

# EDAV fall 2018 Final Project (hnt2107\_nm3086\_sdt2134)

## Student Performance Data Analysis

---

### 1. Introduction

Grades are primary indicators of a student's academic performance. While there may not be a correlation between grades and a student's intelligence, there are definitely multiple external factors which contribute to a student's grade. We wanted to analyze the effect of factors like parents' education, alcohol consumption, free time, etc. on a student's grades.

### 2. Description of Datasets

We chose “*Student Performance Data Set*” which is available on UC Irvine Machine Learning Repository. The dataset was generated by surveying secondary school students studying Mathematics and Portuguese in Portugal. It contains a lot of interesting social, demographic, psychographic and academic information.

(Source) <https://archive.ics.uci.edu/ml/datasets/student+performance>

We categorized 30 data attributes in the dataset into 6 different groups for conducting analysis efficiently. We also performed several data quality checks and identified some issue. For more details, please find the report ([link](#)).

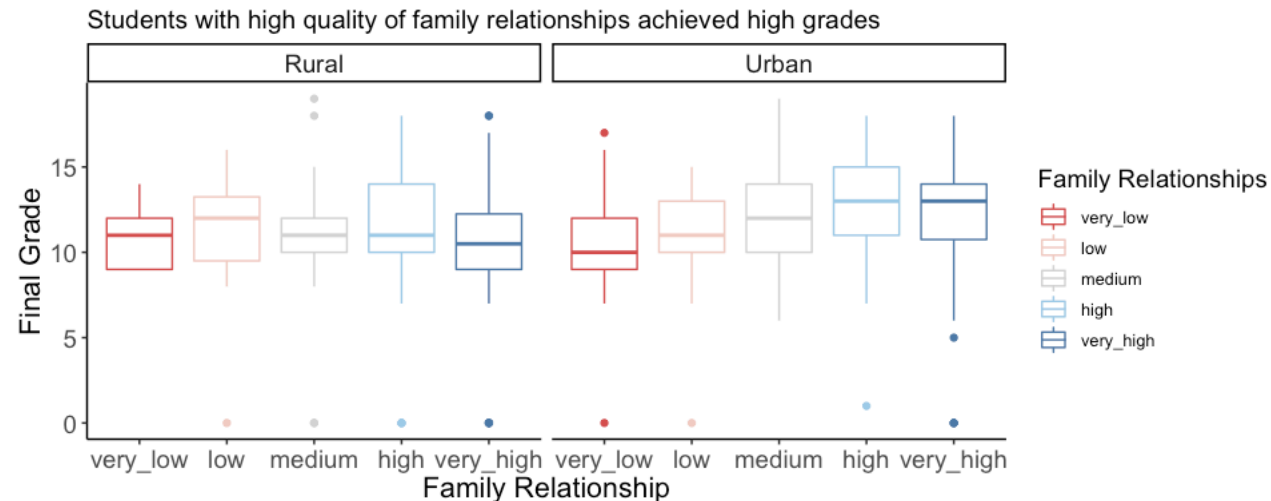
Group	Variables
Home environment	famrel (family relation), internet (internet access), studytime, traveltime, Pstatus (live with parent or not), address (rural or urban), famsize (family size)
Social	freetime, gout, Dalc (weekday alcohol consumption), Walc (weekend alcohol consumption), romantic (with a romantic relationship)
Parents	Medu (Mother education), Fedu (Father education), Mjob (Mother job), Fjob (Father job), guardian (student's guardian)
Profile	sex, age, health (students' health condition)
Academics	G1 (1st period grades), G2 (2nd period grades), G3 (final grades), absences, school (student's school), failures (# of past class failures), activities (extracurricular activities)
Others	reason (reason to choose school), higher (wants to take higher education), schoolsup (school support), famsup (family support), paid (paid classes), nursery (attended nursery school)

