SCHOOL GRADES' ANALYSIS

by Gonçalo Bernardo

May 2023

DATA ANALYSIS PORTFOLIO

I believe everyone, from one way or another, has a story to tell and I'm not an exception. Since I can remember I have a passion for numbers and math in general. When I was a child, I used to spend long hours reading books about math and I really enjoyed applying my knowledge in a real context, making statistics with my toys, analyzing prices and discounts at the supermarkets, and so on.
Later in school I was encouraged to proceed with my studies specifically in math or applied math, statistics, but instead, I decided to take a master in mechanical engineering, which confirmed I should have chosen the "math path" because after working a few years in mechanical industry I had to quit and I jumped to a software developer carrier for other few years when finally I realized I really missed the numbers, the mathematical problems, the graphs, the calculations. I wasn't happy even if programming was giving me joy.
That's when I finally found data analysis and science would fit all my passions at once, so I started a new project to my life creating a learning center to teach math, giving more time to study data analysis, so I could, in a few months, have my dream job: data analyst.

PROFESSIONAL BACKGROUND

Having a degree in mechanical engineering didn't give me access to my dream job and after five years taking a MsC, I have worked for two years as a fellow researcher by programming algorithms in matlab to reconstruct surfaces from point clouds obtained by laser scanning.
Then, for a few years I worked as a quality assurance engineer and mechanical designer, but it didn't make me happy, so I quit my job to do a java and SQL bootcamp, which I finished in 4 months and started working as a IT consultant.
During three years I have worked as a full-stack software programmer and I have participated in the development of some projects, most of the time in the banking field. Those jobs gave me joy, especially because programming fits great in the type of tasks I love to do, and my teams were always supportive and my senior coworkers had taught me a lot.
However, something was missing and I should have been obvious from the beginning, but it took some time to realize that it was math: my biggest passion.
That's why I decided to change my path by opening my own business: a math learning center, which gives me more flexibility to manage my time and study data analysis and science so as soon as possible I proceed to my second part of my plan: become a data analyst.

TABLE OF CONTENTS

1. SCHOOL GRADES PROJECT

In this section, it is explained the description of the project, which can be observed in table 1.1.

Table 1.1 - Project description using SMART Method.

Situation:	Section 3 gives a detailed description about the collected data set, but one can say the main goal is to study the influence of some personal factors in the students' grades of two Portuguese courses existing in two distinct schools. Initially our data set hadn't relevant issues, just a few duplicated pieces of information, which were cleaned and organized using google sheets as well tableau software, mainly in the visualizations.
Task:	Section 2 gives more details about the key questions to be analyzed and answered, which pretend to identify possible trends and patterns related with students' habits, parents' education, intent to proceed to higher education or alcohol consumption habits. Among other key questions, we search for information about the habits and family characteristics of the best and worst students, the influence of parents' relationships and education, and additional school support as well as the access to the internet or the environment where the student lives.
Action:	In order to achieve the final results, and implement our analysis, the data was cleaned using google sheets and to perform the visualizations, it was used google sheets as well as tableau. Section 4 shows the main findings of the project.
Result:	The results are discussed in section 5 with more detail, but, generally speaking, the most relevant findings were the clear influence of parents' education and the intent of proceeding to higher education. The environment where the student lives has a great influence in his/her grades. Additionally we found Gabriel Pereira School students had average grades higher than the Mousinho da Silveira School, where the alcohol consumption is higher.

2. PROBLEM DEFINITION

The problem is more clearly defined if some questions to be answered are defined, so next we have the main questions we want to discuss along the report and try to find patterns and trends to better understand what are the students' habits and family environments that will influence their grades.
So, we have considered the questions below:

- What are the average grades of each school?
- Which school has the higher general rate of alcohol consumption?
- How do fathers' education and the quality of family relationships influence students' average?
- How living in a rural or urban region influences students' average grades?
- Some students intend to get higher education. Does it really have an effect on their grades?
- What are the main characteristics of the best 10 and worst 10 students?

3. DATA DESIGN

Our data was extracted and stored as a csv file, which was read and transformed by importing it with google sheets.
Initially the raw data had some duplicated information and it was corrected by using google sheets tools.
Additionally some visual content was created through tableau software, by importing our cleaned data and analyzing for each key question what would be the best visualizations to better “understand” our data.

4. FINDINGS

4.1 What are the average grades of each school?

Table 4.1 - Average grades by school.

School	Average Grades
MS	10.48
GP	12.24

* Average value of three periods' grades.

