**Evaluation 4**

**Finale**

**Programmation 2**

Par

Frederick De Kinder

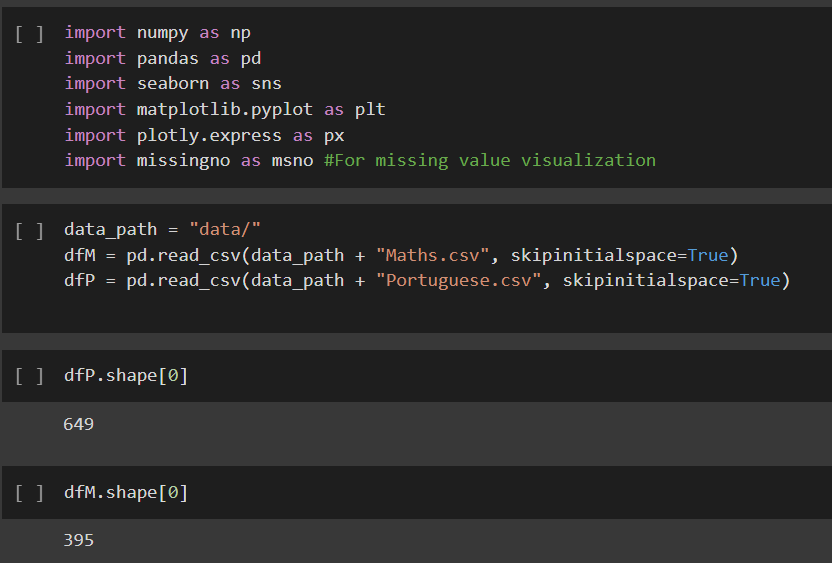
Il fait souvent des fautes d’étourderie. Je ne suis pas etonéprobleme à résoudre : Le dataset choisi comprend des données sur l’effet de l’alcool sur les étudiants. Il y a en tous 2 cours différents sur deux .Csv (math Csv et Portuguese.csv. Tous les deux comprenent des étudiants de 2 écoles différentes ; Gabriel Pereira(GP) et Mousinho da Silveira (MS).

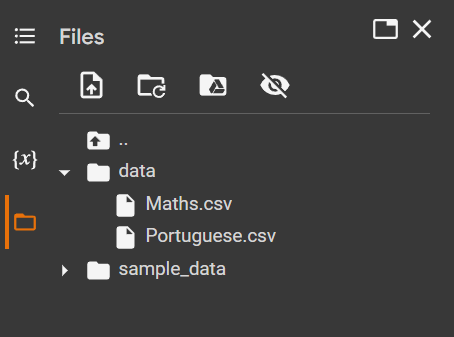
Le but sera d'analyser les différentes combinaisons des caractéristiques disponibles pour comprendre la corrélation entre les données et aussi déchiffrer quelle caractéristique semble influencer les étudiants.

Voici le DataSet choisi : [**https://www.kaggle.com/datasets/whenamancodes/alcohol-effects-on-study**](%20https:/www.kaggle.com/datasets/whenamancodes/alcohol-effects-on-study%20)

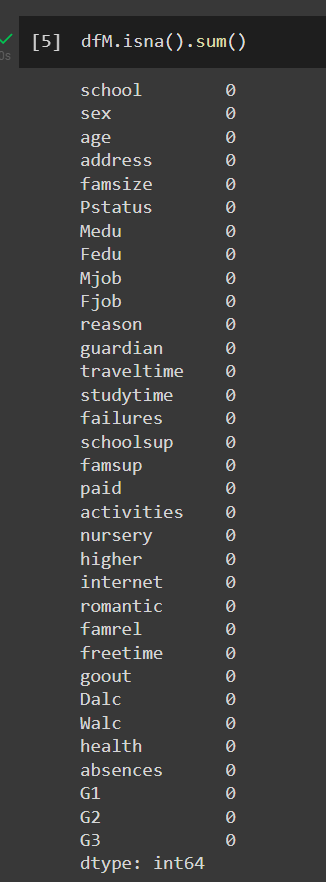
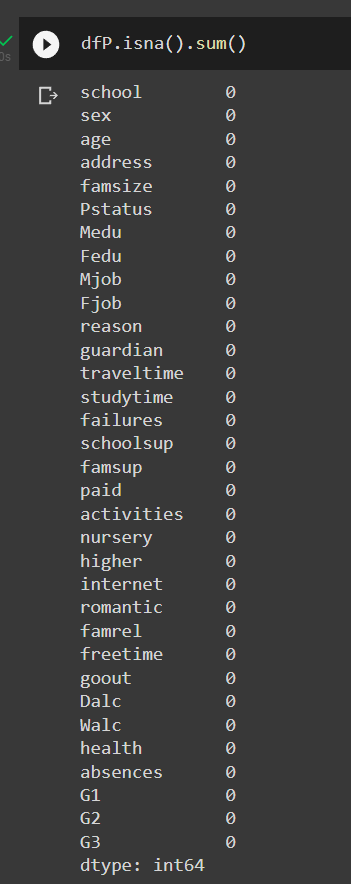
Ces données concernent les performances des élèves de l'enseignement secondaire dans deux établissements portugais. Les attributs des données comprennent les scores des élèves, les caractéristiques démographiques, sociales et scolaires, et ont été recueillis au moyen de rapports et de questionnaires scolaires. Deux séries de données de performance sont présentées dans deux matières séparées: les mathématiques (mat) et le portugais (por). Dans [Cortez et Silva, 2008], les deux ensembles de données ont été modelés en fonction de la classification binaire et des tâches de régression/cinq niveaux. Note importante : L'attribut cible G3 est fortement corrélé aux attributs G2 et G1. Cela s'explique par le fait que G3 est la note de la dernière année (délivrée à la 3e période), tandis que G1 et G2 correspondent aux notes de la 1re et de la 2e période. Il est plus difficile de prédire le G3 sans le G2 et le G1, mais une telle prévision est beaucoup plus utile.

ETL :

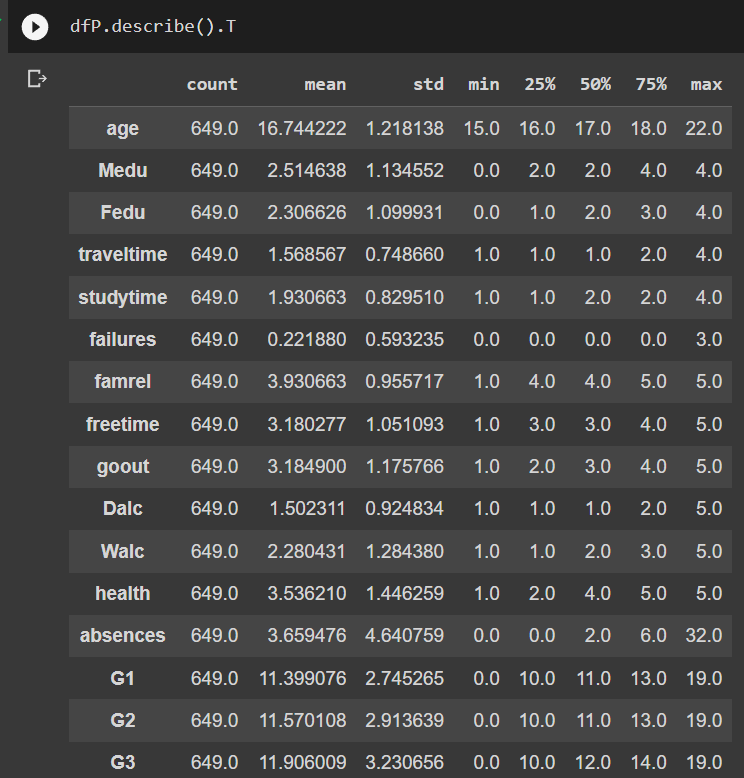




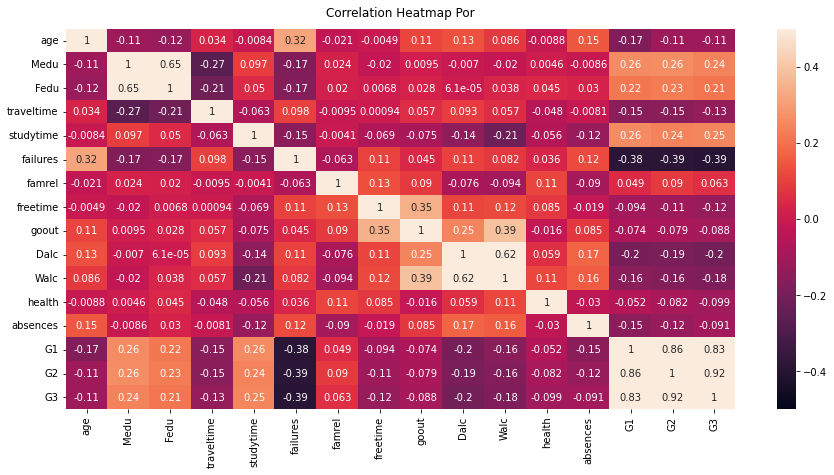
S’assurer qu’il n’est pas de trou dans le dataset : dfM = cours de Math / dfP = Portuguais



Faire un survol des chiffres avec des statistiques de base.



Corrélation avec une Heatmap; NOTE : j’ai mis le vmin et max plus petit pour facilité la visualtion.



Corrélations à noter :

* Les notes ensembles G1 G2 G3
* Les échecs(failures) et les notes
* L’éducation du parent et les notes
* La consommation d’alcools et les sorties
* Le temps d’étude et les notes
* Le temps de voyagement et l’éducation parentale
* La santé et la relation avec la famille
* La relation avec la famille et le temps libre
* Le temps libre et les sorties

Voici la description des colonnes.

| Columns | Description |
| --- | --- |
| school | student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira) |
| sex | student's sex (binary: 'F' - female or 'M' - male) |
| age | student's age (numeric: from 15 to 22) |
| address | student's home address type (binary: 'U' - urban or 'R' - rural) |
| famsize | family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3) |
| Pstatus | parent's cohabitation status (binary: 'T' - living together or 'A' - apart) |
| Medu | mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 5th to 9th grade, 3 secondary education or 4 higher education) |
| Fedu | father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 5th to 9th grade, 3 secondary education or 4 higher education) |
| Mjob | mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other') |
| Fjob | father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other') |
| reason | reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other') |
| guardian | student's guardian (nominal: 'mother', 'father' or 'other') |
| traveltime | home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour) |
| studytime | weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours) |
| failures | number of past class failures (numeric: n if 1<=n<3, else 4) |
| schoolsup | extra educational support (binary: yes or no) |
| famsup | family educational support (binary: yes or no) |
| paid | extra paid classes within the course subject (Math or Portuguese) (binary: yes or no) |
| activities | extra-curricular activities (binary: yes or no) |
| nursery | attended nursery school (binary: yes or no) |
| higher | wants to take higher education (binary: yes or no) |
| internet | Internet access at home (binary: yes or no) |
| romantic | with a romantic relationship (binary: yes or no) |
| famrel | quality of family relationships (numeric: from 1 - very bad to 5 - excellent) |
| freetime | free time after school (numeric: from 1 - very low to 5 - very high) |
| goout | going out with friends (numeric: from 1 - very low to 5 - very high) |
| Dalc | workday alcohol consumption (numeric: from 1 - very low to 5 - very high) |
| Walc | weekend alcohol consumption (numeric: from 1 - very low to 5 - very high) |
| health | current health status (numeric: from 1 - very bad to 5 - very good) |
| absences | number of school absences (numeric: from 0 to 93) |

