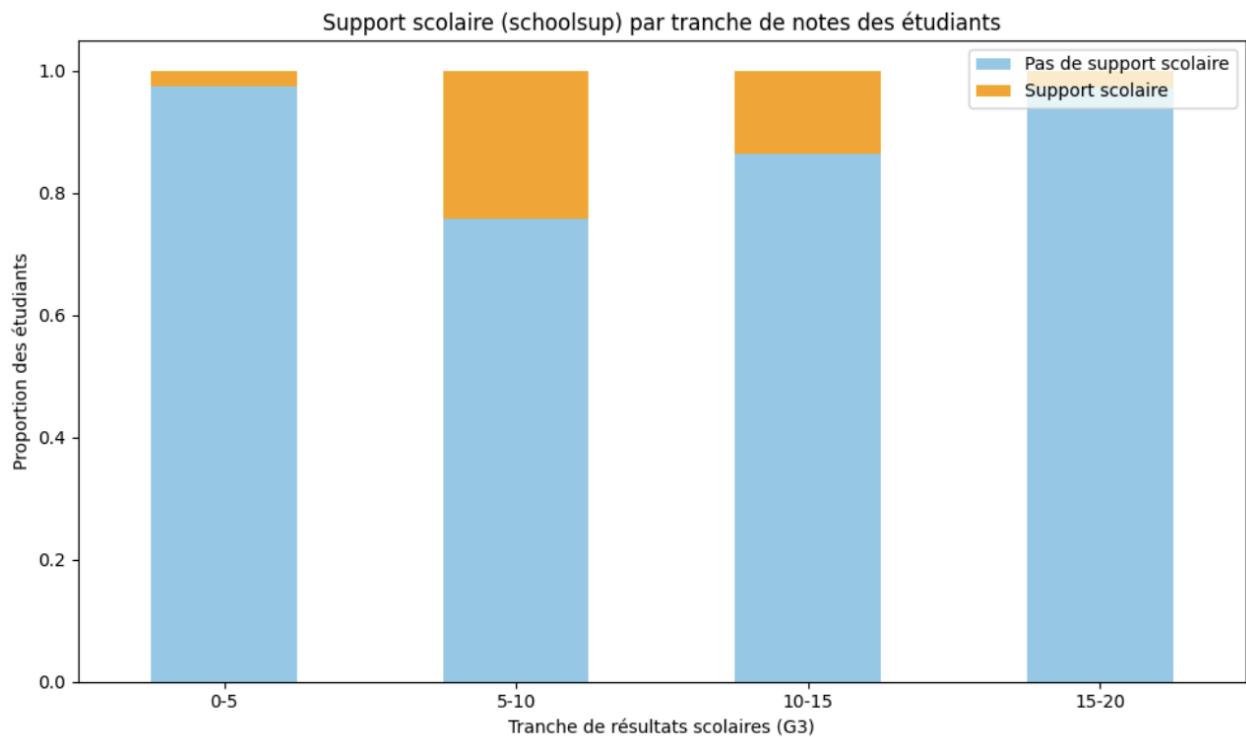


19/12/2024



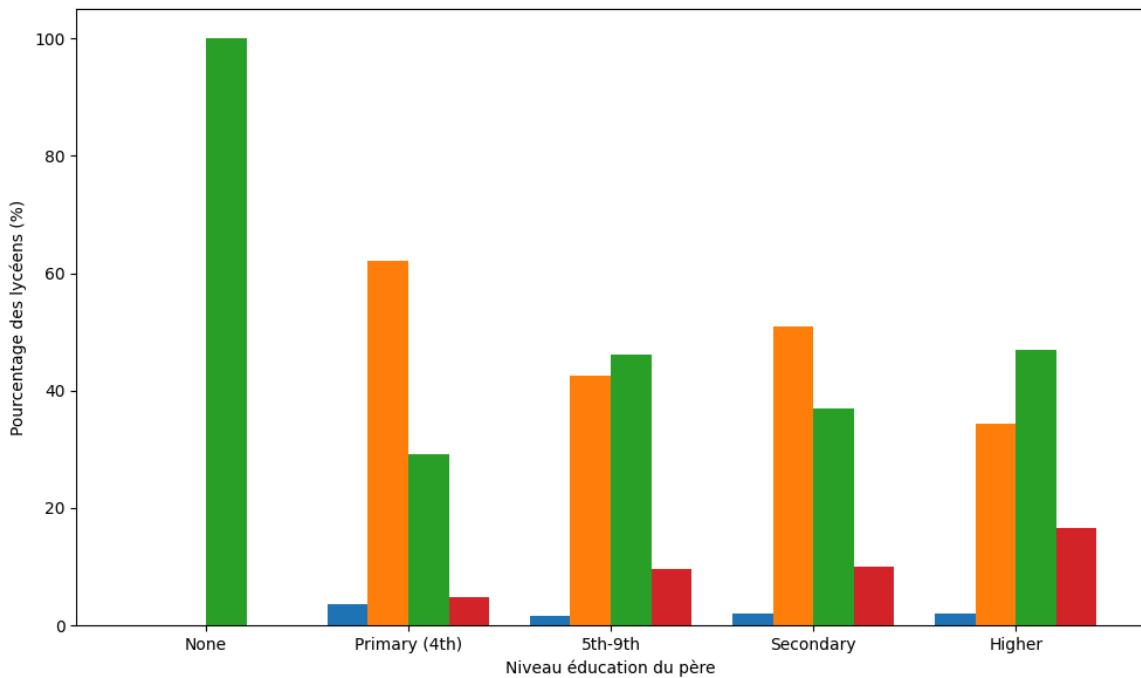
05/01/2025

Le graphique ci-dessus expose une proportion d'étudiants par tranches de notes et suivant une présence ou une absence de soutien scolaire fourni par leurs parents. On peut apercevoir une situation inattendue. Tout d'abord, une tendance émerge : la majorité des étudiants ne reçoivent pas de soutien scolaire, peu importe la tranche de notes. Seule une infime partie des étudiants, de l'ordre de 0.03%, reçoivent un soutien scolaire parmi les notes allant entre zéro à cinq et les notes allant de quinze à vingt. De plus, la proportion d'étudiants recevant le plus un support scolaire, soit 0,25% environ, se trouve parmi les notes allant de cinq à dix. La seconde place revient à la proportion d'étudiants de la tranche de notes allant de dix à quinze, soit 0,15% environ.

Les extrêmes doivent-ils être analysés indépendamment ?

Interprétation possible -> les parents des étudiants avec des notes entre 0 et 5 se disent que c'est peine perdue, d'où les résultats scolaires. les parents des étudiants avec des notes entre 15 et 20 se disent que le soutien scolaire n'est pas nécessaire. Voilà pourquoi il y aurait un plus haut taux d'étudiants bénéficiant d'un soutien scolaire dans la zone entre 5 et 15 (les parents estiment qu'il leur faut un coup de pouce j'imagine).

Peut-on réellement en tirer quelque chose d'utile ?



Ce graphique est à revoir voir supprimer

Le graphique présente la répartition en pourcentage des lycéens selon leurs résultats scolaires (divisés en quatre tranches : zéro-cinq, cinq-dix, dix-quinze, et quinze-vingt) en fonction du niveau d'éducation de leur père. Ces informations permettent d'explorer la relation entre le niveau d'éducation du père et les performances scolaires des étudiants. Une nouvelle tendance générale apparaît : plus le niveau d'éducation du père est élevé, plus les résultats scolaires des lycéens tendent à être meilleurs, avec une transition progressive des notes faibles (zéro-cinq et cinq-dix) vers les notes élevées (dix-quinze et quinze-vingt).

Dans les catégories où le père a un faible niveau d'éducation, comme "Primary (4th)" ou "5th-9th", les résultats scolaires des étudiants sont majoritairement concentrés dans les tranches de notes de cinq à dix et de dix à quinze. Cela indique que les étudiants de ces groupes sont plus susceptibles d'obtenir des résultats moyens, bien que certains parviennent à de bonnes performances ou dans le cas contraire de très mauvaises. La tranche des notes de zéro-cinq, représentant les résultats très faibles, reste minoritaire dans ces catégories mais elle est légèrement plus présente que dans les niveaux d'éducation supérieurs ce qui n'est pas à négliger. À l'inverse, la tranche des notes de quinze-vingt, des notes excellentes, est aussi minoritaire mais ne cesse de grandir avec le niveau d'éducation du père.

Lorsque le niveau d'éducation du père atteint un niveau "Secondary" ou "Higher", la proportion d'étudiants obtenant des notes élevées (dix-quinze et quinze-vingt) augmente considérablement, tandis que les notes faibles (zéro-cinq) restent presque inexistantes. Cela reflète une probable influence positive du niveau d'éducation du père sur les opportunités et le soutien offerts aux étudiants, ce qui contribue ainsi à leur réussite scolaire.

corrélation ?

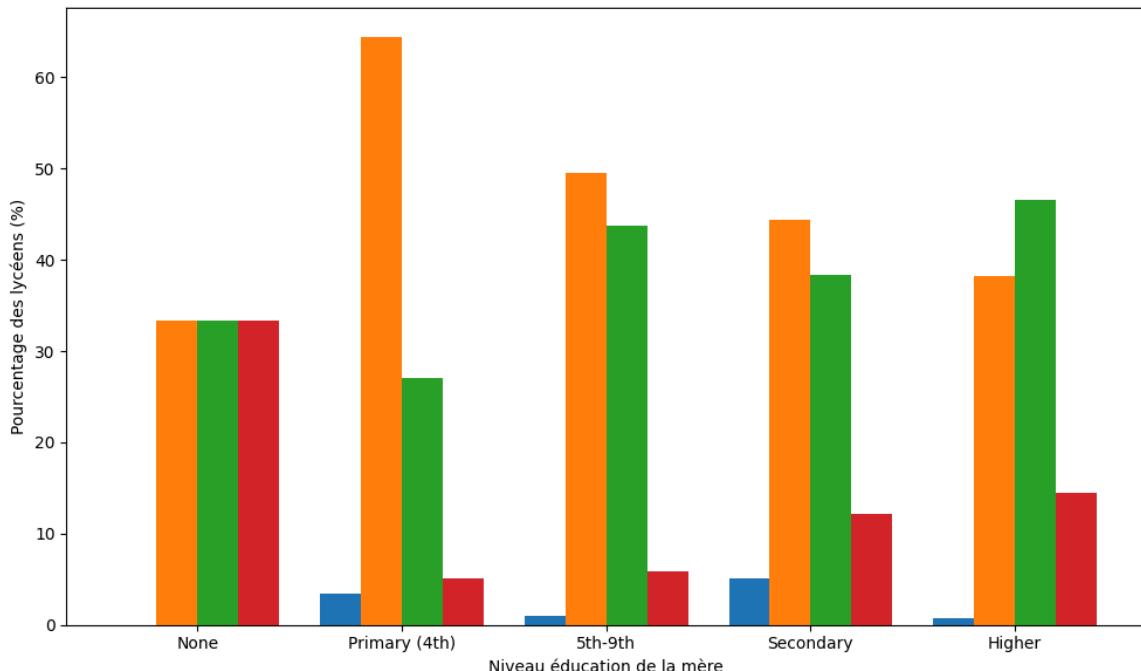
interprétations -> niveau d'études du père influence résultats du fils

on observe une progression claire des résultats scolaires (des notes faibles vers des notes élevées) lorsque le niveau d'éducation du père augmente.

les étudiants dont le père a un niveau d'éducation primaire ou inférieur obtiennent majoritairement des résultats faibles à moyens (tranches bleue et orange).

à l'inverse, les élèves avec un père ayant un niveau d'éducation supérieur (Higher) ont une plus grande probabilité d'obtenir des résultats dans les tranches de 10-15 (vert) et 15-20 (rouge).

pour "none" où les infos du niveau d'éducation des pères manquent, la totalité des étudiants ont des notes entre dix et quinze ce qui est plutôt pas mal mais on ne peut rien en tirer. il faudrait creuser cette piste dans le cas où on aurait l'info suivante -> il nous faudrait le nombre d'étudiants dans cette situation pour pouvoir prétendre en tirer des informations. (car si il y a seulement deux étudiants dans cette situation, aucune corrélation possible car effectif trop faible). un effectif aussi réduit serait insuffisant pour représenter une population.



Dans les catégories où la mère a un faible niveau d'éducation, comme "Primary (4th)" ou "5th-9th", les résultats scolaires des étudiants sont majoritairement concentrés dans les tranches de notes de cinq à dix et de dix à quinze. Cela indique que les étudiants de ces groupes sont plus susceptibles d'obtenir des résultats moyens bien que certains parviennent à de bonnes performances ou même l'inverse. La tranche des notes de zéro-cinq, représentant les résultats très faibles, reste marginale dans ces catégories mais elle remonte légèrement dans la catégorie "Secondary" et redescend ensuite pour atteindre son minimum dans la catégorie "Higher" -> une certaine impression que l'on en tirera pas grand chose et ça n'a pas l'air logique. À l'inverse, la tranche des notes de quinze-vingt, représentant des résultats excellents, est également minoritaire mais montre une progression constante à mesure que le niveau d'éducation de la mère augmente.