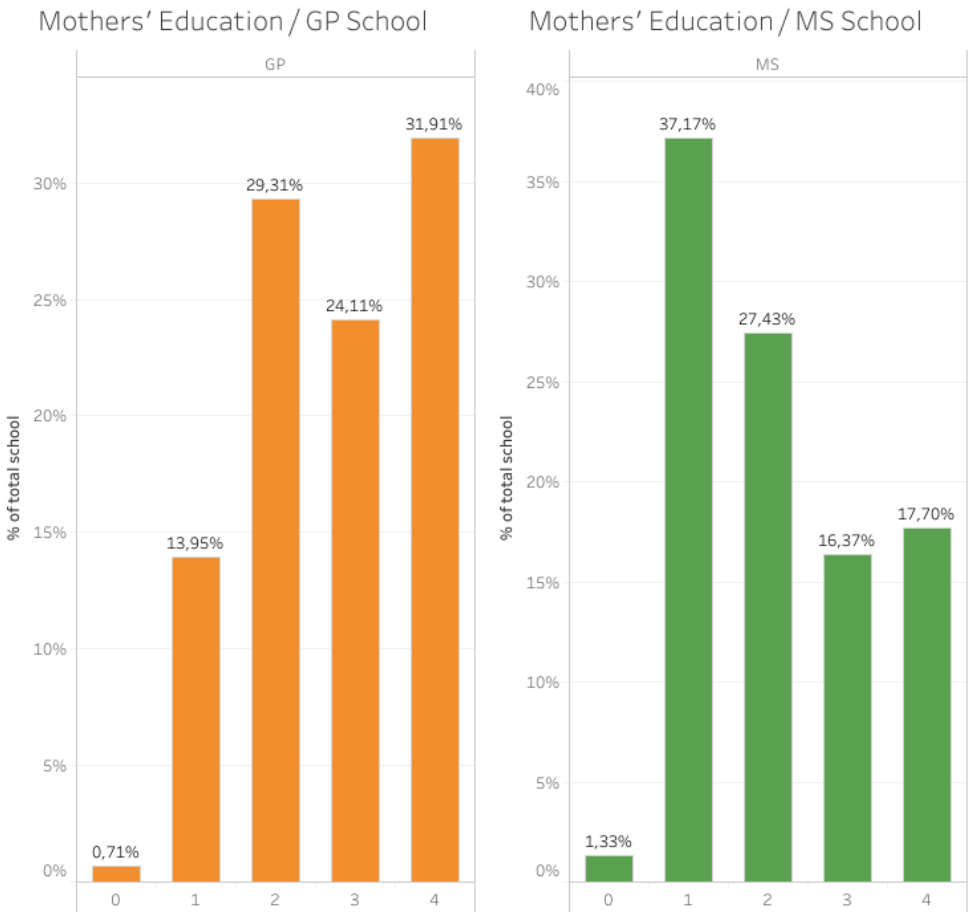

Fig. 4.1 - Mothers' education for each school, in percentage of total students in each school.
*Education level scale: 0 - none, 1 - primary ed, 2 – 5th to 9th grade, 3 – secondary ed or 4 – higher ed