# EDAV fall 2018 Final Project (hnt2107\_nm3086\_sdt2134)

## Student Performance Data Analysis

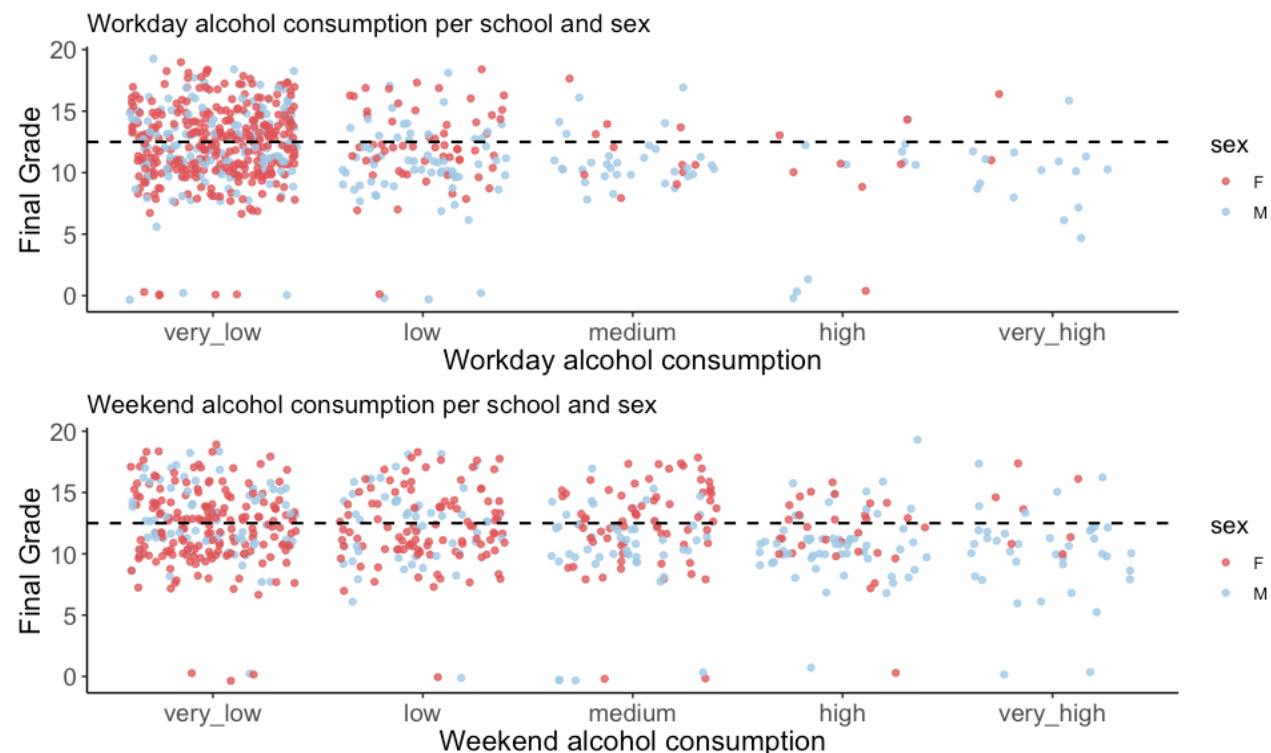
### 3. Key insights and observations<sup>1</sup>

#### 3-1. Home environment group



- Urban students raised in high-quality family relationships earned better grades.
- Urban students<sup>2</sup> used internet more effectively and earned higher grades.
- Surprisingly, male urban students whose parents are separated performed better.
- There were no associations between grades and family size.

#### 3-2. Social Group



<sup>1</sup> It is only association, not causal inference

<sup>2</sup> There are more urban students than rural students in the report. Interestingly, 76% students in urban area study at GP while only 40% of students in rural area study there. GP also has higher median grade than MS.

