

**TASK**

**Exploratory Data Analysis on the Student Alcohol Consumption Data Set**

[](http://www.hyperiondev.com/portal/)

**Introduction**

3 Steps followed:

* Examining the dataset
* Narrowing down the columns for cleaning
* Analysis and visualizations

Libraries of the dataset were imported were numpy, pandas, matplotlib, plotly, scipy, missingno and seaborn. Dataset pertaining to students (alcohol) had to cleaned and analysed. The dataset was in a csv file and needed to be read in. Statistics and data types were done. There was no missing data for this dataset.

**DATA CLEANING**

# SUMMARY OF THE METHODS AND VISUALIZATIONS DONE DURING DATA CLEANING

**Methods**

* Previewed data – head and columns
* Dtypes
* Dataset info
* Missingno matrix
* Described the data
* Value counts

MISSING DATA

# ANY MISSING DATA? HOW DID YOU HANDLE IT

There weren’t any missing data in this data set, but if there was this is a few options how I would have handled it depending on the data types:

* Set the missing value to mean of normalized losses and convert the datatype to integer
* Remove records/data which have punctuations such as “?”
* Convert non-numeric data to null and convert the datatype from object to float
* Drop columns that have too many missing data points and insignificant data

DATA STORIES AND VISUALIZATIONS

**Descriptive Analysis**

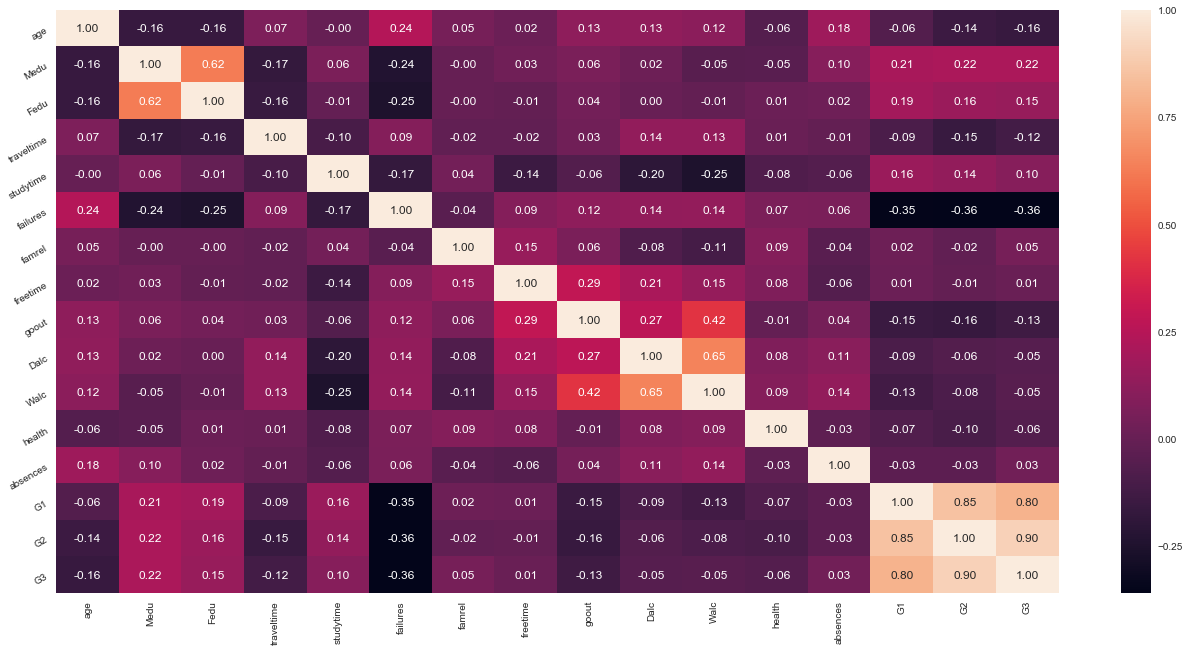
* From the descriptive statistic table, one can see there are 395 students between the ages of 15 to 22 with the average age being 16.7.
* The average level of education of mother's is 2.7 and father’s average education level is slightly lower of 2.5. That is 5th to secondary education for mother’s and slightly lower for fathers. Showing that mothers are slightly more educated than fathers.
* Students' travel time from home to school is between 1 hour to 4 hours, with an average travel time of 1.44 hours.
* Study time ranges from 1 hour to 4 hours, with an average of 2.03 hours.
* The average quality of family relationships is 3.94 signifying that there is a strong relationship with families on average.
* On average students have 3.2 hours free time and Go out with friends on average 3.11 hours. Which is similar to the average free time of students.
* In terms of alcohol consumption students consume alcohol during workdays on average of 1.48 and 2.29 during weekends. That is, low on weekdays and mediocre during weekends.
* Average current health status level is 3.55, that is, students are relatively healthy.

**Visualizations**

**Univariate and Bivariate Analysis were done**

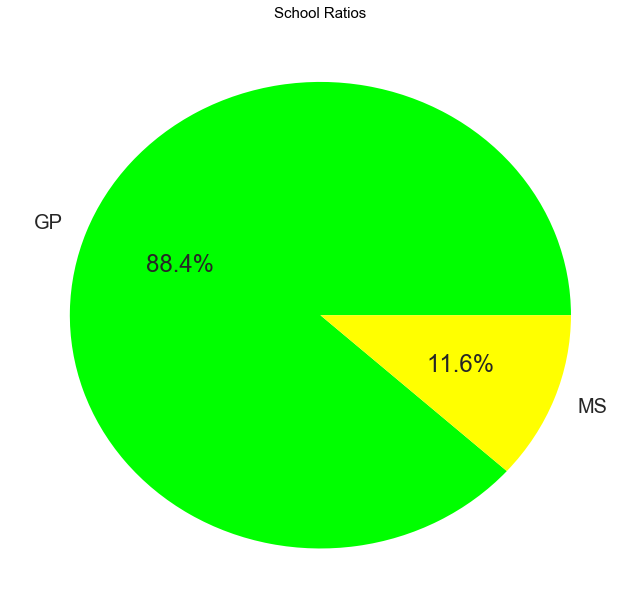
36 Graphs of:

* Correlations
* Bar Charts
* Barh Charts
* Pie
* Boxplot
* Swarmplots
* 3D – Scatter Plot
* Subplots
* Distribution Plots
* Lmplot seaborn