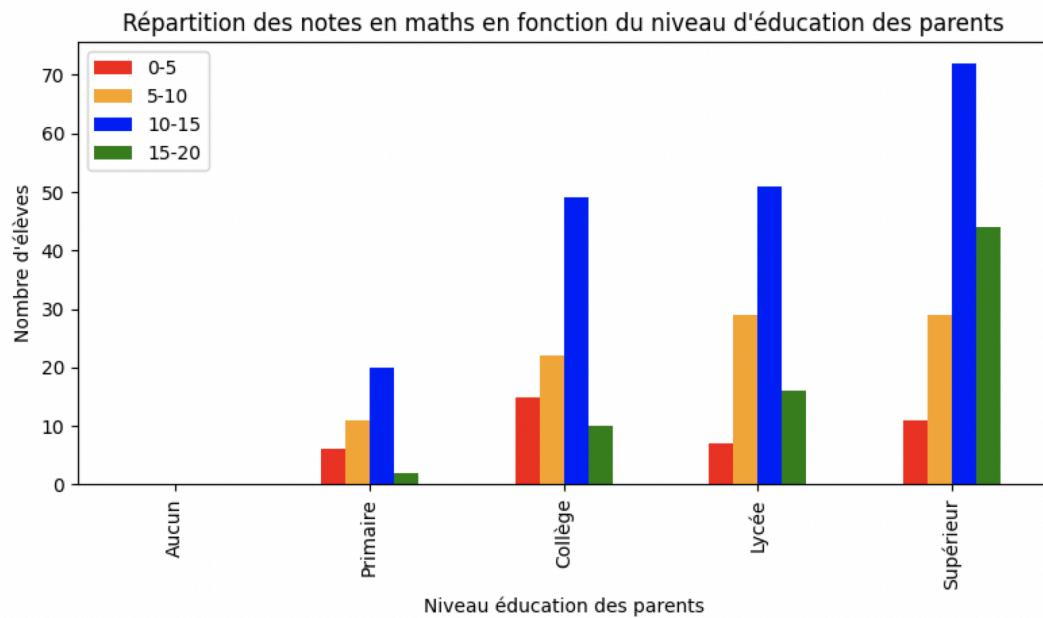
La tranche des notes de quinze-vingt, représentant des résultats excellents montre une progression constante avec l'élévation du niveau d'éducation de la mère. Cette tranche reste faible dans les catégories "Primary (4th)" et "5th-9th", progresse de manière notable dans la catégorie "Secondary" et atteint son maximum dans la catégorie "Higher". Cette tendance est cohérente avec l'idée qu'un niveau d'éducation maternel plus élevé offre aux lycéens un meilleur soutien éducatif ou un environnement propice à la réussite.

difficile d'en conclure quelque chose de cohérent ?

essayer de fusionner les deux graphiques ? -> ne plus prendre séparément le niveau d'éducation des deux parents ?

graphique attendu

07/01/2025

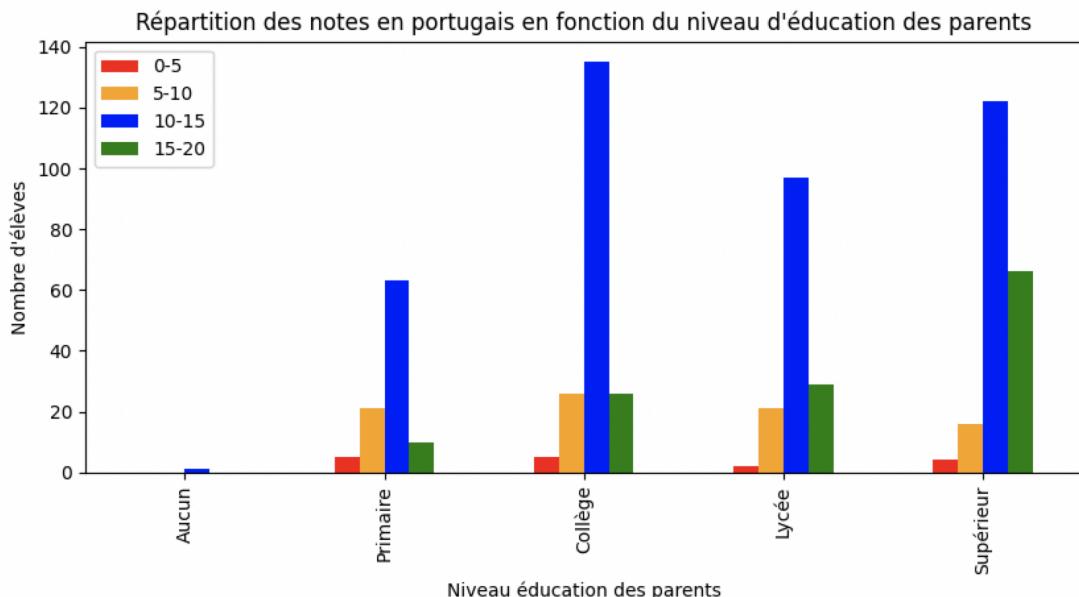


08/01/2025

baisse des notes entre 0-5 qui deviennent minoritaire à partir d'un niveau d'éducation des parents supérieur ou équivalent au lycée bien que ce ne soit pas le cas pour un niveau d'éducation parental de niveau collège ou inférieur

des notes entre 5-10 augmente avec le niveau d'éducation des parents jusqu'au niveau lycée, c'est vraiment le mauvais point ici. (aucune distinction sur ce point là entre le niveau lycée et supérieur.

forte augmentation des notes entre 10-20.



pas d'influence entre les notes de 0-5 peu importe le niveau d'éducation des parents, en effet, les notes entre 0-5 restent négligeables, très minimes tout le long.

bonne baisse des notes entre 5-10 à prendre en compte -> bon point.

très bonne augmentation des notes entre 15-20 -> très bon point comme dans le graphique précédent.

le petit hic -> les notes entre 10-15. On peut dire qu'elles restent constantes quasiment. mais on ne peut pas négliger qu'elles varient positivement à partir d'un niveau d'éducation des parents de niveau primaire vers collège et qu'elles baissent directement après à partir du niveau collège vers lycée.

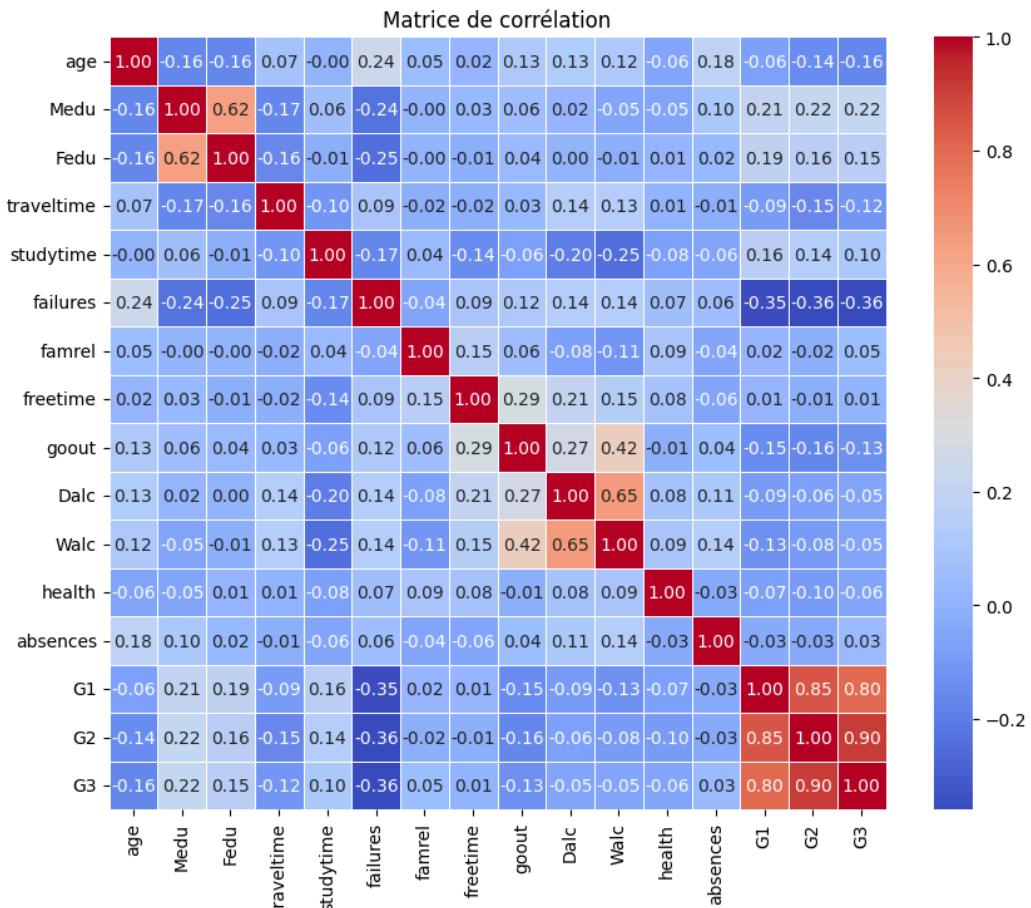
La majorité des élèves peu importe le niveau d'éducation de leurs parents ont des notes entre 10-15.

-> intéressant d'avoir généré des graphiques sans distinction entre le père et ma mère

08/01/2025

corrélations possibles ? liens éventuels ? ce que l'on peut se demander:

- lieu pouvant favoriser le décrochage scolaire (Zone urbaine, rurale, autre ?..)
- le prestige/ la qualité de l'école (privé ? publique ? / simplement la réputation / la qualité)
- la présence / absence des parents
- le niveau d'éducation, d'étude des parents
- problèmes de santé, handicap de l'enfant
- les attentes des parents : trop forte -> stress de l'enfant
trop faible -> manque d'assiduité
- avoir le type de jeu (console / mobile / autre, etc.)
- fréquence et temps de jeu (comment le temps de jeu est reparti)
- le temps de loisir et la proportion du jeu vidéo dedans



Corrélations significatives avec les performances scolaires (G3) (notes de maths)

- Échecs scolaires (failures) :**
 - G3 est négativement corrélé avec **failures** (-0.36).
 - Les élèves ayant accumulé plus d'échecs dans le passé ont tendance à avoir de moins bonnes performances finales.
- Temps d'étude (studytime) :**
 - G3 est faiblement corrélé avec **studytime** (0.10).
 - Les élèves consacrant plus de temps aux études ont des performances légèrement meilleures.

Facteurs comportementaux et sociaux :

- Alcool (Dalc et Walc) :**
 - Dalc (consommation d'alcool en semaine) et Walc (consommation d'alcool le week-end) sont faiblement mais négativement corrélés avec G3 (-0.05 et -0.06 respectivement).
 - Une consommation d'alcool plus élevée est associée à de moins bonnes performances scolaires.
- Sorties (goout) :**
 - goout (fréquence des sorties) est faiblement négativement corrélé avec G3 (-0.13).
 - Les élèves qui sortent plus souvent tendent à avoir des notes légèrement inférieures.

■

1. Relation entre Dalc et Walc :

- Une forte corrélation (0.65) indique que les élèves consommant de l'alcool pendant la semaine ont aussi tendance à consommer davantage le week-end.

1. L'impact du temps d'étude et de l'absentéisme :

- Même si la corrélation est faible, il serait pertinent d'explorer comment les comportements liés à l'étude influencent les résultats finaux.