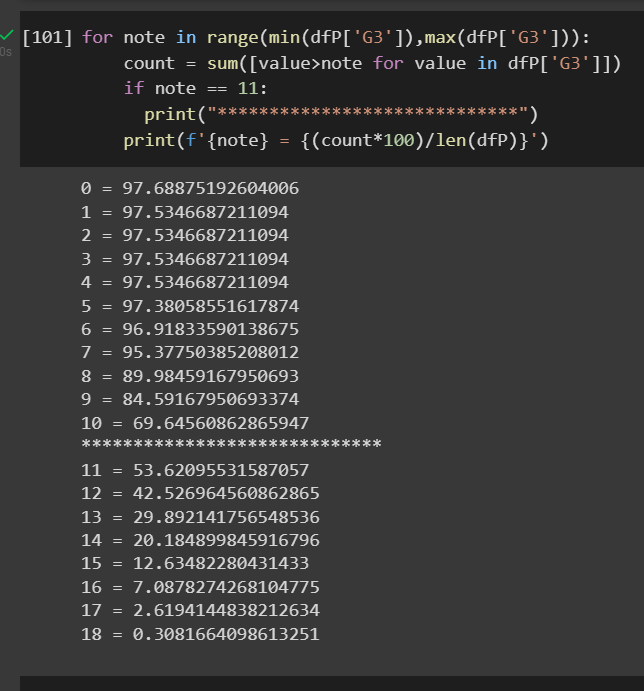
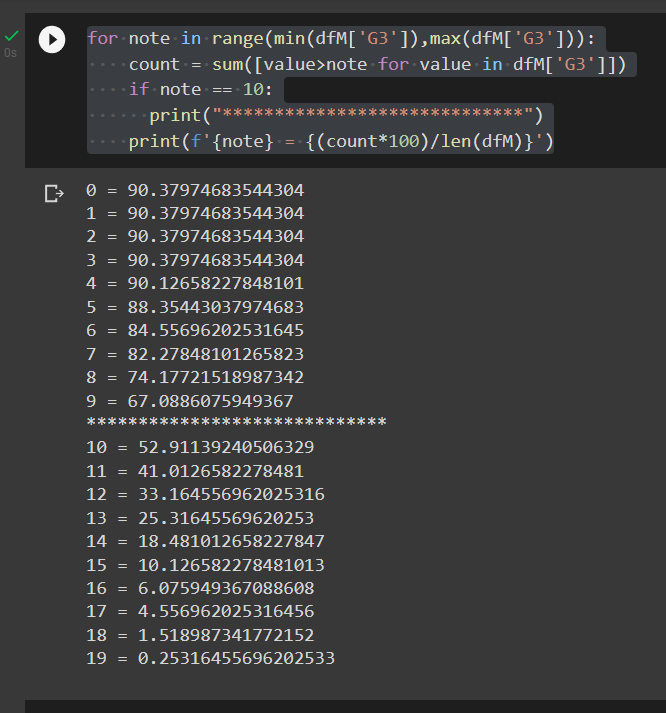
\*source https://www.kaggle.com/datasets/whenamancodes/alcohol-effects-on-study

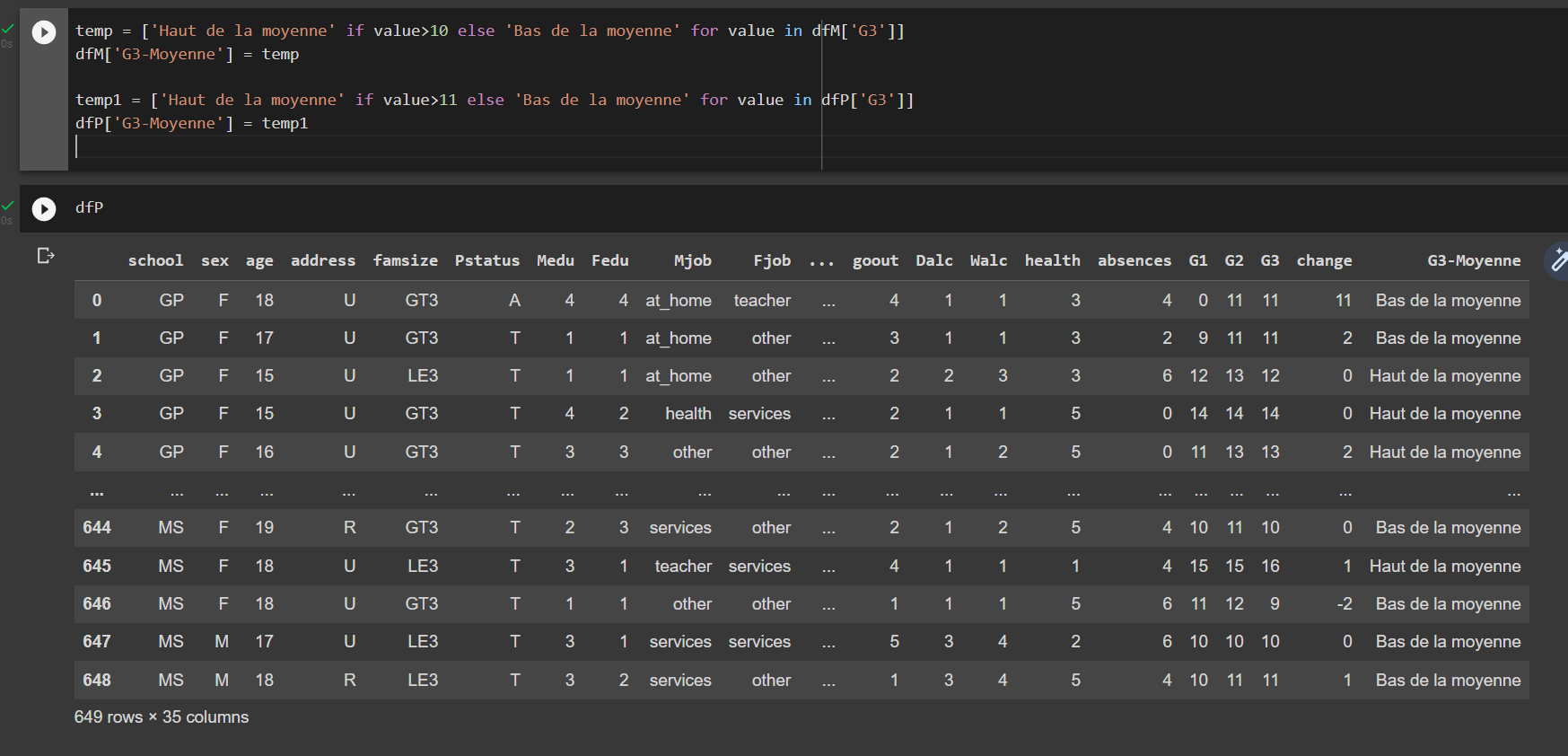
| Grade | Description |
| --- | --- |
| G1 | Première partie des notes (num: De 0 à 20) |
| G2 | Deuxième partie des notes (num: De 0 à 20) |
| G3 | Note finale (num: De 0 à 20, cible de sortie) |

Pour déterminer si un eleve a bien fait ou pas il est primordiale de voir s’il est au-dessus de la moyenne.

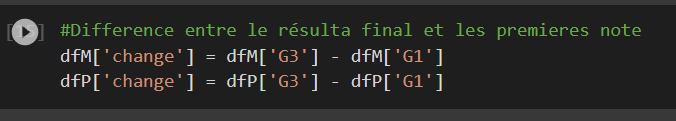
Math à gauche et Portugais à droite. Nous pouvons constater que la moyenne est plus élevée cotée du cours portugais cest pourquoi quand dans la future création de colonne, nous allons créer deux algorithmes un avec 10 comme échelle pour la moyenne et l’autre a 11.

NOTE : (dfM = MATH / dfP = PORTUGUESE)

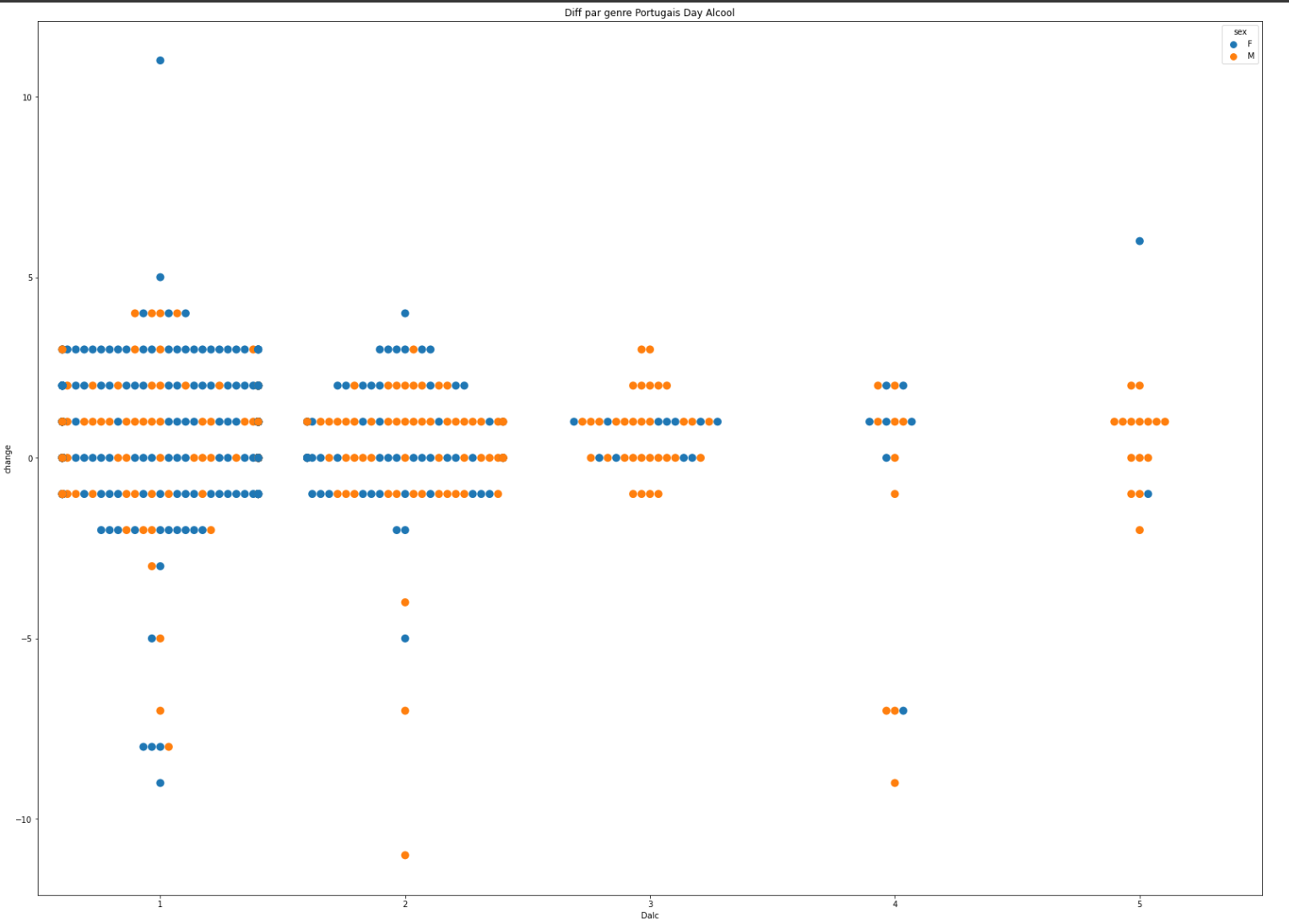


Pour voir l’effet positif ou négatif dans un ordre chronologique. Je soustrais le G3 du G1 pour voir la différence entre le début des notes et la fin.

