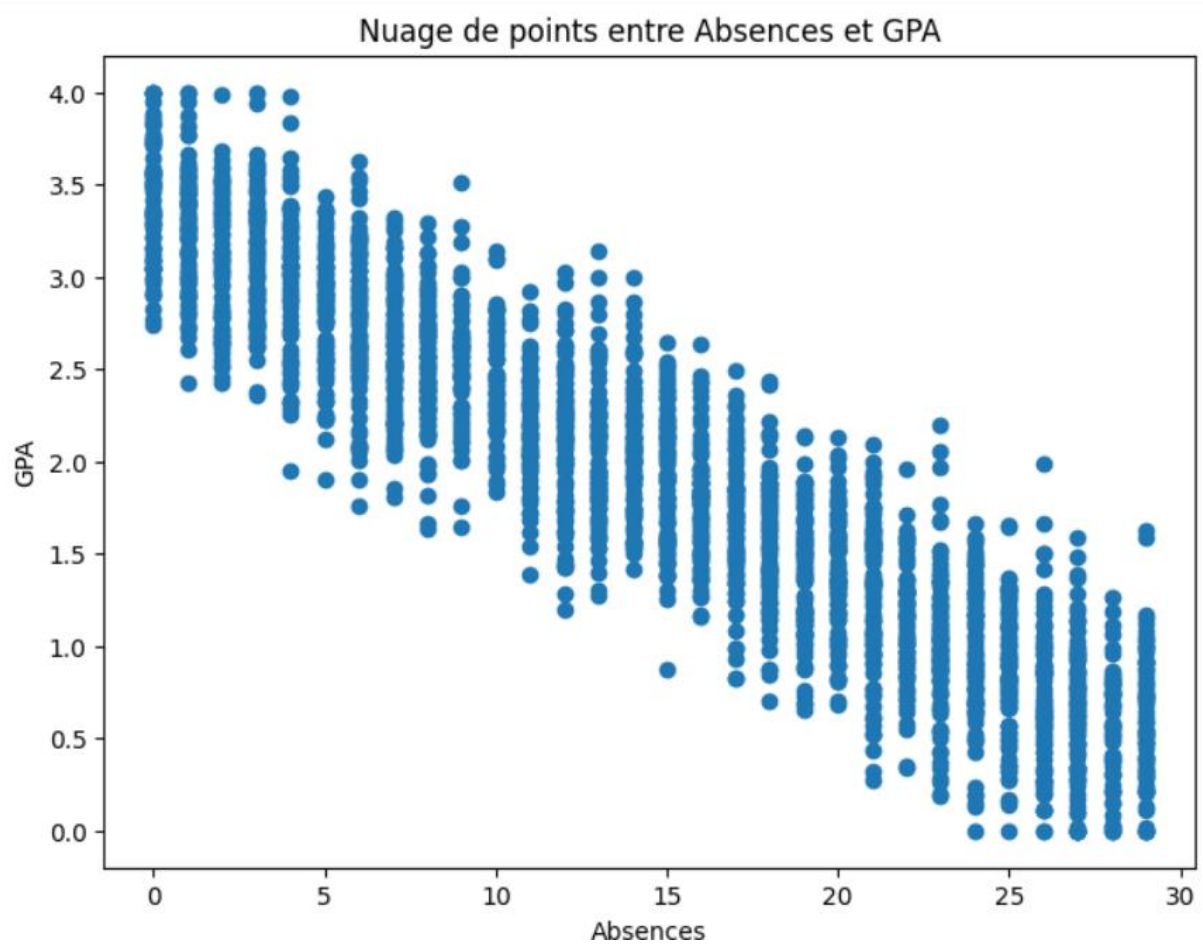
R2: 0.8653312812729919

Forêt aléatoire:

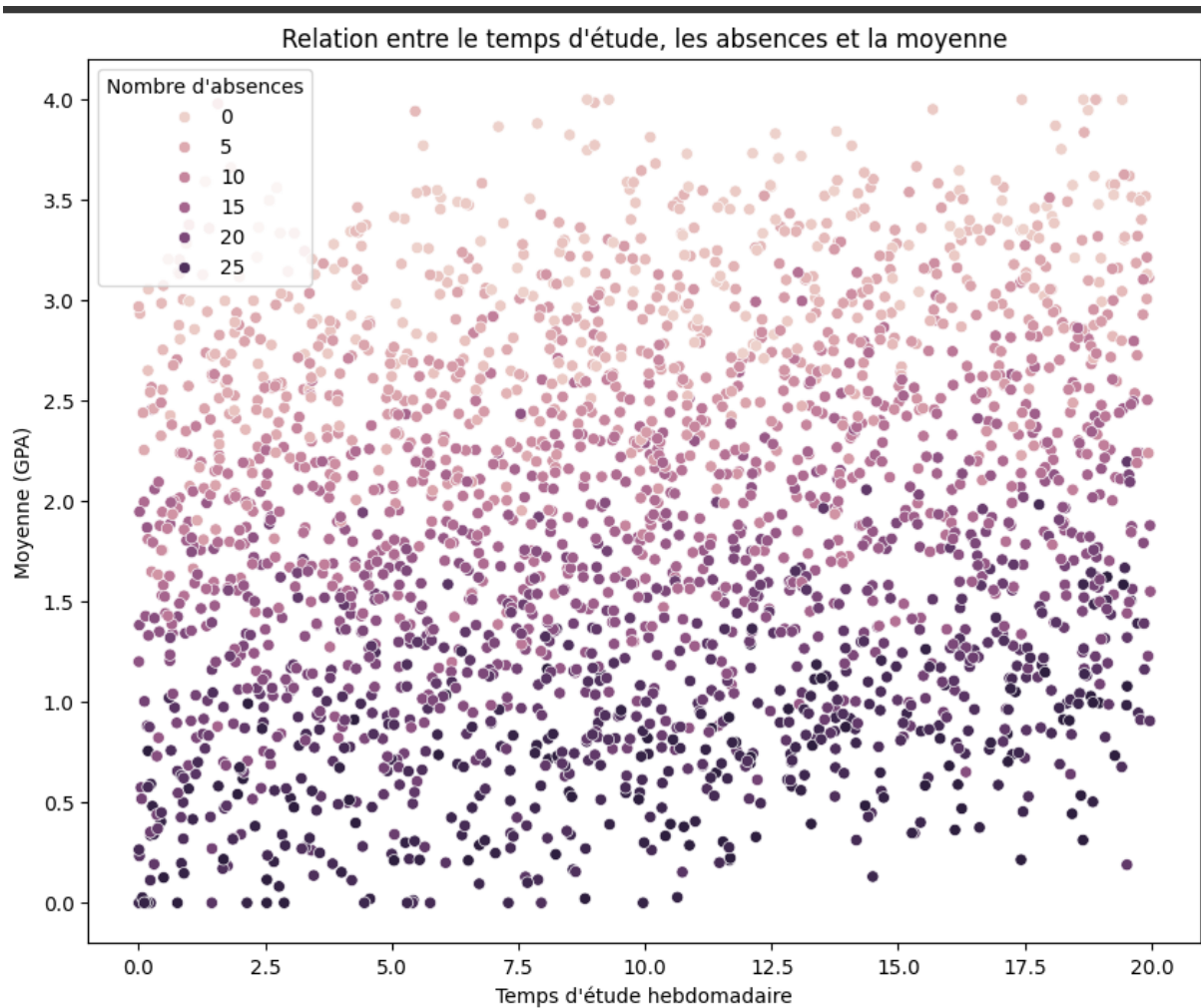
MSE: 0.058024550643929286

R2: 0.929831585235161

Pour mieux comprendre la tendance dans les données de notre data frame on pourrait utiliser la barre à moustache (complot), qui montre quasiment très bien la distribution des données les unes par rapport aux autres. Le nuage de points (scatterplot) pour bien aussi montrer la distribution des données mais juste que sur quelles qu'unes « telles que les absents par rapport à la moyenne des notes ». Par exemple ci-dessous :



Exemple de représenter visuellement la relation entre plusieurs variables :



En conclusion, les absences influe énormément sur la note, ainsi il est mieux d'être le plus présent sur en cours, et pour les parent, il est important de suivre l'enfant à la maison, on l'aide à réviser et on lui prenons un tuteur pour les cours(répétiteur).

L'utilisation des résultats peut alimenter un chatbot; en utilisant les modèles de régression linéaire pour prédire la performance des étudiants en fonction de leurs données; en fournissant des conseils personnalisés basés sur les résultats de l'analyse, en fonction des variables clés comme le temps d'étude, les absences, le support parental, le tutoring, et les activités extrascolaires; et également informer les étudiants sur les facteurs qui influencent leur performance.