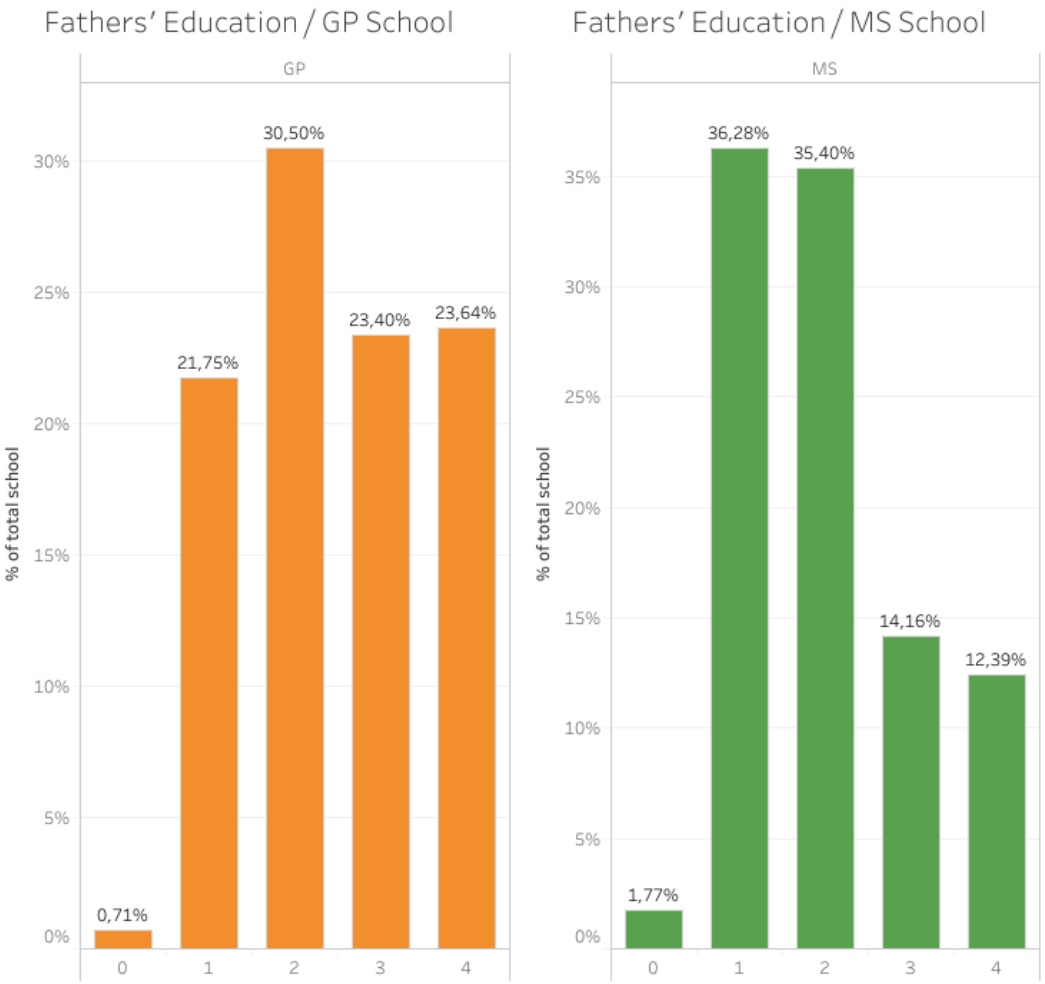

Fig. 4.2 - Fathers' education for each school, in percentage of total students in each school.
*Education level scale: 0 - none, 1 - primary ed, 2 – 5th to 9th grade, 3 – secondary ed or 4 – higher ed

4.2 Which school has the higher general rate of alcohol consumption?

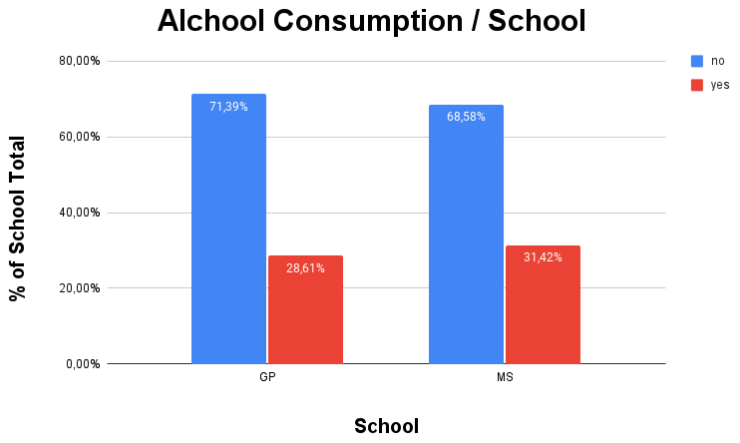

Fig. 4.3 - Alcohol consumption in percentage of total students, by school.

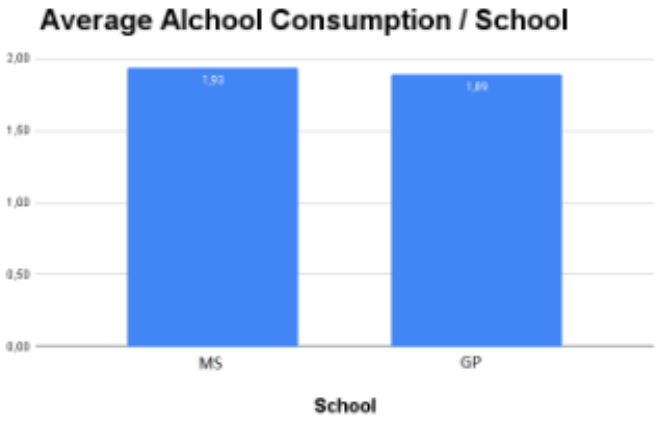

Fig. 4.4 - Average alcohol consumption by school.

