



PREDICTING STUDENT
PERFORMANCE:
AN APPLICATION OF MACHINE
LEARNING

BY MINH H. LE (2020)

TOPIC OVERVIEW

1. Background
2. Problems
3. Description of Datasets
4. Exploratory Analysis
5. Machine Learning
6. Conclusion

BACKGROUND

Education is the passport to the future. My goal is to help educational institutions, and parents to improve the student academic performance. Based on my analysis, schools and parents will know which factors and reasons affect the student academic performance. With that knowledge, they can help the students to prepare for the future.

PROBLEMS

1. Find out which features (attributes), including student grades, demographic, social and school related features, have effect on high school student performance.
2. Predicting the student performance.

DESCRIPTION OF DATASETS

- Two datasets:
 - Mathematics (CSV) - 395 observations (students)
 - Portuguese (CSV) - 649 observations (students)
- Total 32 features in both datasets.
- Most of the features are categorical (nominal and ordinal) data.
- Numerical features:
 - Absences (number of school absences),
 - G1 (first period grade)
 - G2 (second period grade)
 - G3 (final grade, also the output target)

DESCRIPTION OF DATASETS (DATA WRANGLING)

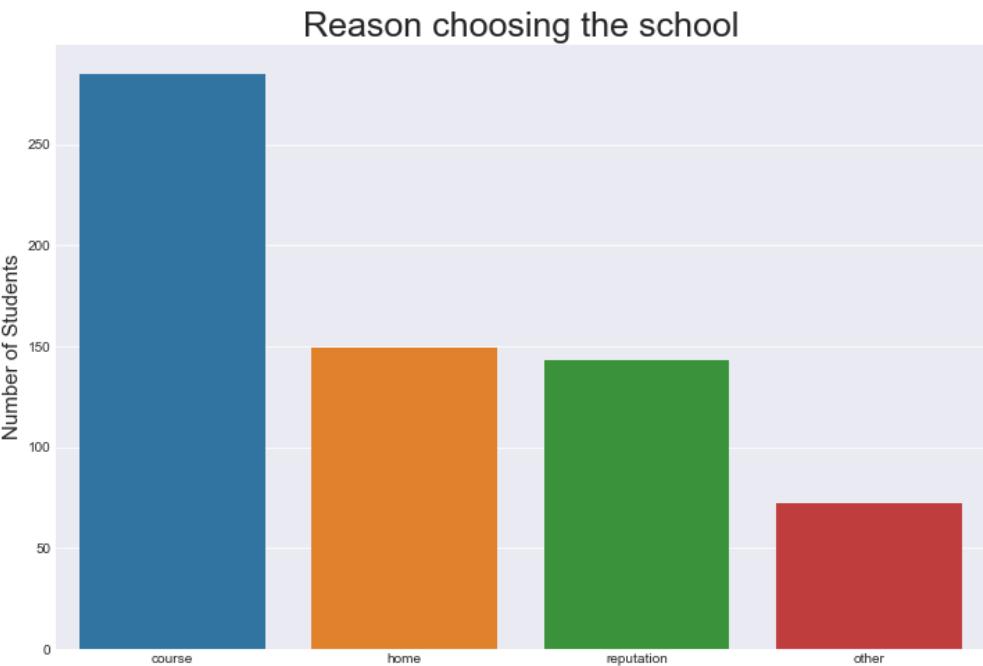
- No missing value in both datasets.
- For numerical features, there are outliers in both datasets. However, based on information from the dataset description, since none of the outliers are due to incorrectly entered or measuring data, the outliers are kept in the datasets.

EXPLORATORY ANALYSIS

QUESTION 1:

What is the main reason why students choose their current school?

