

EDAV fall 2018 Final Project (hnt2107_nm3086_sdt2134)

Student Performance Data Analysis

1. Introduction and Description of Datasets

Grades are obviously one of biggest concern during student lives. To make the most of our lives at Columbia DSI, we wanted to analyze what factors are associated with better grades. 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>

2. Categorization of Datasets

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

- Firstly, we found that students who live in urban areas earn higher GPA than students in rural area. There are more students from urban than from rural in the report. Students in urban areas have more access to the internet than students in rural area.
- We analyzed if a particular school is more lenient and urban people go to that school. School GP indicated more higher GPA. Interestingly, 76% students in urban area study at GP while 40% of students in rural area study there.
- We don't see any difference regarding grades due to family size. Most of the families have sizes greater than 3.

EDAV fall 2018 Final Project (hnt2107_nm3086_sdt2134)

Student Performance Data Analysis

- Regarding whether parents live separately or not, it does not have impact on students' grades. However, when we look at this variable faceting by if students live in urban or rural area, we see students living separately from parents in urban area perform better.
- When we look into the combination of gender and parents' status, male students whose parents were separated performed better.
- Regarding family relationship, students with higher quality of family relationships earned better grades. Such tendency is seen for urban families.

3-2. Social Group

- In social group, Freetime was the most interesting variable which affects grade. We can see lower grade students have higher free time. Then, we came up with the question - "So what do they do during their free time?". Interestingly, students with high free time reported participation extracurricular activities.
- Compared to males, females frequently reported they have lower free time. Students with a lot of free time to go out with friends a lot more. For the students with very low free time, their health condition is worse. Students have less free time when they take extra paid courses.
- Regarding workday/weekend alcohol consumption, we don't see any significant difference among the groups. However, we found something interesting when we also observe whether there are any association between workday/weekend alcohol consumption and grades, splitting data by gender. Males who consumes alcohol weekend very highly tend to earn worse grades.
- This is for fun. We found a surprising coincidence. Look at where alcohol consumption is very high.

3-3. Parents Group

- First of all, students whose parents' job is teacher performs better than other students. This tendency can be shown regardless of father's job or mother's job.
- Secondly, parents' education is also associated with students' grade. When father took higher education, students perform better in their grades. The same thing holds true of mother's case. We observed students' grades become better, as parents' education goes higher.
- It appears to be important that at least either mother or father took higher education. 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. Having said that, we will describe what we analyzed as follows.

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. Therefore, motivation for higher education is significantly associated with grades performance.

EDAV fall 2018 Final Project (hnt2107_nm3086_sdt2134)

Student Performance Data Analysis

- Compare to students in rural area, urban students values school reputation as well as perform better in grades. School GP has a better reputation over MS.
- Students who values school reputation tend to have higher study time.
- Students who received educational support from school showed lower grades. We don't have any reason to assume causality in either direction. Also, very few students getting education support from school.
- As for family support and if students are in romantic relationship, we don't see any grades difference due to those variables.