4.3 How do fathers' education and the quality of family relationships influence students' average grades?

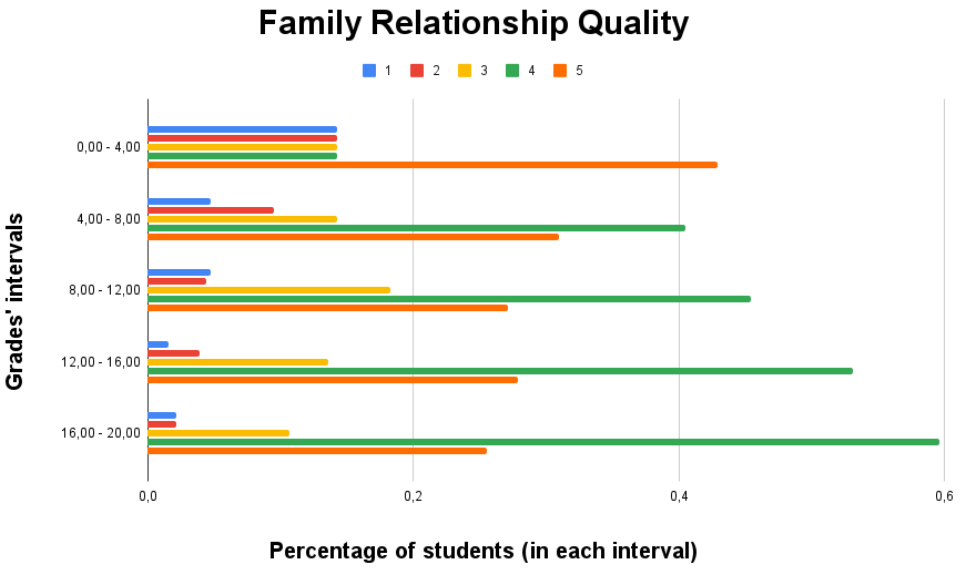

Fig. 4.5 - Percentage of students Vs Grades' intervals, for each family relationship classification (from 1 to 5).

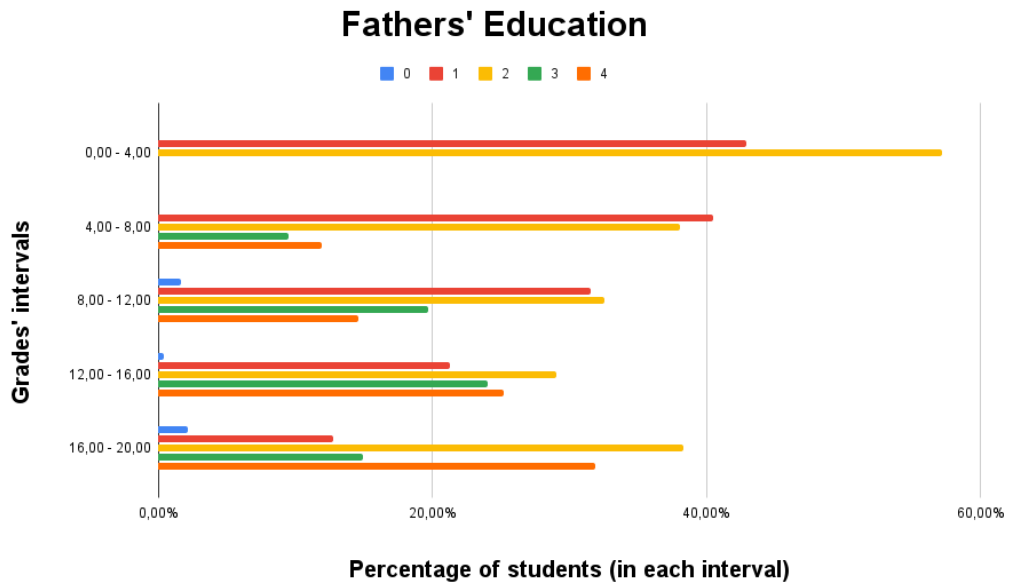

Fig. 4.6 - Percentage of students Vs Grades' intervals, for each mothers' education degree (from 0 to 4).