2. Les comportements sociaux (alcool, sorties) :

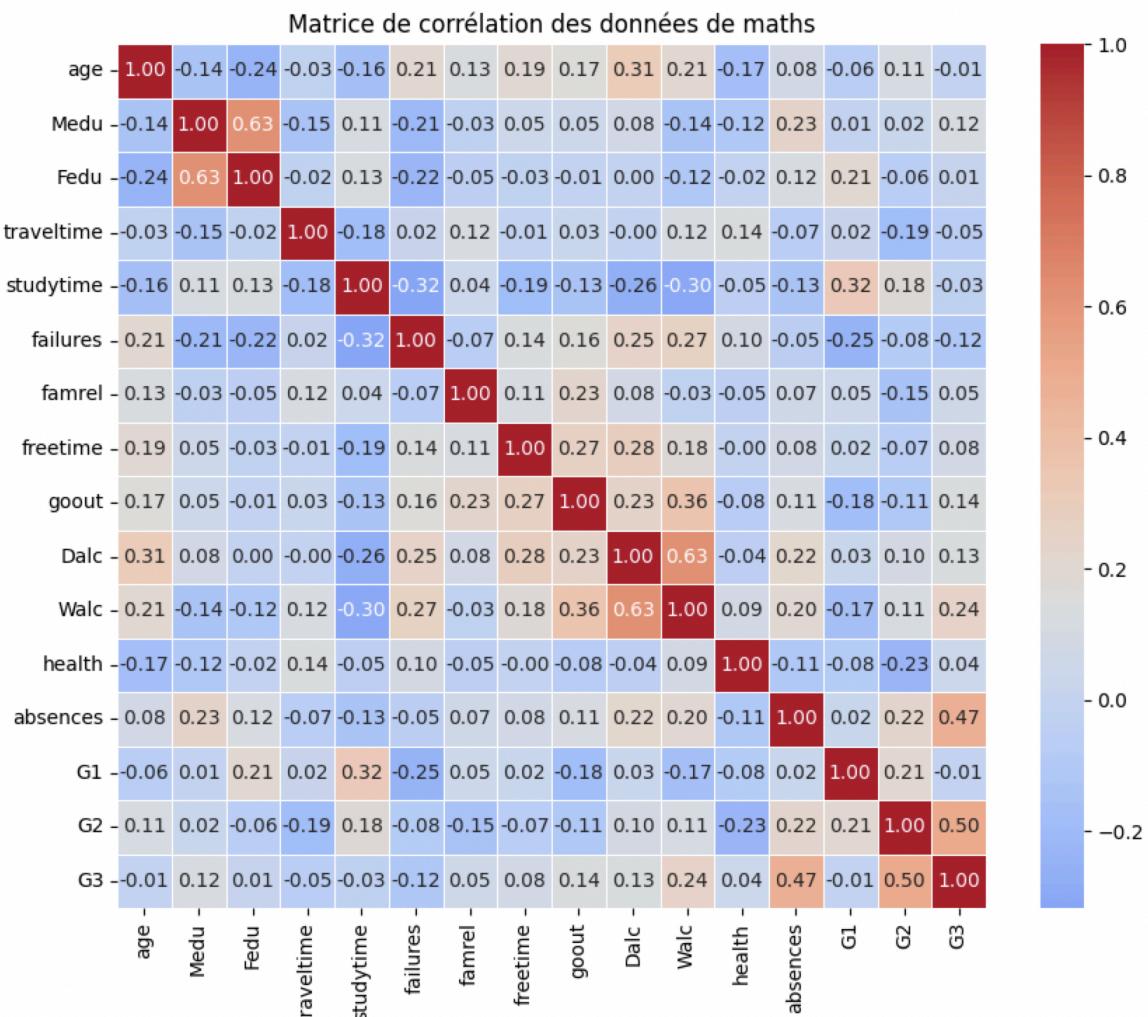
- La consommation d'alcool et les sorties, bien que faiblement corrélés, pourraient être étudiés davantage comme des indicateurs de dépendance sociale.

3. Niveau d'éducation des parents :

- Une dépendance entre l'éducation des parents et les performances des élèves peut être explorée davantage.

10/01/2025

matrice de corrélation découpée en trois groupes de notes, 0-8, 9-13, 14-20



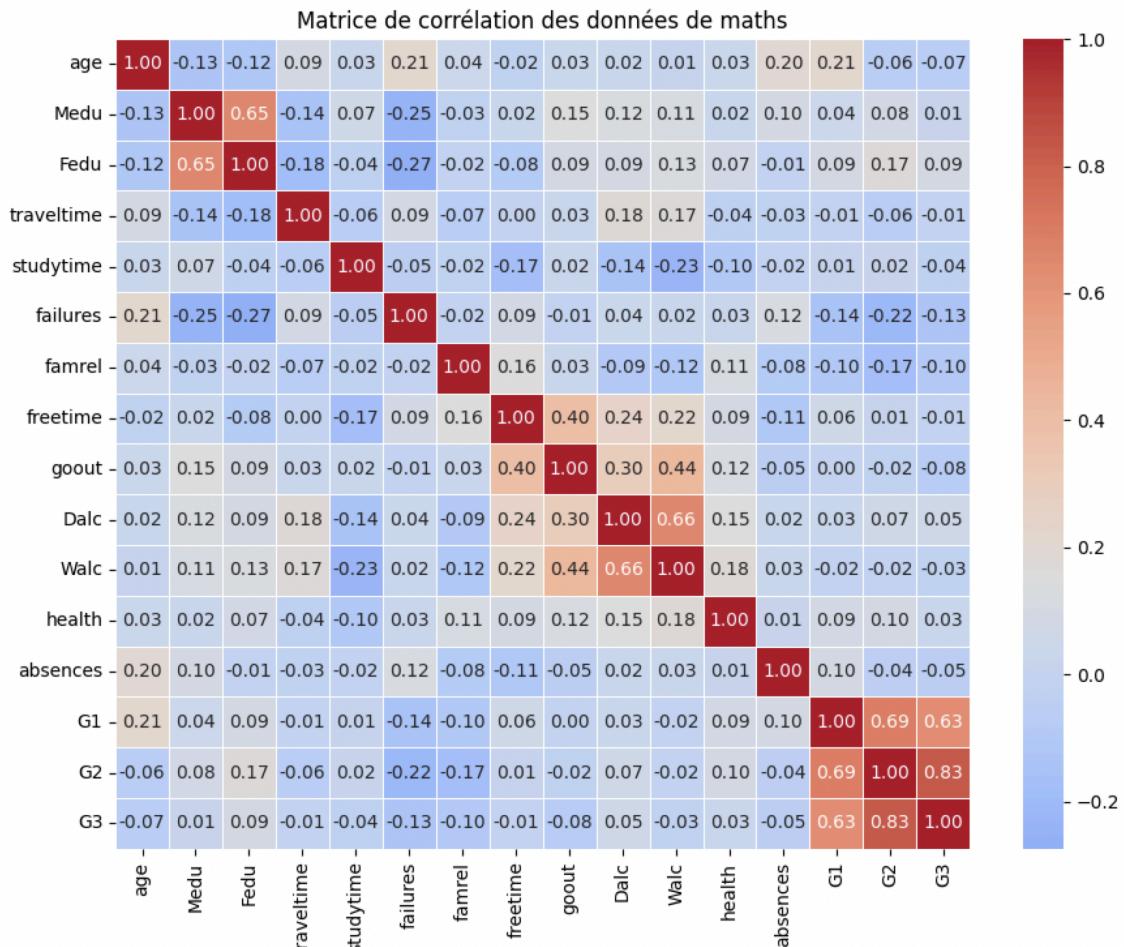
Corrélations avec G3:

[0-8]

- G3 montre une corrélation très faible avec Medu (≈ 0.12) et Fedu (≈ 0.09), suggérant que l'éducation des parents n'a pas un impact significatif pour ce groupe de faibles notes.
- Une corrélation modérée avec absences (≈ -0.23) : davantage d'absences tend à réduire les notes finales.

Autres observations importantes :

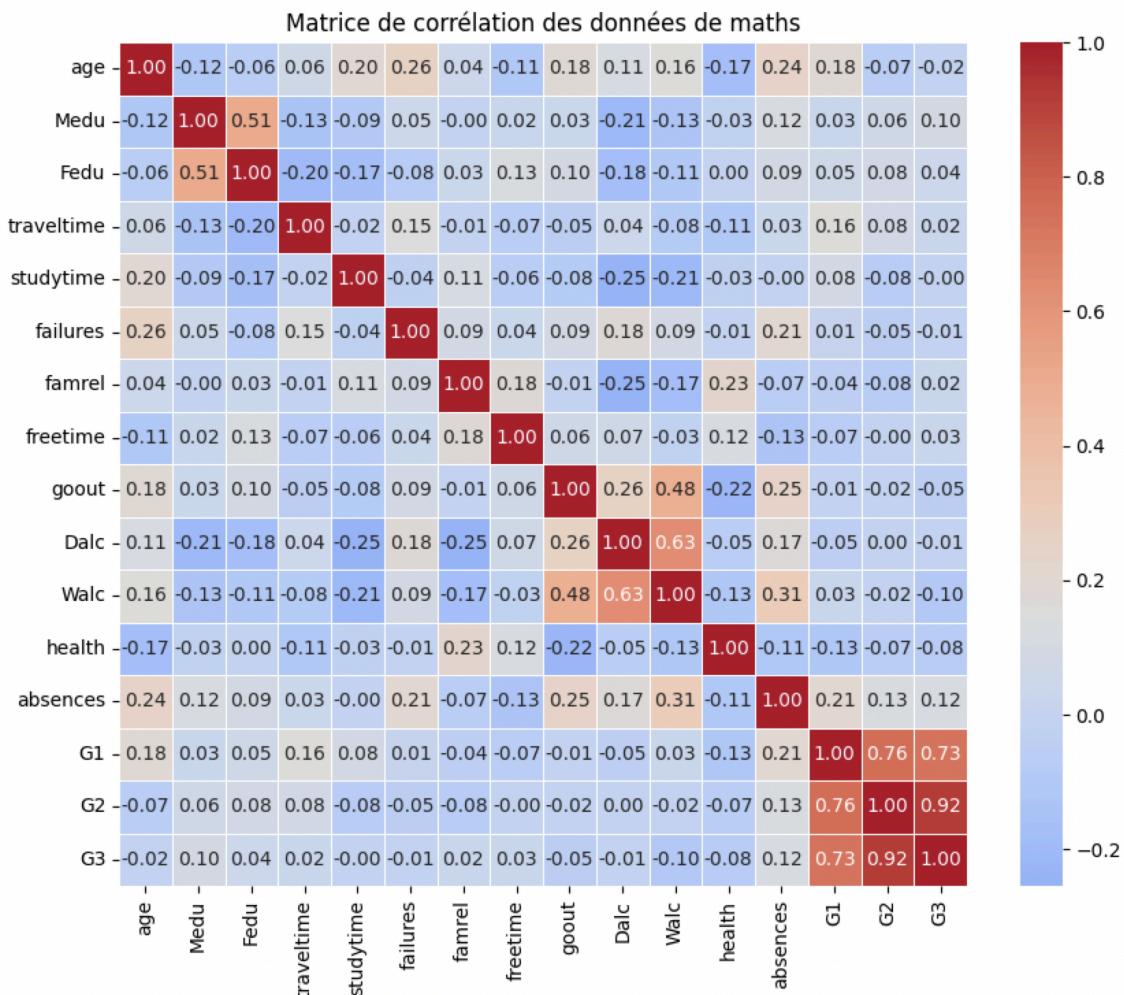
- Dalc (consommation d'alcool en semaine) et Walc (consommation d'alcool le week-end) ont une corrélation modérée (≈ 0.63). Les comportements liés à l'alcool sont généralement cohérents.
- failures a une corrélation négative significative avec studytime (≈ -0.30) et une corrélation modérée avec G3 (≈ -0.25). Cela montre que les échecs passés sont liés à des habitudes d'étude moins productives.
- Les faibles performances dans ce groupe semblent être influencées par des facteurs comme les absences et les échecs passés.



observations importantes :

[9-13]

- **studytime** a une corrélation plus positive avec **G3** (≈ 0.22), montrant que le temps d'étude devient un facteur plus déterminant pour ce groupe.
- **failures** montre une corrélation négative modérée avec **G3** (≈ -0.22), mais moins forte que dans le groupe 0-8. Cela suggère que les échecs passés sont moins préjudiciables dans cette tranche de notes.
- **Dalc** et **Walc** restent fortement corrélés (≈ 0.66), montrant des habitudes cohérentes dans la consommation d'alcool.



