THE RELATIONSHIP BETWEEN STUDENT SUCCESS AND ALCOHOL CONSUMPTION

GROUP 3: **RYAN PERMENTER**
HOA ROACH
ABBY HERRUP
CLARENCE ROBINSON

1. Overview

The aim of this report is to detect if there is any correlation between the students environmental and personal factors and their academic performance. The analysis also focuses on student's alcohol consumption, the factors impact on alcohol consumption and how it in turn affects student's grades. We done exploratory data analysis of data set collected by a survey. Our method of analysis was done by implementing a cross examination using qualitative variables, in which we evaluated demographic and non-demographic information such as education level, parent's education and their current jobs, parents support in education, if the students want to achieve higher education, amount of absences, etc. Subsequently, this report aims to see the impact of student alcohol consumption on their overall examination performance, if these students were to drink during the weekday, rather than the weekend.

2. Methods

The methods used in our analysis is exploratory data analysis (EDA). We intensively employed the visual to summarize the data and to tell the story behind the dataset. Modeling and hypothesis testing also are used to test for the correlation and confirm the relationship between different factors.

3. Data

Data set used was sourced from Kaggle, but was originally obtained through a survey of students in math class. The dataset includes 395 students from ages 15-22. All students involved in the survey went to either Gabriel Pereira or Mousinho sa Silveira (secondary schools). Information provided:

Research Questions

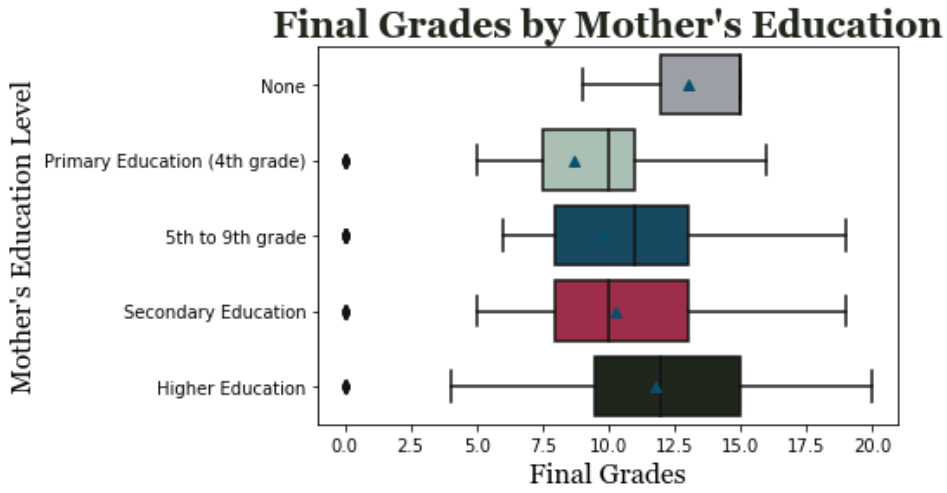
We used the following 7 questions to frame our research to come up with most applicable data.

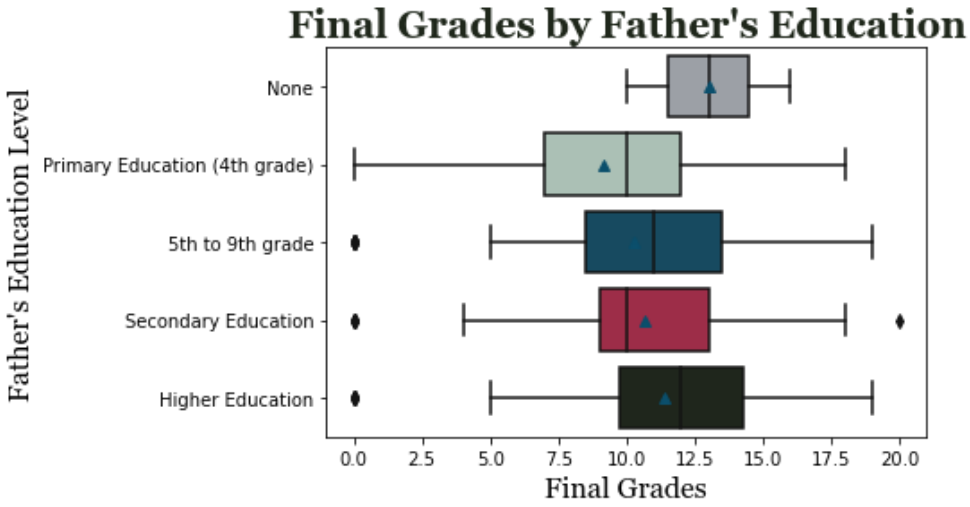
- Is student's educational success predictive based on parent's education?
- Does travel time or study time have an impact on the student's grade?
- Does internet access or romantic relationship have an impact on student's grades
- Does alcohol consumption differ based on the student's grades or gender?
- Are student grades predictive based on the amount of alcohol they consume?
- Is there correlation between demographics, grades or extracurriculars and alcohol consumption?
- Does alcohol consumption have an effect on the number of absences students have?
- Does alcohol consumption differ based upon student support or amount they study?
- Does travel time or study time have an impact on the student's grade?