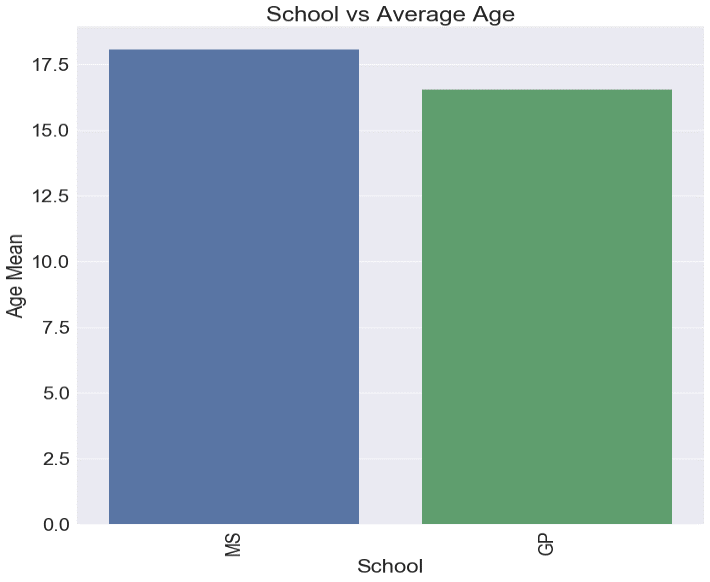
**Figure 1:** Correlation

From the correlation map in Figure 1, it can be seen there is a high positive correlation between mothers (Medu) education level and fathers (Fedu) education level. Also, there is a very high positive correlation between the grades of the students (G1, G2, G3). A moderate (below average) positive correlation exists between workdays (Dalc) and weekend (Walc) alcohol consumption. Also, a moderate positive correlation exists between going out (gouout) and weekend alcohol consumption (Walc) and there is moderate positive relation between workdays and weekend alcohol consumption.



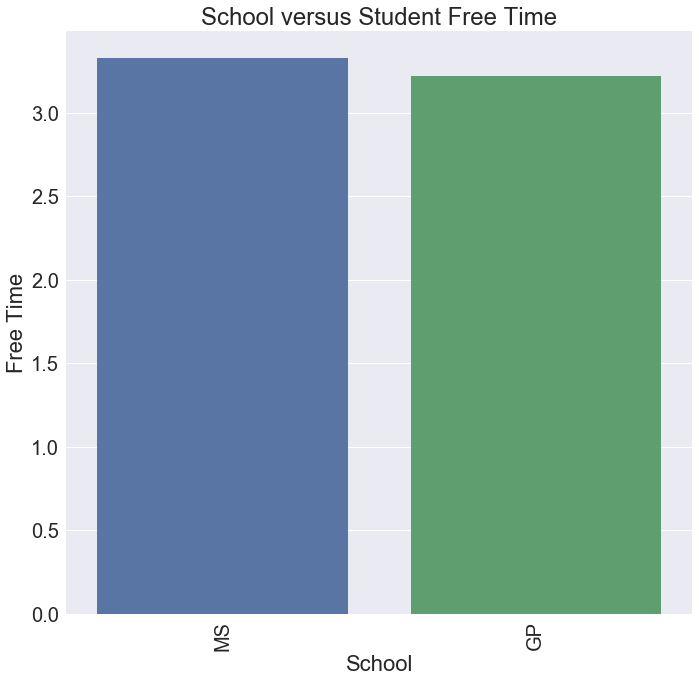
**Figure 2:** School Ratios

88.4% students go to GP (Gabriel Pereira) school and 11.6% go to MS (Mousinho da Silveira) school.



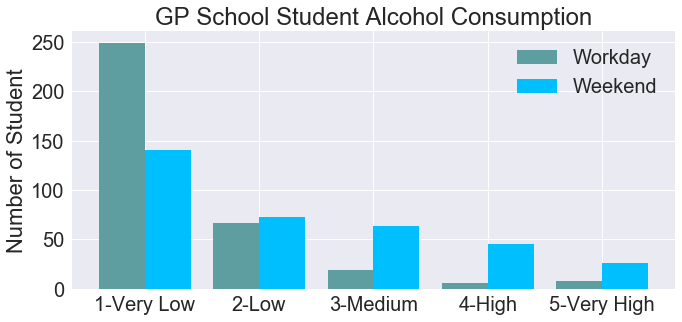
**Figure 3:** School vs Average Age

School MS has a student average age of 18.02 years and school GP has a student average age of 16.52 years.



**Figure 4:** School vs Student Free Time

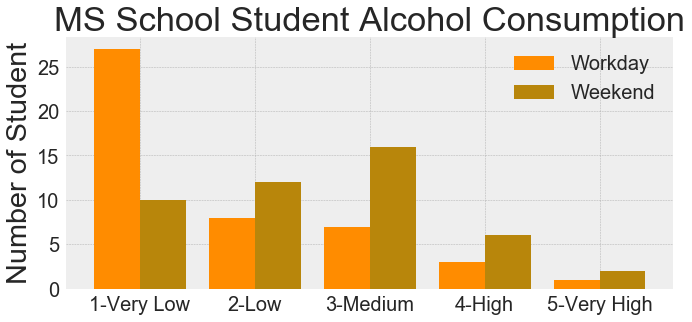
MS school students on average have about 3.32 hours free time and GP school students have an average of 3.22 hours free time.



**Figure 5:** GP School Student Alcohol Consumption

There are 349 students from GP school and their workday alcohol consumption is on average 1.44, which is moderately low and their weekend end consumption on average is 2.26 which is moderately medium.

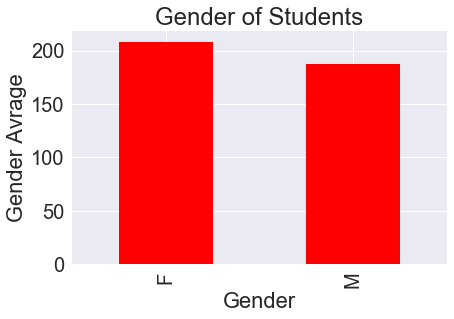
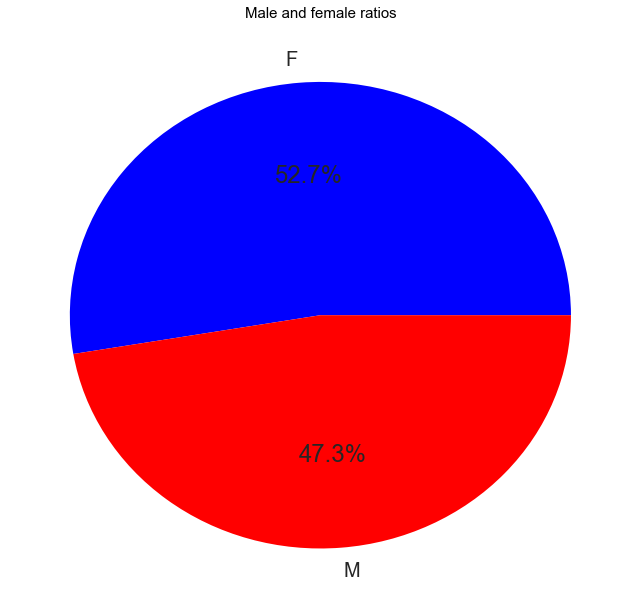
Students from school GP have a low (2-low) to very high (5-Very high) alcohol consumption in weekends in comparison to workdays.



