# IE6600_Sec05_Group18_Hackathon

Simran Bhatia, Yanming Liu, Yash Nema

10/12/2021

## Introduction and Problem Statement

Education is an integral factor in development or upliftment of any social group. It is also an essential part of any humans' life: that shapes their personal outlook and opinions and helps them progress. That is why the dataset "Student Alcohol Consumption" was chosen by this group for our hackathon. This data comprises of drinking habits, grades, demographic information and other social factors (parents' education, family size, relationship status, etc) for school-going students in Portugal from 2005-2006. Eurostat's "Early leavers from education and training by sex and labour status" data from 2007 shows Portugal's early leaver from education rate as 36.5% against European Union's average of 15%, thus emphasizing the significance of this subject and its analysis.

Even though Portugal has since pulled through their education crisis exhibiting dropout rate of 8.9% against EU's average of 9.9% in 2020, we aim to explore causation, relations and insights from students' alcohol consumption habits, their grades, and social factors they may be affected by. This report is divided into three sections: `Understanding Student Audience in Mathematics class of Portuguese Schools`, `Analyzing factors impacting alcohol consumption among math students in Portuguese Schools`, `Deep dive into factors impacting grades among mathematics students in Portuguese Schools`.

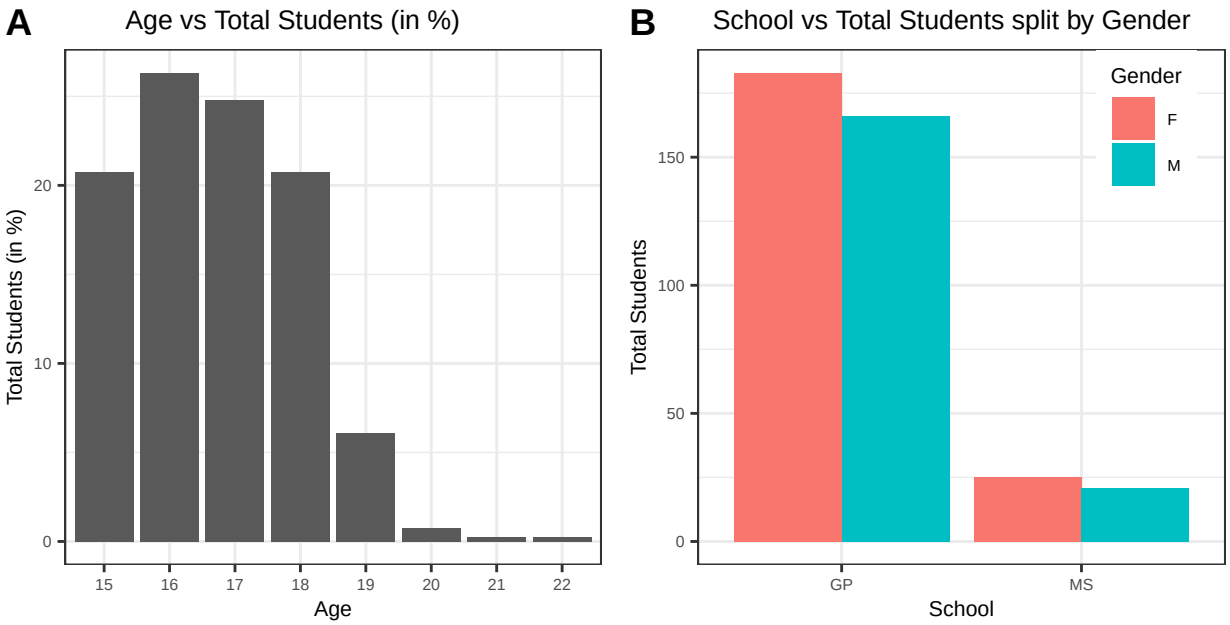
In `Understanding Student Audience in Mathematics class of Portuguese Schools`, exploratory data analysis is conducted on the dataset to understand the dataset, look for points of interest and determine what portion of the data should be explored next. Alcohol consumption habits of Portuguese students are compared against demographic and social indicators such as age, parents' jobs, social life in `Analyzing factors impacting alcohol consumption among math students in Portuguese Schools`. Finally, geographical, social and demographic data points like study time, travel time, health, and internet access are compared to find trends affecting students' grades and the aforementioned analysis on alcohol consumption is used to determine relationship between students' grades and alcohol consumption in `Deep dive into factors impacting grades among mathematics students in Portuguese Schools`.

## Section 1: Understanding Student Audience in Mathematics Class of Portuguese Schools

Q1: Who are the students taking mathematics classes in Portugal?

Insight: Math class consists of ~93% students in age group of 15-18 years. School GP consists of significantly higher enrollment of students in Math class compared to school MS, but both contain higher no. of female enrollment compared to male enrollment.