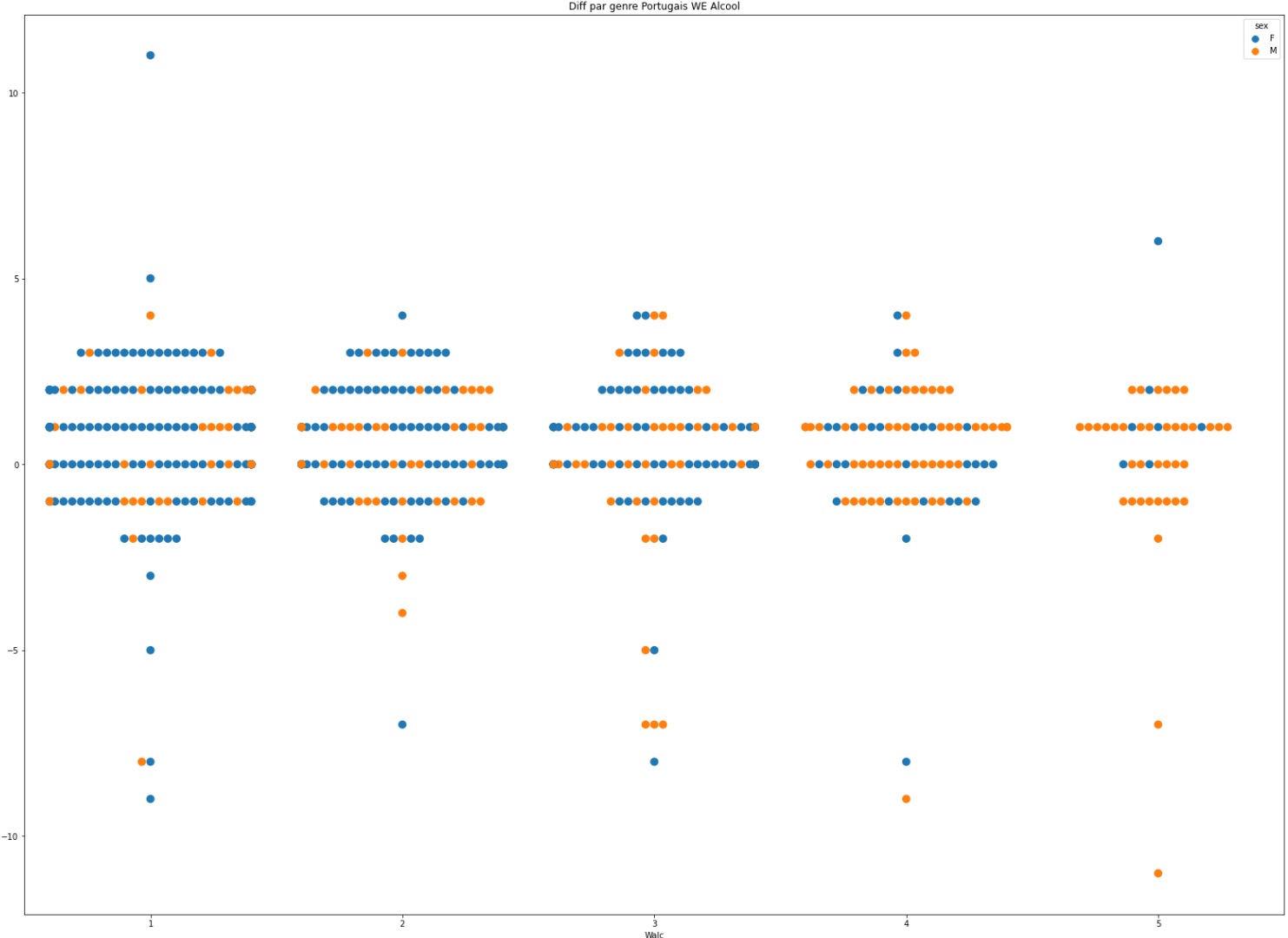


Premièrement je voulais la **distribution des genres** dans difference entre le G3 et G1

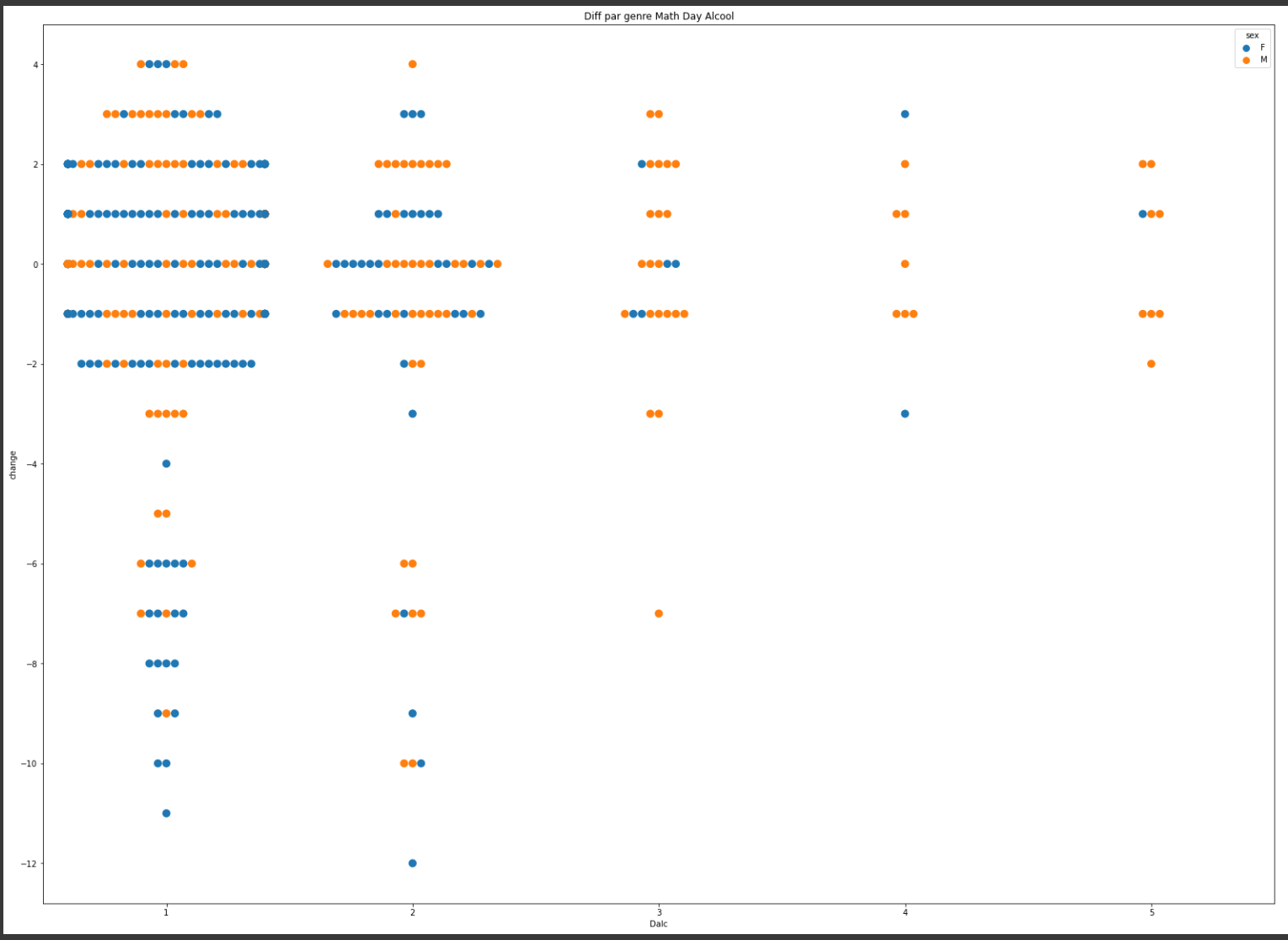
Alcool Semaine Portuguais



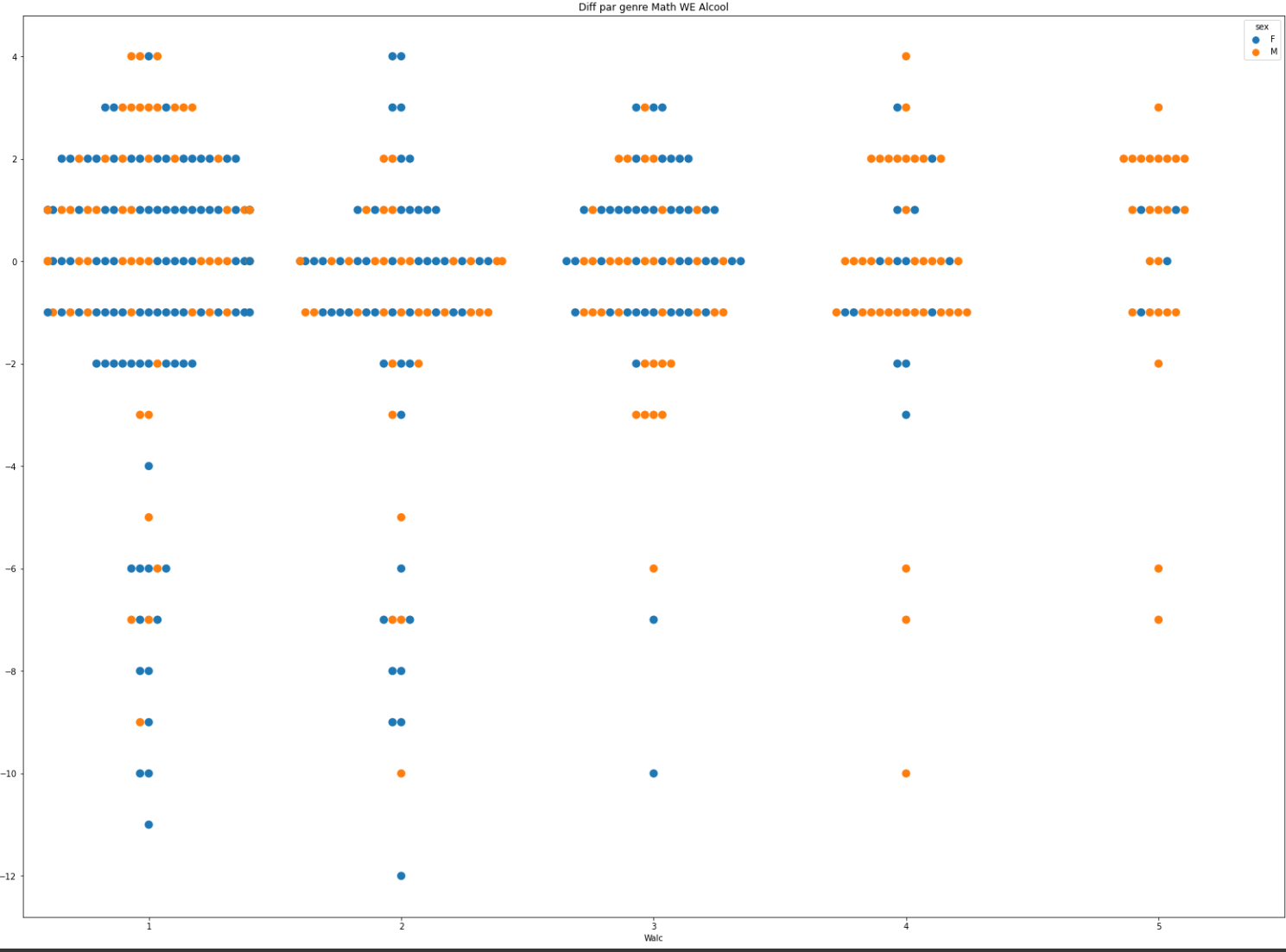
Alcool Week-end Portuguais



Alcool Semaine Math



Alcool Week-end Portuguais

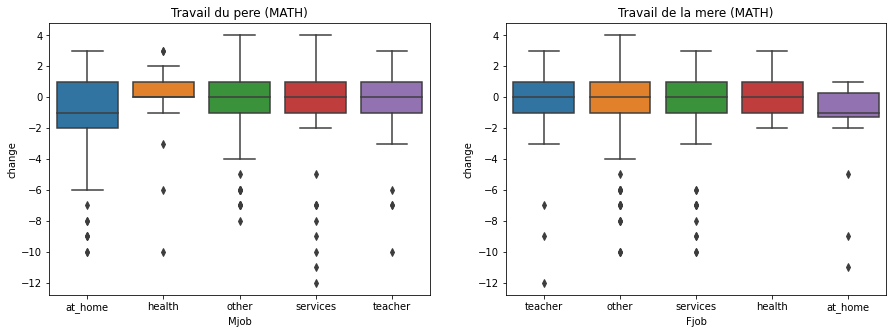


Nous pouvons constater que le niveau 4-5 Walc est plus rependu chez les femmes. Que ce soit pour le cours de math ou portuguais. Il faut aussi noter que l’alcool ne semble pas influencer les notes à la baisse et on contraire en moderation semble être bénéfique.