**Figure 6:** MS School Student Alcohol Consumption

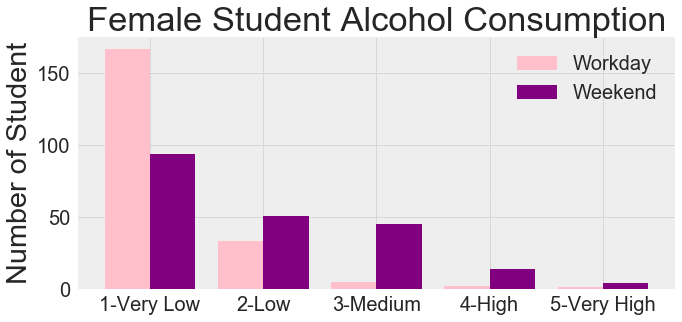
There are 46 students from MS school and their workday alcohol consumption is on average 1.76, which is moderately low and their weekend end consumption on average is 2.52 which is moderately medium.

Students from school MS have a low (2-low) to very high (5-Very high) alcohol consumption in weekends in comparison to workdays.

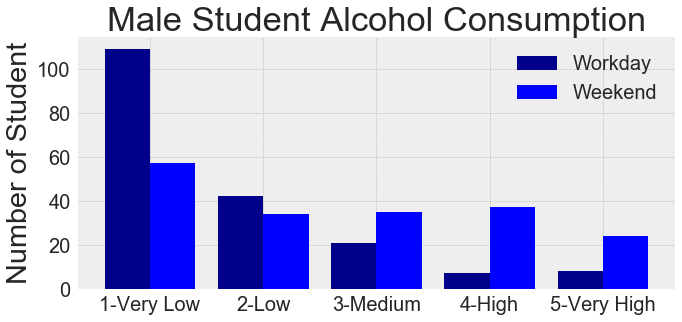
**Figure 7:** Gender of Students

There are more female students than male students in both schools combined. 52.7% make up female students and 47.3% make up male students.



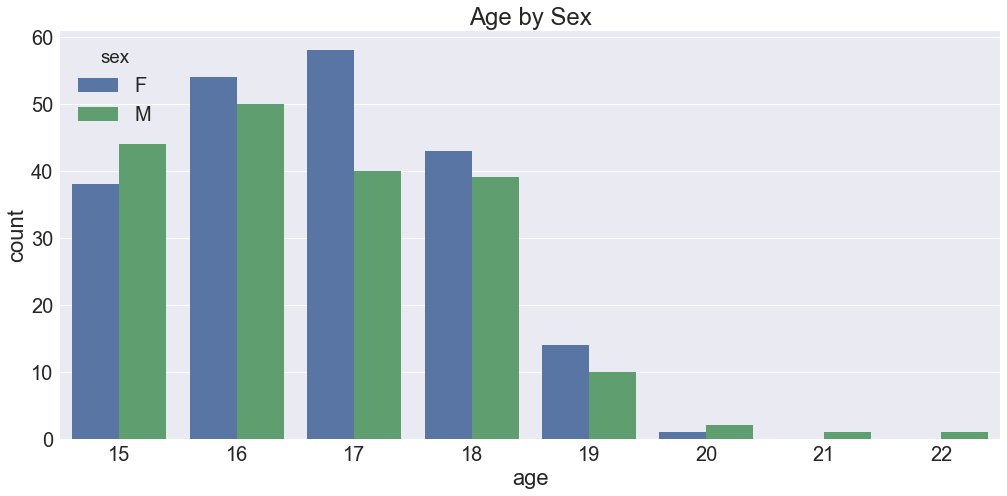
**Figure 8:** Female Student Alcohol Consumption

On the working days, female student’s alcohol consumption level is generally very low, however on the weekends they consumes more alcohol from medium to high level.



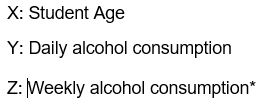
**Figure 9:** Male Student Alcohol Consumption

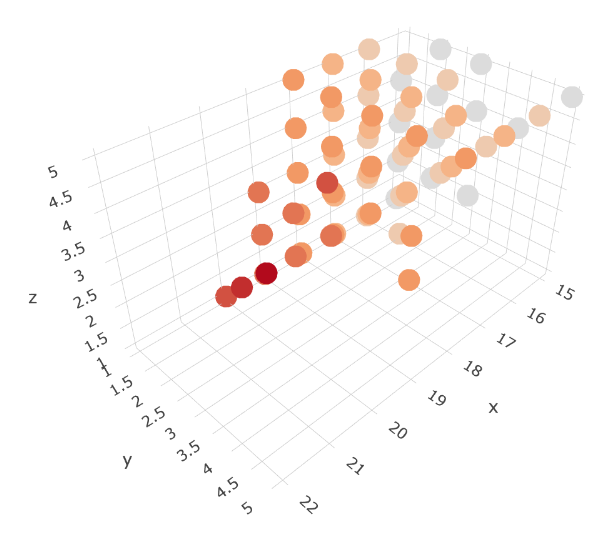
The male students consume much more alcohol than the female students. The level of alcohol consumption of male students is much higher on other working days (2-Low) and especially high on weekends.

**

**Figure 10:** Students Age by Sex

Between the ages of 16-19 majority are females, at age 15 there are more males than females and the same holds true from ages 20 to 22 years. Collectively between males and females the average age for students from both schools is 16.69 years old.





**Figure 11:** Alcohol Consumption by Student Age

(P.s this looks better in the notebook 😊)

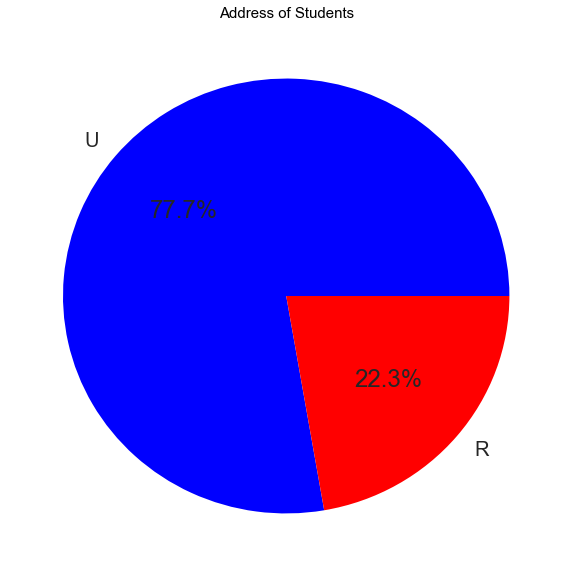
From Figure 11 students of all age groups consume more alcohol on weekends than on workdays. Alcohol consumption of 15-year-olds students is generally very low, but few students consume alcohol moderate and high levels on weekends. Alcohol consumption of 16-year-old students is low, but on weekends a handful of students consume more alcohol from the medium to very high levels. 17-year-old students have a low level of alcohol consumption on workdays, but the alcohol consumption level is increasing compared to the than other age groups, the same is seen for the 18-year-old students. The 19 to 22-year-old students, alcohol level consumed on workdays and weekends decreased compared to other age groups.