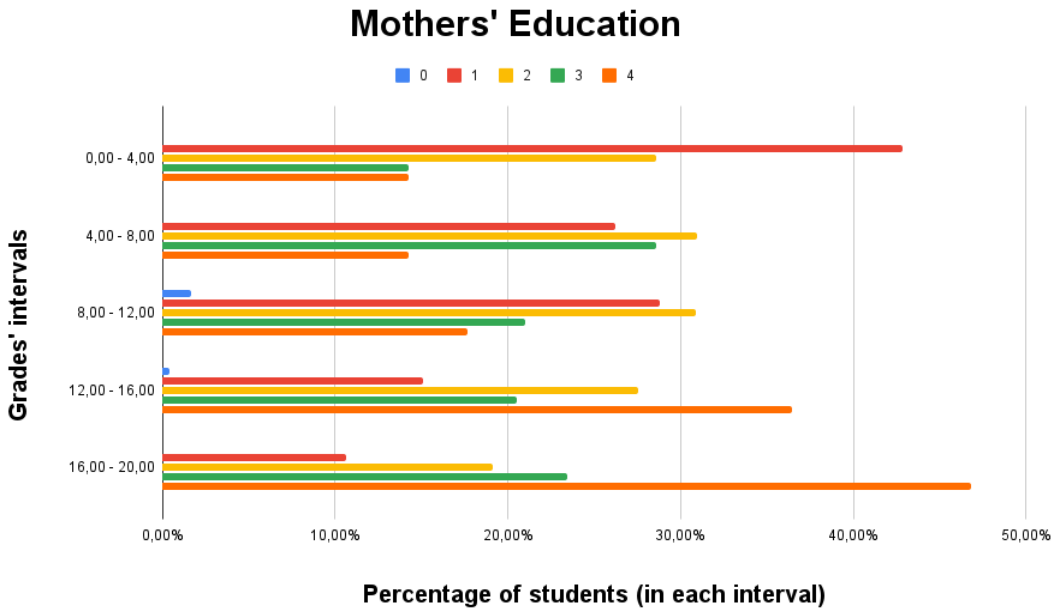

Fig. 4.7 - Percentage of students Vs Grades' intervals, for each fathers' education degree (from 0 to 4).

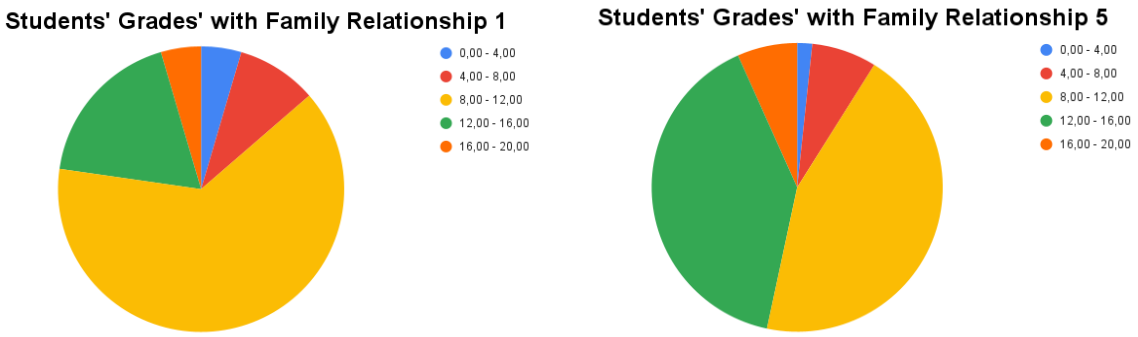

Fig. 4.8 - Students' grades distribution for family relationships classified with 1 (left) and 5 (right).

4.4 How living in a rural or urban region influences students' average grades?

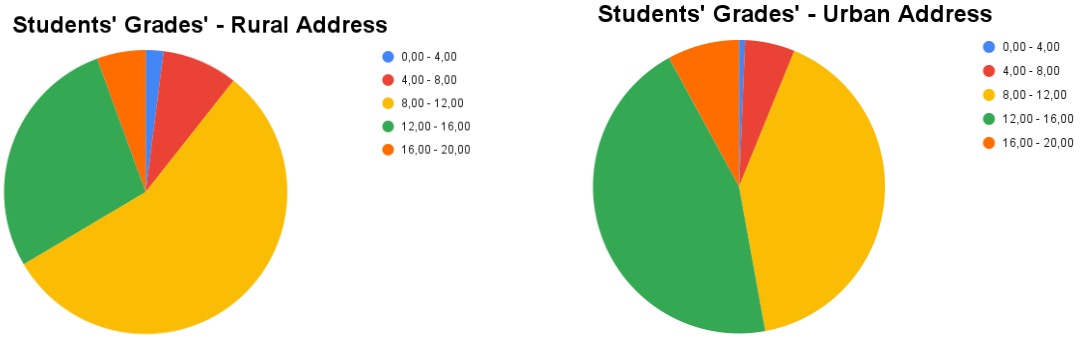

Fig. 4.9 - Students' grades distribution for rural address (left) and urban address (right).

4.5 Some students intend to get higher education. Does it really have an effect on their grades?

Table 4.2 - Comparison of grades intervals with intent of higher education.