4. Analyze Data

Final Grades by Parents' Education

We explored how parent's education level affected Final Grades of the students, as well as, comparing weekend and daily alcohol consumption by gender to see how the Final grade was impacted. The box plot graphs depict the mother's and father's education levels independently in correlation to final grades. We found that a greater proportion of mother's achieved higher education levels than father's and that parents' education level has a positive correlation to the academic performance of the student. The data suggests that you're more likely to have a higher final grade if your parent's attended higher education.

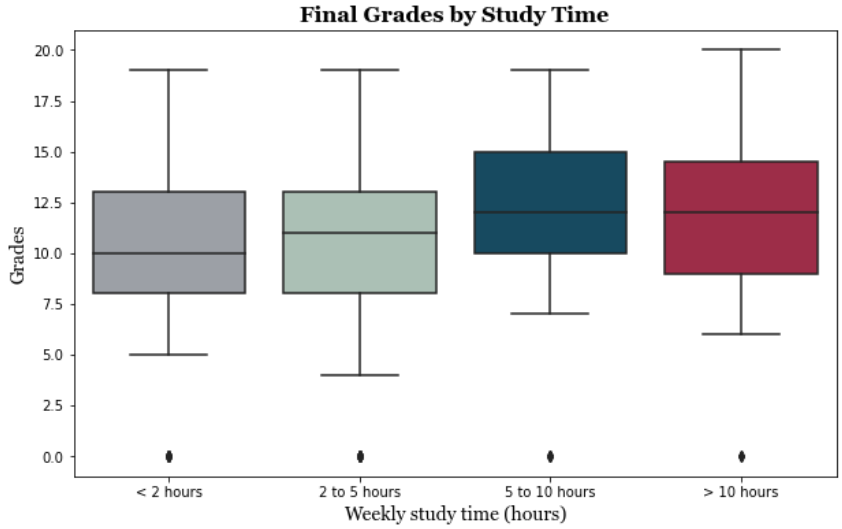




Grades based on Study Time

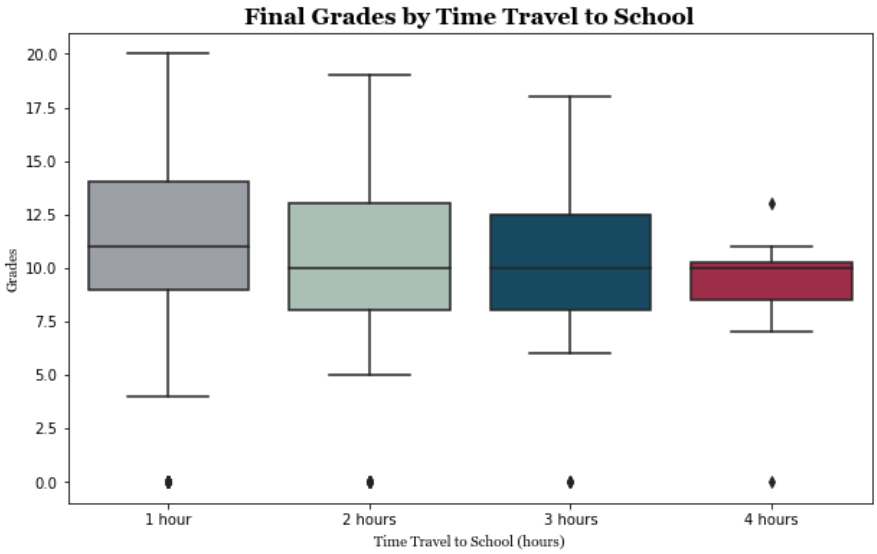
The boxplot graph shows that the lowest average grades are from the group who spend least amount of time studying (less than 2 hours). This totally makes sense because 2 hours studying per week is not enough to get good grades especially for mathematics. The students who spend between 2 to 5