**Figure 12:** Workday and Weekend Alcohol Consumption and Going out with friends

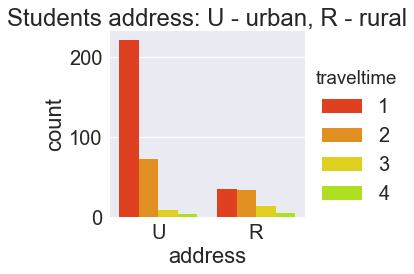
\*N. B The Q1(lower quartile) does not appear because the alcohol consumed on a workday is the same as the median and the lower quartile.

Weekend alcohol consumption is higher than workday and the going out with friends is higher. Therefore, in essence one can say that most students when going out with friends (which is usually on weekends) consume the most of alcohol during the weekends as opposed to the workdays.



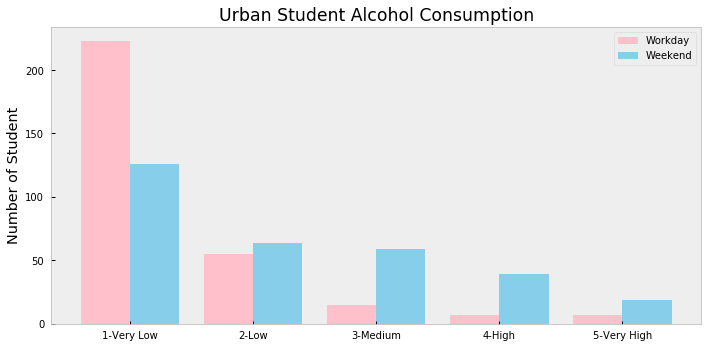
**Figure 13:** Address of Students

77.7% of students live in the urban district whereas 22.3% travel from rural areas to get to school.



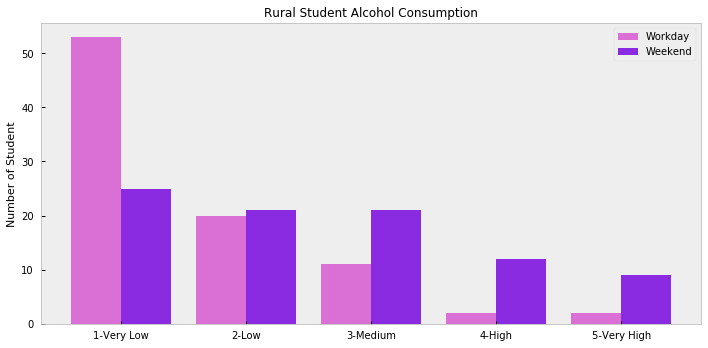
**Figure 14:** Count ofStudents Address by Travel Time

Most students take a travel time of 1 to 2 hours for both urban and rural areas. However, there are more students from the rural areas that take 3 to 4 hours to get to school in comparison to students from urban areas.



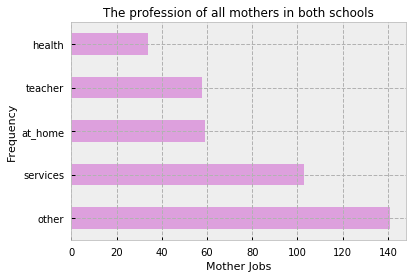
**Figure 15:** Urban Students Alcohol Consumption

There are 307 students traveling from urban areas with an average alcohol consumption level of 1.56 on workdays and 2.31 on weekends. Most students consume alcohol in the weekend from 2-Low to 5-Very High.



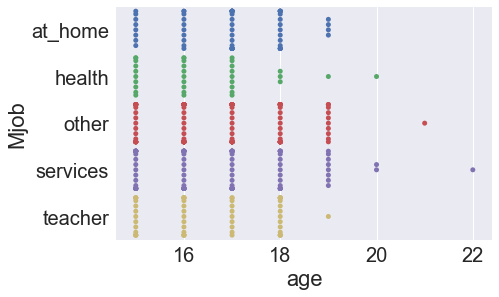
**Figure 16:** Rural Students Alcohol Consumption

88 students live in rural area. Their average alcohol consumption level is 1.63 on workdays and 2.53 on weekends. Most students consume alcohol in the weekend from 2-Low to 5-Very High.



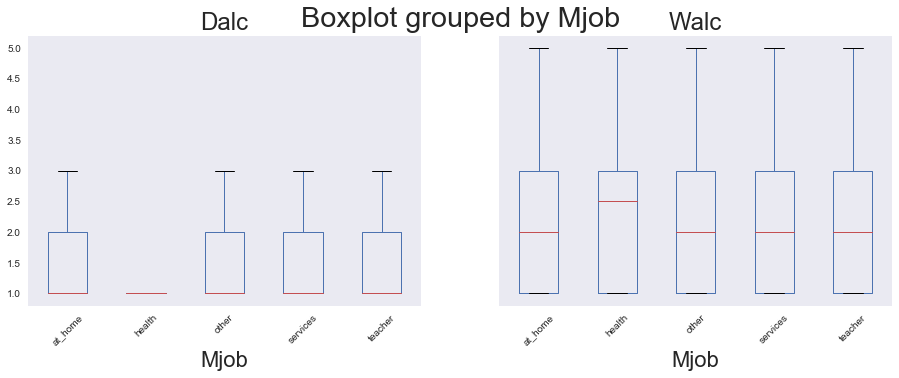
**Figure 17:** Professions of all mothers

Most of the mother’s professions is in the other industry sectors,141. 103 of mother’s work in services, 59 stay at home, 58 are teachers and 34 are in the health sector.



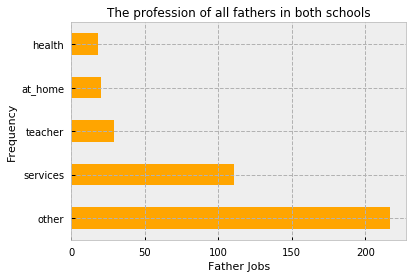
**Figure 18:** Mothers Profession by Student Age

Most of the students whose mothers have professions are between the ages of 15 and 19 years old and starts decreasing with increasing age. Majority of mothers’ work in the services sector across all ages (15-22).



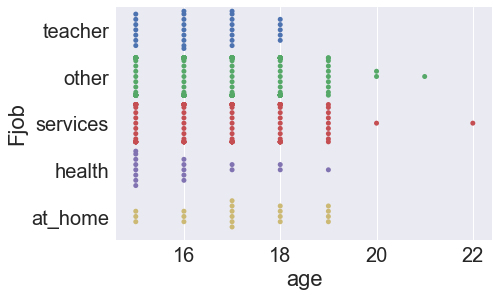
**Figure 19:** Mothers Profession by Student Alcohol Consumption

From Figure 19 it can be seen that student’s alcohol consumption is similar to that of the mother’s professions. On workdays the alcohol consumption by students are same to the mothers professions, students whose mothers working in the health sector consume very low levels of alcohol during the week, interestingly for students whose mothers that work in the health sector have a high median for alcohol consumption compared to the rest of the mothers professions and alcohol consumption by students during the weekends.



