Social, gender and study data from secondary school students The data was obtained in a survey of students math and Portuguese language courses in secondary school. It contains a lot of interesting social, gender and study information about students. You can use it for some EDA or try to predict students ﬁnal grade. The problem here is to determine the impact of alcohol and other variables on the grades of students and also investigate any other variables which may impact the grades. My client would ideally be the school authorities who could capture this information when doing entry screening of students. The captured demographics could help them decide which kind of students could be most successful and plan for those, based on the model, who may not be and plan remedial actions eg extra classes etc I assume the data was captured via student survey of math and Portuguese students and contains information on relationships, drinking patterns, demographics and course grades.

The important fields here are mainly demographics and the weekly and alcohol consumption levels of the students. There are 3 different grades (1st, 2nd and final grades). I only used the final grades but I have included average grades for analysis. Limitations are mainly data sizes as I had to merge both datasets and obtained 959 unique observations. The data was clean with no missing values and the only data cleaning I had to do was merge the data, create a pass/fail grade and also convert that to a factor.

Some of my initial observations after EDA were;

1. slight decrease in grades as weekly alcohol consumption increases.
2. Grades are fairly constant between ages 15-18 but slightly dip from then. Maybe there are other factors which come into play from that age (dating, alcohol, late nights?)
3. School Support did not impact median scores however those without school support seem to have scores skewed to the right (higher scores) while those with school support seem to have scores skewed to the left ( lower scores). This tells me that school support had a negative impact on grades distribution.
4. There was no perceived impact of romance of average grades within this dataset.
5. Absences increased as daily alcohol consumption increases