	Intent of higher education		
Average Grades	No	Yes	Total
0-4	3	4	7
4-8	17	25	42
8-12	47	248	295
12-16	2	256	258
16-20	0	47	47
Total	69	580	649

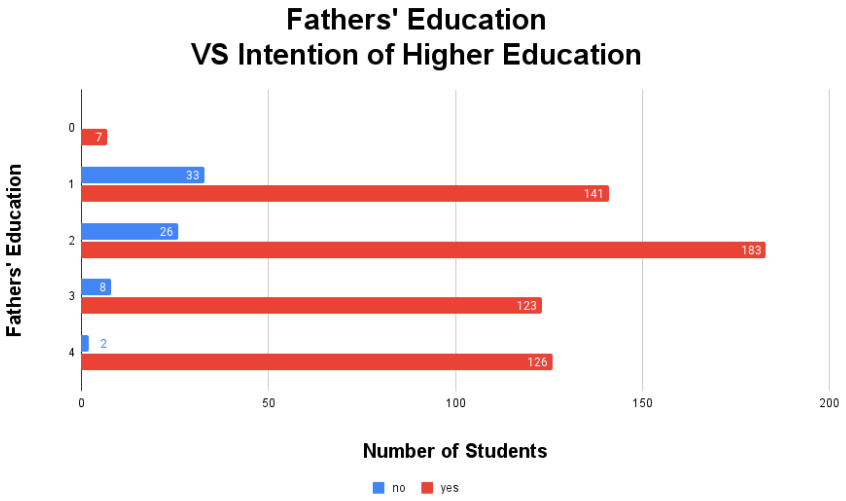

Fig. 4.10 - Fathers' education VS Intention of higher education.

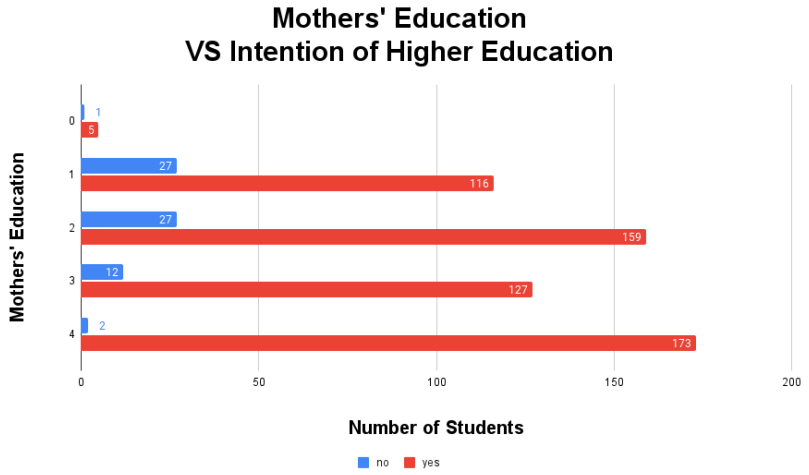

Fig. 4.11 - Mothers' education VS Intention of higher education.

4.6 What are the main characteristics of the best 10 and worst 10 students?

Tables 4.3-4 - Detailed information about the worst and best students (from left to right).

	School	Address	Study time*	Higher Education	Average Grade
1	MS	R	1	no	4.67
2	MS	R	1	yes	4.33
3	MS	R	2	yes	4.00
4	MS	R	2	yes	3.00
5	MS	R	2	no	2.67
6	MS	U	2	yes	2.33
7	MS	U	1	yes	2.33
8	MS	R	2	yes	1.67
9	MS	U	1	no	1.67
10	MS	R	1	no	1.33

	School	Address	Study time*	Higher Education	Average Grade
1	GP	R	4	yes	18.67
2	GP	U	3	yes	18.00
3	GP	U	4	yes	18.00
4	MS	U	2	yes	18.00
5	MS	U	4	yes	18.00
6	MS	U	3	yes	18.00
7	MS	U	2	yes	18.00
8	GP	U	1	yes	17.67
9	GP	U	3	yes	17.67
10	GP	U	1	yes	17.67

* Study time: 1- <2 hours, 2- 2 to 5 hours, 3- 5 to 10 hours, or 4 - >10 hours

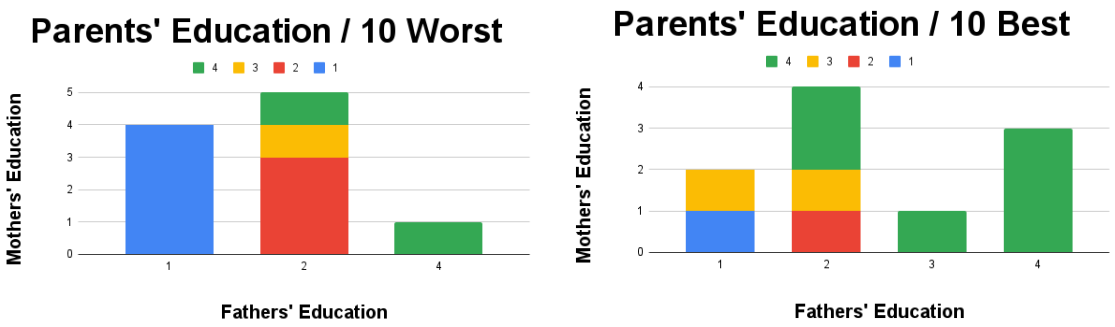

Fig. 4.12 - Parents' education for 10 worst (left) and 10 best (right) students.