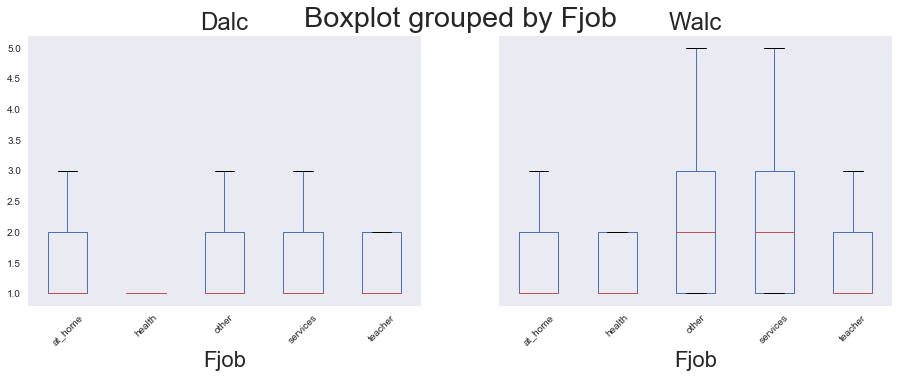
**Figure 20:** Professions of all fathers

Most of the father’s professions is in the other industry sectors, 217. 111 of fathers work in services, 29 are teachers, 20 are at home and 18 are in the health sector.



**Figure 21:** Fathers Profession by Student Age

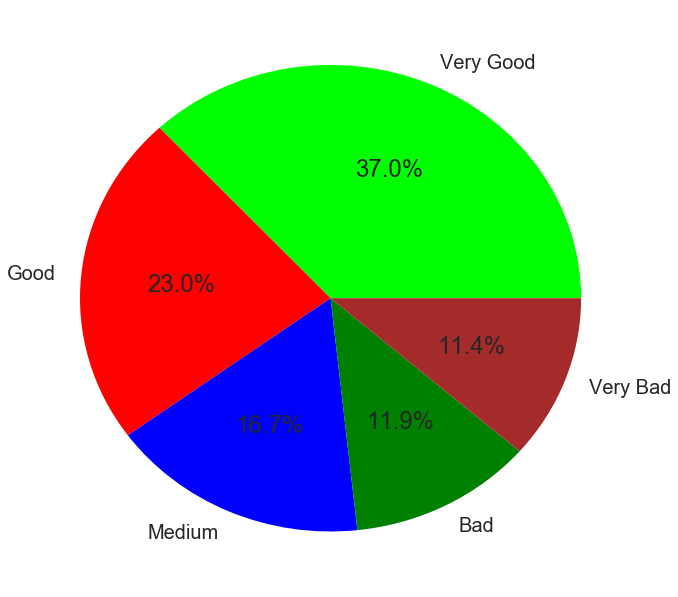
Most of the students whose fathers have professions are between the ages of 15 and 19 years old and starts decreasing with increasing age. Majority of fathers work in the services sector across all ages (15-22).



**Figure 22:** Fathers Profession by Student Alcohol Consumption

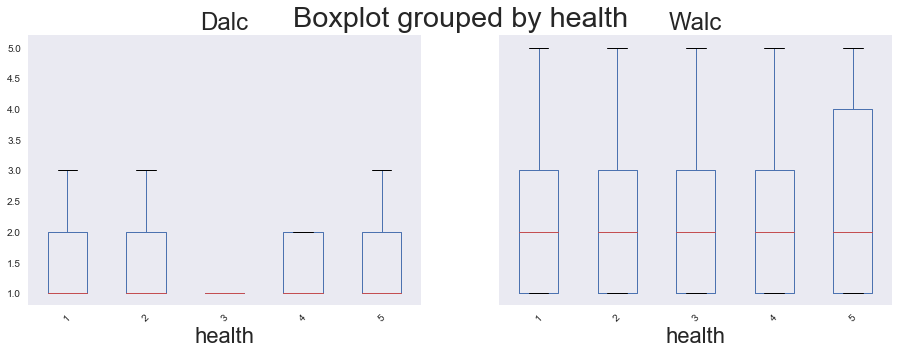
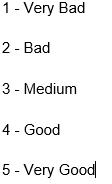
During the working days, students whose fathers are at home, work in other industry sectors and are in services consumes alcohol at most a level of 3 and a level of 2(low) for students whose fathers are teachers.

During the weekends, students whose fathers work in the services and other industry sectors consume a lot more alcohol than fathers in the other professions. The health profession is still low.



**Figure 23:** Health Status of Students

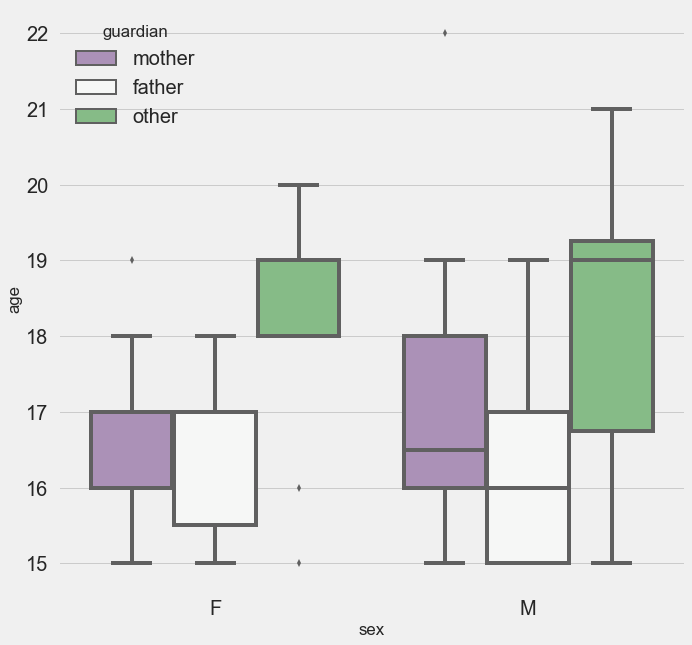
37% of students are in very good health, 23% good health, 16.7% in medium health, 11.9% in bad health and 11.4% in very bad health conditions. Students have good health overall.