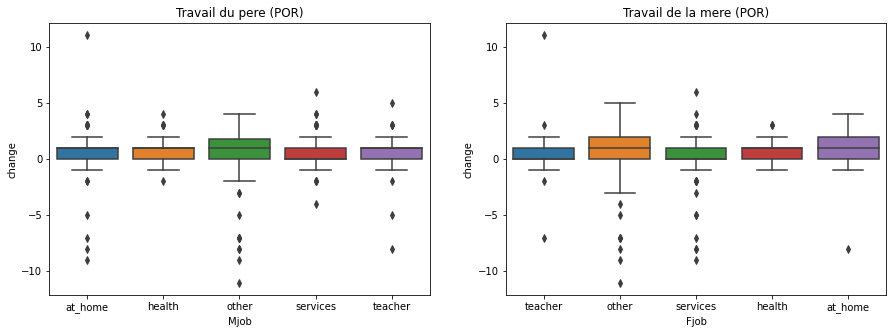
Influences sur la note finale.

Mother or father Jobs.

MATH



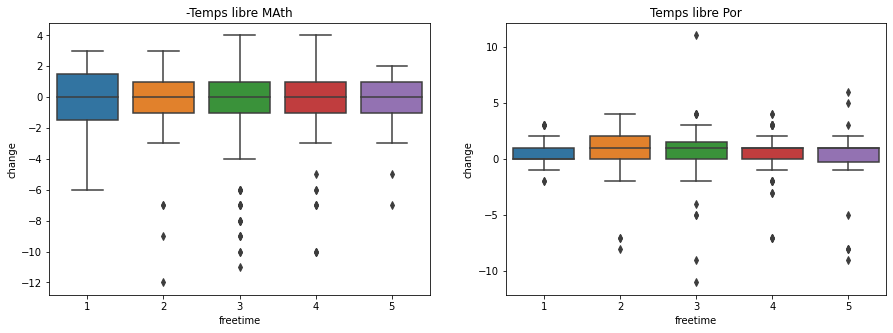
Portuguais



Conclusion : Les différents type de travail ne semble pas influencer les notes en general. Il n’y a pas vraiment de travail )même enseignant qui semble faire une différence majeur.

FREETIME :

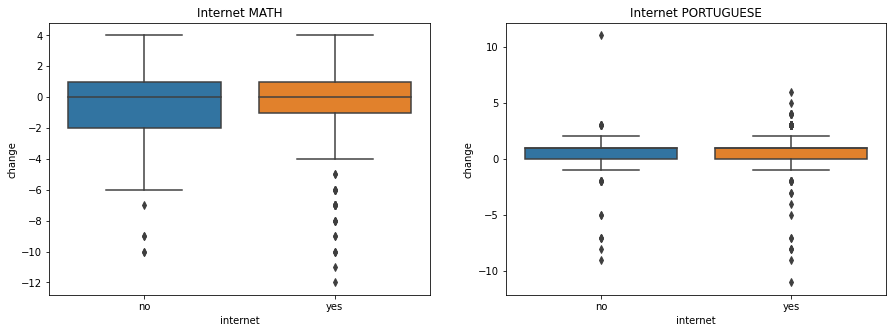
MATH VS PORTUGUAIS



Conclusion : Mathematique semble avoir plus de gens avec des difficultés. Parcontre la moyennes semble être la même pour chaque categorie de 1 à 5 (math et por).

INTERNET :

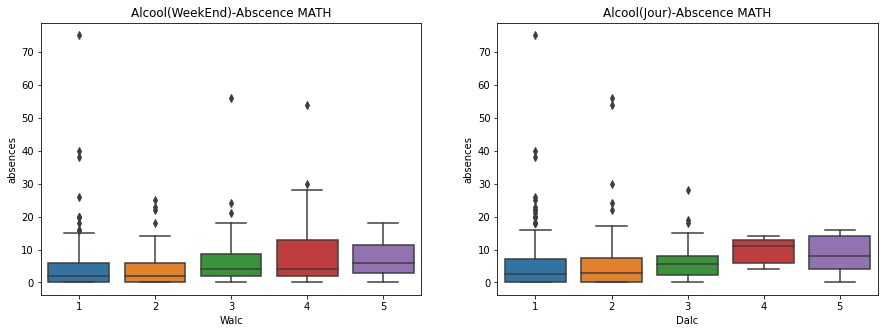
Math VS Portuguais



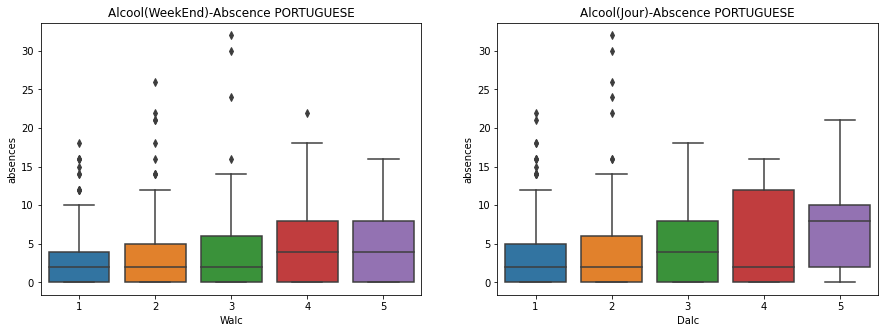
Conclusion : L’internet n’est pas guise de réussite. Nous pouvons voir qu’en mathematique il peut y avoir une certaine difference entre les gens qui ne l’ont pas. Parcontre pour le Portuguais il ne semble pas davoir un effet direct pour une meilleur note.

ABSENCE VS ALCOOL

Math



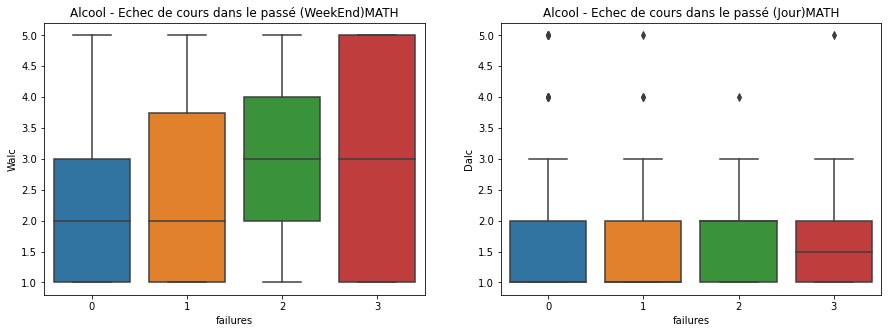
Portuguais

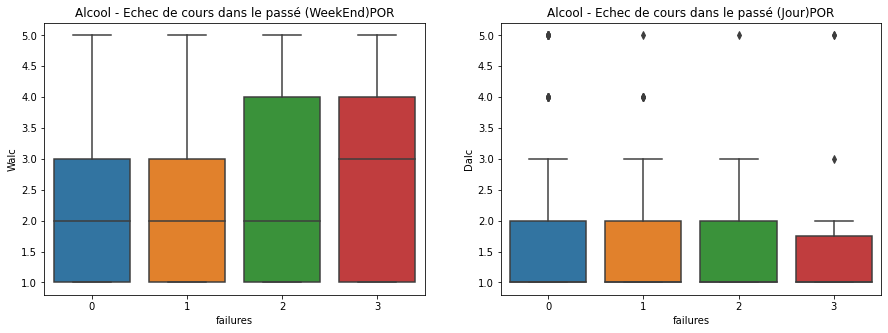


Conclusion : nouspouvons voir une difference evidente entre les personnes qui boivent plus fréquemment que ce soit la semaine ou les weekend, bien que la semaine a un peu plus d’influence sur les absences.

FAILURES (Echecs passés)

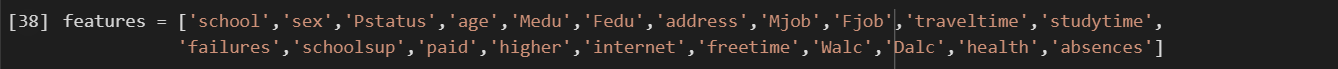
Il est important de savoir qu’ici nous parlons d’historique des gens avec echec et non des échecs du a la présente consommation d’alcool.

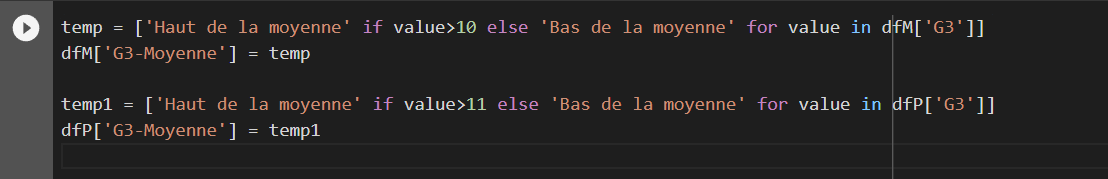


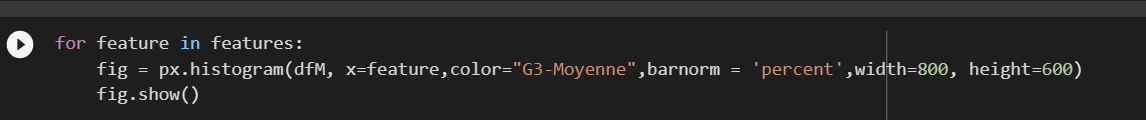


Conclusion : Nous pouvons constater que les gens plus habitués au échecs vont aussi plus consommé le Week-end.

Moyenne Haut-Bas et leurs corrélations.