Corrélations avec **G3** :

[14-20]

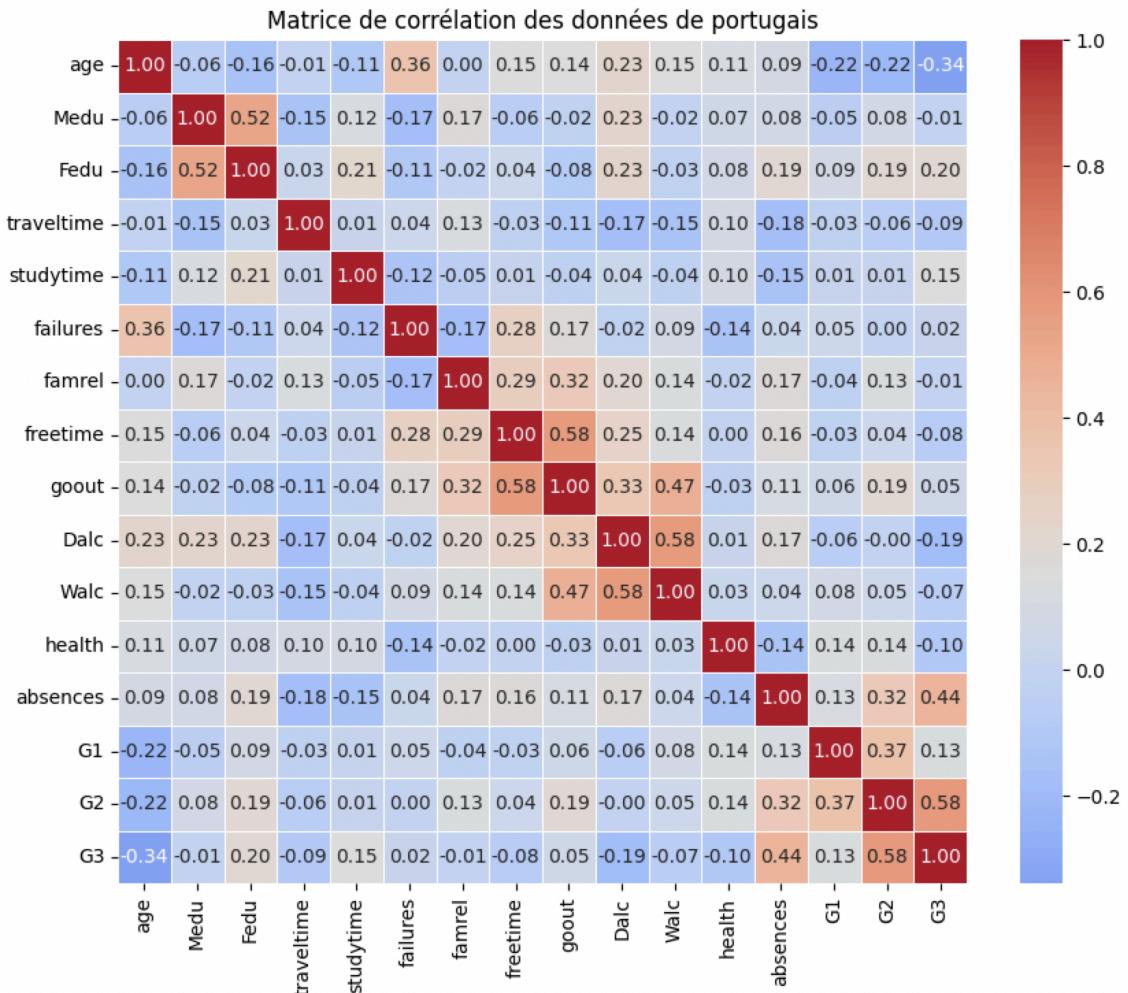
- **Medu** (≈ 0.10) et **Fedu** (≈ 0.04) ont une influence marginale. Les élèves performants semblent moins dépendants du niveau d'éducation des parents.
- **studytime** montre une corrélation faible mais positive (≈ 0.12), ce qui reflète que le temps d'étude n'est pas le seul facteur pour les excellents élèves.

Autres observations importantes :

- Les corrélations avec les absences sont faibles ou inexistantes (≈ 0.12), suggérant que les absences ne pénalisent pas les meilleurs élèves.

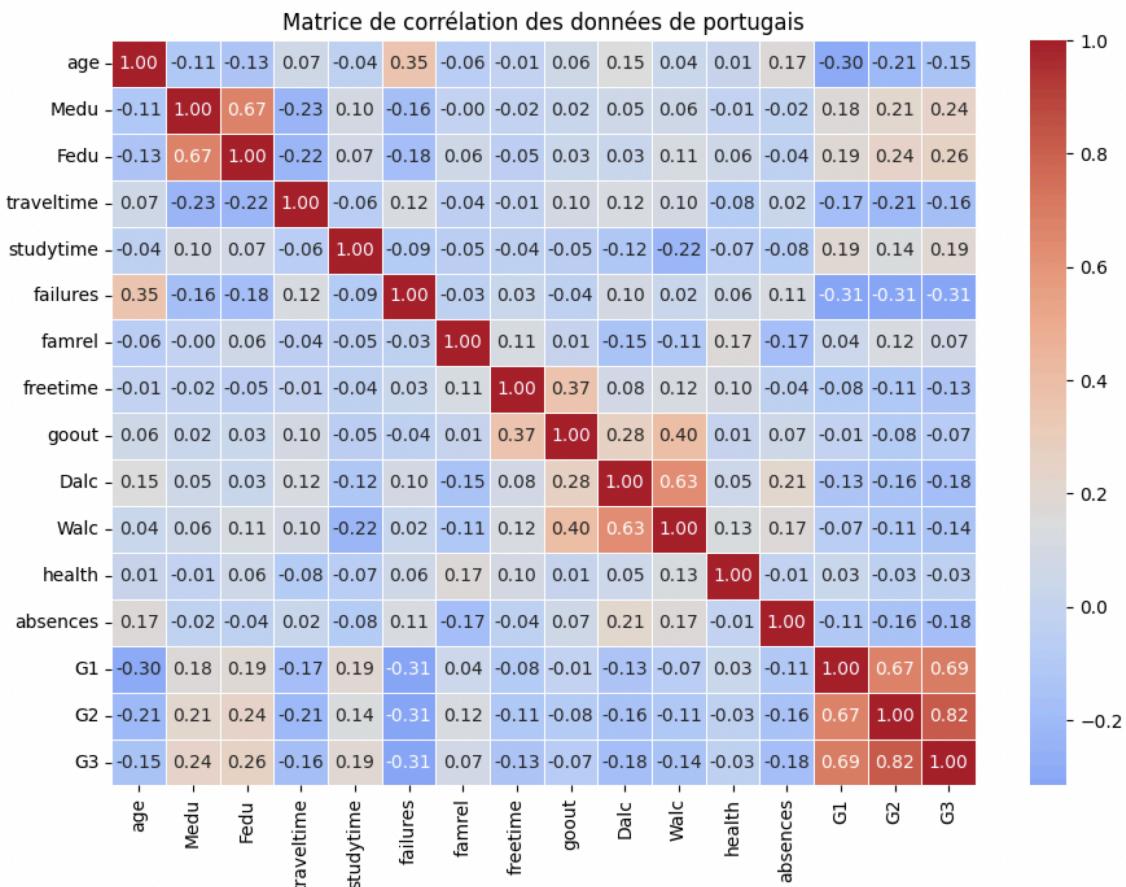
- **failures** a une très faible corrélation négative avec **G3** (≈ -0.08), indiquant que les échecs passés ont un impact presque nul dans ce groupe.
- Les variables sociales (**freetime**, **goout**) et les habitudes liées à l'alcool (**Dalc**, **Walc**) ont peu d'influence sur les notes, avec des corrélations proches de zéro.

pareil pour portugais:



- Corrélations avec G3: [0-8]**
 - Corrélation modérée avec **Fedu** (+0.20) : l'éducation du père a une influence positive sur les performances, mais l'effet est faible.
 - Corrélation faible avec **Medu** (+0.08) : l'éducation de la mère a un rôle encore plus limité pour ce groupe.
 - Corrélation faible avec **absences** (+0.44) : les absences augmentent légèrement dans ce groupe, mais cela n'a pas un fort impact négatif direct sur les performances.
- Autres observations :**
 - failures** est corrélé négativement avec **studytime** (-0.14) : les échecs passés sont associés à moins de temps passé à étudier.
 - Les corrélations entre **Dalc** et **Walc** (consommation d'alcool) restent modérées (+0.58).

- Les relations sociales (`freetime`, `goout`) ne montrent pas d'impact significatif sur les notes.

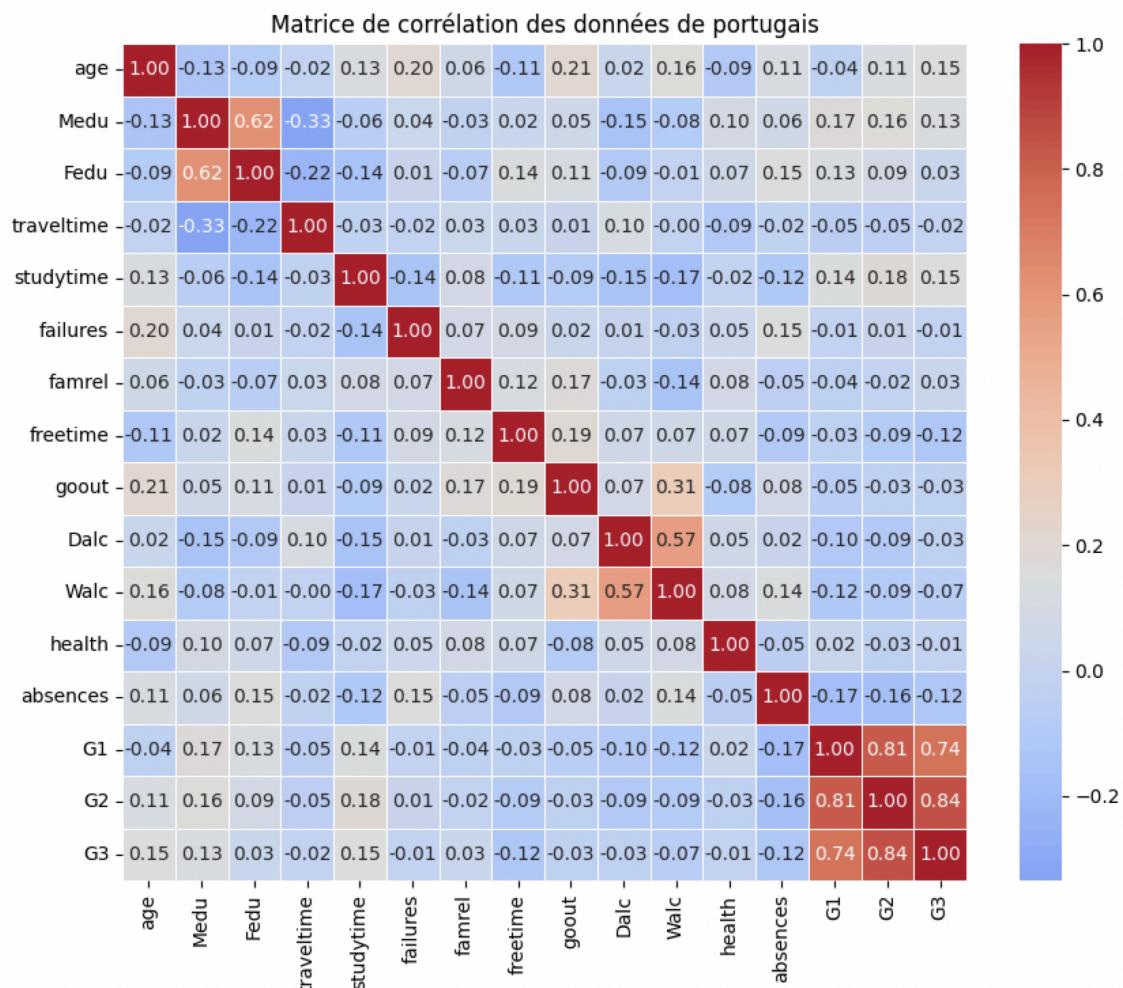


Corrélations avec `G3` : [9-13]

- Corrélation modérée avec `Medu` (+0.24) et `Fedu` (+0.26) : l'éducation des parents a un impact plus fort dans ce groupe que dans le précédent.
- Corrélation négative avec `failures` (-0.31) : les échecs passés influencent négativement les performances.

Autres observations :

- Les absences ont une corrélation négative avec `G3` (-0.18), mais l'effet est réduit par rapport au groupe précédent.
- Les relations sociales (`freetime`, `goout`) continuent de montrer des corrélations faibles ou inexistantes avec les performances.



Corrélations avec G3 : [14-20]

- Corrélation faible mais positive avec **Medu** (+0.13) et **Fedu** (+0.03) : l'éducation des parents joue un rôle mineur pour ce groupe de bons élèves.
- Très faible corrélation avec les absences (+0.12) : contrairement aux groupes précédents, les absences n'ont quasiment aucun effet sur les notes.

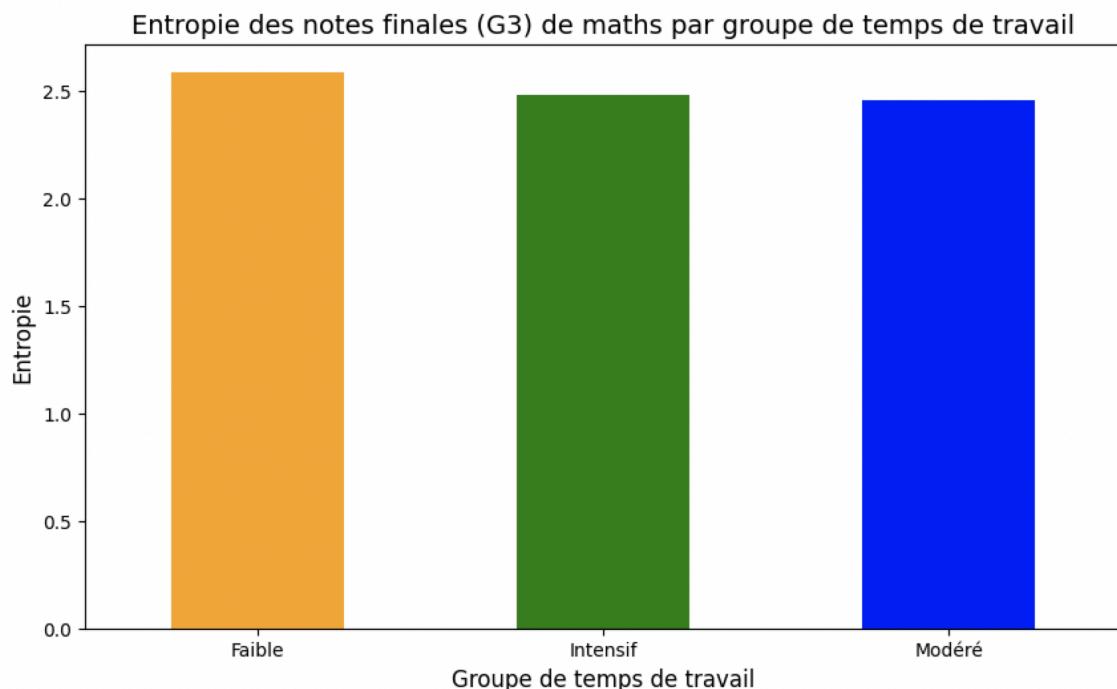
Autres observations :

- failures** a une corrélation quasi nulle avec **G3** (-0.01), indiquant que les échecs passés n'impactent pas significativement les excellents élèves.
- Les corrélations entre les variables sociales (**freetime**, **goout**) et les notes restent faibles ou nulles.
- Les comportements liés à l'alcool (**Dalc** et **Walc**) ont une corrélation encore plus faible avec les notes, proches de zéro (-0.09 et -0.07 respectivement).