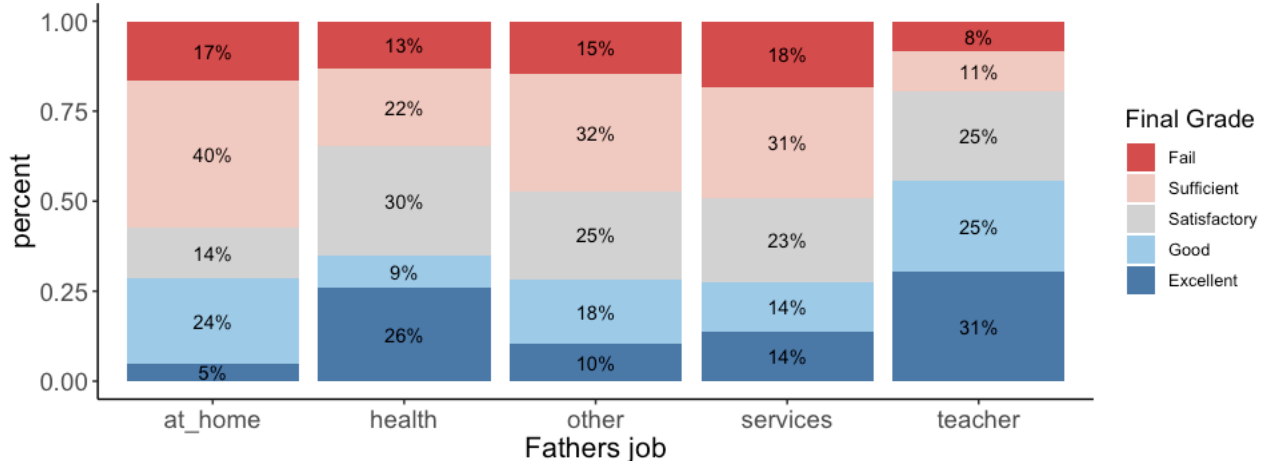
# EDAV fall 2018 Final Project (hnt2107\_nm3086\_sdt2134)

## Student Performance Data Analysis

- While moderate to low consumption of alcohol during weekends may not affect the final grade, students consuming alcohol during weekday or high levels of alcohol during the weekend received lower grades.
- Students performing poorly reported higher free time, such students spent a large chunk of their time on extracurricular activities
- Some other observations for this group:
  - Students with a lot of free time frequently went out with friends.
  - Students taking extra paid courses reported lower free time.
  - Several students with very low free time had subpar health.
  - Female students frequently reported lesser free time.

### 3-3. Parents Group

Students whose father is a teacher performs well in grades



- Students whose parents worked as teachers tended to outperform their peers in terms of the final grade.
- Parents' education level is positively correlated with academic performance.
- Students whose parents did not pursue higher education tended to have poor final grades compared to students whose parents were well educated.