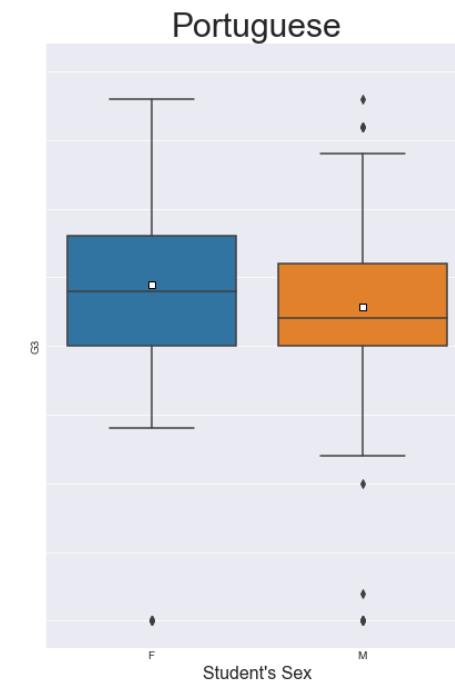
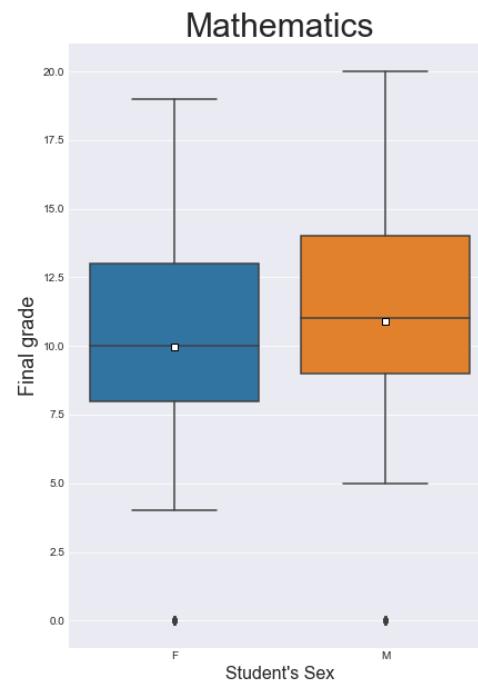
Most of the students choose their school is because of the course preference that the school offers.



QUESTION 2:

There is this stereotype that boys are better than girls at S.T.E.M. (Science, technology, engineering, and mathematics), and girls are better than boys at humanity. Is this true?

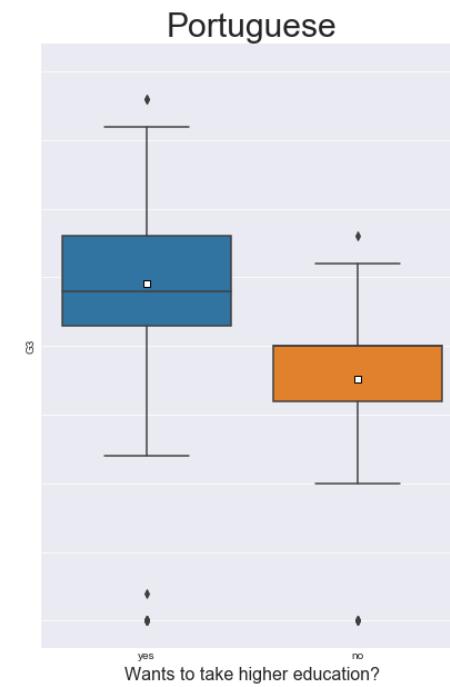
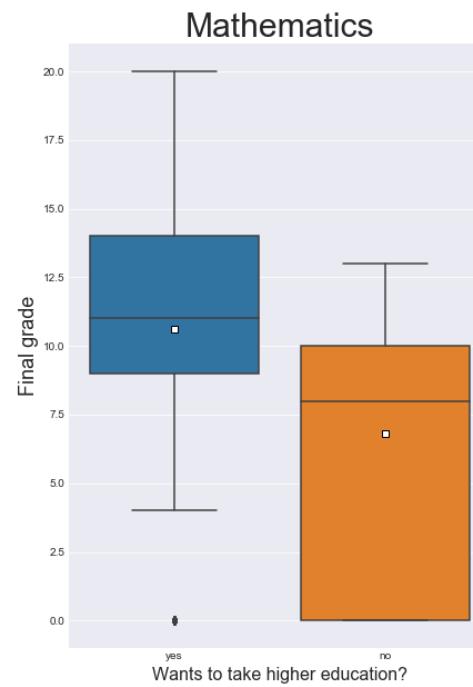
- Based on both data sets, the average and median mathematics final grades of male students are higher than those of female students.
- In contrast, the average and median Portuguese final grades of female students are higher than those of male students.