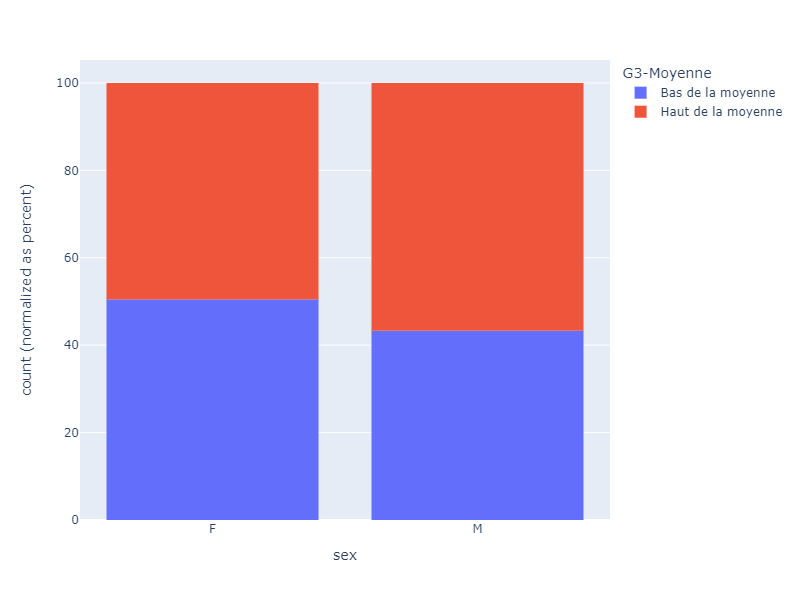
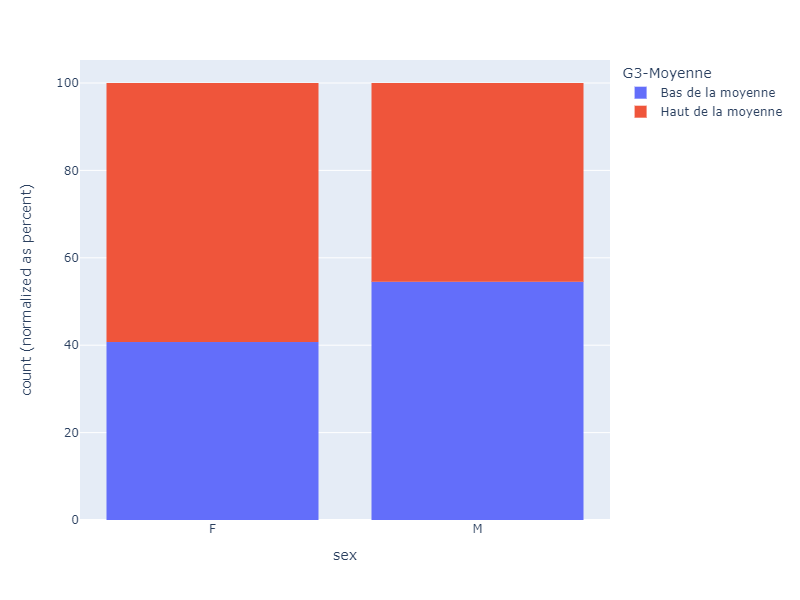




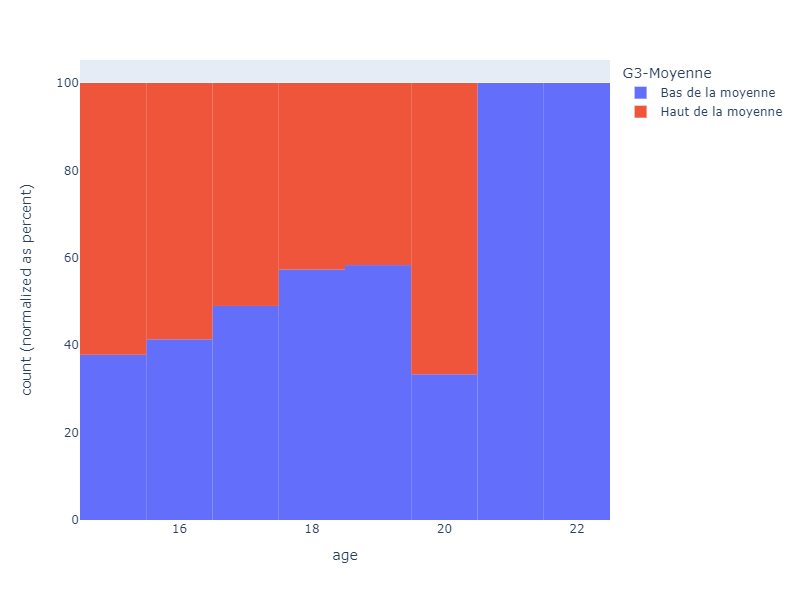
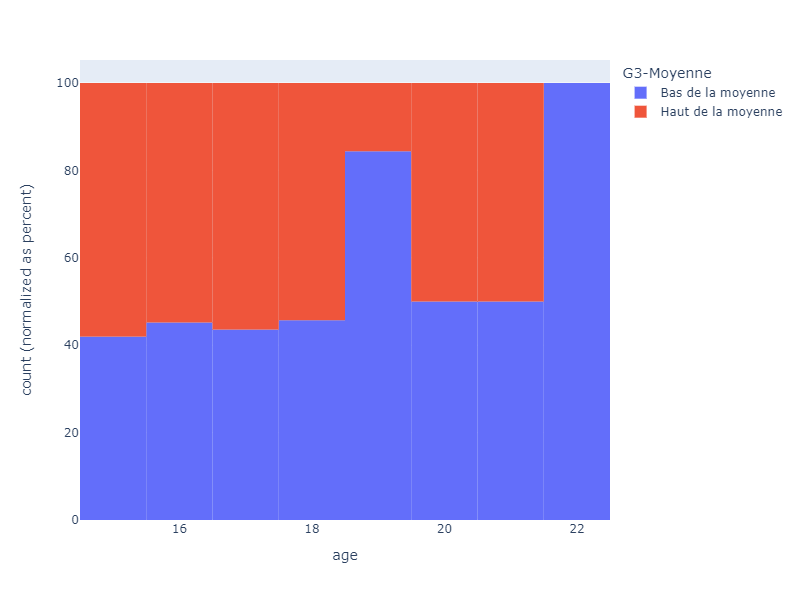
MATHEMATIQUE (dfM) / PORTUGUESE (dfP)

MATH VS PORT

Hommes ont plus de facilité en mathématique / les femmes en portuguais

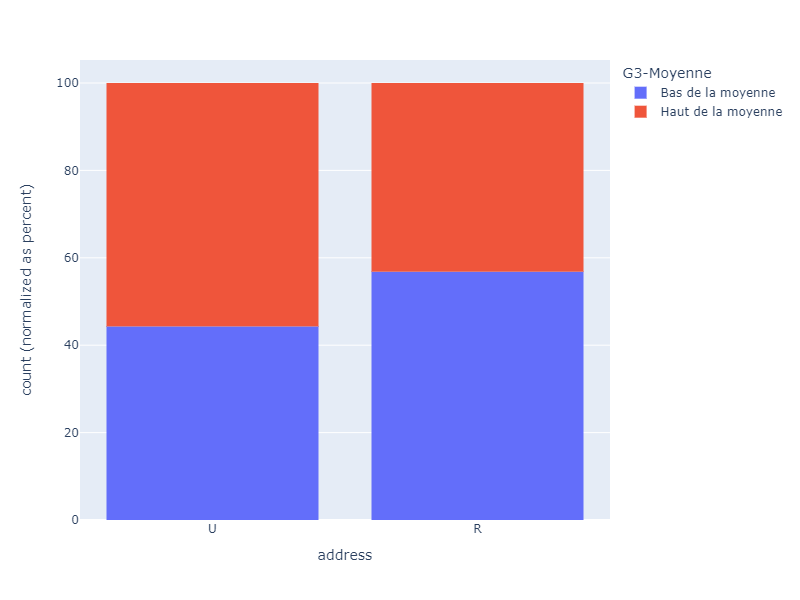
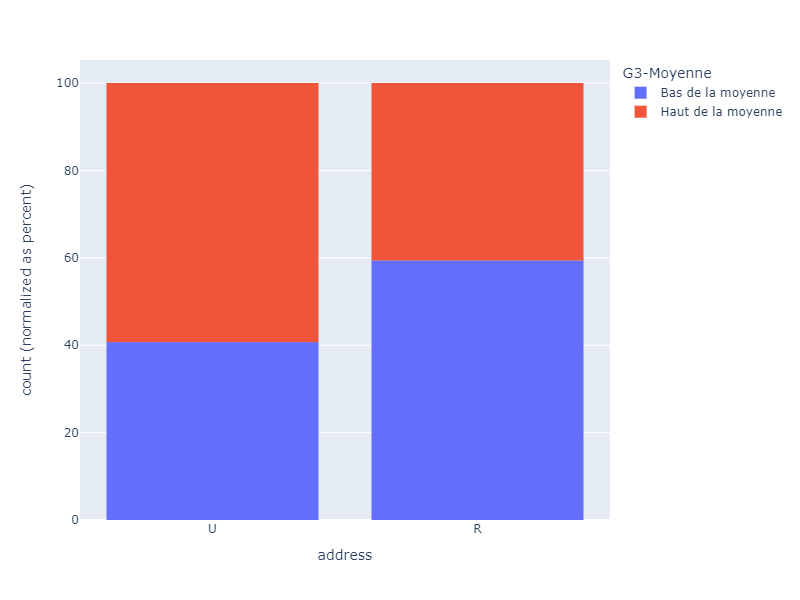


Il semble que les examens finaux en math fuent plus facile / Por. Plus difficile 18-19 ans

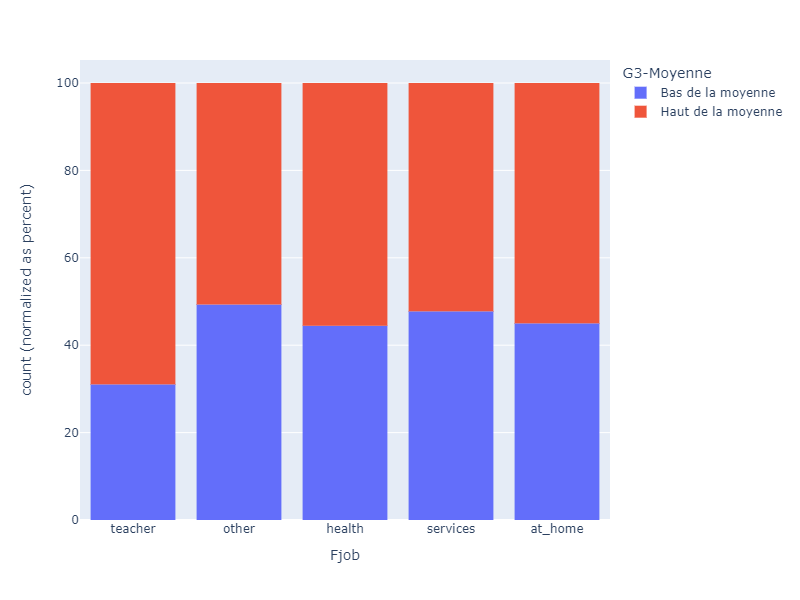
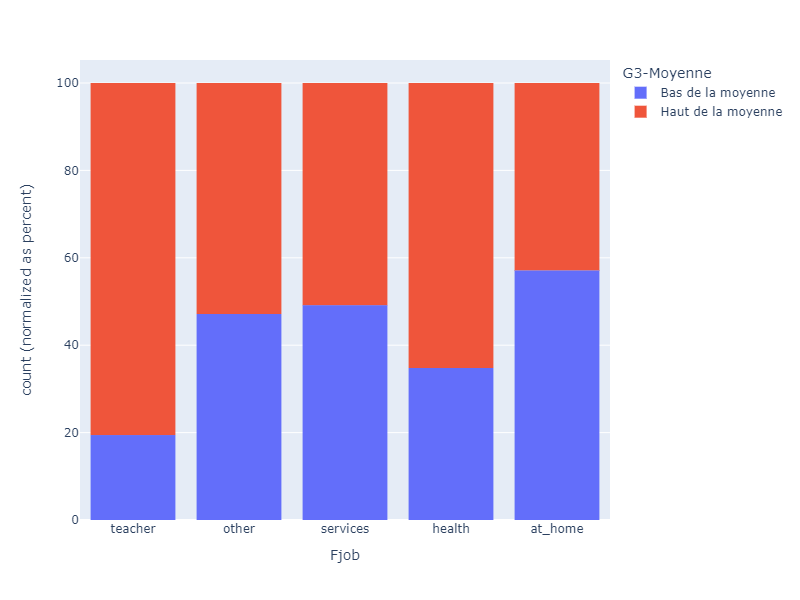