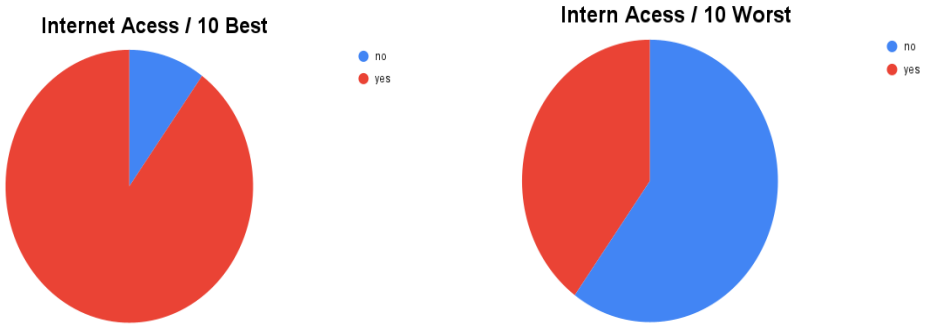

Fig. 4.13 - Internet access for 10 best (left) and 10 worst students (right).

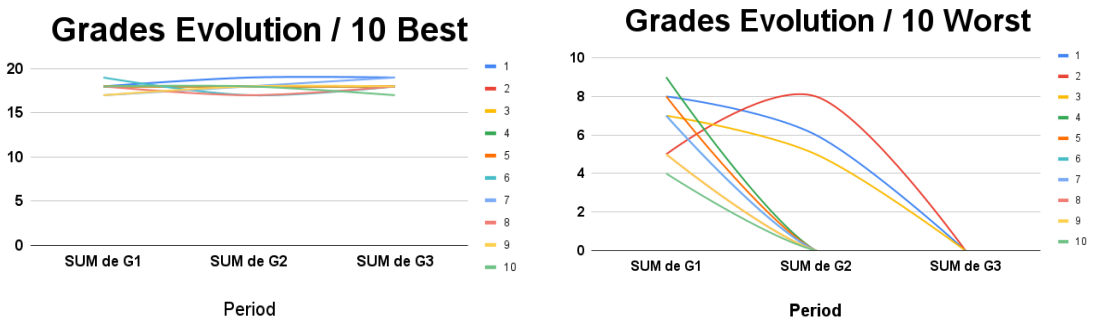

Fig. 4.14 - Grades' evolution for 10 best (left) and 10 worst students (right).

5. ANALYSIS

By the values shown in table 4.1, one can conclude that Gabriel Pereira School has a higher average grades compared with the other school, and from the two graphs, it's interesting to see that, in general terms, parents' education tends to a higher level precisely in GP school. Later we'll analyze it in more detail.

Each student was asked about his alcohol consumption habits using a scale between 1 and 4 (one and four corresponding to low and high consumption respectively), during work days and weekends. The graph 4.3 on subsection 4.2 shows the percentage for each school who have an alcohol consumption higher than one in general. The figure 4.4 on subsection 4.2 shows the average alcohol consumption level for each school.

The values are similar and we can't have any relevant conclusion, but MS school has shown to have a higher alcohol consumption in general.
Additionally, students were asked to classify how good family relationships are at home, from 1 to 5 (being 1 the worst classification).
By grouping the students by 5 intervals of grades, the column graph 4.5 on subsection 4.3 shown on the left, we can show that, compared with the total of students in each group, the students with higher quality relationships tend to have better grades, in general.

One should highlight that the percentage of total students of [0;4] interval is 1.08%, so it is not relevant to make conclusions with that group.
The group with more students is [8;12], corresponding to 45.5% of total students.
Similar approach performed with familiar relationships was made with parents' education level and the results showed that the percentage of students with lower grades that have parents with low education level is high, contrary to the students with better grades, whose parents have higher levels of education. Almost 50% and 40% of mothers in grade's groups [12;16] and [16;20], respectively, have a level 4 education.