QUESTION 3:

Do students who want to take higher education have better grades than those who do not?

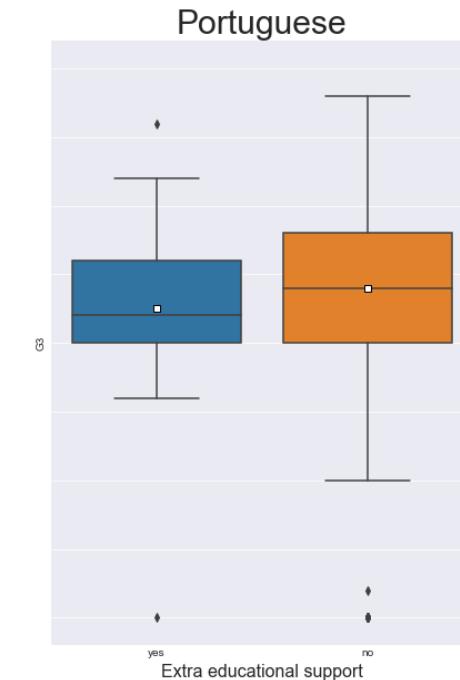
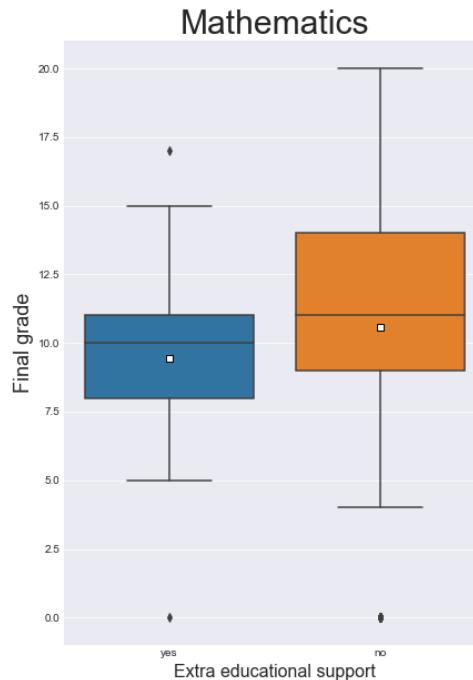
- The average and median final grades in both mathematics and Portuguese of students who want higher education are higher than those of students who do not want a higher education.
- This seems to be right. Students who want to go to colleges or universities will need good grades to be accepted; therefore, they will have higher grades than those who don't want to.



QUESTION 4:

Do students who have extra educational support have better grade than those who do not?