L'**entropie** est une mesure de la dispersion ou de l'incertitude dans la distribution des notes (G3). Une entropie élevée signifie que les notes sont réparties de manière plus imprévisible (plus de variabilité), tandis qu'une entropie faible indique que les notes sont concentrées autour de quelques valeurs.

l'entropie est un facteur d'homogénéité/hétérogénéité

petit exemple pour le temps de travail et les notes:



La fonction `classify_study_time` transforme la colonne `studytime` en trois groupes distincts (**Faible**, **Modéré**, **Intensif**), basés sur le nombre d'heures d'étude : **Faible** : `studytime` ≤ 2 , **Modéré** : $3 \leq \text{studytime} < 4$, **Intensif** : `studytime` ≥ 4 . Cela permet de regrouper les étudiants en fonction de leur temps d'étude.

Pour chaque groupe (**Faible**, **Modéré**, **Intensif**), on calcule la distribution des notes finales. Par exemple, dans le groupe **Faible**, on compte combien d'étudiants ont obtenu chaque note (de 0 à 20), puis on calcule les **proportions** :

- Si 10 étudiants ont une note de 10 et qu'il y a 50 étudiants dans ce groupe, la proportion pour la note 10 est de $10 / 50 = 0.2$.

Le résultat est une distribution normalisée des notes finales pour chaque groupe.

L'entropie mesure l'incertitude ou la diversité dans la distribution des notes. Voici comment cela fonctionne :

Formule de l'entropie de Shannon :

$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i)$$

- x_i est la proportion d'une note i dans le groupe.
- H est l'entropie.

Interprétation de l'entropie :

- **Faible entropie** : Une distribution concentrée sur quelques valeurs (les notes sont similaires ou prévisibles).
- **Haute entropie** : Une distribution répartie sur de nombreuses valeurs (les notes sont variées et imprévisibles).

On applique cette formule en utilisant la fonction `scipy.stats.entropy`. Par exemple :

- Si un groupe a des proportions [0.5, 0.5], l'entropie est élevée.
- Si un groupe a des proportions [1.0, 0.0], l'entropie est faible.

Les proportions permettent de mesurer la diversité des valeurs dans un groupe.

- **Quand toutes les proportions sont égales (par exemple, [0.5, 0.5])**, cela signifie qu'il y a autant d'occurrences de deux valeurs. Cela reflète une plus grande incertitude ou diversité.
- **Quand une proportion est dominante (par exemple, [1.0, 0.0])**, cela signifie qu'une seule valeur est présente. Cela reflète une faible incertitude ou diversité.

Exemple :

Supposons que les notes dans le groupe **Faible** soient réparties ainsi :

- Notes=[10,12,15,10,10]Notes=[10,12,15,10,10].
- Les proportions : [0.6 (pour la note 10), 0.2 (pour la note 12), 0.2 (pour la note 15)].
- Entropie: $H=-[0.6 \cdot \log_2(0.6)+0.2 \cdot \log_2(0.2)+0.2 \cdot \log_2(0.2)]$
- $H=-[0.6 \cdot \log_2(0.6)+0.2 \cdot \log_2(0.2)+0.2 \cdot \log_2(0.2)]$.

pour en revenir à notre graphique:

Groupe "Faible" (Temps de travail ≤ 2)

- **Entropie la plus élevée (~2.58) :**
 - Ce groupe présente la plus **forte variabilité** des notes.
 - Les étudiants dans ce groupe semblent avoir des performances très hétérogènes, allant des échecs à de bonnes notes.
 - Cela pourrait indiquer que **travailler peu n'aboutit pas nécessairement à des échecs** mais peut entraîner une grande incertitude dans les performances.

Groupe "Modéré" (Temps de travail = 3)

- **Entropie la plus faible bien que élevée (~2.46) :**
 - Ce groupe a les notes les **plus concentrées**, indiquant une meilleure homogénéité.
 - On ne peut pas dire que les étudiants dans ce groupe soient plus constants dans leurs performances que les deux autres groupes car l'entropie reste élevée.

Groupe "Intensif" (Temps de travail = 4)

- **Entropie modérée (~2.48) :**
 - La variabilité des notes reste présente, mais légèrement moins que dans le groupe "Faible" même si c'est légèrement plus que dans "Modérée".
 - Les étudiants qui travaillent intensément tendent à obtenir des performances un peu plus homogènes que ceux qui "ne travaillent pas".

14/01/25

d'après étude avec l'entropie -> entropie n'est pas applicable dans notre situation car notre jeu de données ne possède pas une quantité de données suffisantes. on peut notamment l'apercevoir dans les résultats que l'on obtient dans le graphique ci-dessus où il y a 300 notes pour le groupe "faible", 60 notes pour le groupe "modéré" et 30 notes pour le groupe "intensif" -> ce n'est pas assez.

14-15/01/2025

Le **test d'homogénéité du khi-deux** (χ^2) est utilisé pour déterminer si deux ou plusieurs échantillons proviennent de populations ayant la **même distribution**. C'est un outil important pour comparer les données catégoriques (fréquences) issues de groupes différents.

H0 (hypothèse nulle) : Les distributions des échantillons sont identiques (homogènes).

H1 (hypothèse alternative) : Les distributions des échantillons diffèrent (hétérogènes).

Conditions d'application

1. Les données doivent être des fréquences ou des comptages.
2. Les échantillons doivent être indépendants.
3. Chaque cellule du tableau doit avoir une fréquence théorique d'au moins 5 pour garantir la validité du test.

Les variables **G3** (notes entre 0 et 20) et **studytime** (temps d'étude) sont toutes deux numériques ou ordinales. Cependant, le test de khi-deux nécessite des **catégories distinctes** pour créer un tableau de contingence. Pour utiliser le test, nous devons :

1. **Regroupier les notes (G3)** en intervalles catégoriques (par exemple : 0-5, 6-10, 11-15, 16-20).
2. Conserver **studytime** comme variable catégorique, si elle contient des valeurs discrètes, ce qui est le cas ici (probablement des niveaux tels que 1, 2, 3, 4).

La p-valeur indique la probabilité d'obtenir une statistique χ^2 aussi élevée (ou plus) sous l'hypothèse nulle (H0). Si cette valeur est inférieure au seuil de significativité (α , souvent fixé à 0.05), on rejette H0.

g3/studytme	un-deux	trois	quatre	Total
zéro-cinq	8	0	0	8
six-dix	115	18	7	140
onze-quinze	125	32	12	169
seize-vingt	26	9	5	40
Total	274	59	24	357

g3/studytme	un-deux	trois	quatre	Total
zéro-cinq	6,14	1,32	0,53	8
six-dix	107,45	23,1	9,41	140
onze-quinze	129,7	27,9	11,36	169
seize-vingt	30,7	6,61	2,68	40
Total	274	59	24	357

le second tableau est obtenu en calculant la fréquence de chaque cellule.

$$\rightarrow 6,14 = 274 \times 8 / 357$$

ça ne vérifie pas une des conditions: - Chaque cellule du tableau doit avoir une fréquence théorique d'au moins 5 pour garantir la validité du test.

pour y remédier, il faudrait jouer avec les échantillons et par exemple dans cet exemple, jouer avec les intervalles de notes pour pouvoir retrouver pour chaque cellule des fréquences supérieures à 5.

j'ai donc décidé de regrouper les deux intervalles de notes les plus faibles et voici ce que ça donne et on va fusionner les colonnes trois et quatre:

g3/studytme	un-deux	trois-quatre	Total
zéro-dix	123	25	148
onze-quinze	125	44	169
seize-vingt	26	14	40
Total	274	83	357

g3/studytme	un-deux	trois-quatre	Total
zéro-dix	113,59	34,41	148
onze-quinze	129,71	39,29	169
seize-vingt	30,7	9,3	40
Total	274	59	357

on vérifie maintenant les conditions nécessaires à son utilisation.

Maintenant que toutes les fréquences observées et attendues sont prêtes et respectent les conditions (toutes les fréquences attendues sont supérieures ou égales à 5), nous allons calculer la statistique du test de khi-deux et sa p-valeur pour tirer des conclusions.

Je procède avec ces données et voici ce que l'on obtient:

g3/studytme	un-deux	trois-quatre
zéro-dix	0,79	2,57
onze-quinze	0,17	0,56
seize-vingt	0,72	2,38

pour chaque cellule on a appliqué la formule, soit: $0,79 = (123 - 113,59) * (123 - 113,59) / 113,59$

$$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

en faisant la somme de chaque valeur on obtient 7,18

Les degrés de liberté sont calculés avec la formule :

$df = (\text{nombre de lignes} - 1) \times (\text{nombre de colonnes} - 1)$ Nombre de lignes : 3 (0-10, 11-15, 16-20). Nombre de colonnes : 2 (1-2, 3-4).

$$df = (3-1) \times (2-1) = 2$$

pour calculer p-valeur il faut utiliser une fonction de la librairie suivante:

```
from scipy.stats import chi2
p_value = 1 - chi2.cdf(chi2, df)
```

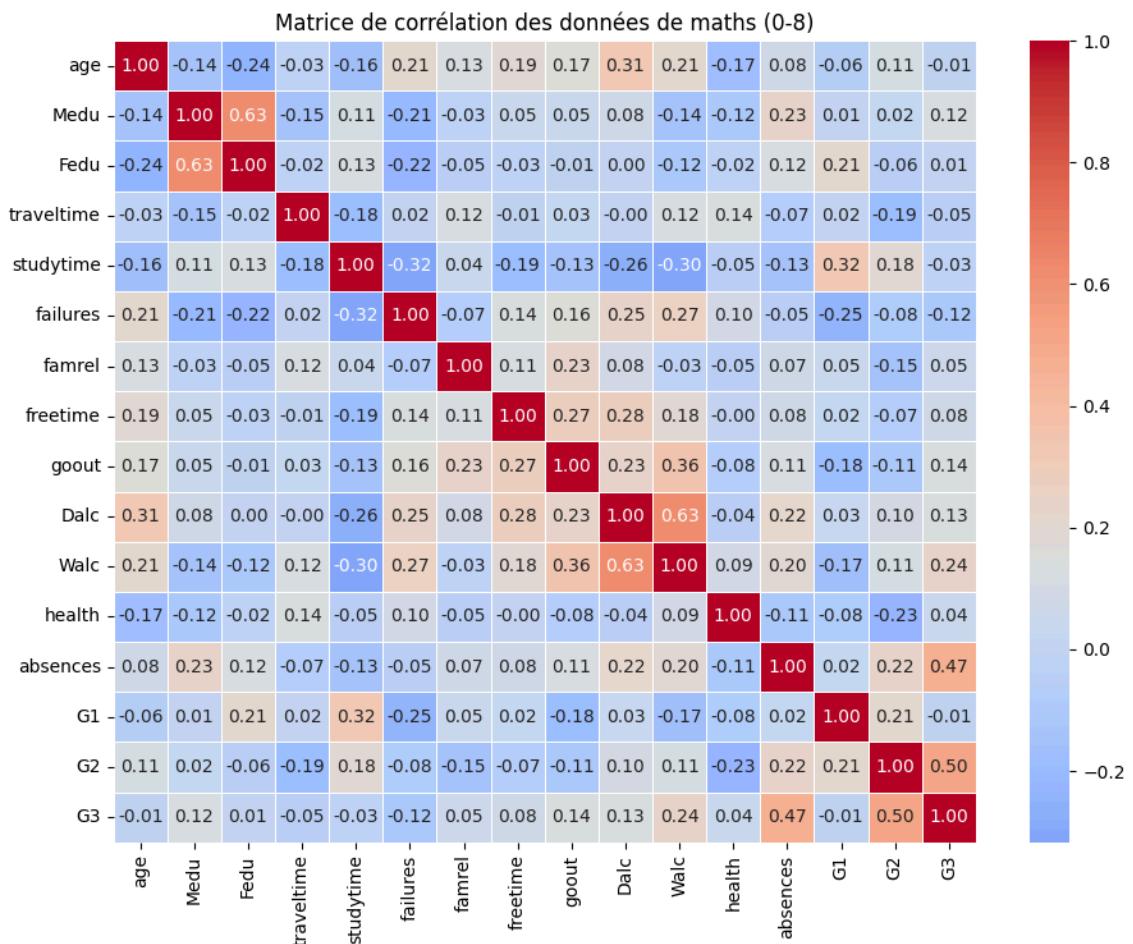
Résultats du test de khi-deux

- **Statistique du khi-deux (χ^2)** : 7.18
- **Degrés de liberté (df)** : 2
- **p-valeur** : 0.0276