hours to study per week have higher grades than group who spend less than 2 hours. But the difference is significant between students who spend more than 5 hours to study vs less than 5 hours studying per week. In general, the more time students studying, the better they perform academically. This finding suggests that students should invest enough time studying for good grades at school.

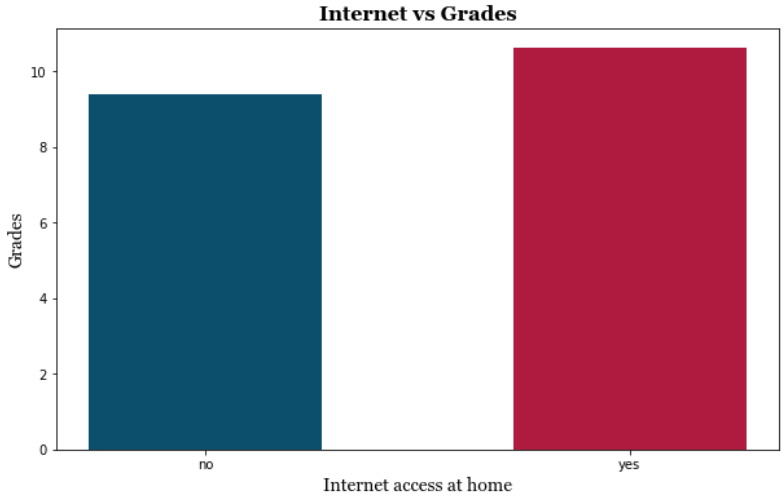


Grades based on Travel Time

The boxplot graph of final grades in corresponding to travel time to school shows the opposite relationship. Overall, the more time students spend to travel to school, the lower their grades are. The group of students who spend 4 hours to travel to school daily have the lowest grades. Students who spend only 1 hour or less have highest average final grades.

