

EDAV fall 2018 Final Project (hnt2107_nm3086_sdt2134)

Student Performance Data Analysis

1. Introduction

Grades are the primary indicator 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 multiple factors like parents' education, alcohol consumption, free time, etc. on students' grades.

2. Description of Datasets

We chose Student Performance Data Set which is provided on UC Irvine Machine Learning Repository. The data were obtained in a survey of students at math and Portuguese language courses in secondary school. It contains a lot of interesting social, gender and study information about students.

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

We categorized 30 data attributes in the dataset into 6 different groups so that we can conduct analysis efficiently. We also conducted data quality analysis and identified some issue. For more details, please find the report.

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)

3. Main Results

3-1. Home environment group

- Urban students use the internet more effectively to earn higher grades.
- We don't see any difference regarding grades due to family size.
- Surprisingly, male urban students living separately from parents perform better.
- Urban students with higher quality of family relationships earned better grades.

3-2. Social Group

- Lower grade students have higher free time. Interestingly, students with high free time reported participation in extracurricular activities.
- Males who consumes alcohol weekend very highly tend to earn worse grades.

EDAV fall 2018 Final Project (hnt2107_nm3086_sdt2134)

Student Performance Data Analysis

- Some other observation than grade-related association are below:
 - Females frequently reported they have lower free time.
 - Students with a lot of free time to go out with friends a lot more.
 - Students with very low free time have worse health condition.
 - Students have less free time when they take extra paid courses.
 - We found a surprising coincidence. (Look at where alcohol consumption is very high.)

3-3. Parents Group

- Students whose parents' jobⁱⁱⁱ is teacher performs better than other students.
- Students' grades become better, as parents' education^{iv} goes higher.
- When both parents did not take higher education, students grades appear to be worse than other case.

3-4. Profile group

- Overall, this group appears to be not so related to grades performance.

3-5. Academics Group

- Final grade is strongly correlated with 1st period grade and 2st period grade.
- Students most frequently absent from school have worse grades performance than student less absences.

3-6. Others Group

- 86% of the students who were not interested in studying further performed poorly in grades.
- Urban students values school reputation^v more as well as perform better in grades.
- Students who received educational support from school^{vi} showed lower grades.
- We don't see any grades difference due to those variables for family support and if students are in romantic relationship.

ⁱ There are more urban students than rural students in the report. Interestingly, 76% students in urban area study at GP while 40% of students in rural area study there.

ⁱⁱ It is only association, not causal inference

ⁱⁱⁱ This association was seen regardless of mother's job or father's job

^{iv} This association was also seen regardless of mother's education or father's education

^v School GP has a better reputation over MS. Also, students who values school reputation tend to have higher study time.

^{vi} Very few students getting education support from school.