Interprétation

1. **Hypothèses :**
 - H_0 : Les distributions des notes (G3) sont **identiques** entre les groupes **studytime (1-2)** et **studytime (3-4)**.
 - H_1 : Les distributions des notes diffèrent entre les groupes **studytime (1-2)** et **studytime (3-4)**.
2. **Seuil de significativité (α)** : 0.05 (valeur standard).
3. **Conclusion :**
 - Comme $p=0.0276 < 0.05$, nous **rejetons l'hypothèse nulle** (H_0).
 - Cela signifie qu'il existe une différence significative entre les distributions des notes (G3) en fonction des catégories de temps d'étude (**studytime_final**).

partie sur les matrices de corrélation spearman:



Analyse de la matrice de corrélation (notes de maths 0-8)

L'analyse de la matrice révèle plusieurs observations clés :

- Corrélations entre les notes G1, G2 et G3 :**
La corrélation entre G2 et G3 est modérée (0.50), indiquant que les résultats intermédiaires influencent en partie les notes finales. En revanche, G1 ne présente presque aucune corrélation avec G3 (-0.01), suggérant que les performances initiales ne sont pas un bon indicateur des résultats finaux.
- Absentéisme et performances scolaires :**
Une corrélation de 0.47 entre les absences et les notes finales montre que l'absentéisme est un facteur clé dans les faibles résultats scolaires. Une meilleure assiduité pourrait améliorer les performances.
- Consommation d'alcool :**
Une forte corrélation (0.63) entre la consommation d'alcool en semaine et le week-end suggère des habitudes de consommation régulières, pouvant potentiellement affecter la concentration et le rendement scolaire.
- Temps d'étude et performances :**
Des corrélations négatives (-0.26 à -0.30) entre le temps d'étude et les notes

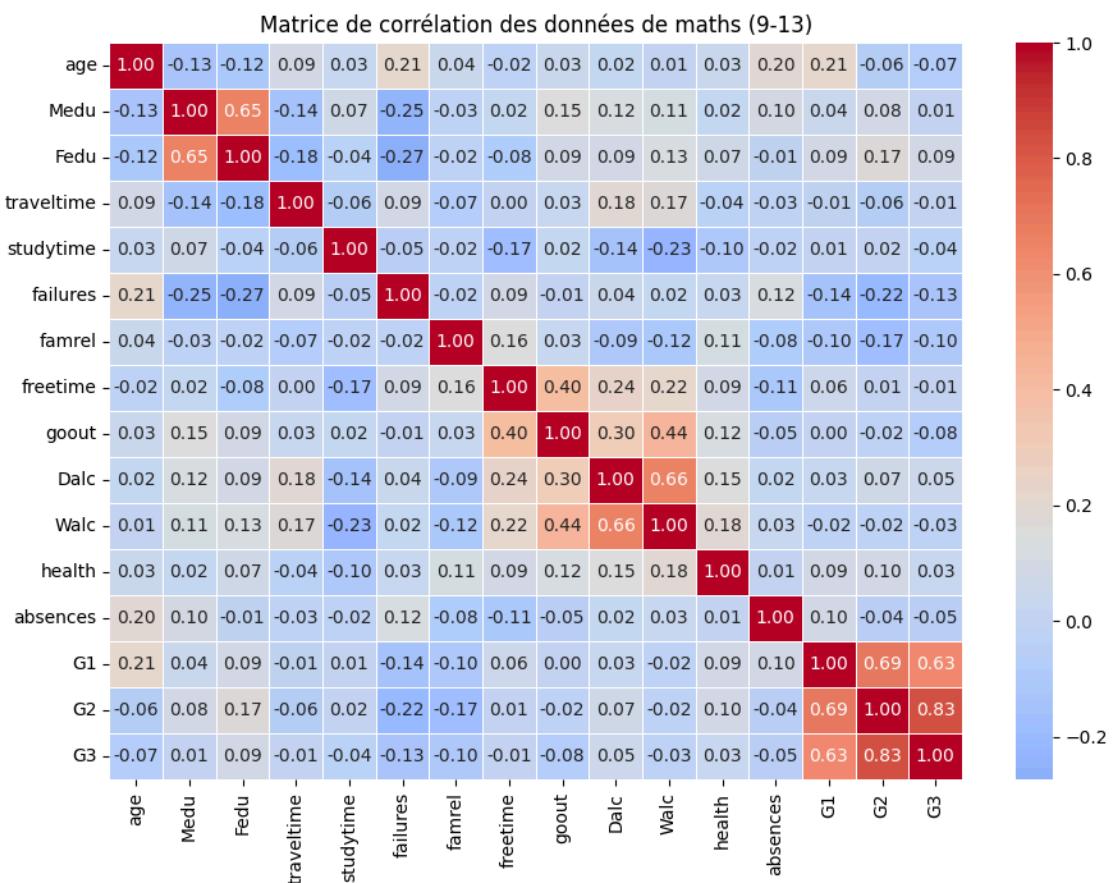
indiquent que les élèves en difficulté étudient davantage sans pour autant améliorer leurs résultats, suggérant un besoin de méthodes d'apprentissage plus efficaces.

- **Niveau d'éducation des parents :**

Avec des corrélations faibles (entre 0.01 et 0.12), l'impact de l'éducation parentale sur les résultats des élèves en difficulté semble limité, mettant en évidence l'importance de facteurs tels que la motivation personnelle et le soutien externe.

Conclusion

L'absentéisme et les mauvaises stratégies d'étude apparaissent comme les principaux facteurs contribuant aux faibles performances. Des actions ciblées, telles que l'amélioration de l'encadrement pédagogique et la sensibilisation à l'importance de l'assiduité, sont essentielles pour aider ces élèves.



Analyse de la matrice de corrélation (notes de maths 9-13)

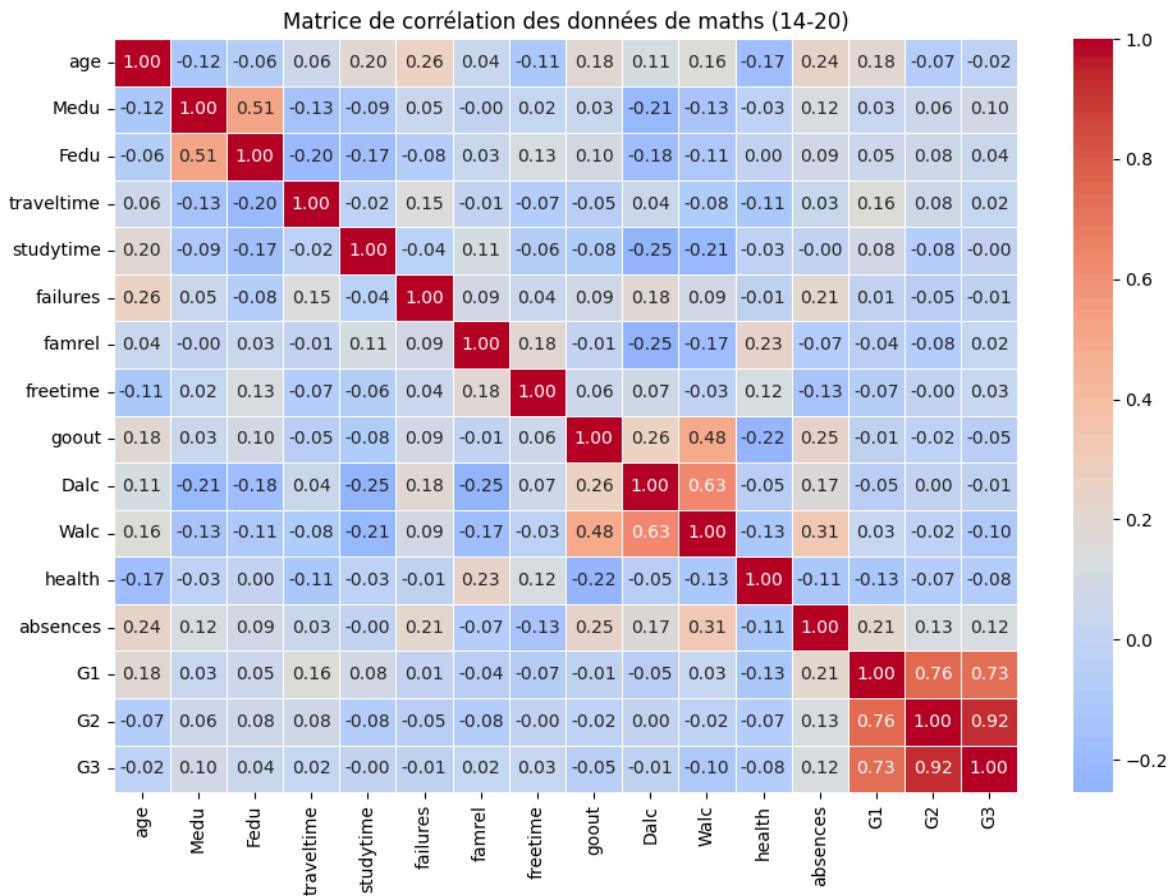
- **Corrélation entre les notes G1, G2 et G3**

Les corrélations sont relativement fortes : G1-G2 (0.69), G1-G3 (0.63), G2-G3 (0.83). Cela indique une continuité des performances scolaires, suggérant que les résultats intermédiaires sont de bons indicateurs des résultats finaux pour cette tranche d'élèves.

- **Consommation d'alcool (Dalc et Walc)**
Une corrélation élevée (0.66) entre la consommation d'alcool en semaine et le week-end montre des habitudes de consommation cohérentes. Cependant, l'impact direct sur les résultats scolaires est faible, ce qui suggère que ces élèves parviennent à gérer leur consommation sans compromettre leur rendement académique.
- **Échecs scolaires et performances académiques**
Les corrélations négatives entre le nombre d'échecs et les notes sont modérées (jusqu'à -0.22 pour G2), indiquant que ces élèves parviennent à surmonter leurs échecs mieux que les élèves en difficulté.
- **Temps d'étude et résultats**
Des corrélations négatives (jusqu'à -0.23) suggèrent que l'augmentation du temps d'étude ne se traduit pas nécessairement par de meilleures performances, mettant en avant l'importance de la qualité des méthodes d'apprentissage.
- **Influence parentale limitée**
Le niveau d'éducation des parents présente des corrélations faibles avec les notes finales (Medu : 0.01, Fedu : 0.09), indiquant que dans cette tranche, l'influence familiale est moins marquée sur les résultats.

Conclusion

Les résultats des élèves dans cette tranche de notes sont relativement stables au fil du temps, suggérant qu'ils maintiennent leurs performances sans subir d'impact majeur de facteurs externes comme l'absentéisme ou l'environnement familial. L'optimisation des méthodes d'apprentissage et de gestion du temps pourrait être une piste plus efficace pour améliorer leur progression académique.



Analyse de la matrice de corrélation (notes de maths 14-20)

- Corrélation entre les notes G1, G2 et G3**

Les corrélations élevées entre les notes intermédiaires et finales (G1-G2 : 0.76, G1-G3 : 0.73, G2-G3 : 0.92) indiquent une forte stabilité des performances académiques. Cela confirme que les résultats intermédiaires sont d'excellents prédicteurs des résultats finaux pour ces élèves performants.

- Influence parentale**

Le niveau d'éducation des parents est modérément corrélé entre eux (0.51), mais présente une faible influence sur les résultats des élèves (corrélation de 0.10 et 0.04 avec G3). Ces élèves réussissent donc principalement grâce à leurs propres efforts et stratégies.

- Absences et résultats scolaires**

Contrairement aux élèves en difficulté, les absences ont une légère corrélation positive avec les notes (0.12), suggérant que ces élèves peuvent gérer leur emploi du temps efficacement et compenser leurs absences.

- Temps d'étude et performances**

La corrélation négative avec G3 (-0.08) indique que ces élèves n'ont pas besoin de consacrer un temps excessif à l'étude pour obtenir de bons résultats, ce qui souligne l'importance de méthodes d'apprentissage efficaces plutôt que la quantité de travail.