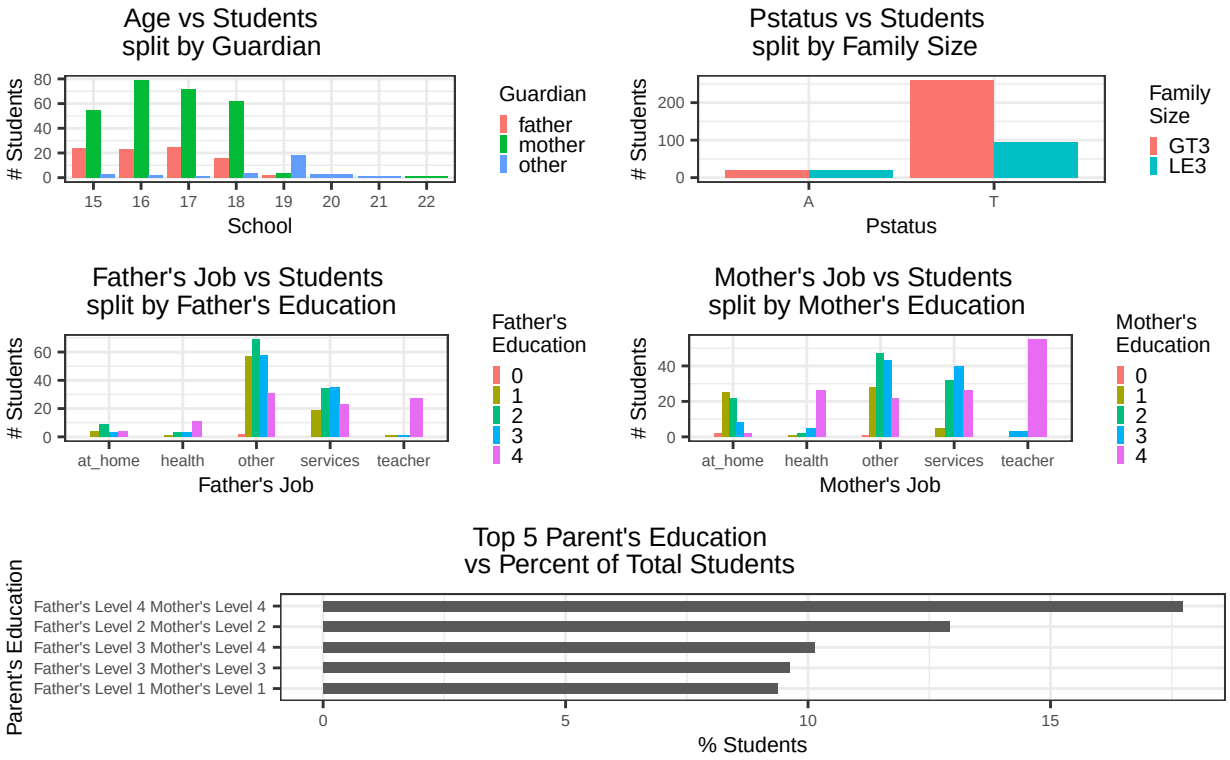
## Who are the students taking mathematics classes in Portugal?




Q2: What is the family background for mathematics students in Portugal?

Insight: Mothers are the primary guardian for most students across different ages. So factors related to mother's education and job could impact future analysis. Most students belong to families that are together. These families tend to have high (Greater than 3) family members. For separated families, we see number of students to be approximately equal for greater than and less than 3 family size. In both Father and Mother Jobs, we can see teacher and health roles have highly educated fathers and mothers. For other jobs we see education levels between 1 and 3. We also observe that both parents generally have same education level.

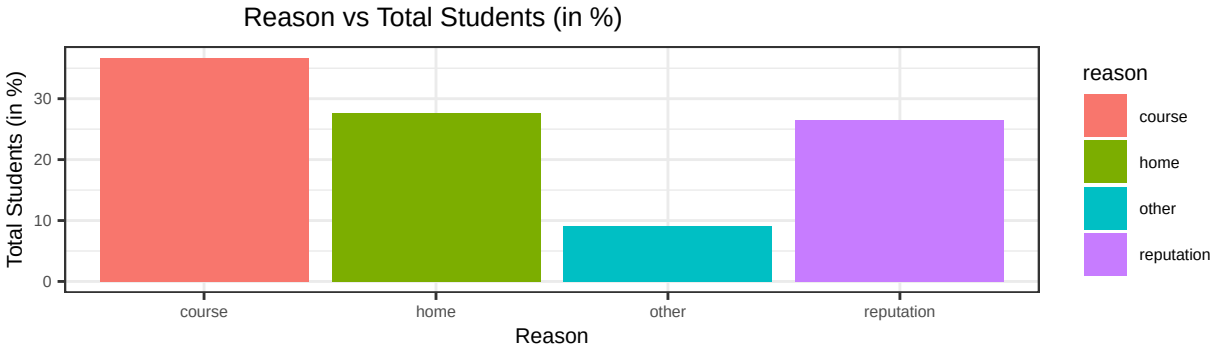
## What is the family background for mathematics students in Portugal?