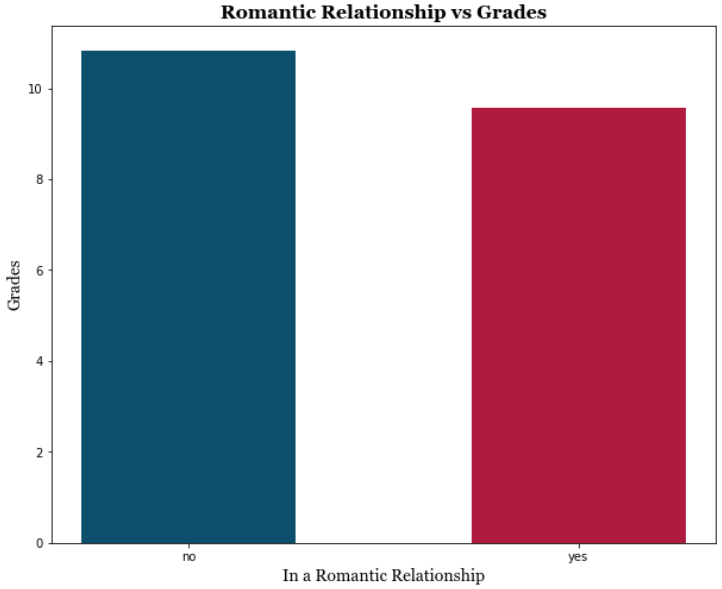


Grades based on Internet Accessibility

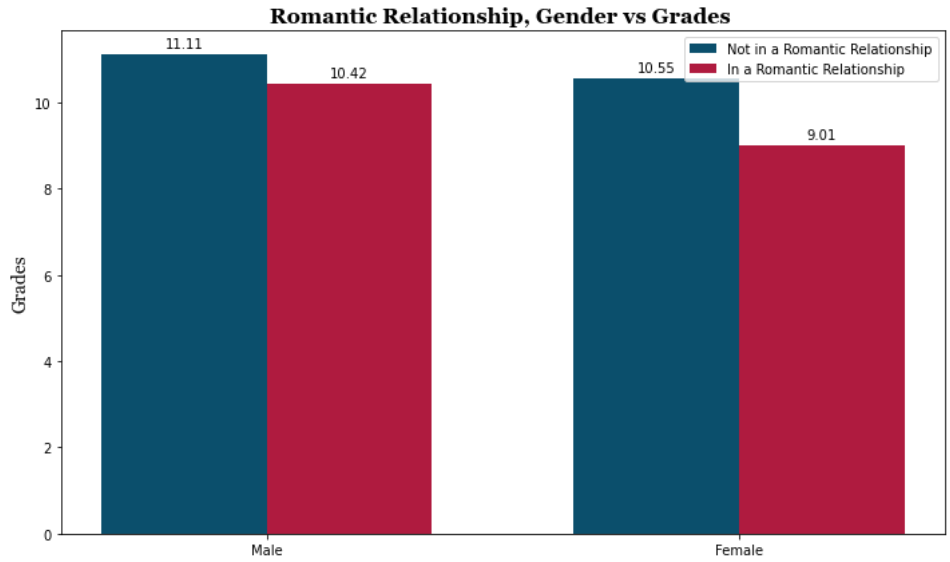


The above graph shows the average grade of student who have internet access at home is higher than the average grade of the student who do not have internet at home. This implies that internet has positive impact on student's academic performance.

Grades based on Romantic Relationship

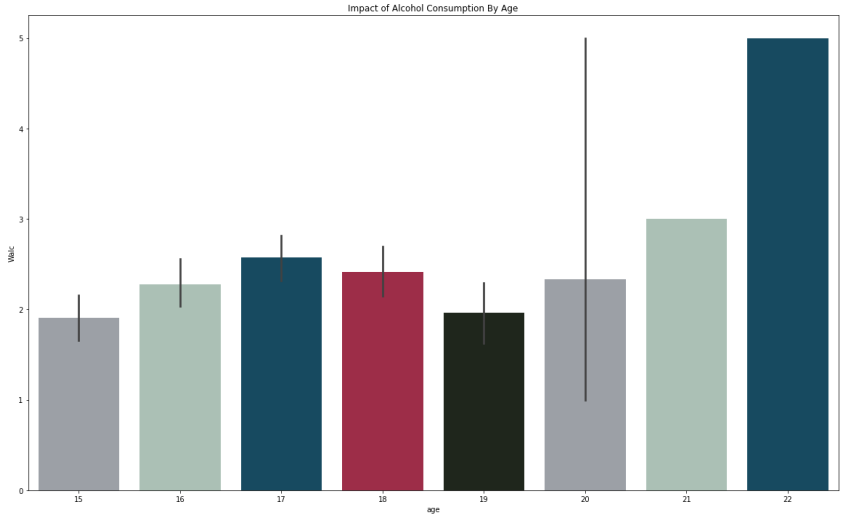


The graph above shows the difference in grades between the group of students who are in romantic relationship vs the group of students who are not in romantic relationship. We can see that those in romantic relationship has lower average grades then those do not. In other words, having boyfriend/girlfriend when in school distract students from studying! However, we decided to break this down more to see how the romantic relationship impacts differently on male vs female students and find interesting finding from the below graph below. Female students on average have lower grades than male students. But also, the having boyfriends affect female student's grades worse than male students having girlfriend. Their average grade drops dramatically from 10.55 to 9.1 while the male student's average grade only drops slightly from 11.11 to 10.42.



