Secondary Education
Portuguese & Mathematics
Student Performance

Mo Alkhudhayri, Max Lake, Ibrahim Shubayli, Pranav Vogeti, Kendall Walch

Factors in Study

Academics

- Student's school
- Weekly study time
- Number of past class failures
- Extra educational support
- Family educational support
- Extra paid classes within course subject
- **First period grade**
- **Second period grade**
- **Final grade**

Family

- Family size
- Parent's cohabitation status
- Mother's education
- Father's education
- Mother's job
- Father's job
- Student's guardian

Demographics

- Student's sex
- Student's age
- Student's home address type
- Reason to choose this school

Lifestyle

- Extra-curricular activities
- Internet access at home
- In a romantic relationship
- Quality of family relationships
- Free time after school
- Going out with friends

Health

- Workday alcohol consumption
- Weekend alcohol consumption
- Current health status
- Number of school absences

Research Question

What academic, familial, cultural, and environmental factors influence final exam (G3) performance across the Mathematics and Portuguese subject studies?

Why Target The Final (G3) Only?

- **G1** (first period grade) and **G2** (second period grade) reflect similar influencing factors
- Focusing on **G3** avoids redundancy and directly evaluates final performance
- Our objective is to identify the factors that influence student's final grades, focusing on elements that directly impact their outcomes rather than intermediate assessments

Final Grade Vs. Failures

Multiple failures don't significantly lower average grades, suggesting other factors may play important roles
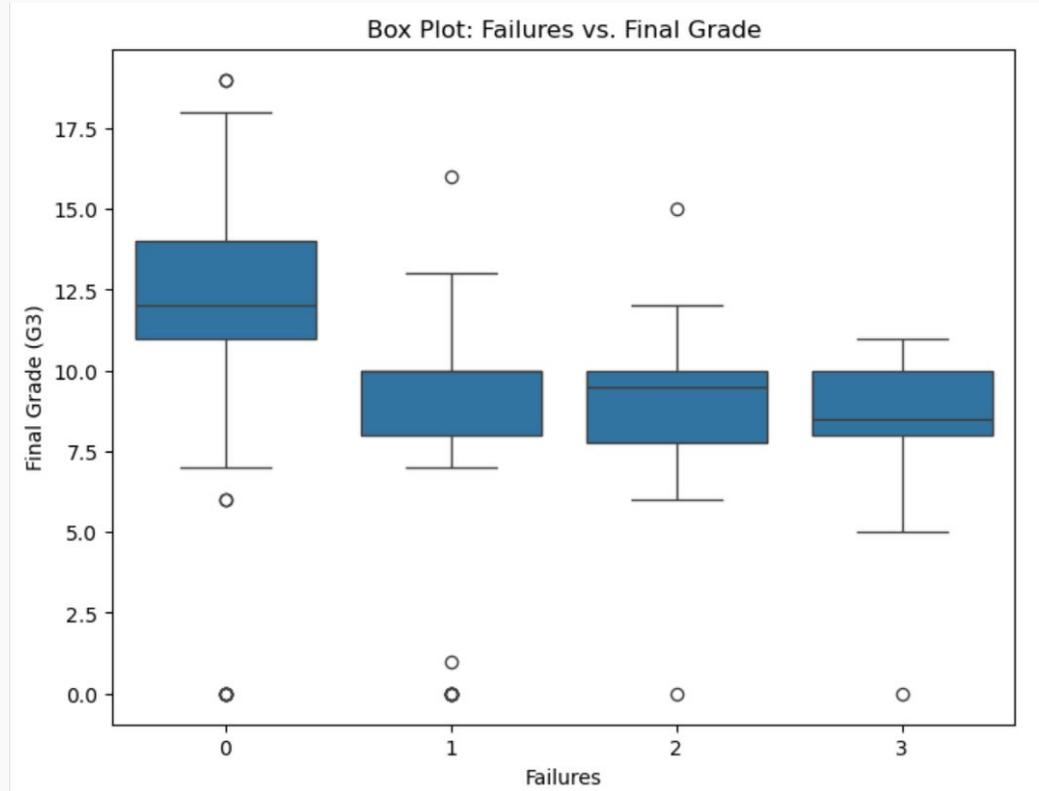
Final Grade Vs. School