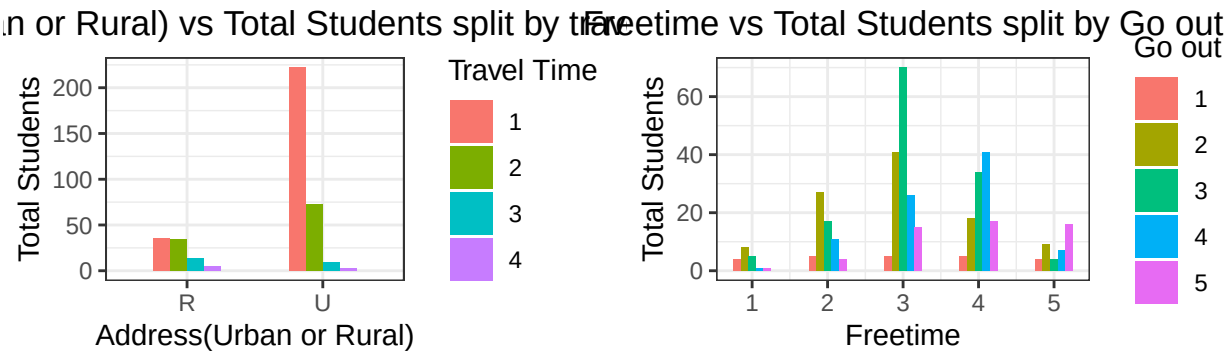



Q3: How do mathematics students behave/think based on the subject data set in Portugal?

Insight: ~37% math students in Portugal chose their school by course preference. Urban address students generally have a lower travel time, compared to travel time for students living in rural areas. We also see that the level of going out depends on the amount of free time of a student. Higher the free time level, higher the going out level.

## How do mathematics students behave/think based on the subject data set in Portugal?



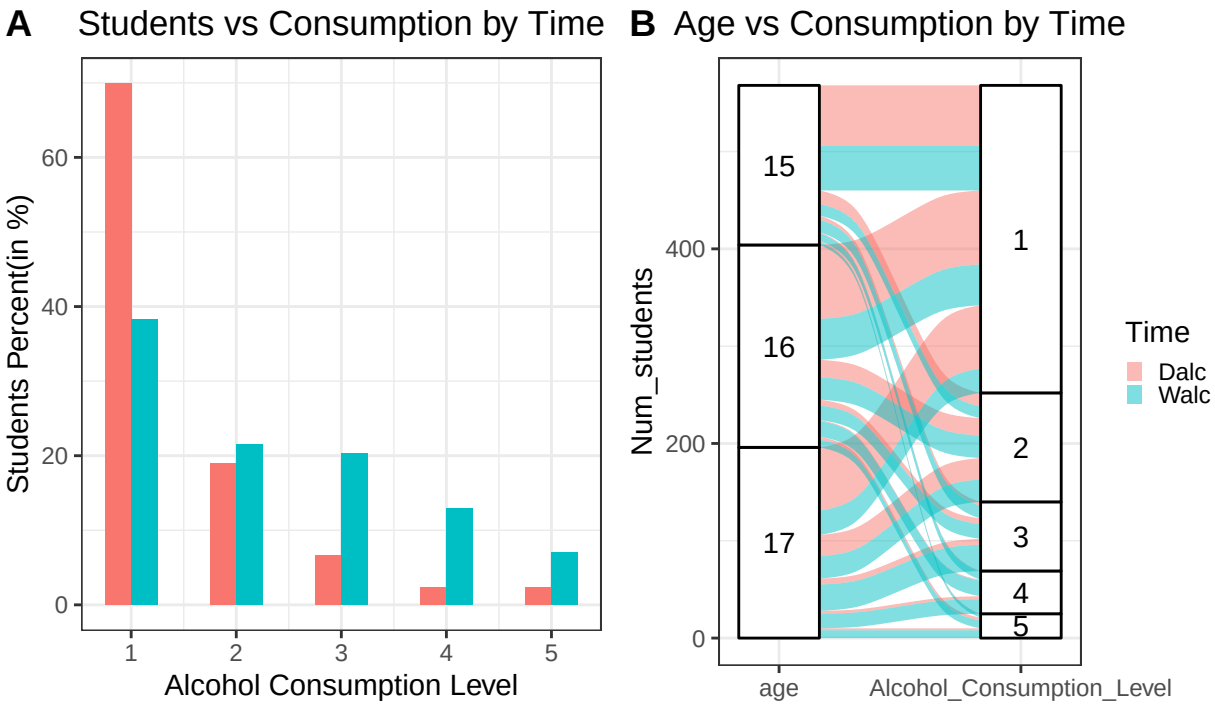




## Section 2: Analyzing factors impacting alcohol consumption among math students in Portuguese Schools

Q1: When and what age groups consume alcohol in Portugal?