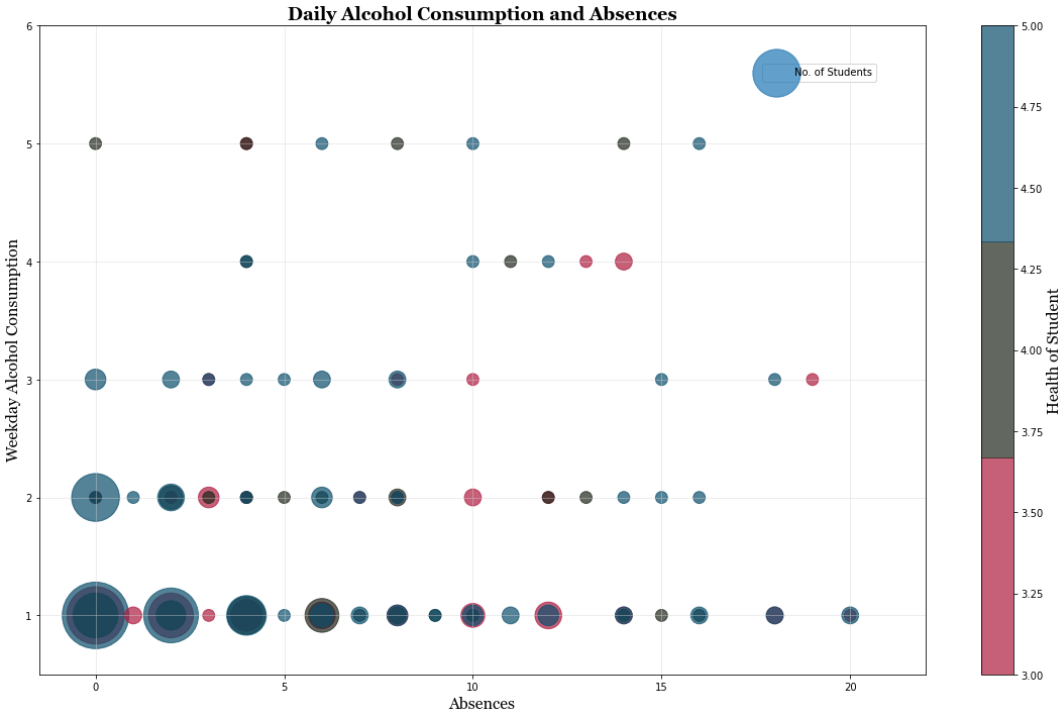
Alcohol consumption by age

We graphed the average alcohol consumption (weekday alcohol consumption) based on the student's age. In general, students tend to drink more when they get older



Alcohol consumption and absences

The average Alcohol consumption self-report on a scale of 1-5 was 1.481. When looking at alcohol consumption and absences we found that the health had a more significant impact on absences than alcohol consumption. For all students the average absences was 5.709. There is some promise in that students that reported that they consume 4 on a scale of 1-5 amount of alcohol during the week had on average more absences than the rest at 9.778. However, the sample size is way too small at only nine data points to determine a strong positive relationship, and the students that reported 5 receded a little back towards the mean at 8.666. Most importantly though is that a few of these students reported poor health which also could account for the increase in absences (Could be because of the hangovers!)

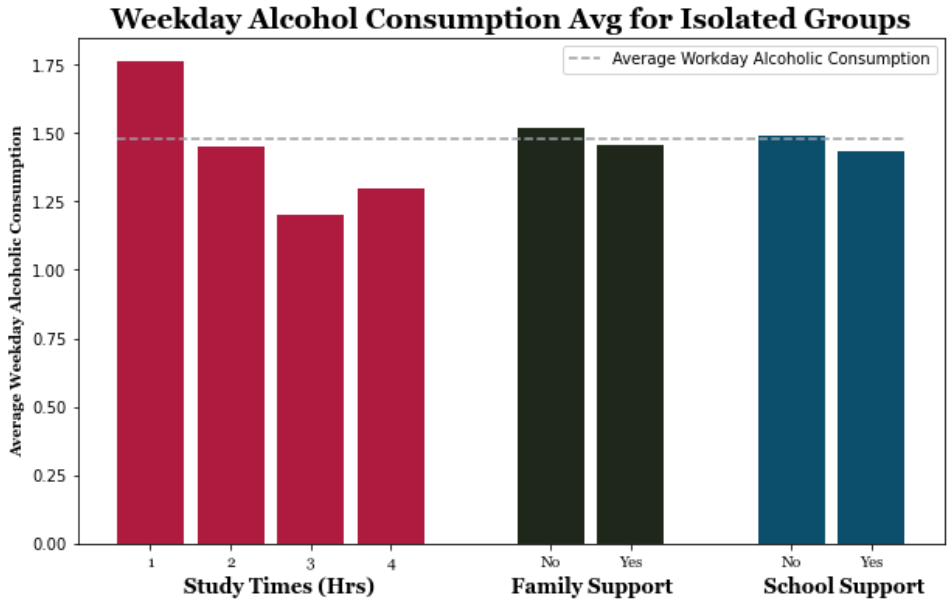


When taking these students out the sample size just becomes way too small to evaluate for meaningful numbers. For the students above 20 absences the average reported health was 3.133 compared to the average health of everyone at 3.554. Reported alcohol consumption only fell to 1.467 for that same group. So, we factored all the students who self-reported as having bad health (reported a