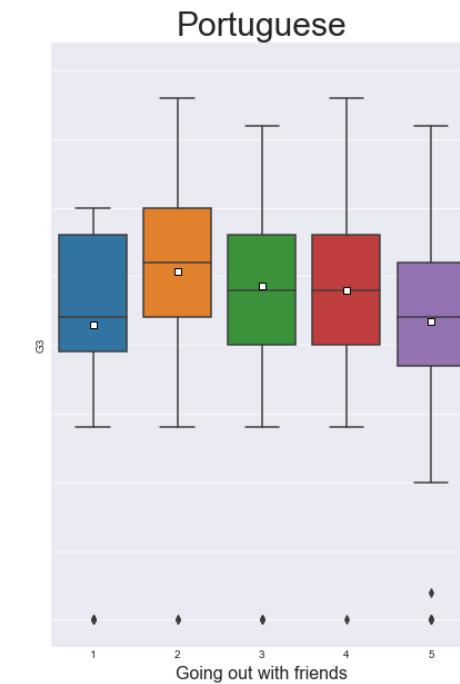
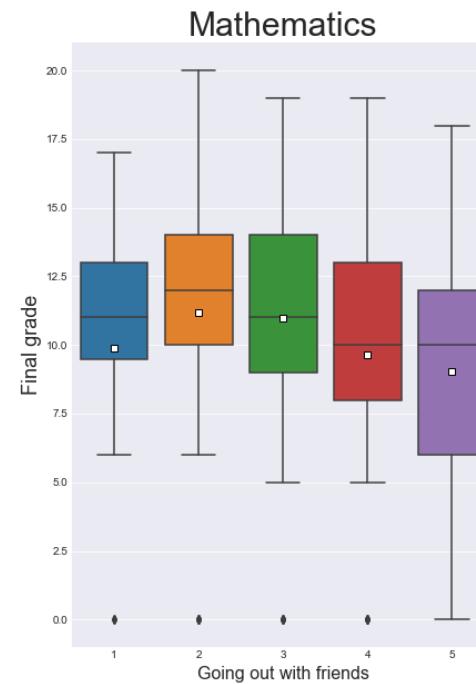
- The average and median final grades in both maths and Portuguese of students who have extra educational support are lower than those who do not have.
- The reason can be because students who are struggling with the subject will need extra support than those who don't. Therefore, their grades will lower than those who are already good at the subject.



QUESTION 5:

Does going out with friends affect final grades?

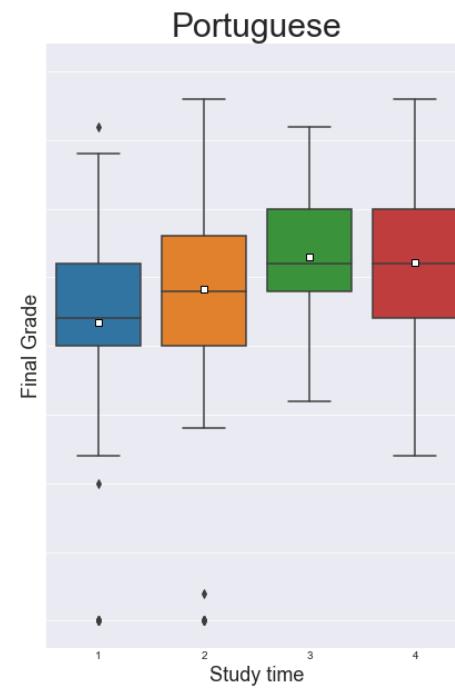
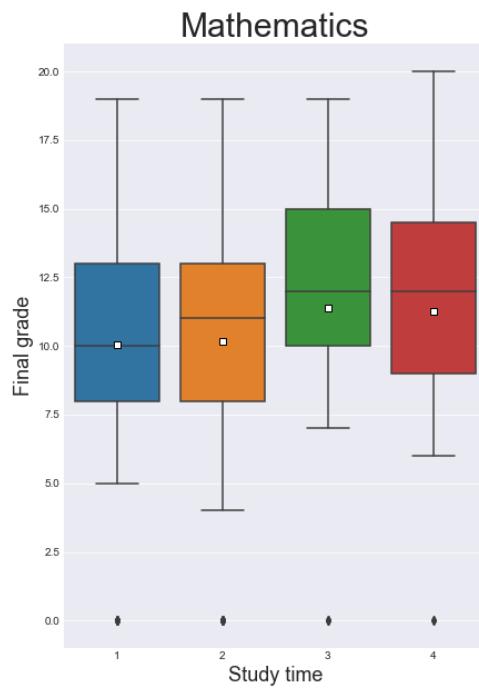
- Surprisingly, people who don't go out with friends much (1) and people who are always going out with friends(5) have the lowest average and median final grades in both maths and Portuguese.
- People who casually go out with friends (2) have the highest average and median final grades in both courses.