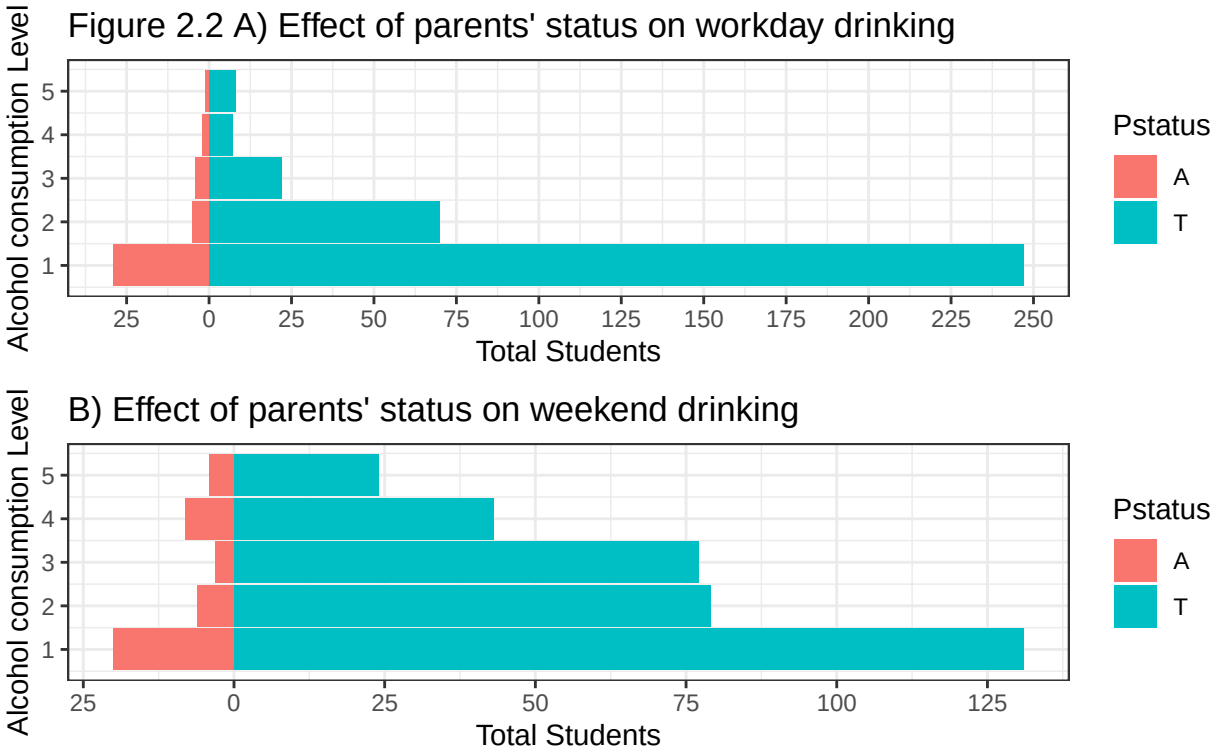
Insight: While alcohol consumption is generally on the lower side(Level 1 consumption dominates), a shift is observed in drinking habits from weekdays to weekend. More students from various Walc alcohol consumption group are consuming alcohol on the weekends than weekdays, and a higher quantity of alcohol is consumed on weekends compared to weekdays.

## Figure 2.1 When and What Age Consume Alcohol?




Q2: How does Pstatus impact alcohol consumption?