1 or 2). Even though the new data is controlled for health there was still no relationship between alcohol and absences. Alcohol reported consumption rose to 1.5 and absences dropped to 4.609

Alcohol consumption for different student demographics

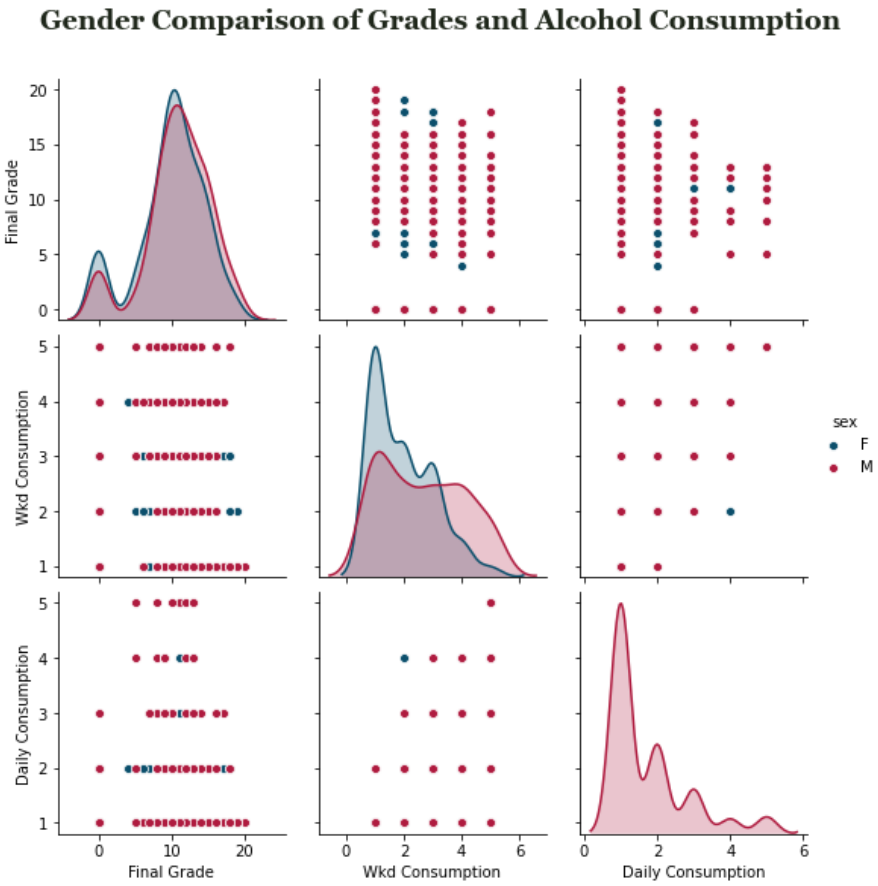
Study times for students reflected our initial assumption. The students that studied less consumed more alcohol. Zero to two hours reported an average of 1.762, two to five hours reported an average of 1.449, five to ten hours reported an average of 1.200, over ten hours reported an average of 1.296. Students that had student support reported 1.431 while those that did not have school support reported 1.488. Similarly, students that had parent support consumed 1.459 while those that did not reported 1.516. There is the slightest of decreases for reported alcohol consumption by students that had support but this data is well within the margin of error and prone to the bias that students with support could be more likely to underreport.