The distribution of grades among the 2 extreme levels of family relationships were analyzed with more detail, and it was created those 2 circular graphs (figure 4.8 on subsection 4.3) that show us the percentage of students with better grades tend to grow with the family relationship, i.e., the portion of green and orange sections in the graphs is higher in the right graph.

Another interesting characteristic to study is the address of each student. They were asked if they were living in an urban or rural region. It was found that approximately 30% of students were living in a rural region, and another 70% in an urban region.
Figure 4.9 on subsection 4.4 shows the distribution of students in each grade group for rural (left) and urban (right) addresses. Among urban students, approximately 53% have an average grade higher than 12, being 34% the percentage of rural students with such grades.

Clearly, we can say the region where students live has an impact on their grades.
Concerning the intent of each student in getting higher education, from table 4.2 , we can see that it affects the grades of the students, especially when only 3% of

students who do not intend to take higher education have grades greater than 12. Among the other group of students 52% have grades above 12.

Continuing the analysis of the intention of higher education, it is interesting to compare parents' education and conclude that the higher is parents' education, especially in case of mothers' education, the bigger is the number of students who intend to proceed their studies to higher education. For example, 300 of students with mothers' education levels 3 and 4 intend to take higher education.
Considering our last key question on subsection 4.6, among all the data, one can find several interesting patterns in the two groups and the above tables show some of them. All the worst 10 students study in MS school and 40% of the best 10 students study in that school. Only one of the best students lives in a rural region, and it's the best student of all.

All of the best students want to take higher education, and 60% of the worst intend to do the same. It is clear that the study time, which was not analyzed in general terms, is clearly higher among the best students.
By observing both graphs in figure 4.12, we can see that four of the worst students have both parents with education level 1 and none of them have both parents with higher education.

In contrast, among the best students, 3 of them have both parents with level 4 education, one with father with level 4 and mother level 3 and only one has both parents with no education.
In conclusion, once again, it's clear that a parent's education is related to the student's success in school.

Analyzing the access to the internet of both groups is another interesting question.
From the circular graphs in figure 4.13, one can conclude that almost all the best students have access to the internet and more than half of the worst students don't have access. It could influence the access of students to information, making the difference in their grades.

Finally, we have another interesting fact: the consistency of best students' grades compared with the worst, i.e., all the best students during the 3 periods maintain their grades above 17/18, but all the worst students tend to decrease their grades to 0 in the last period.

Figure 4.14 shows that since the beginning of the course it is important to have good grades because most of those worst students started to demotivate and gave up the course (70% of them gave up the course in the second period).

6. CONCLUSIONS

❖ Best and worst students:

By comparing the worst and best students, we can say that parents' education levels are higher among the best students (40% have both parents with high education). The average study time is higher for those students and 90% of them live in urban regions. All of the best students study in GP school.

❖ Influence of family education and relationships:

The majority of students claimed to have a good family relationship at home, and most of them are among those who have grades greater than 12. By comparing the students with lower level of family relationships with those who have the higher, we see that the percentage of students with higher grades tend to be higher among the students with good family relationships. Like we see in the analysis of best and worst students, in general, parents' education has a considerable influence on the grades as well as the intention of taking higher education.

❖ Access to internet:

The analysis of best and worst students led to the conclusion that best students tend to have access to the internet, contrary to the worst.

❖ Alcohol Consumption:

Concerning the alcohol consumption, in general, it's not clear if it has a great influence on the grades, but one can say that the average consumption seems to be higher among the 10 worst students, especially at weekends. Comparing both schools, the difference is not significant, but students of MS tend to have a higher average alcohol consumption.

7. RECOMMENDED ACTIONS

❖ Habits recommendations:

Our data tell us the importance of parents' education on the students' interest to take higher education and have better grades, so it is important to sensitize our society for such facts, as well as the quality of relationships at home.

Important actions must be taken to motivate the students to focus on their studies, as to reduce the alcohol consumption, or have interest in taking higher education.

❖ Support recommendations:

Concerning specific actions, from the data we can conclude that some actions could have a positive impact in students' life and consequently in their grades.

We recommend implementing some actions to give access to all students to the internet, and to support all the students who live in rural environments.

Another recommendation is to create, in each school, a vocational team who could give support to the students, giving them opportunity to explore, discuss and explore their skills in order to make them understand themselves better and gain some motivation to choose a path where they have to take higher education.

❖ Further work:

Additional data was collected, but not analyzed in the context of this presentation, but it would be interesting to study the influence of extra help and support from the school as well as the influence of hobbies and extracurricular activities on the motivation and students' grades

DATASET SOURCE

https://www.kaggle.com/code/zelzebu/students-habits-and-grades-prediction