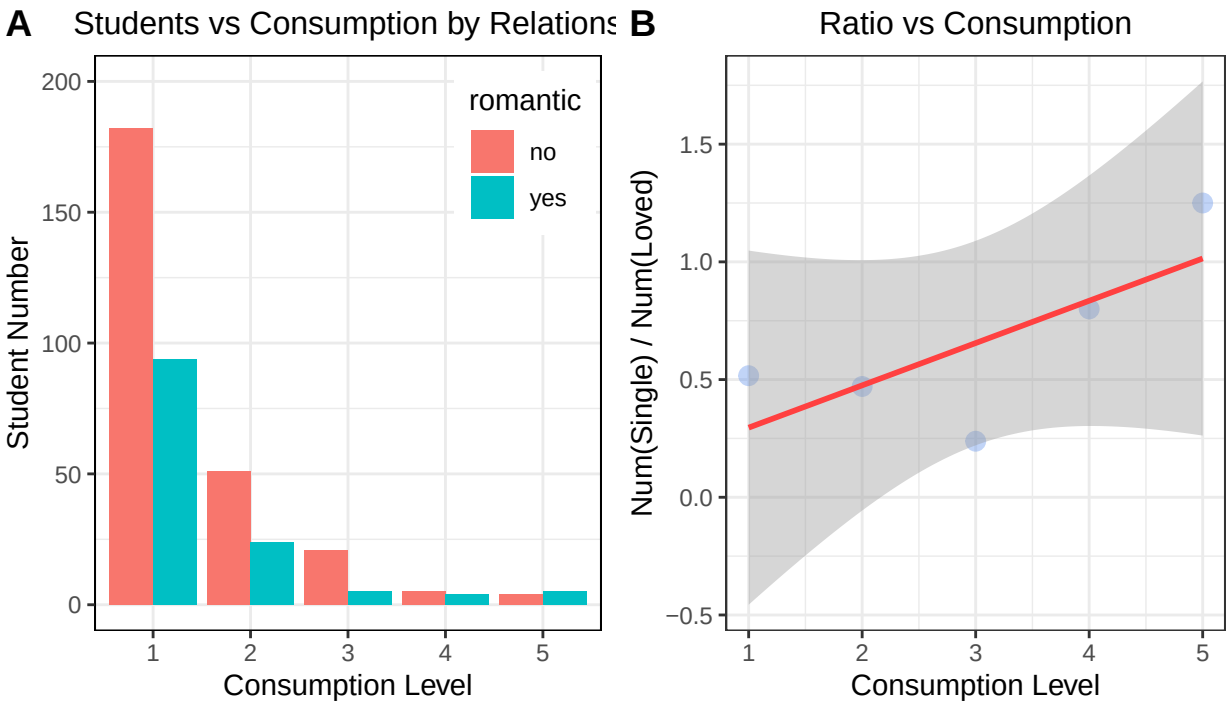
Insight: Observing family status of students ie whether parents are together or apart, shows most students live with both their parents. The same trend is observed for weekday vs weekend alcohol consumption for students regardless of their family status. Students are consuming more alcohol over weekends than weekdays and a larger number of students is consuming alcohol over the weekend than weekdays.




Q3: How does romantic involvement impact alcohol consumption?

Insight: As alcohol consumption increases, the ratio of the number of single students to committed ones increases in both daily and weekend alcohol consumption. As can be seen for weekday data below, students involved with romantic partners are more likely to over indulge in weekday drinking.

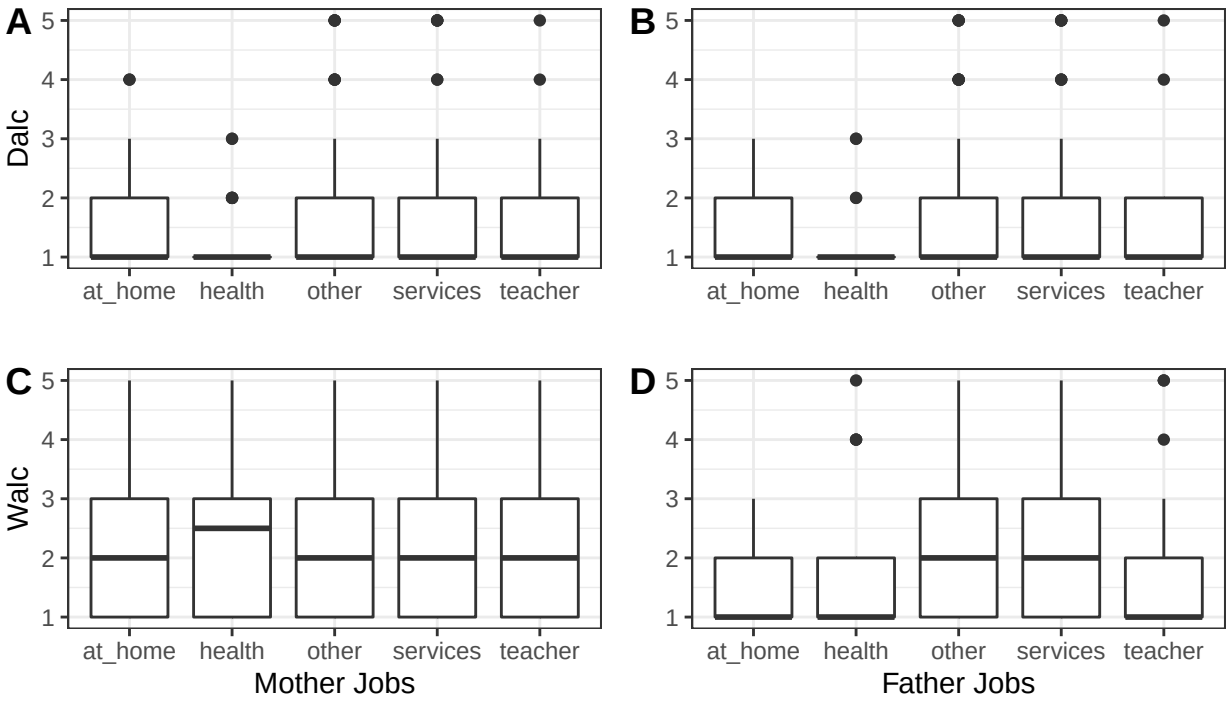
## Figure 2.3 Romantic Relationship vs. Weekday Alcohol