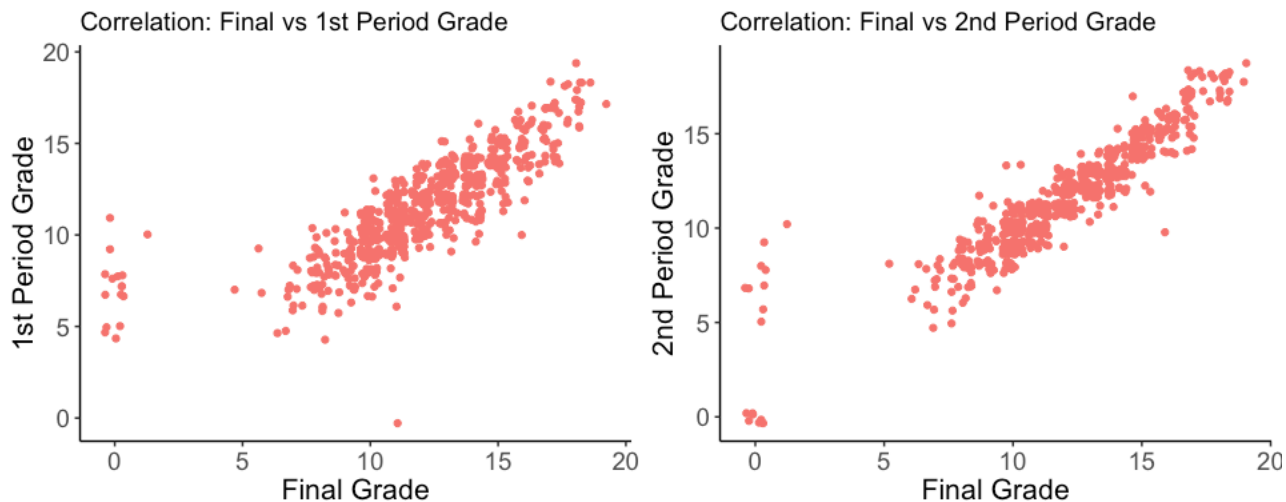
### 3-4. Profile group

- Overall, this group appears to be uncorrelated with final grades.

# EDAV fall 2018 Final Project (hnt2107\_nm3086\_sdt2134)

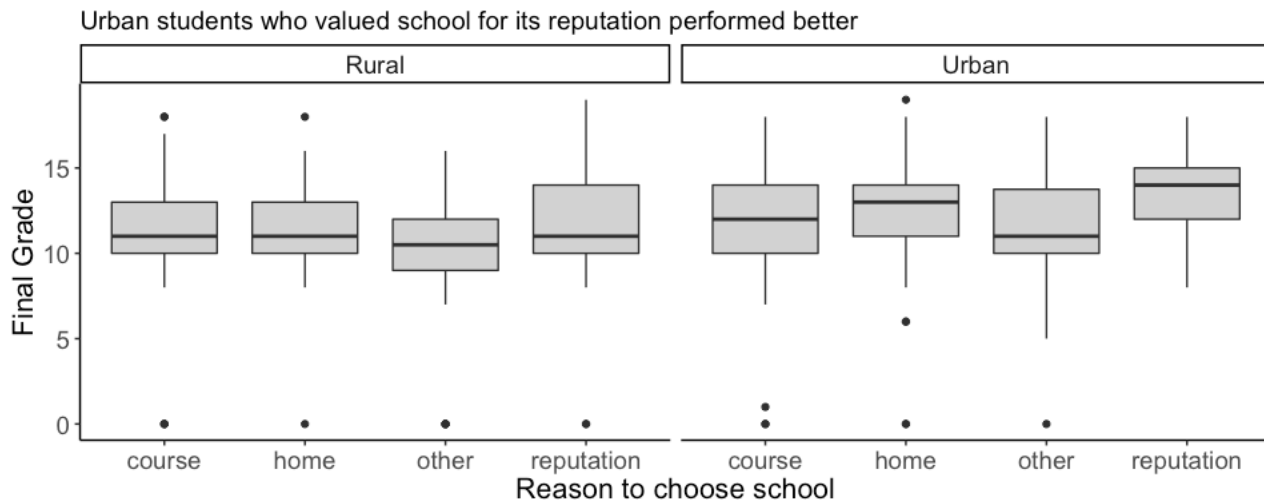
## Student Performance Data Analysis

### 3-5. Academics Group



- Final grades are strongly correlated with grades received in period 1(G1) & 2(G2).
- Students with frequent absences from school recieved lower grades.

### 3-6. Others Group



- ~86% of the students who showed a lack of interest in higher education fell in two the bottom-most bucket of the Erasmus scale<sup>3</sup>.
- Urban students value their school for its reputation<sup>4</sup>, study longer and perform better academically.
- Negative correlation between receiving educational support from school and final grade. Possibly, academically weak students receive support from the school in the form of tutoring.
- Other factors like being in a relationship, receiving family support etc. didn't affect the final grade.

<sup>3</sup> As described on page 4 in the pdf: <https://repositorium.sdum.uminho.pt/bitstream/1822/8024/1/student.pdf>

<sup>4</sup> School GP has a better reputation over MS.