The MS school had a smaller sample size compared to GP school.
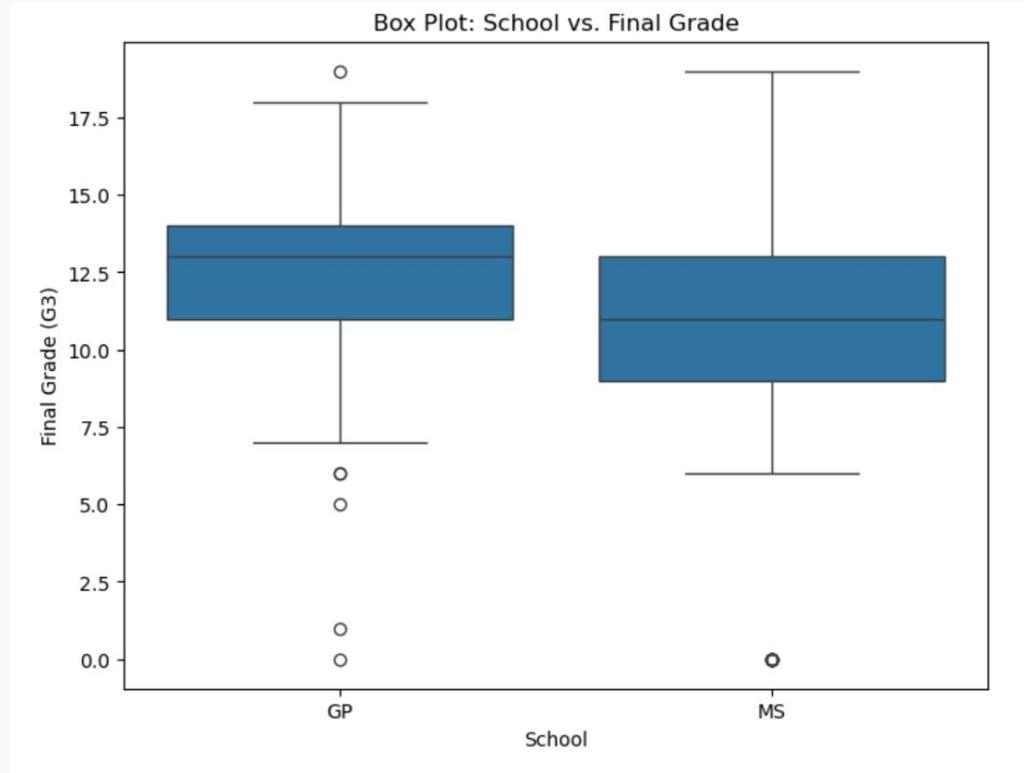
Final Grade Vs. Sex
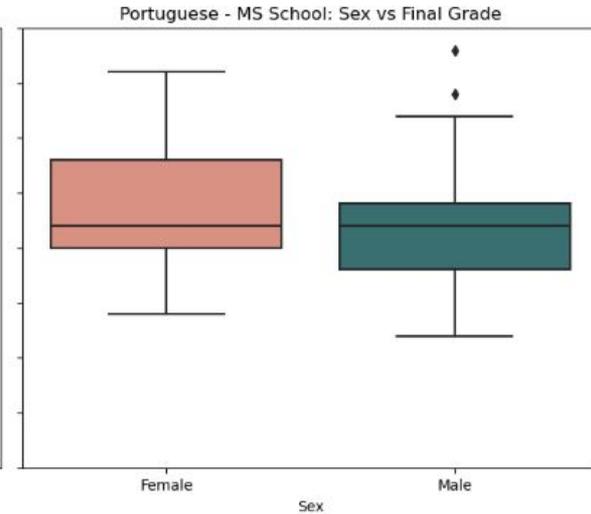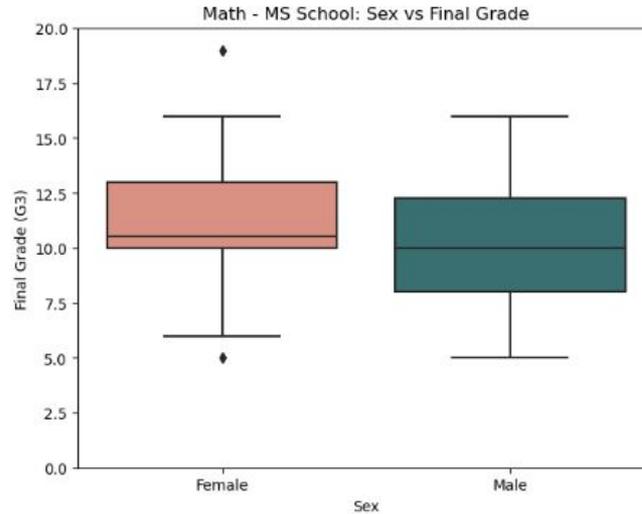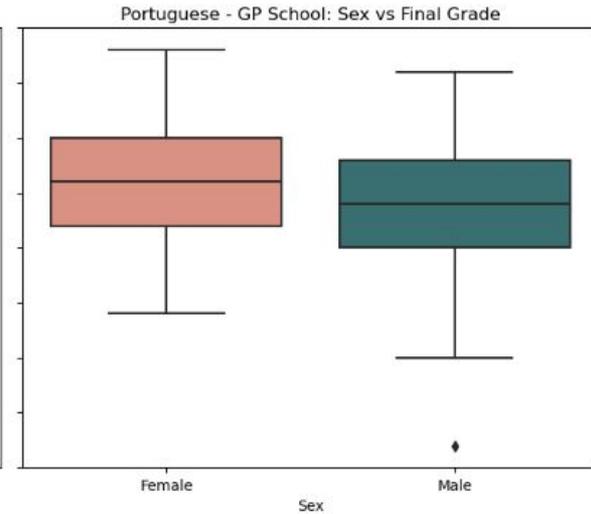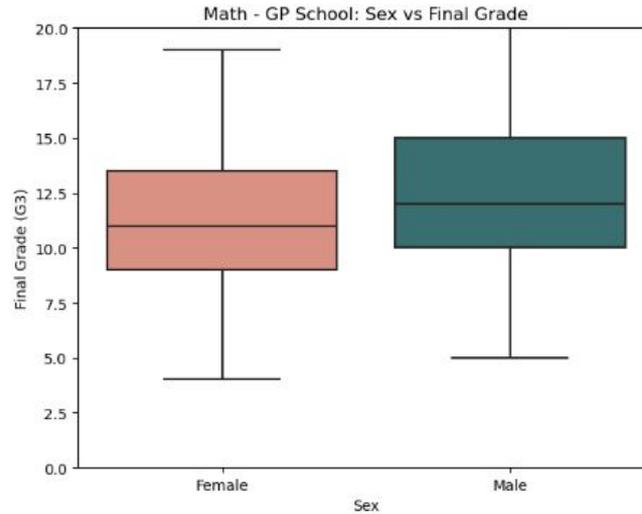
Final Grade Vs. Parent Occupation