**Figure 24:** Health Status of Students by Alcohol Consumption

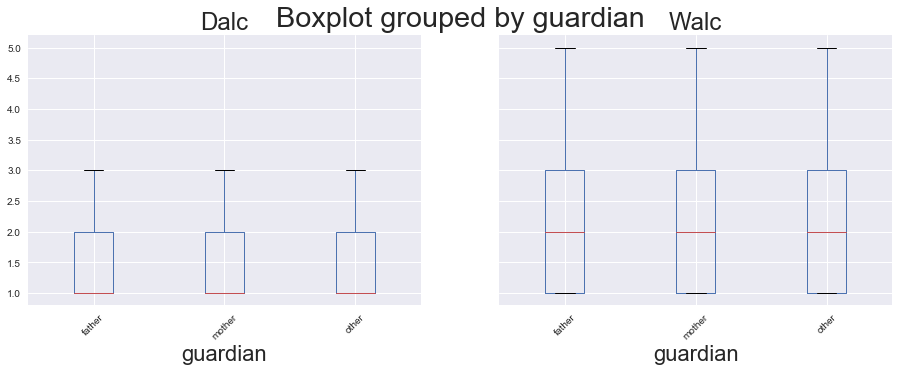
During the workday week students of all health status ranging from very bad(1) to very good(5) health consumes low amounts of alcohol with students with medium health consuming the least.

Whereas, during the weekend students with very good health consume the most amount of alcohol with Quartile 3 being the highest. Students ranging from very bad to good consume the same level of alcohol during the weekends (Medians of 2).



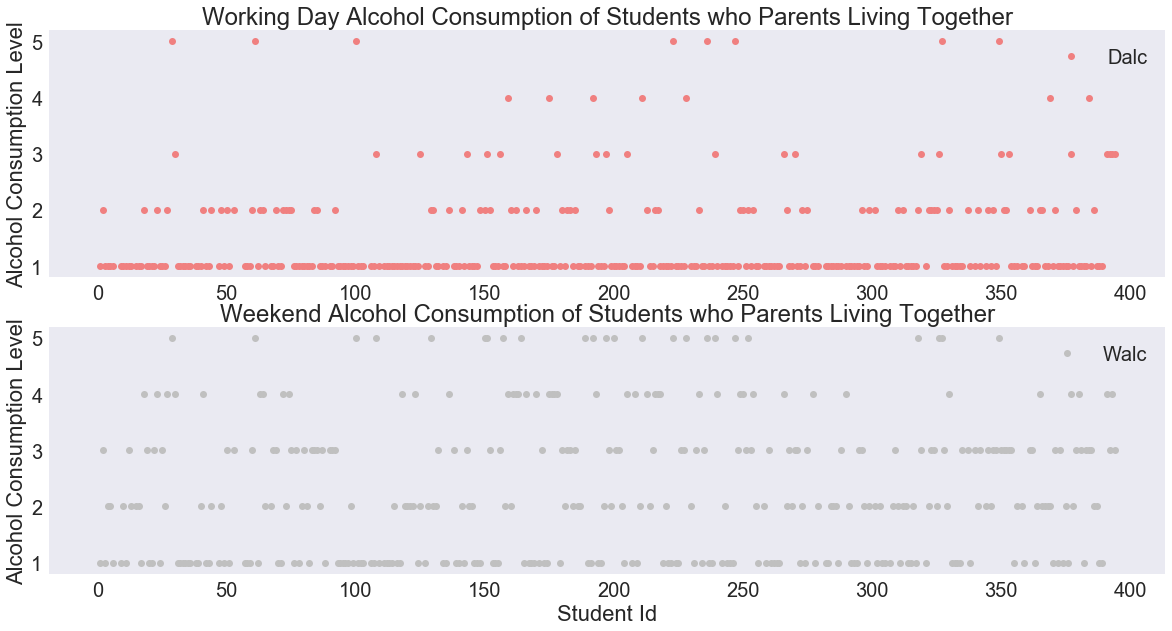
**Figure 25:** Students Guardian

Female students (18-20 years) have a lower guardian support structure than male students (15-21 years of age). Male students have more of a mother presence as opposed to a father presence in their lives.



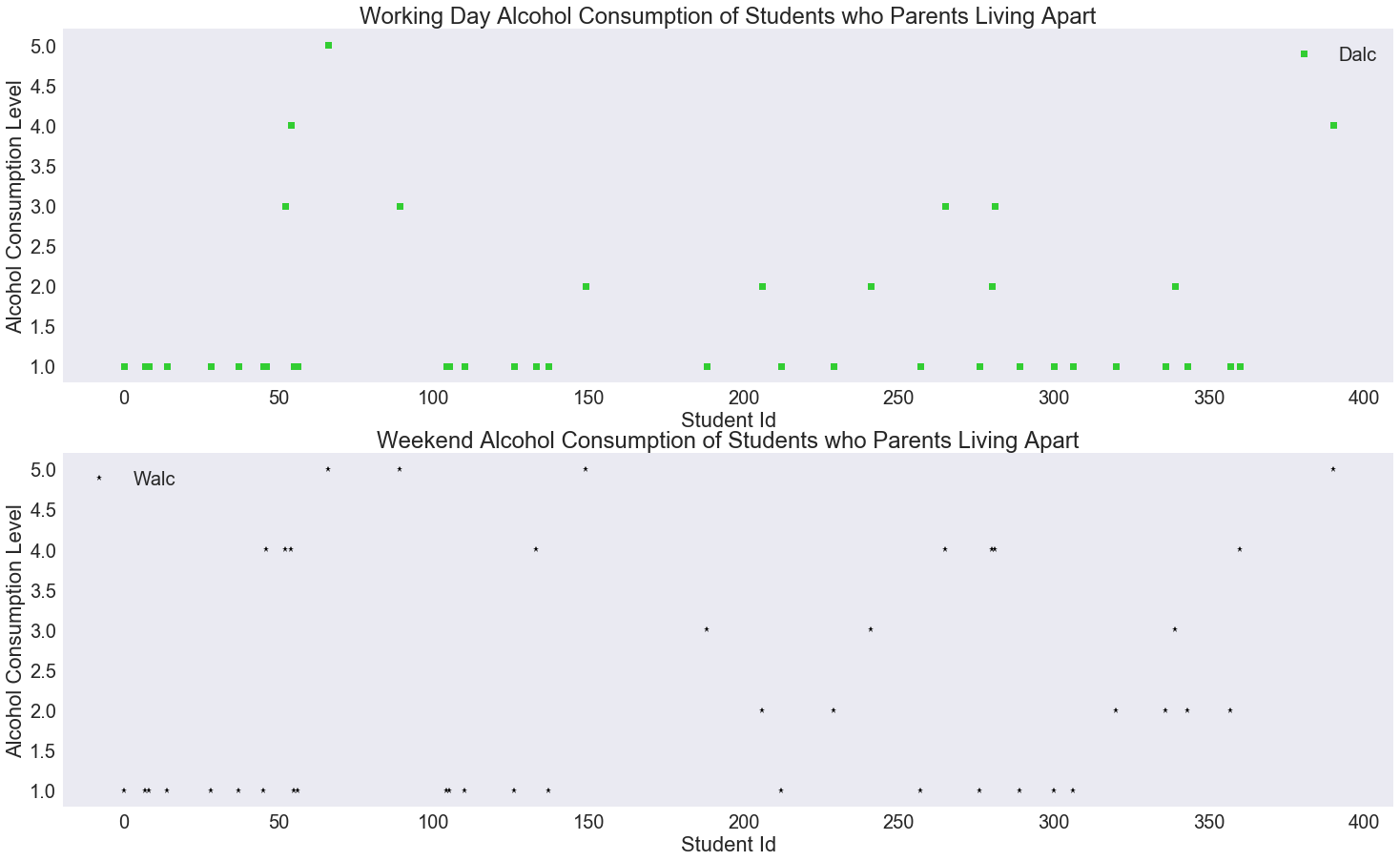
**Figure 26:** Guardian by Students Alcohol Consumption

The amount of alcohol consumed during the workday week and weekend is the same for student with mother, fathers or other(guardians).