MATH VS PORT

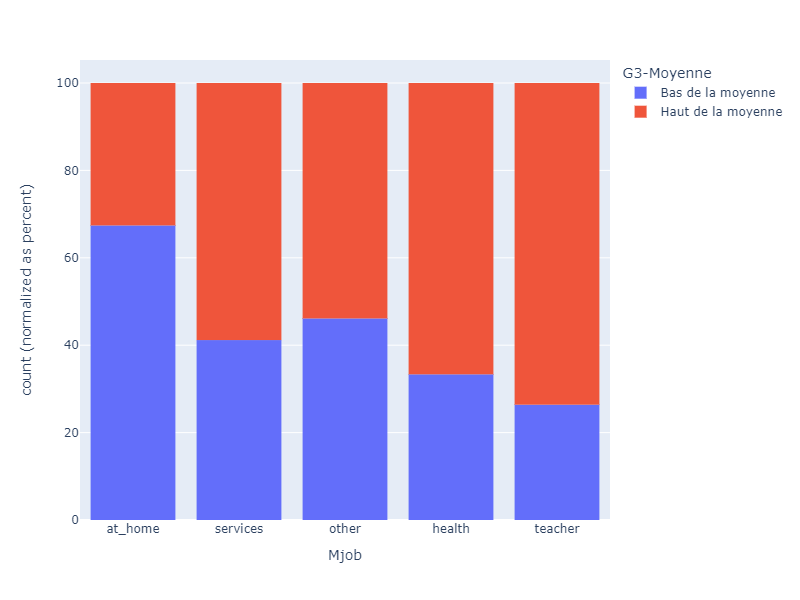
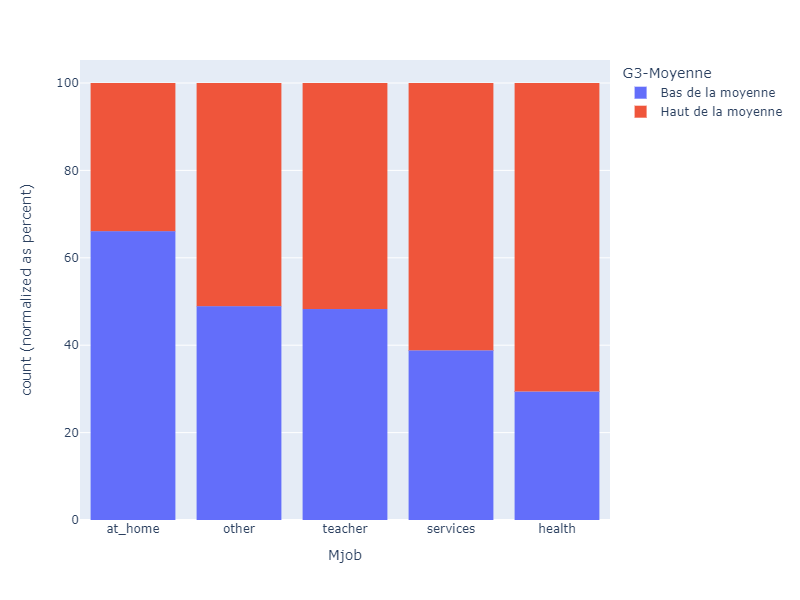
Habiter en ville semble une meilleur option pour être en haut de la moyenne.



Un parent professeur n’ameliore pas la note mais s’ assure d’être en haut de la moyenne

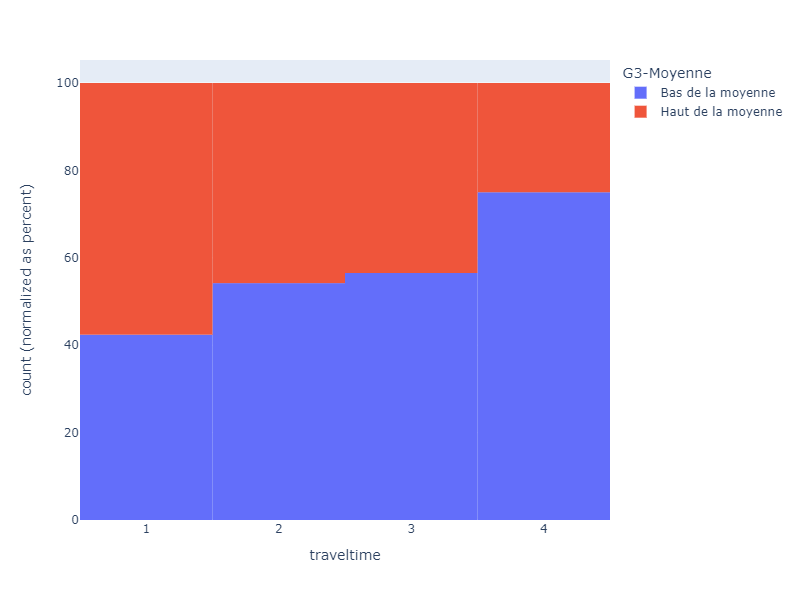
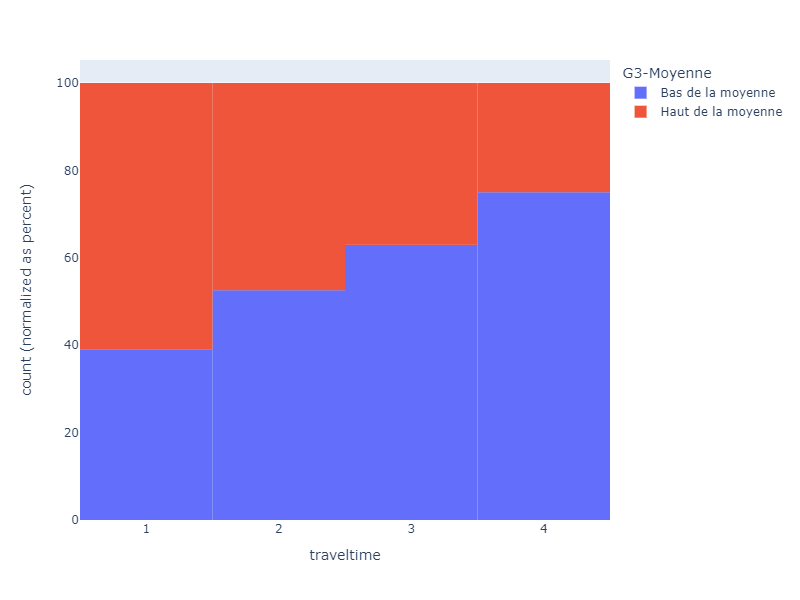


Même chose ici.

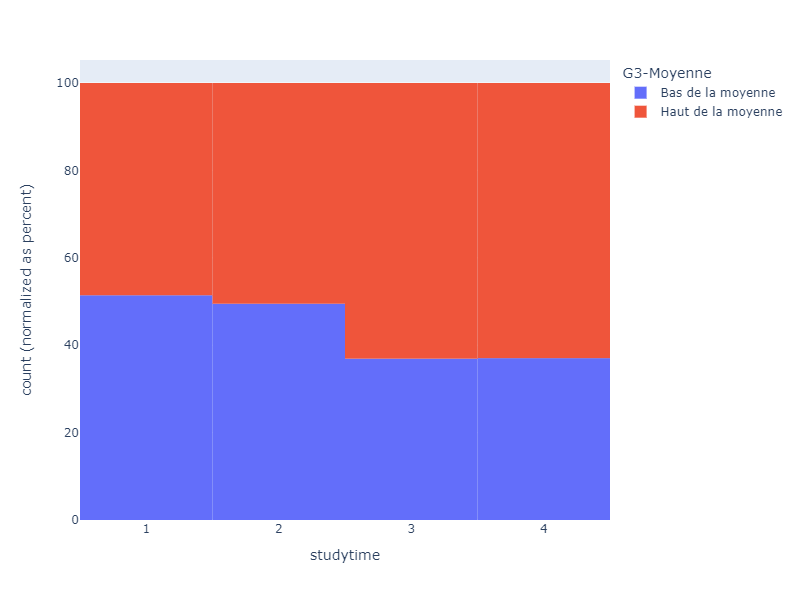
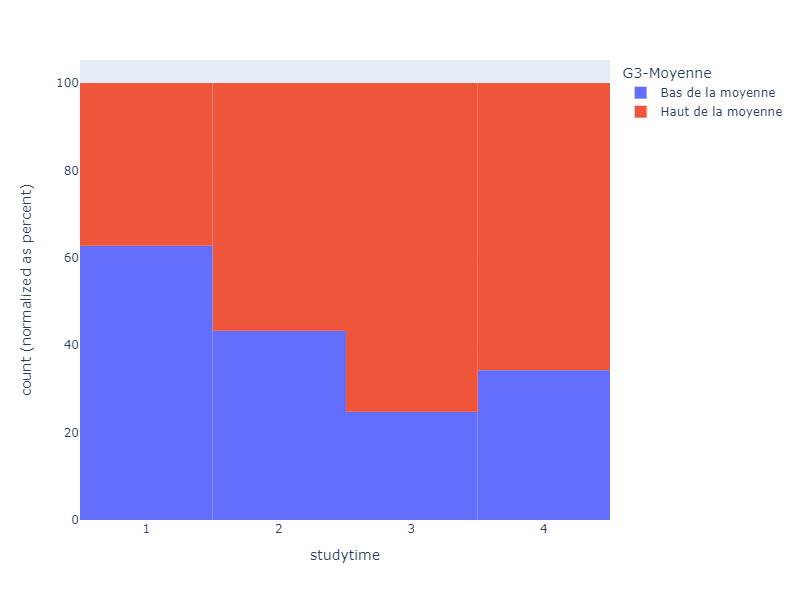


MATH VS PORT

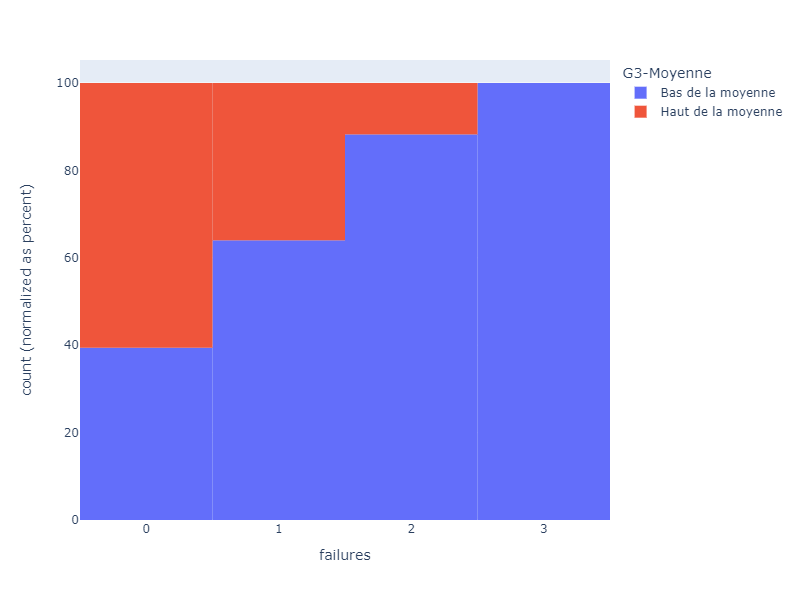
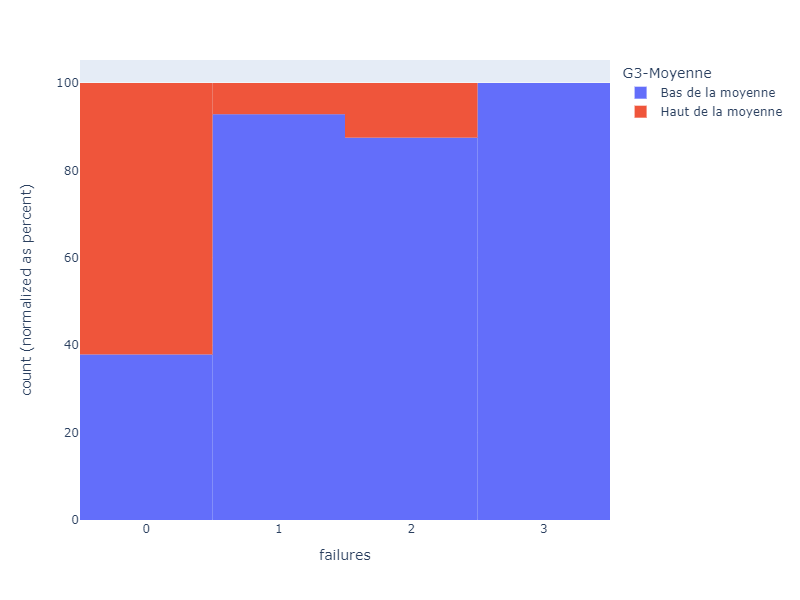
Avoir plus de distance semble affecter les notes des étudiants. Plus loin plus difficile.