QUESTION 6:

Normally, people believe that study time is a good indicator of student performance. Students who spend more time studying will have better grades than those who don't. Is this true?

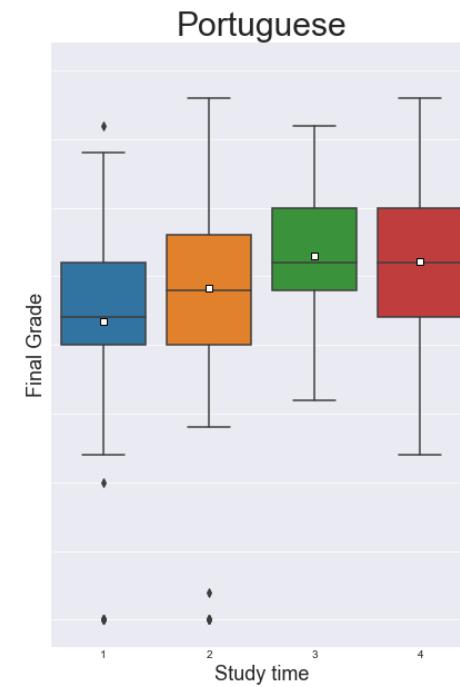
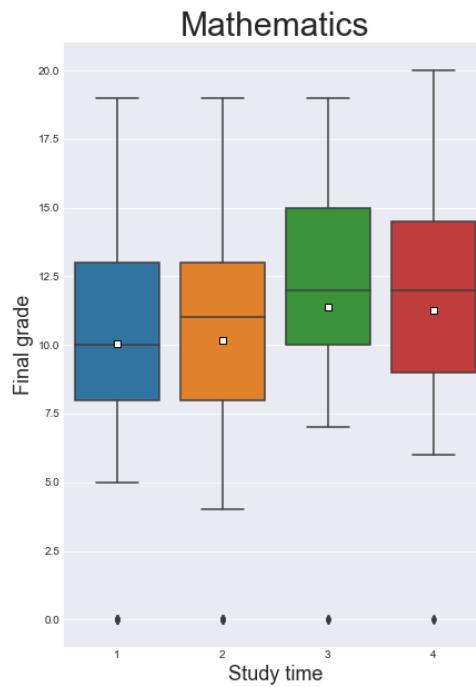
On average, students who spend 5 to 10 hours and more than 10 hours studying have higher grades than those who spend less.



QUESTION 6:

Normally, people believe that study time is a good indicator of student performance. Students who spend more time studying will have better grades than those who don't. Is this true?

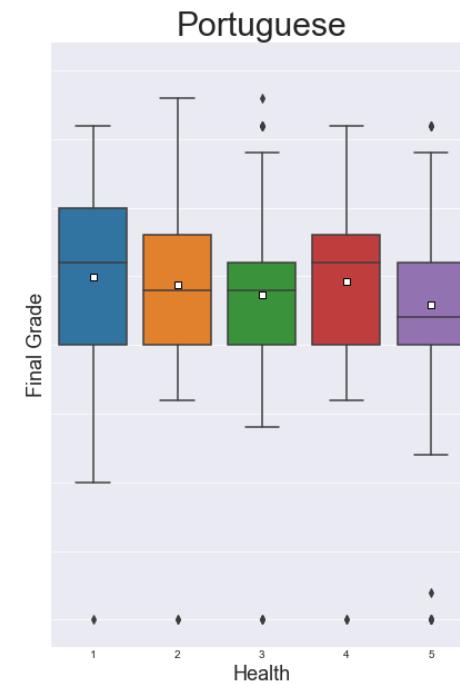
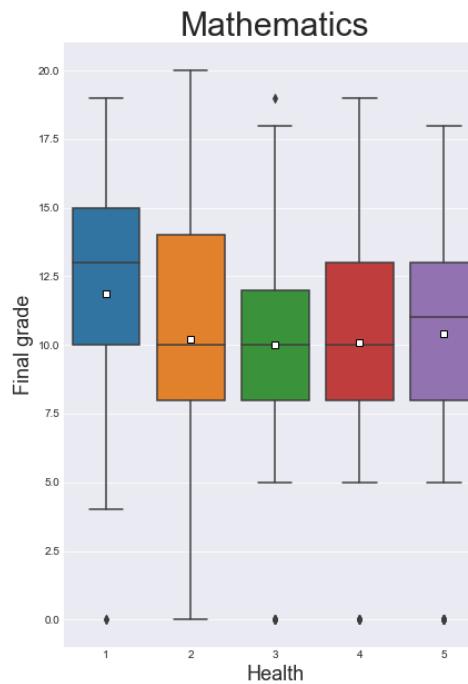
On average, students who spend 5 to 10 hours and more than 10 hours studying have higher grades than those who spend less.