Gender Comparison of Grades and Alcohol Consumption

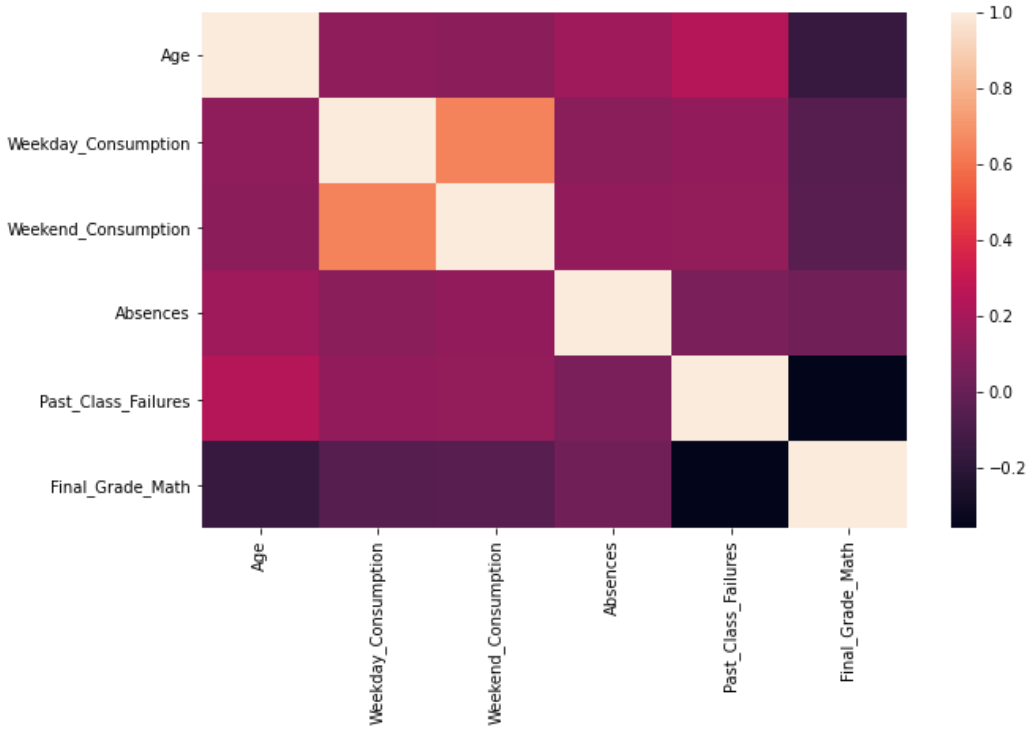
The pairplot below was used to show the relationship between a combination of variables. The variables we chose to analyze were final grades when daily alcohol consumption occurs, weekend only consumption occurs, and how that compares among male and females. The additional layer of comparison, gender, is considered a hue in a pairplot and is indicated by red being male and blue being female. The pairplot uses scatterplots for each pairing of the variables and typically histogram for the diagonal, with the additional hue variable it changed the histoplot to a layered kernel density estimate. We could have forced a traditional histogram but this method makes it easier compare distributions between males and females.

Overall, this graph suggests that students who drank less on the weekend and daily had better final grades. It's important to remember that this does not prove that one causes the other. We also see that females had a higher overall final grade then males.