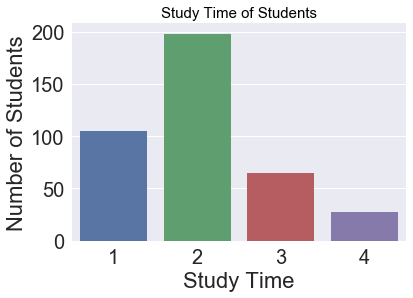
**Figure 27: Students Alcohol Consumption of Parents Living Together**

The average alcohol consumption of students with parents that live together is 1.47 on the workday week and 2.29 on the weekend. This can be seen with the density of the graphs especially during weekend alcohol consumption 😊.



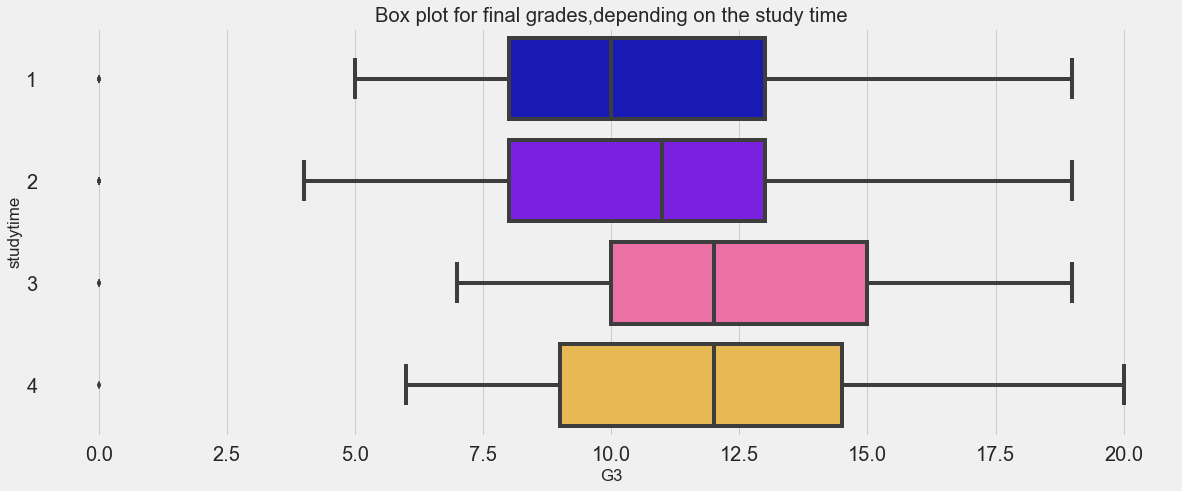
**Figure 28: Students Alcohol Consumption of Parents Living Apart**

The average alcohol consumption of students with parents that live together is 1.56 on the workday week and 2.26 on the weekend. Interestingly, the workday alcohol consumption of students with parents living apart is higher than parents living together – sense some issues (shame).



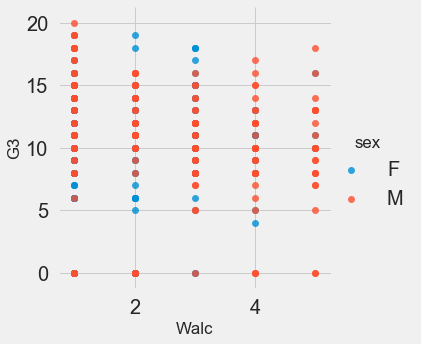
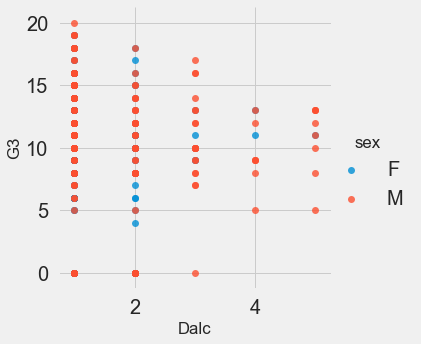
**Figure 29:** Study Time of Students

Most students study for 2 hours at a time, followed by 1 and 3 hours. Lastly a very small number of student’s study for 4 hours.