QUESTION 7:

Normally, people believe that student health is a good indicator of student performance. Students who have good health will perform than those who don't. Is this true?

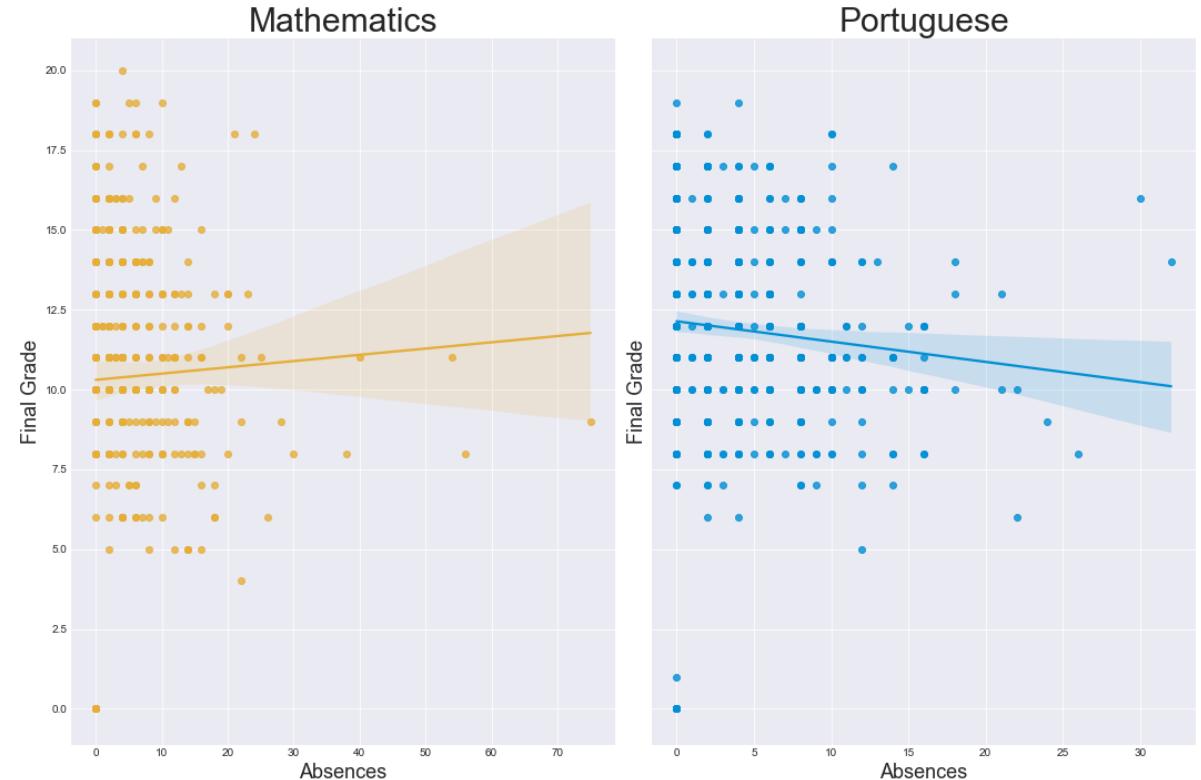
- It seems to be false. Students who are the least healthy have the highest average and median final grades in both subjects.
- This can be because healthy students are more likely to do more extracurricular activities than those who are not healthy. Therefore, they will spend less time studying, which can affect their final grades.