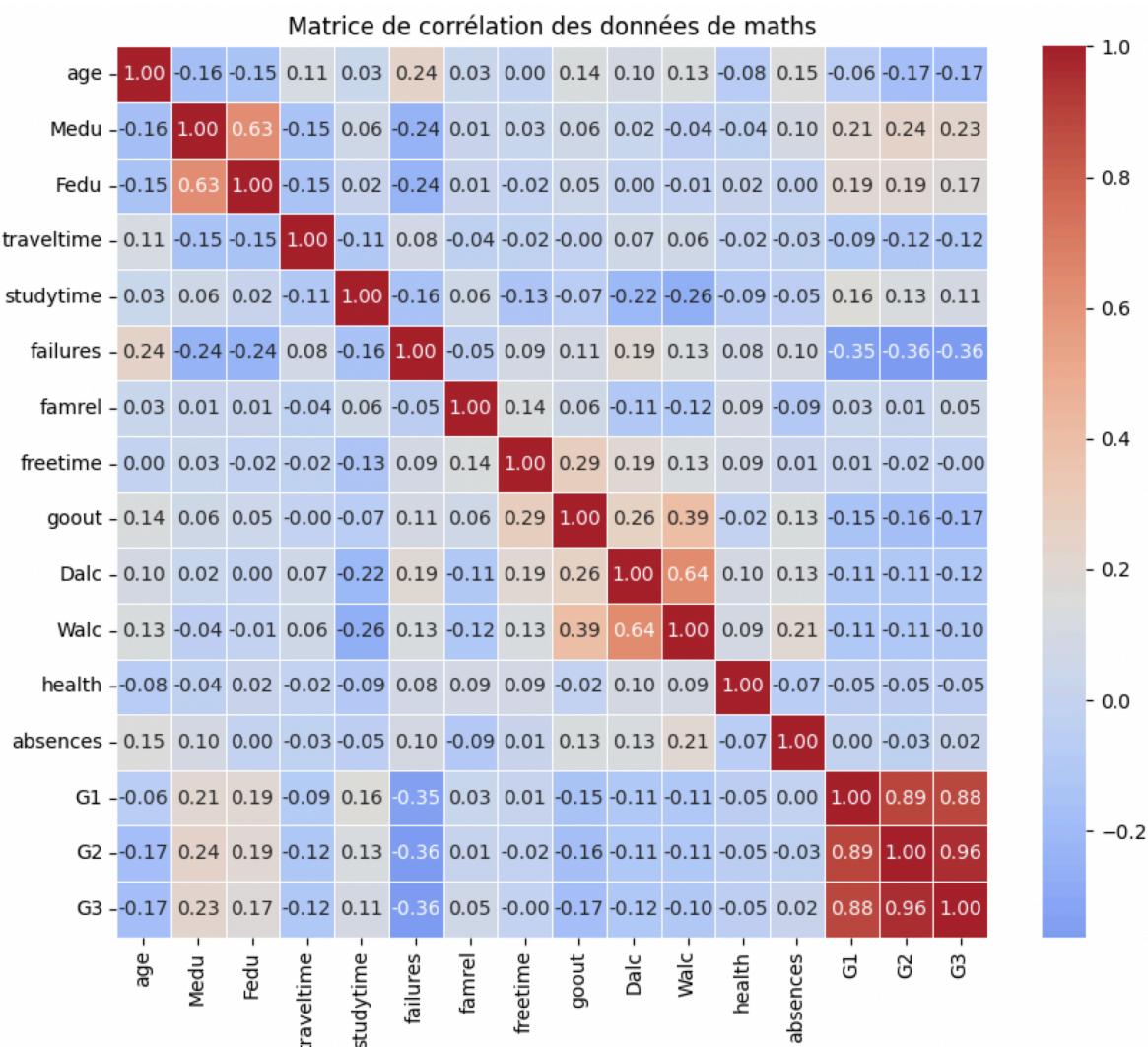
- Consommation d'alcool**

Une forte corrélation entre la consommation d'alcool en semaine et le week-end (0.63) est observée, mais son impact sur les résultats est négligeable, montrant que

ces élèves parviennent à concilier vie sociale et études sans affecter leurs performances.

Conclusion

Les élèves ayant des notes élevées présentent une progression stable et prévisible tout au long de l'année, avec une autonomie et une capacité d'organisation importantes. Leur réussite est principalement due à des facteurs internes tels que la discipline personnelle et l'efficacité des stratégies d'apprentissage, plutôt qu'à des influences externes comme l'environnement familial ou le temps d'étude.



Analyse de la matrice de corrélation des données de mathématiques

- Corrélation entre les notes G1, G2 et G3**

Les corrélations élevées (G1-G2 : 0.89, G1-G3 : 0.88, G2-G3 : 0.96) montrent une forte stabilité des performances scolaires. Cela suggère que les évaluations intermédiaires sont de bons prédicteurs des résultats finaux.

- **Éducation parentale et performances scolaires**

Une corrélation modérée est observée entre le niveau d'éducation des parents et les notes finales (Medu-G3 : 0.21, Fedu-G3 : 0.23), ce qui suggère une influence positive, bien que non déterminante, de l'environnement familial.

- **Impact des échecs scolaires**

Les corrélations négatives avec les notes finales (failures-G1 : -0.35, failures-G2 : -0.36, failures-G3 : -0.36) confirment que l'accumulation d'échecs passés affecte significativement les résultats futurs, soulignant l'importance de la prévention et du soutien scolaire.

- **Consommation d'alcool et performances**

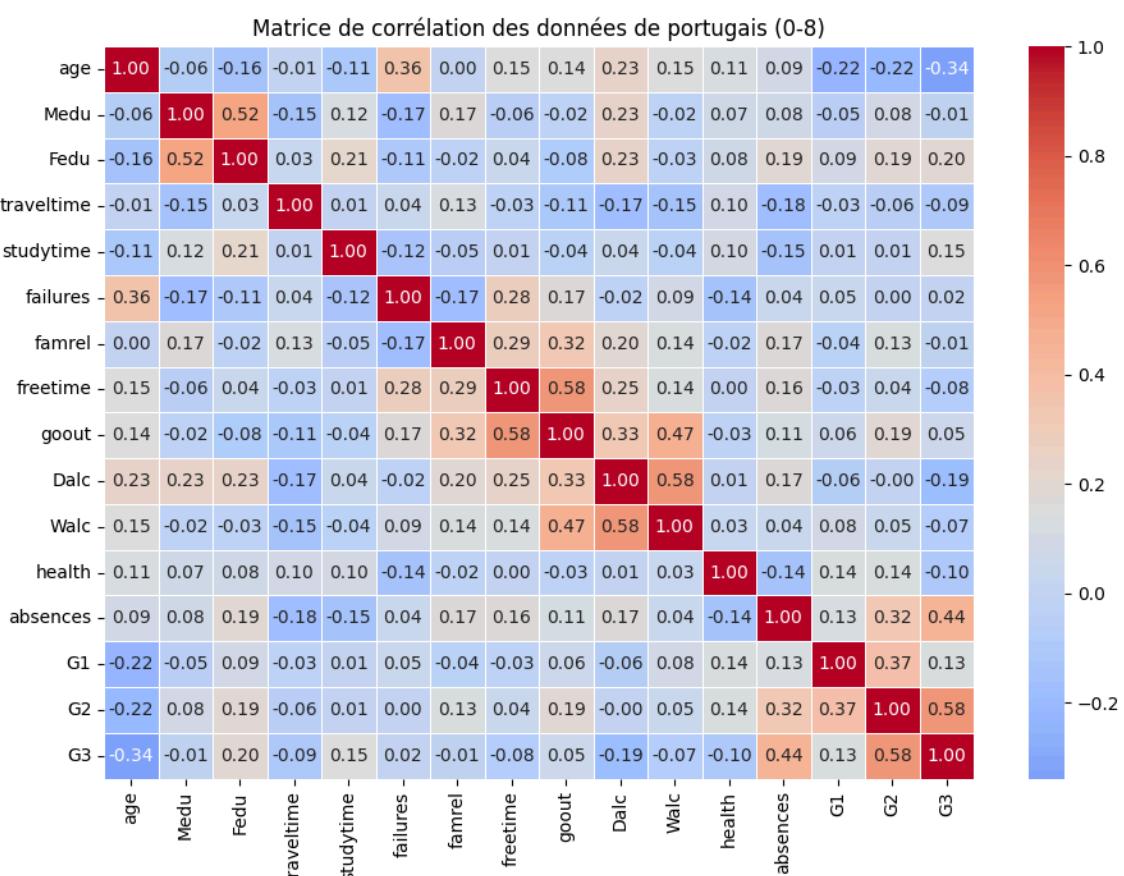
Une forte corrélation (0.64) est constatée entre la consommation d'alcool en semaine et le week-end, montrant une constance dans les habitudes. Cependant, son impact sur les résultats reste légèrement négatif, indiquant une possible influence sur les performances.

- **Temps d'étude et efficacité**

Les corrélations légèrement négatives (studytime-G3 : -0.26) suggèrent que le volume d'étude seul n'est pas un facteur clé de réussite, mettant en avant l'importance de la qualité de l'apprentissage.

Conclusion

Les résultats mettent en évidence que la réussite scolaire repose principalement sur la constance des performances, l'historique des échecs et, dans une moindre mesure, le niveau d'éducation des parents. Une attention particulière doit être accordée aux élèves en difficulté pour éviter l'accumulation d'échecs et améliorer l'efficacité du temps d'étude.



Analyse de la matrice de corrélation des données de portugais (notes 0-8)

Cette matrice de corrélation met en lumière certains facteurs influençant les performances des élèves en difficulté scolaire.

Corrélations significatives

- **Corrélation entre les notes G1, G2 et G3**

Les corrélations entre G1 et G2 (**0.37**) ainsi que G2 et G3 (**0.58**) indiquent une relation modérée, suggérant une certaine stabilité des performances au fil des évaluations. En revanche, la faible corrélation entre G1 et G3 (**0.13**) souligne que les premières évaluations ne permettent pas de prédire avec précision les résultats finaux.

- **Absences et performances académiques**

Une corrélation modérée (**0.44**) avec G3 met en évidence l'impact négatif de l'absentéisme sur les résultats scolaires, suggérant que la présence régulière en classe est un levier d'amélioration important.

Corrélations négatives notables

- **Âge et résultats scolaires**

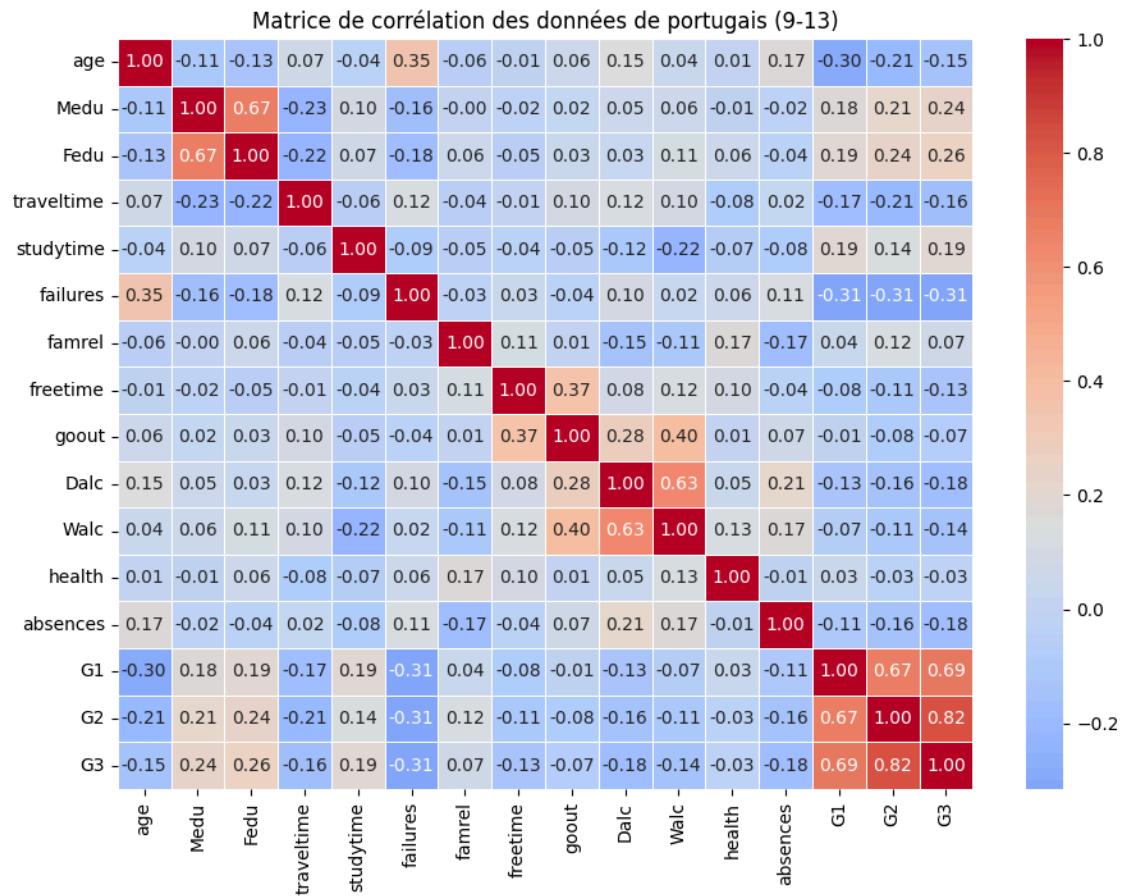
Une corrélation négative de **-0.34** montre que les élèves plus âgés obtiennent des résultats inférieurs, ce qui pourrait être lié à des retards scolaires ou un décrochage progressif.