Le temps d’étude fait une légère difference sur la moyenne.

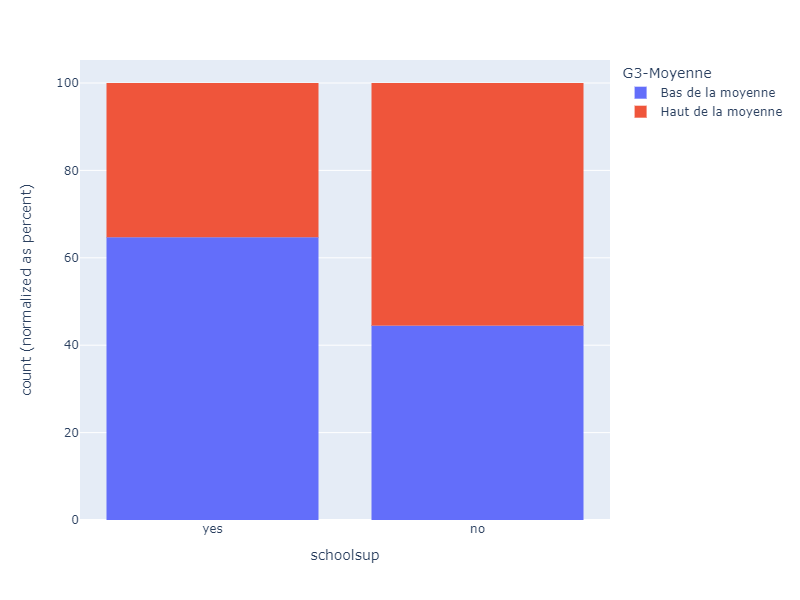
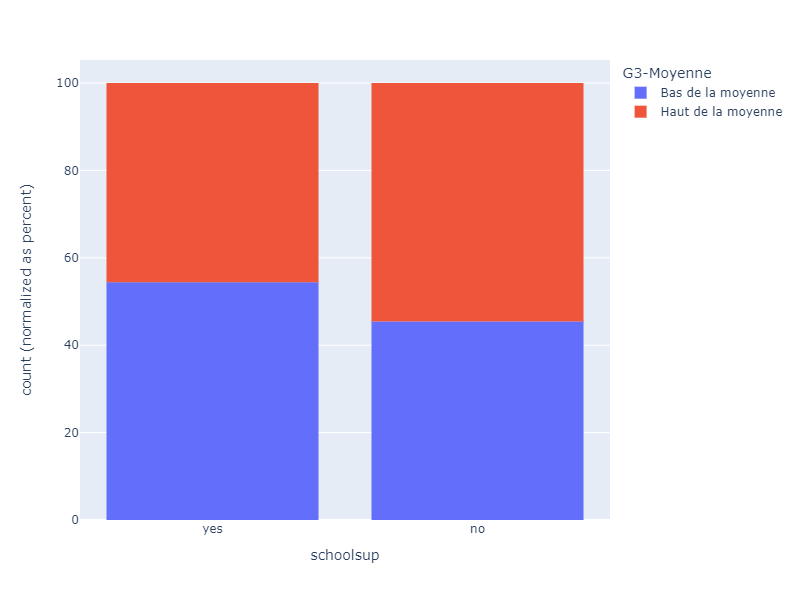


Les gens qui n’ont pas eu d’echec sont plus aptent a être au dessus le moyenne.

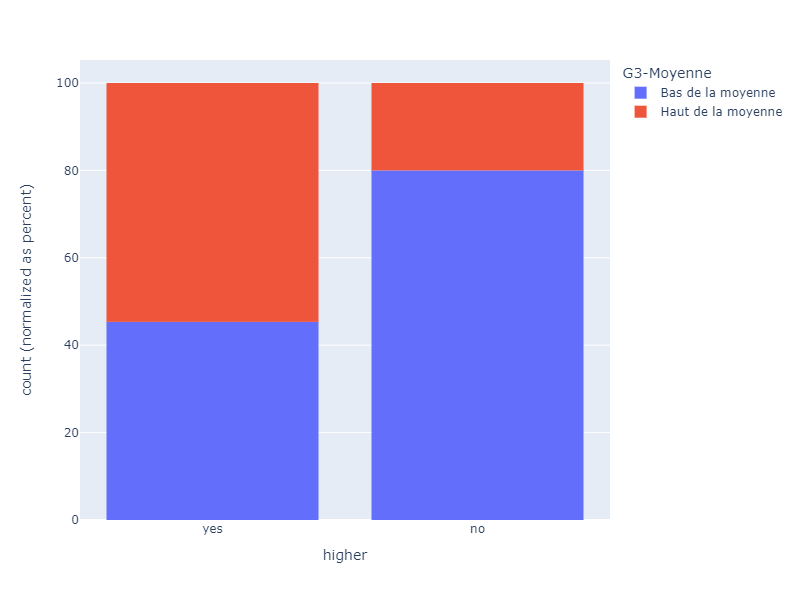
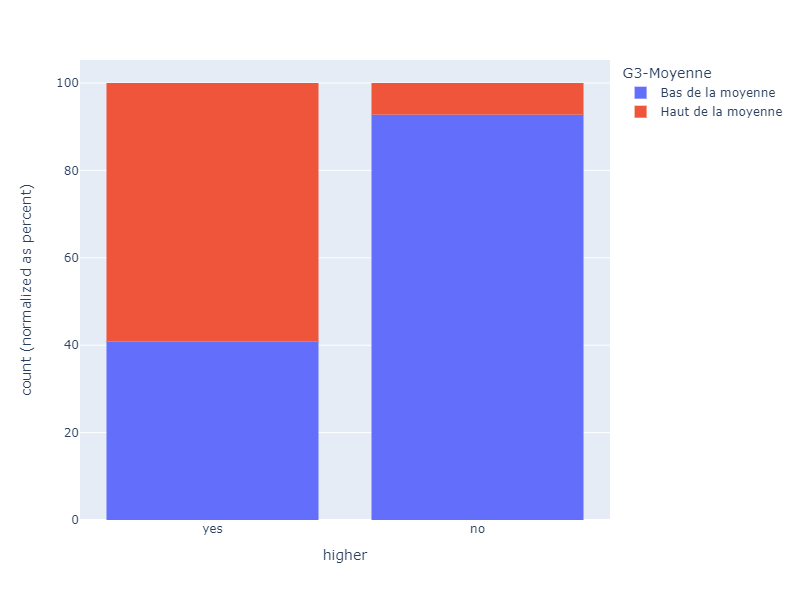


MATH VS PORT

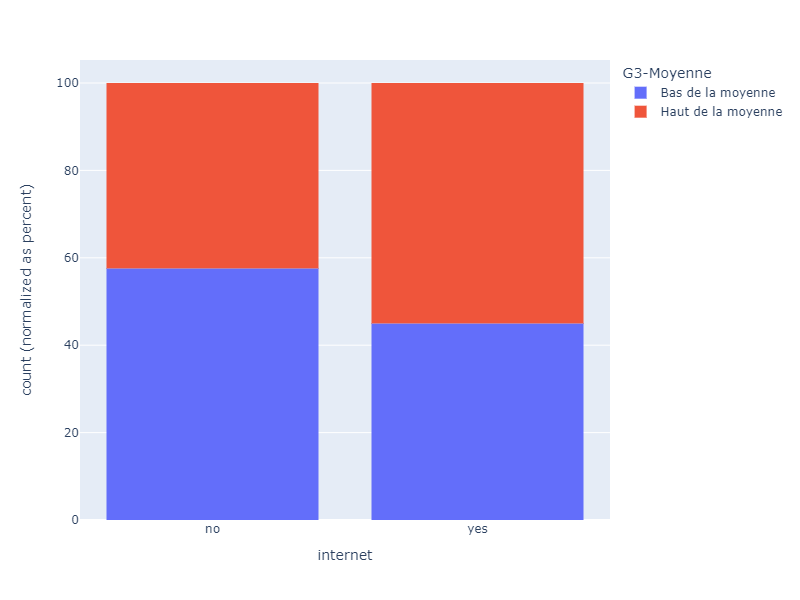
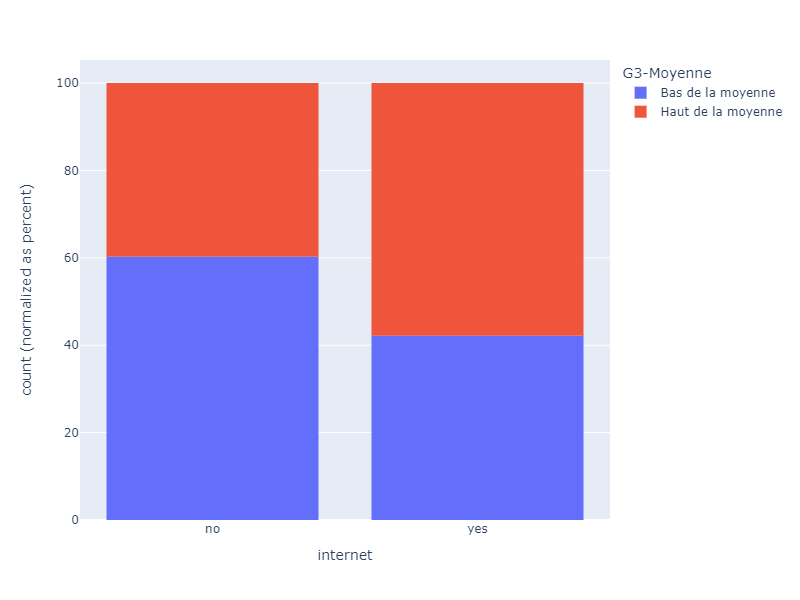
Ceux qui ont eu besoin d’aide a l’ecole sont majoritement en bas de la moyenne.



Prévoit l’université + sont academiquement prêt. La plus part en haut de la moyenne

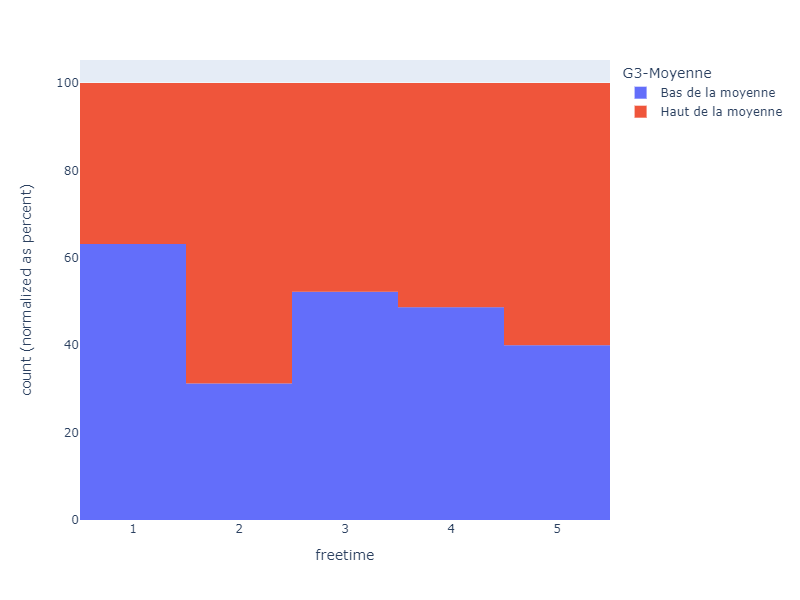
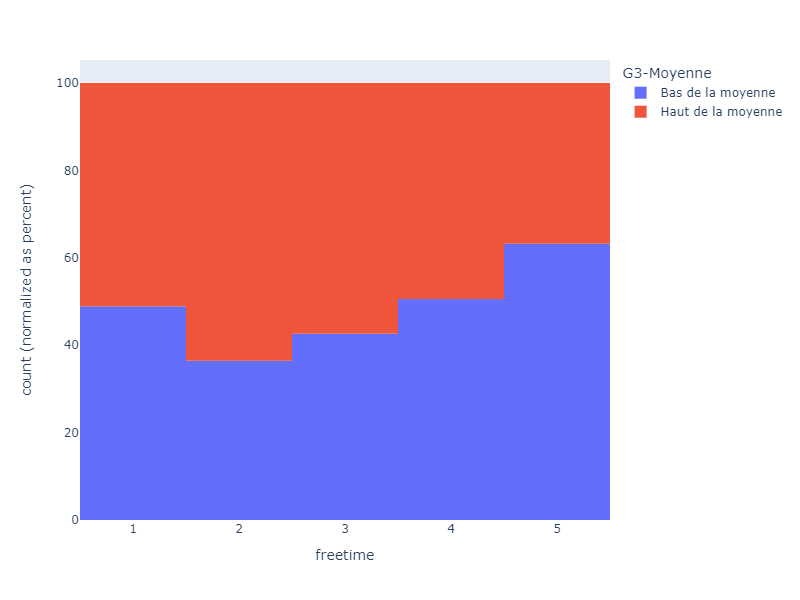


Internet donne un avantage au étudiants.

