discuss why the data source is **interesting** and **relevant** to **data science.**

The dataset we are using, displays Portuguese secondary students and different factors which might affect their academic progress. The dataset includes 395 students, and has 30 independent variables and 3 dependent variables (G1, G2, and G3). This dataset is interesting because it shows how different factors affect the students' performance in school. It's relevant to data science because we can use this information to observe what steps we can take to help a student have a successful academic career.

In this analysis, we will be going through comparisons between hours spent studying, number of absences, having a tutor, coming from an educated family and its effect on grades. We also go analyse whether having a longer commute to school affects final grades and previous failed classes.

As a result, if teachers noticed a student is performing poorly, they can use our analysis to effectively assist the student live to his or her fullest potential.

(G1+G2+G3)/3 = final which is the final mark for the student

The Portuguese education system divides each semester into 3 terms, G1, G2, and G3. They use a 20-point grading system where a score above 10 is passing, 10-13 is a C, 14-15 is a B, 16-17 is a A, and an 18-20 is an A+. In the below cell, we figure out the final mark by calculating the average of all the grades per student.

Hours spent studying to final

In this cell below, we compare the number of hours spent studying with the students' final grade. We took the students who studied less than 2 hours, 2-5 hours, 5-10 hours, and greater than 10 hours a week and grouped them accordingly. Then we calculated the mean grade for each group to determine whether the number of hours spent studying per week affects the students' final grade.

	studytime	final
0	1	10.253968
1	2	10.442761
2	3	11.651282
3	4	11.72

Conclusion:

Based on the results that we acquired, we can conclude studying does have an impact on academic performance. As seen in the table, groups that performed well have a longer study period. Thereafter, we can conclude studying more does indeed affect the final grade of the students.

Absences vs final

We were curious what the relation between absences and final grades were, so we decided to plot the number of absences on the y-axis and their final grade on the x-axis. Our hypothesis is that the number of absences are going to have a strong correlation with the final grades.

Conclusion:

We were expecting a strong correlation between the two but surprisingly we discovered a Pearson correlation coefficient of -0.0059, meaning there was no relationship between the two. The scatter plot we made allows us to see there was an even spread of absences throughout the range of the final grade. There were a couple of outliers which may have occurred. There was no significant information here so we could not use it to support our initial objective.

Parents education vs FG

In the following cell, we grouped the students according to their parent's education level. We had a group for each possible combination of the student's parent's education. We are testing this out because we suspect the education level of the parents would reflect their children's academic success. On the chart, 0 is no education, 1 is primary education, 2 is middle school, 3 is secondary school, 4 is post secondary education. We then ordered the parents from the least educated to most educated. Then we plotted the graph with the index on the x-axis and the final grades on the y-axis. We used the index because we couldn't plot two numbers on one x-axis, so we used indexes to represent the combination of the mother's and father's education.

Conclusion:

From the scatter plot we can see the data is fairly scattered, there are high and low grades at both ends of the x-axis, we can conclude there is no significant relationship between parents being educated and the student's final grades.

Paid tutor

We hypotheses the students who go to tutoring are better off academically. To test this hypothesis, we grouped the students into two groups, those who go to a paid tutor and those who don't. Then, we found the average of all the students.

Conclusion:

From the results we can see the tutored group performed better with a mean score of 11.04, whereas the group which wasn't tutored has an average of 10.38. This is not a substantial difference however it does still support our hypothesis.

Final grade vs commute vs failures

We tested whether students who had a long commute to school had poor performance at school. We suspect students who commute longer times spend less time studying. We grouped the students according to the amount of time they spend commuting. Then we find the average final grade for each group and the average number of failed classes.

Conclusion:

In the chart we generated, we can see the grades decrease as the time for commuting increases, the number of failures also increases.

Conclusion:

Our objective for these analyses is to help struggling students improve. We intend to do this by isolating factors that may contribute to better grades. From all of our analyses, we found many interesting connections in our dataset. We combined the 3 terms to get our final grade. We compared the number of hours spent studying with the final grade, and we came to the conclusion students who studied for a longer period of time had a higher grade. From our analysis, we found studying between 5 and 10 hours per week provided the best ratio of improvement. When we were testing the correlation between the number of absences and final grade, we found a Pearson correlation coefficient of -0.0059 which proved there was no significant correlation between the two. The next test was with the student's parent's level of education and the student's final grade. Our hypothesis was the parent's level of education would reflect on their child's performance at school. After our analysis we concluded there was no significant relationship between parents being educated and the students' final grades. This can be motivational for students who think because they come from an uneducated family, they won't be able to succeed academically, this data shows them there is nothing holding them back. After, we analysed students who went to tutors and those who didn't, we came to the conclusion that going to a tutor is a factor which helps improve a student's grade. We also analysed the effect of different commute times on final grades and classes failed. We found as the commute time goes up the student's final grade drops and failed classes increases. This might be because the student spends more time commuting, resulting in having less time for other activities including studying.

The most successful student in our dataset follows many of our hypotheses, such as studying for many hours a week, having highly educated parents, commuting the minimal amount

From the two cells we were able to depict that the most successful student and the least successful student followed many of our hypotheses. Ranging from the hours spend studying on a weekly basis to their parents' education, many of the characteristics are identical.

Using our analysis of this data, we can apply it to improve quality of life for a student who is performing poorly in school. It gives a confidence boost to those students, showing them there is an academic future for them. For students who are not academically inclined it opens up a plethora of career paths, an important decision we all have to make in life. Perhaps allowing them to pursue a career they otherwise might have not. If this analysis can change the life of even one student then we know we have completed our jobs.