Q4: How does parents' education impact alcohol consumption?

Insight: Parents, especially with health related careers bear a strong influence on students' alcohol consumption on weekdays i.e, students with parents in health care tend to drink less during the week (as can be seen in below Fig A&B). Also, we see fathers play a stricter role in limiting students' drinking during weekend if they have a job at home or are either teachers or in health industry. We can see the decrease in mean value of alcohol consumption in Fig D compared to mean consumption seen in Fig C for mother's jobs.

## Figure 2.4 Alcohol Consumption vs Jobs Type

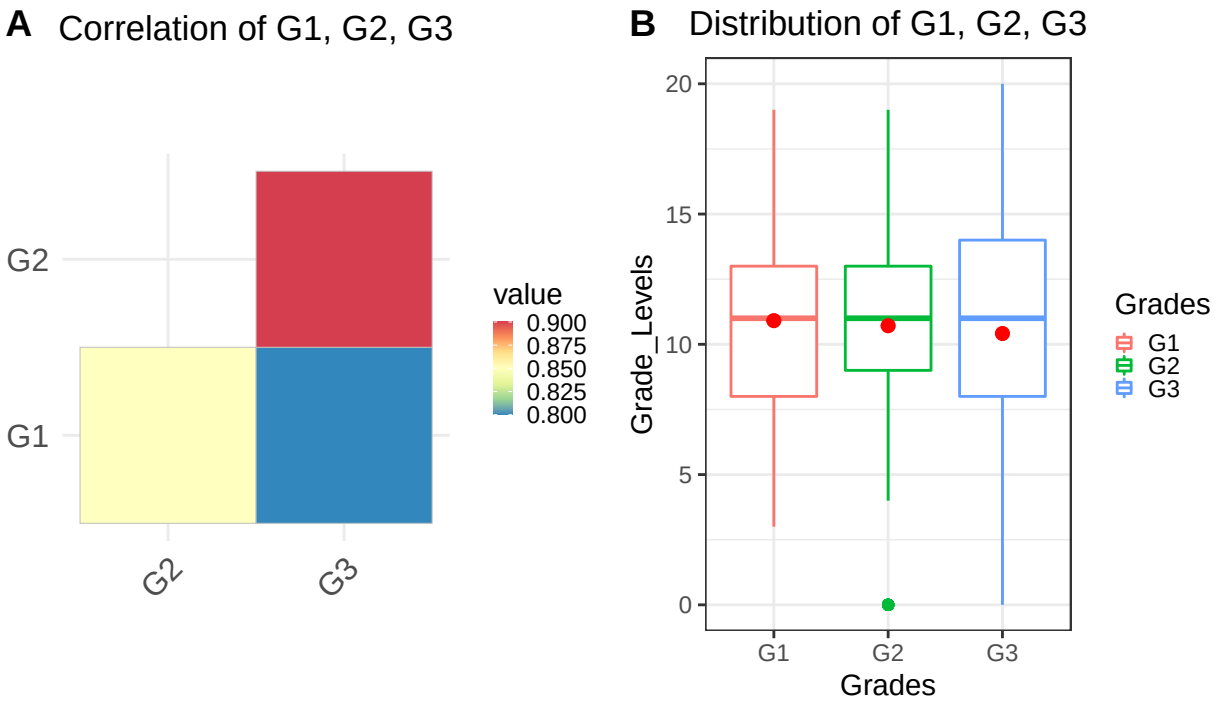



## Section 3: Deep dive into factors impacting grades among mathematics students in Portuguese Schools

Q1: How does performance in G1 and G2 impact the final grade G3?