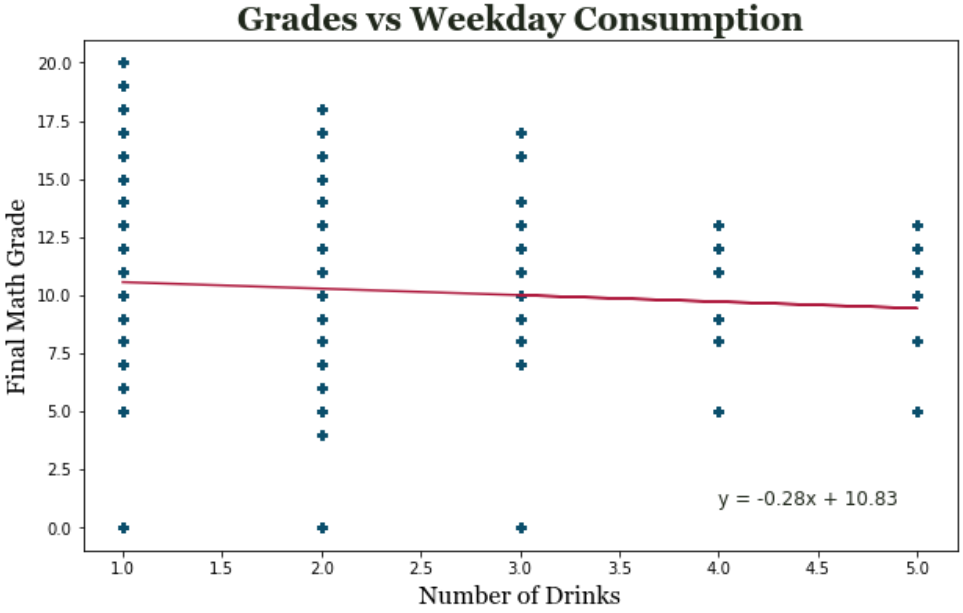


Correlation

The heatmap below shows the correlation among the quantitative data in our dataset. This map shows that there is not any strong correlation between any of these variables besides weekend alcohol consumption and weekday alcohol consumption. This explains the fact that if someone drink more during the week, they are more likely to drink more during the weekends and vice versa. We also can see that the alcohol consumption and absences are not much correlated with the final grades of students.



Relationship between Grades and Alcohol Consumption



The linear regression line between number of drinks and final grades is very flat suggested there is no correlation between these two.

We test our regression model with final math grades as dependent variable and independent variables include weekend alcohol consumption, weekday alcohol consumption, age, number of

absences and number of failed class in the past. Our OLS regression result gives very high P-values for alcohol consumption as well as for absences. Therefore, alcohol consumption and absences are not good predictive measure for how the student would perform in the class

OLS Regression Results

Dep. Variable:	Final_Grade_Math	**R-squared:**	0.141
Model:	OLS	**Adj. R-squared:**	0.130
Method:	Least Squares	**F-statistic:**	12.74
Date:	Wed, 11 Nov 2020	**Prob (F-statistic):**	1.76e-11
Time:	18:45:02	**Log-Likelihood:**	-1131.2
No. Observations:	395	**AIC:**	2274.
Df Residuals:	389	**BIC:**	2298.
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	16.3260	2.914	5.602	0.000	10.597	22.055
Weekend_Consumption	0.0013	0.221	0.006	0.995	-0.433	0.435
Weekday_Consumption	-0.0235	0.318	-0.074	0.941	-0.650	0.603
Age	-0.3240	0.177	-1.830	0.068	-0.672	0.024
Absences	0.0414	0.028	1.505	0.133	-0.013	0.096
Past_Class_Failures	-2.1098	0.301	-7.011	0.000	-2.701	-1.518

Omnibus:	27.418	**Durbin-Watson:**	2.034
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	31.108
Skew:	-0.660	**Prob(JB):**	1.76e-07
Kurtosis:	3.388	**Cond. No.**	247.

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

5. Conclusions

Based on our analysis we have found the following…

- That on average if a student's mother or father has more education that the student's grade was slightly better.
- On average, students who studied 5+ hours had the best math grades.
- The less the students have to travel, the better their grades are.
- Students who can access internet from home have higher grades

- Being in romantic relationship has negative impact on student's performance at school. This effect is stronger on female students than male students
- The older the students were, the higher their alcohol consumption was on the weekend.
- Number of drinks has no effect on student's absences
- On average, males had higher alcohol consumption, on weekdays and weekends, than females
- There is no correlation between math grades and alcohol consumption.
- The amount a student drinks during the week is found to not be a good predictive measure for what their final math grade is.

The results gathered by this report would be most useful for parents, teachers, and students because it provides insight into any underlying or non-direct causes of underage drinking. As an outcome of this research, more efficient teaching and parenting can be developed, improving the quality of education for students overall.

6. Limitation

Though we have found some useful insights from this dataset, there are some limitations we want to point out about our findings and suggestions for possible further research. First and for most, the sample data we used are from only two schools in Europe. The student's alcohol's consumption is very different in Europe comparing with others areas especially with US where the underaged drinking law is totally different. It will be biased to conclude from this sample to the whole population. Secondly, though the dataset has grades for math and Portuguese, we only took grades from math class and the affect of factors on mathematic performance could be different from the language performance. This will need further analyzing to conclude. Thirdly, the students may down play the amount they drank when being asked to answer the survey. This could lead to the bias in our findings. Students could have elected to not take the survey this also a possibility of bias in our dataset. Last but not least, the dataset does not include students who did not drink. This limits our analysis to find the difference between drinking vs not drinking students.

7. References

Dataset taken from Kaggle

https://www.kaggle.com/uciml/student-alcohol-consumption

Sample analysis from Kaggle: What is the secret of academic success?

https://www.kaggle.com/hely333/what-is-the-secret-of-academic-success