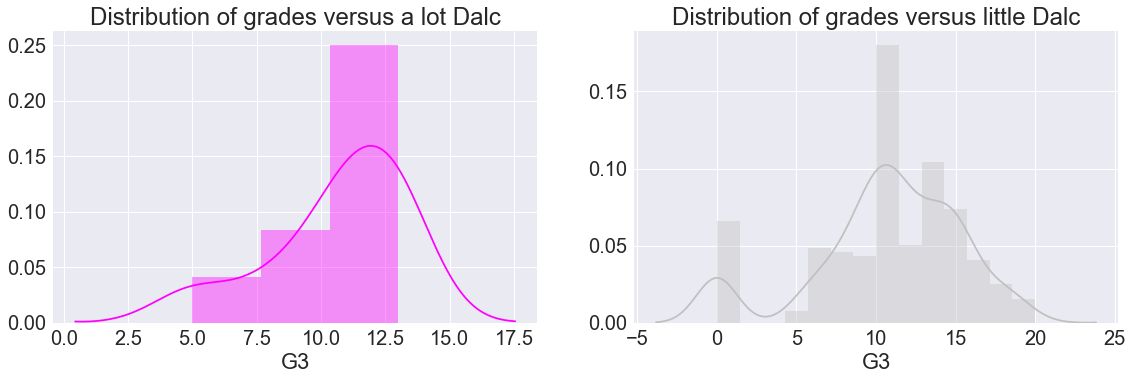


**Figure 30:** Final Grade (G3) by Study Time

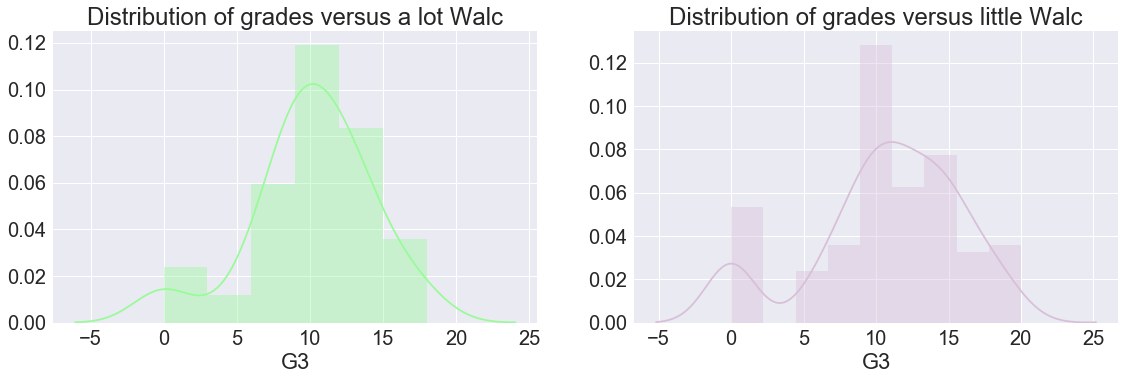
Interesting to note the students who studied the longest in terms of time had higher final grades than students who studied for the least amount of time.



**Figure 31:** Final Grade (G3) by Alcohol Consumption

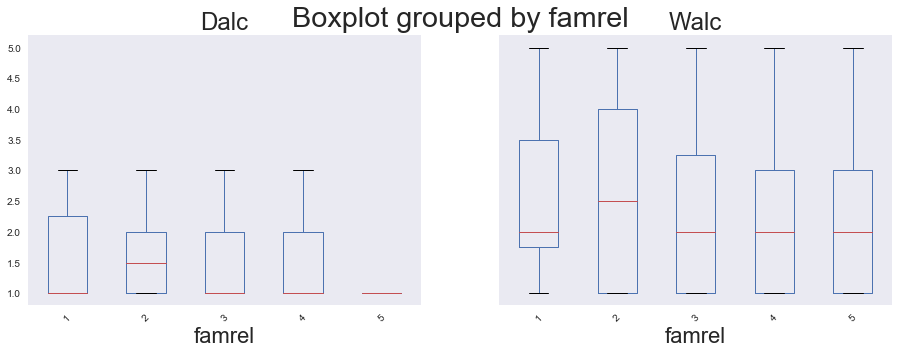
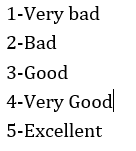


**Figure 32:** Distribution of Grades vs Dalc



**Figure 33:** Distribution of Grades vs Walc

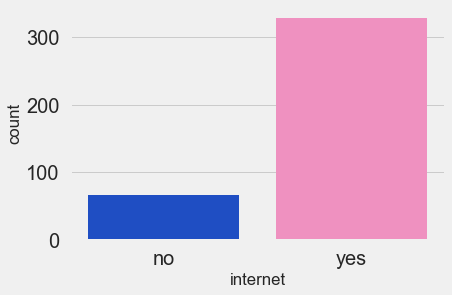
As seen from Figure 31 to Figure 33 alcohol consumption on workdays and weekends don’t seem to be an influencing factor to low final grades (G3).

**Figure 34:** Students Alcohol Consumption based on Family Relationship

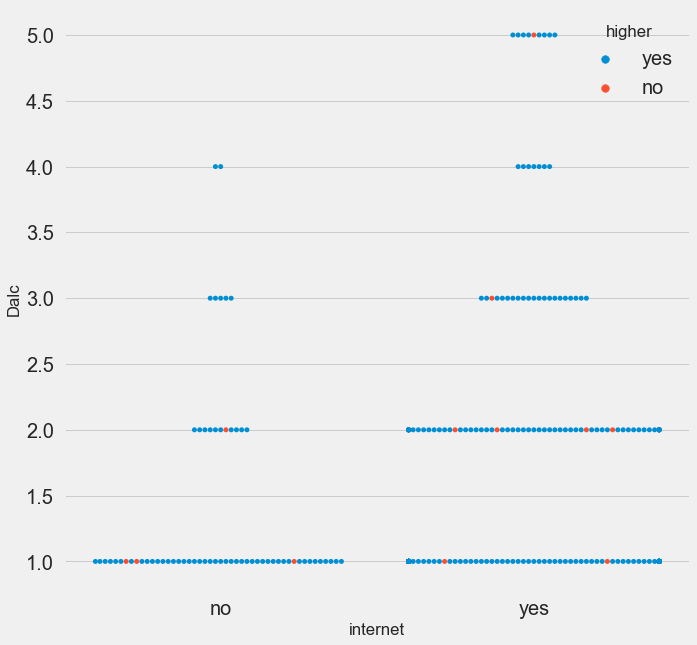
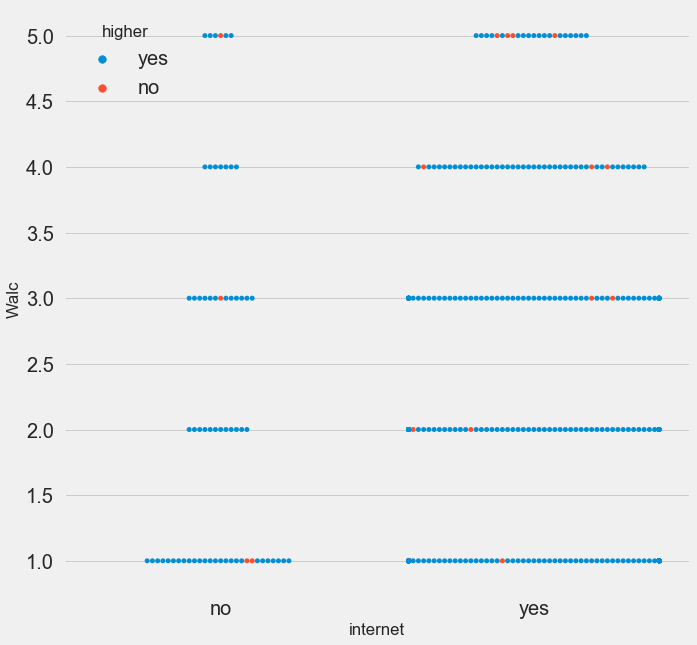
During the workday week students that have poor to good family relationships consume low levels of alcohol, but students who have bad family relationships consume slightly more during the week.

Students during weekends despite their family relationships consume alcohol at a level of 2, but again those students that have ‘bad (2)’ family relationships consume much more alcohol.



**Figure 35:** Students with and without Internet Access

329 students have access to the internet and 66 students do not.

**Figure 36:** Student Alcohol Consumption with and Without Internet

For both workday and weekends students with Internet access at home consume more alcohol than students that don’t have internet access at home.

**THIS REPORT WAS WRITTEN BY : Kamil Sewsunker**

**Appendix**

# 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)

# addressS-tudent'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' - living 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)

# G1-First period grade (numeric: from 0 to 20)

# G2-Second period grade (numeric: from 0 to 20)

# G3-Final grade (numeric: from 0 to 20, output target)