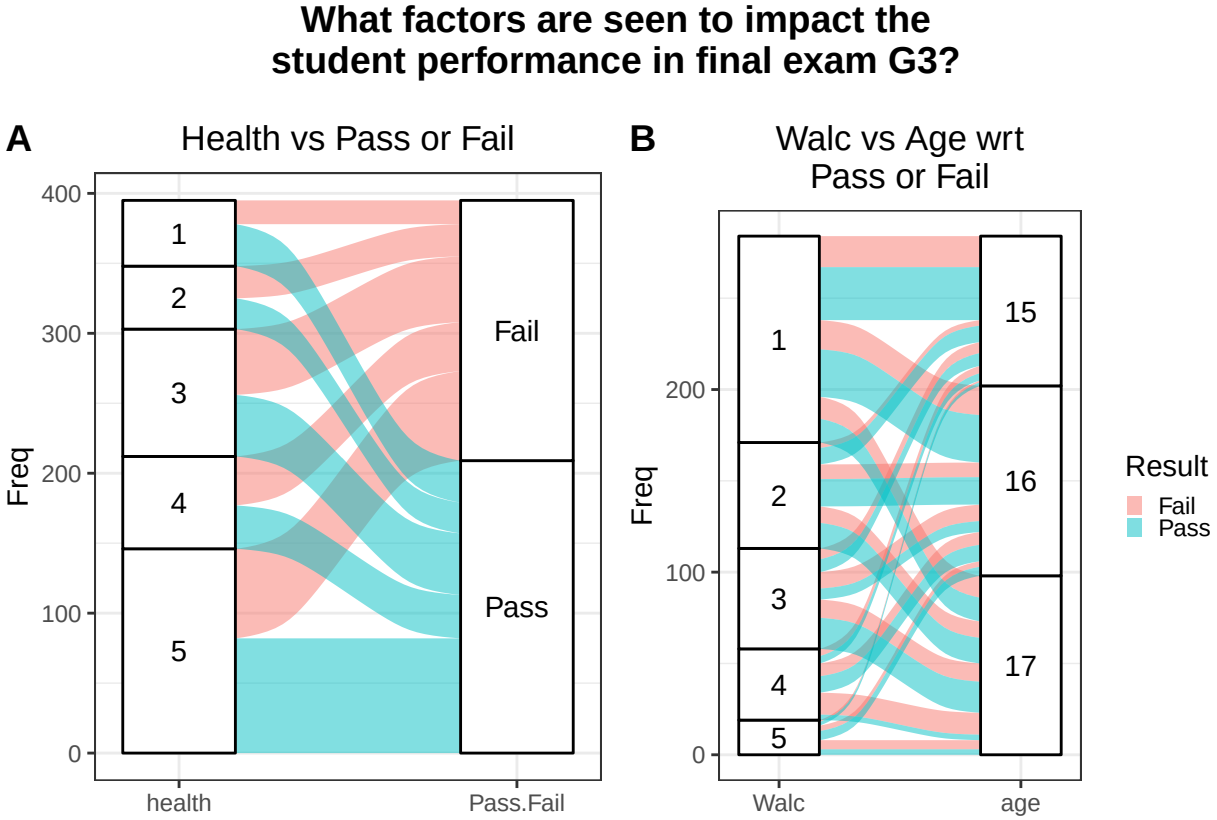
Insight: We see that G1 and G2 are very closely related to G3. A strong correlation of +0.8 is seen between G1 and G3, and +0.9 between G2 and G3. We also see that G1 and G2 have a correlation of +0.85. It is possible that doing well in G1 improves chances of performing in G2 which ultimately results in good G3 performance. We can also see that even though the mean of the grades are very close in the box plot, the spreads are different for G1, G2 and G3 (widest spread)

## Figure 3.1 G1 and G2 Impact the Final Grade G3




Q2: What factors are seen to impact the student performance in final exam G3?