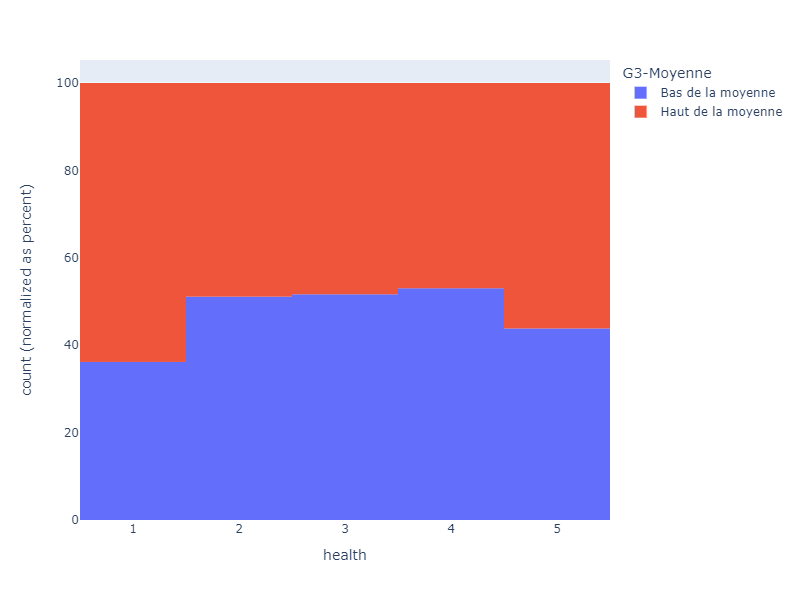
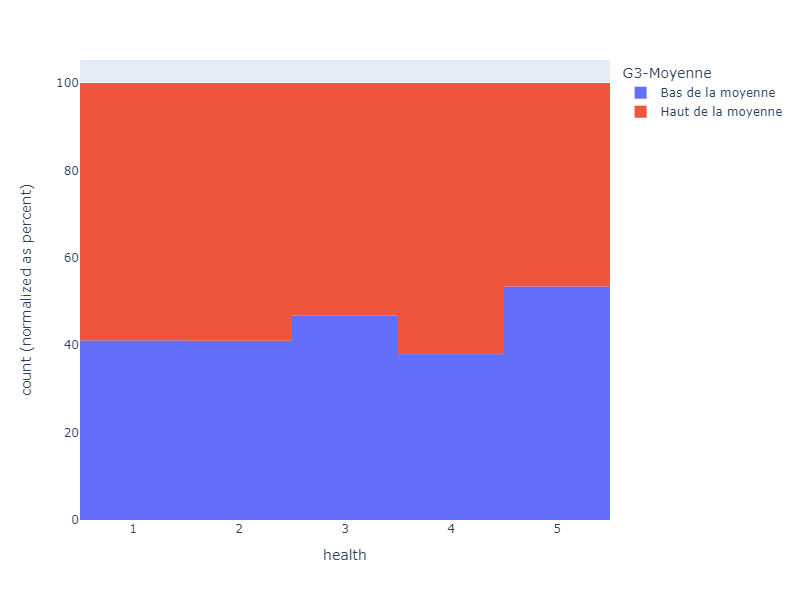


MATH VS PORT

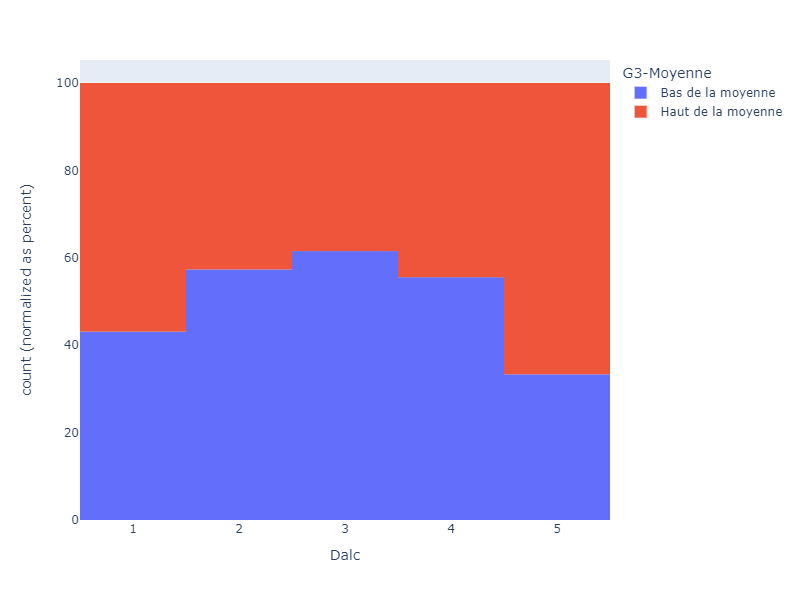
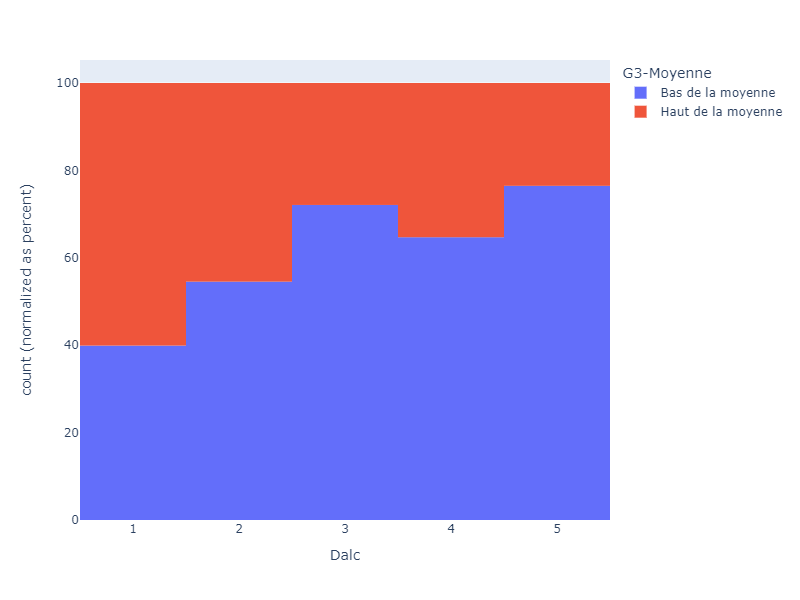
Les personnes n’ayant pas de temps libre sont majoritairement en bas de la moyenne



La santé ne semble pas être un facteur de changement pour les notes.

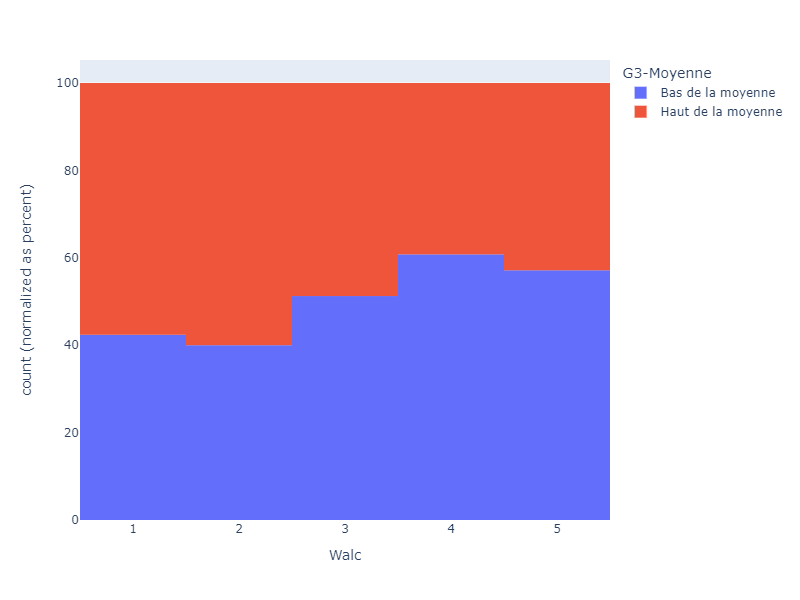
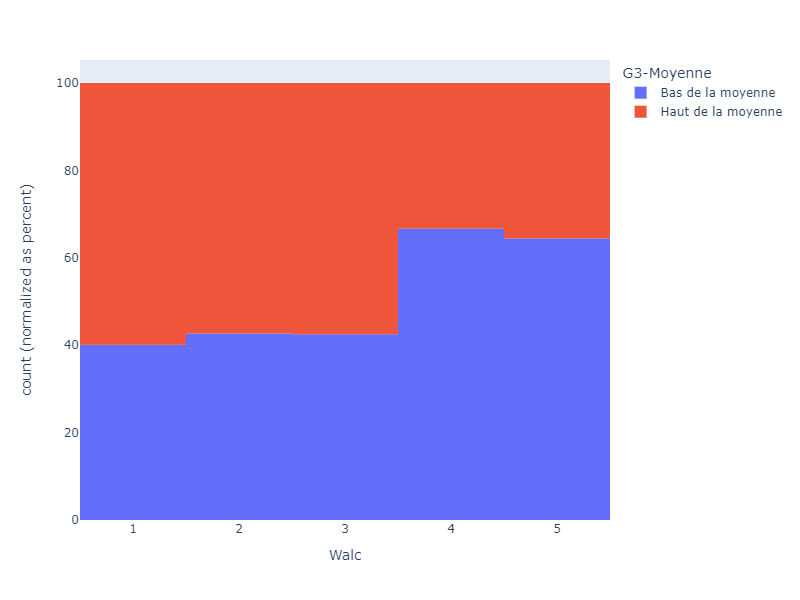


On peut voir que boire la semaine est facteur de risque surtout pour la langue.



MATH VS PORT

Il semble que boire le weekend aille moins d’impact en math mais toujours pour la langue



**Remarque et analyse :**

* **L’*alcool* a clairement un effet sur les notes des étudiants.**
* **La *santé* n’affecte pas les résultats des étudiants.**
* **La modération fait en sorte que les élèves performent mieux.**
* **Le *type de travail des parents* n’influence pas l’amélioration des notes mais la performance de celui-ci.**
* **L’*internet* aide à avoir une note stable mais aide rarement à s’améliorer.**
* **Les *échecs passés* exercent une grande influence sur les notes reçues par rapport à la consommation d’alcool.**
* **Habiter en à *l’extérieur de la ville* semble désavantager les élèves.**
* **Les étudiants aillant comme but d’aller *plus loin dans leurs* études sont génialement en haut de la moyenne.**
* **Le *temps libre* est nécessaire pour performer à l’école.**