Points à retenir

- **L'influence parentale est limitée**, avec des corrélations faibles entre le niveau d'éducation des parents et les notes finales, indiquant que les performances des élèves en difficulté sont davantage influencées par des facteurs individuels et scolaires.
- **Le temps d'étude montre une faible relation avec les résultats**, suggérant que la qualité de l'apprentissage est plus déterminante que la quantité d'heures étudiées.

Conclusion

Les résultats soulignent l'importance de lutter contre l'absentéisme et d'apporter un soutien pédagogique adapté aux élèves plus âgés. L'influence des parents étant limitée, des interventions spécifiques en milieu scolaire apparaissent comme la meilleure solution pour améliorer les performances des élèves en difficulté.



L'analyse de cette matrice de corrélation pour les élèves ayant des notes moyennes (9-13) en portugais met en évidence plusieurs éléments pertinents.

Corrélations significatives :

- Notes intermédiaires et résultats finaux**
Les corrélations entre G1 et G2 (**0.67**) et entre G2 et G3 (**0.82**) indiquent une relation relativement forte, confirmant que les notes intermédiaires sont de bons indicateurs des performances finales. Toutefois, la corrélation entre G1 et G3 (**0.69**) suggère des fluctuations potentielles dans les résultats.
- Impact des échecs scolaires**
La corrélation négative entre les échecs passés et les notes finales (G3 : **-0.31**) montre que les antécédents scolaires influencent négativement la progression des élèves, nécessitant un accompagnement spécifique pour éviter la stagnation.
- Âge et résultats scolaires**
Une corrélation négative avec G3 (**-0.15**) suggère que les élèves plus âgés ont tendance à obtenir des résultats légèrement inférieurs, potentiellement en raison d'un retard académique ou de difficultés d'adaptation.

Points d'attention :

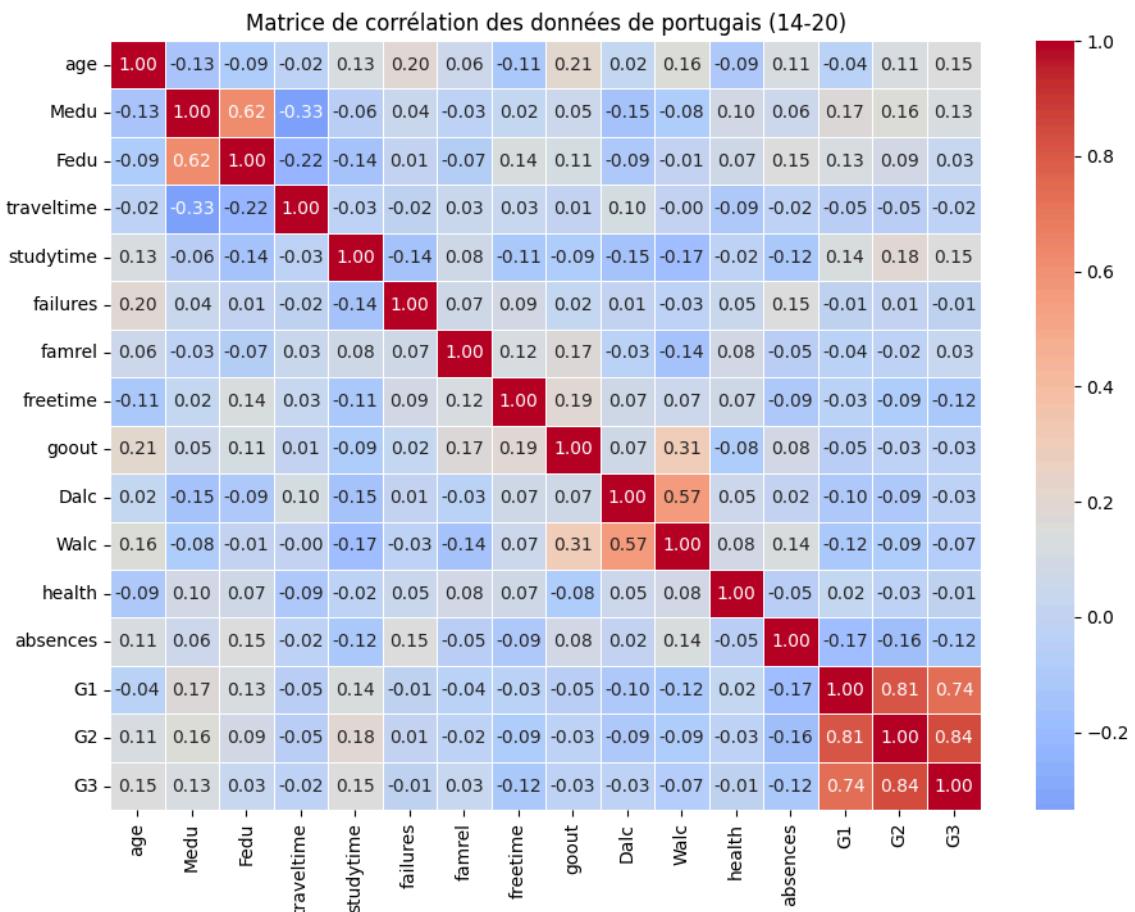
- Absences et performances**
La corrélation modérée entre les absences et G3 (**-0.18**) indique que l'assiduité a un impact limité sur cette tranche d'élèves, contrairement aux élèves en difficulté.

- **Consommation d'alcool**

Une forte corrélation entre la consommation en semaine et le week-end (**0.63**) montre des habitudes cohérentes, mais l'impact négatif sur les notes reste faible (**-0.18**), suggérant un contrôle relatif de cette consommation.

Conclusion :

Les performances des élèves de ce groupe sont influencées principalement par les résultats intermédiaires et les antécédents scolaires. Des actions ciblées pour éviter l'accumulation d'échecs et un suivi personnalisé pourraient contribuer à l'amélioration de leurs résultats.



L'analyse de cette matrice de corrélation pour les élèves ayant des notes élevées (14-20) en portugais met en évidence plusieurs points essentiels.

Corrélations clés :

- **Cohérence des performances scolaires :**

Les corrélations entre G1, G2 et G3 sont fortes (**0.81, 0.74, 0.84**), indiquant une stabilité académique des élèves tout au long de l'année. Les premières évaluations sont de bons prédicteurs des résultats finaux.

- **Niveau d'éducation des parents :**

Bien que Medu et Fedu soient corrélés entre eux (**0.62**), leur influence sur G3 reste

faible (0.13 et 0.03), suggérant que ces élèves s'appuient davantage sur leur autonomie et leur motivation personnelle.

- **Absences et performances scolaires :**

Une corrélation légèrement négative avec G3 (**-0.12**) indique que les absences ont un impact limité sur ces élèves, qui parviennent à maintenir de bons résultats malgré des absences occasionnelles.

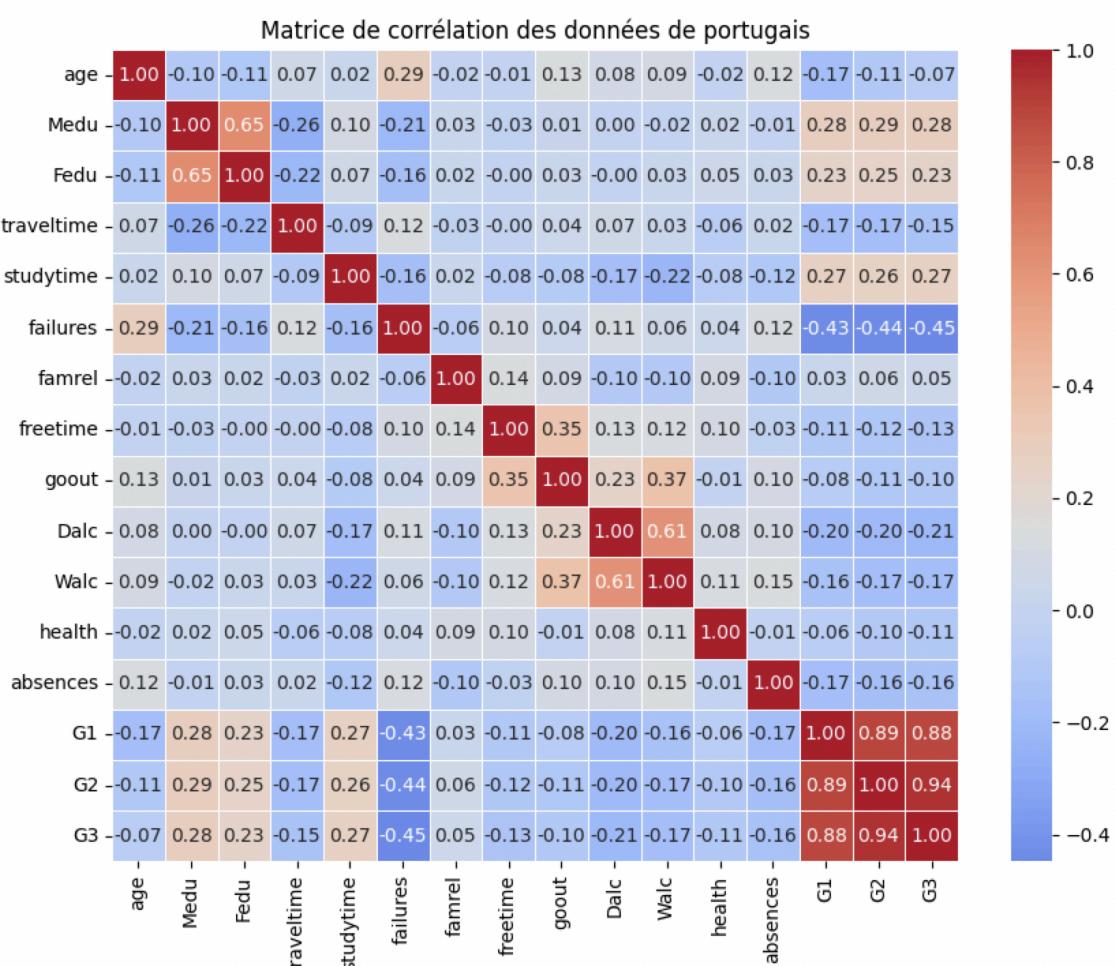
Interprétations :

Les résultats montrent que ces élèves sont caractérisés par une grande régularité dans leurs performances et une bonne capacité à concilier études et vie personnelle.

L'absentéisme et le temps d'étude n'ont pas d'effet significatif, indiquant une gestion efficace du travail et une autonomie renforcée.

Conclusion :

Ces élèves bénéficieraient davantage de stratégies visant à perfectionner leurs compétences et à optimiser leur potentiel plutôt qu'un encadrement strict



L'analyse de cette matrice de corrélation des données de portugais met en évidence plusieurs points clés :

Corrélations significatives :

- **Performance académique stable :**

Les corrélations élevées entre G1, G2 et G3 (**0.89, 0.88, 0.94**) montrent une forte cohérence des performances scolaires, indiquant que les notes intermédiaires sont de bons prédicteurs des résultats finaux.

- **Impact de l'éducation parentale :**

L'éducation des parents (Medu et Fedu, corrélation de **0.65**) influence modérément les résultats scolaires (corrélation avec G3 de **0.28** et **0.23**), suggérant que l'environnement familial peut jouer un rôle positif mais non déterminant.

- **Échecs scolaires et résultats finaux :**

Une corrélation négative marquée entre le nombre d'échecs scolaires et les notes finales (G3 : **-0.45**) montre que les difficultés antérieures impactent significativement la réussite.

Points d'attention :

- **Consommation d'alcool :**

La corrélation entre la consommation d'alcool en semaine et le week-end (**0.61**) est forte, mais son impact négatif sur les résultats scolaires reste modéré (-0.20 à -0.21), suggérant que les élèves concernés pourraient maintenir un équilibre entre études et loisirs.

- **Temps d'étude inefficace :**

La corrélation négative entre le temps d'étude et les notes finales (G3 : **-0.15**) suggère que les efforts en termes de quantité d'étude ne se traduisent pas forcément par une amélioration des résultats, mettant en avant l'importance des méthodes d'apprentissage.

Conclusion :

Les résultats scolaires sont fortement influencés par les performances antérieures et l'éducation parentale, tandis que les échecs passés restent un frein majeur. Une attention particulière aux méthodes d'étude et un accompagnement personnalisé pourraient être bénéfiques pour ces élèves.