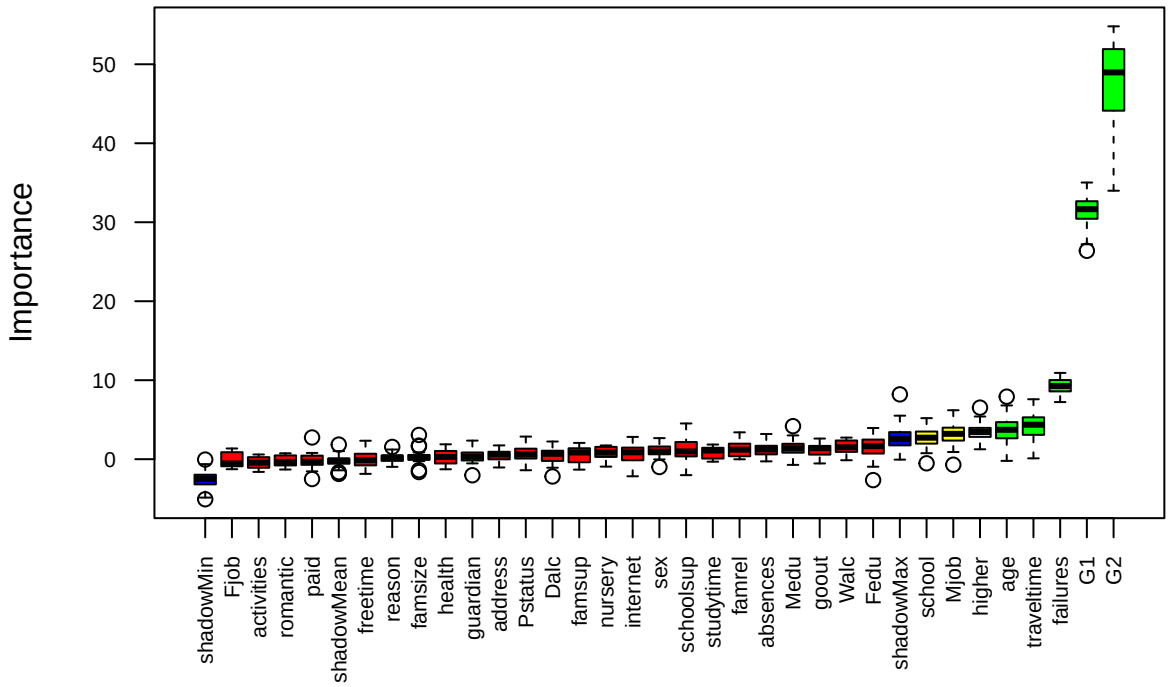
Insight: Study time has a strong correlation (0.55) with final performance in G3, ie more study time translates to higher G3 score. If a student passes, they are 1.03x more likely to have spent a high amount of time studying. Also, if a student fails they are 1.08x more likely to have taken absences in the past; absences have a correlation of -0.41 with G3 grade. Health has correlation of -0.46 with G3, this is contrary to what was hypothesized earlier. Since correlation doesn't mean causation, there can be correlation that doesn't impact G3 performance. Finally, it's observed that if a student fails, they are 1.06x more likely to have higher travel time. There's a weak negative correlation -0.36 between travel time and G3 performance, ie more travel time leads to lower G3 score.




Q3: What are the most important variables that can be used to predict final grade (G3) performance?

Insight: Higher the variable importance, the more strongly the variable impacts "Pass" or "Fail" metric in the final G3 exam. Most important metrics in green color like G2, G1, failures, travel time etc can be used in combination to predict the final grade G3 for any student.

**Figure 3.3 Variable Importance**




## Summary and Conclusion

To summarize, the following analysis was conducted:

Section 1: Understanding Student Audience in Mathematics class of Portuguese schools Q1: Who are the students taking mathematics classes in Portugal? Q2: What is the family background for mathematics students in Portugal? Q3: How do mathematics students behave/think based on the subject data set in Portugal?

Section 2: Analyzing factors impacting alcohol consumption among math students in Portuguese schools Q1: When and what age groups consume alcohol in Portugal? Q2: How does Pstatus impact alcohol consumption? Q3: How does romantic involvement impact alcohol consumption? Q4: How does parents' education impact alcohol consumption?

Section 3: Deep dive into factors impacting grades among mathematics students in Portuguese schools Q1: How does performance in G1 and G2 impact the final grade G3? Q2: What factors are seen to impact the student performance in final exam G3? Q3: What are the most important variables that can be used to predict the final grade G3 performance?

In Section 1, we tried to understand demographic and psychological behavior of students enrolled in Mathematics class in Portuguese schools. We saw that mathematics classes majorly consisted of students of ages between 15-18 and higher percentage of females were enrolled in the course. We also saw that having mother as primary guardian impacts the alcohol consumption and final G3 grades. The job and education of the student's mother played a crucial role in predicting the final G3 performance. Also, we found that larger number of students live in urban areas and have a lower travel time to schools due to this.

Then, in section 2, we tried to see what factors lead to alcohol consumption by students. Here we explored that most students have low alcohol consumption on workdays and weekend and are likely to drink more on

weekends compared to weekdays. We didn't see a very significant impact of Pstatus (Parents separated or Together) on their alcohol consumption habits. We then found that being in a relationship resulted in more alcohol consumption on a workday.

Finally in section 3, we wanted to see if the alcohol consumption impacted final exam performance, and how we can use few variables from the entire dataset of 33 columns to predict the likely grade of a student. Here, we saw that G1 and G2 performance was crucial for good G3 final exam performance. Past failures, large travel times, Mother's job play an important role in predicting the final G3 performance. Using just these 5-6 variables we can predict the final G3 grade without using all the 33 variables in the complete dataset.

## References


1. P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.

2. Eurostat, 2007. Early school-leavers. http://epp.eurostat.ec.europa.eu/