QUESTION 8:

Is there any correlation between the student's absence and the final grade?

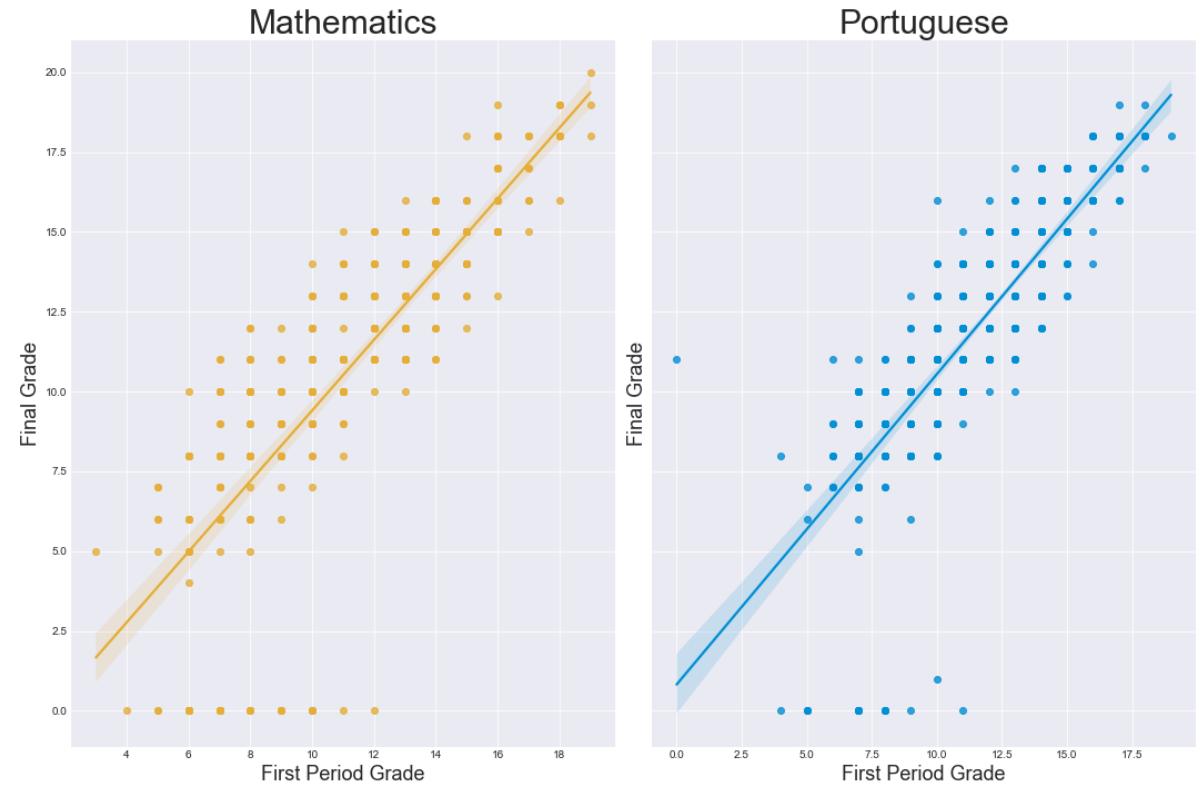
- Normally, people believe that student attendance (or absence) is a good indicator of student performance.
- However, based on both data sets (Mathematics and Portuguese), there seems to be no correlation between students' absence and final grade.



QUESTION 9:

Is there any correlation between student's midterm grade and final grade?

There is a positive linear correlation between first-period grade and final grade.



There is a positive linear correlation between second-period grade and final grade.