-1.00 indicates the occupation combination does not exist
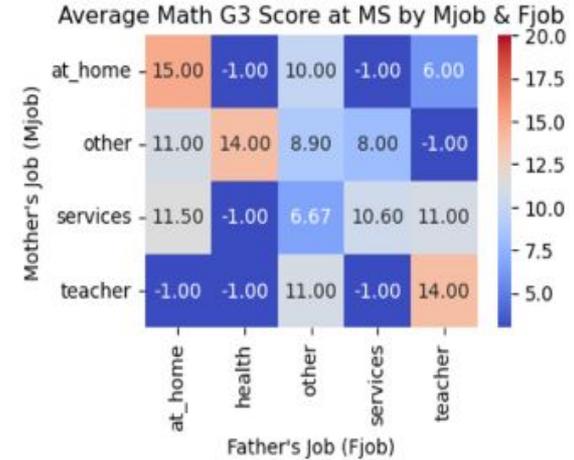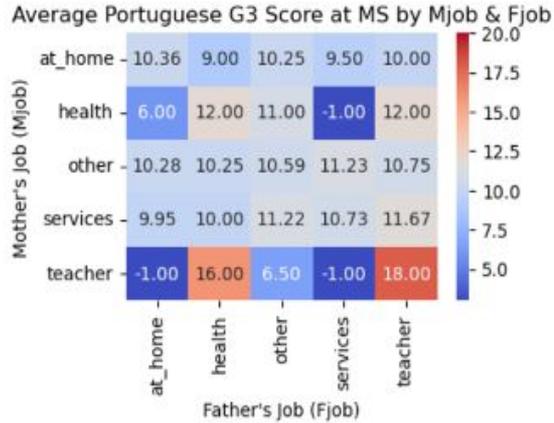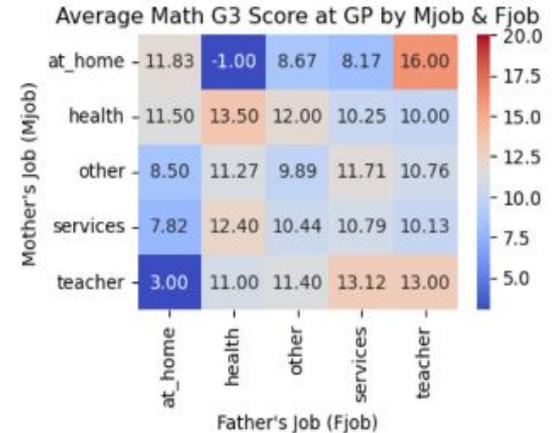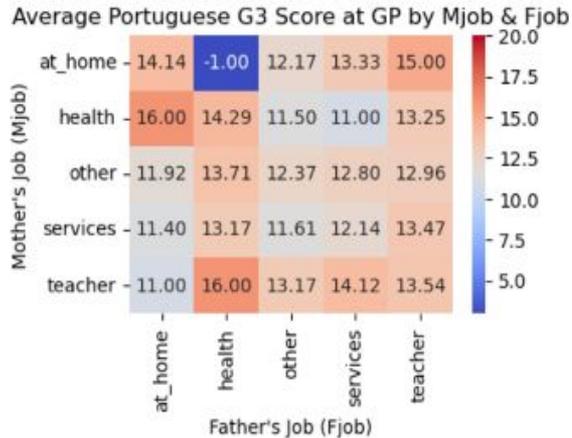
Combined Permutation Feature Importance

Top Features by Permutation Importance
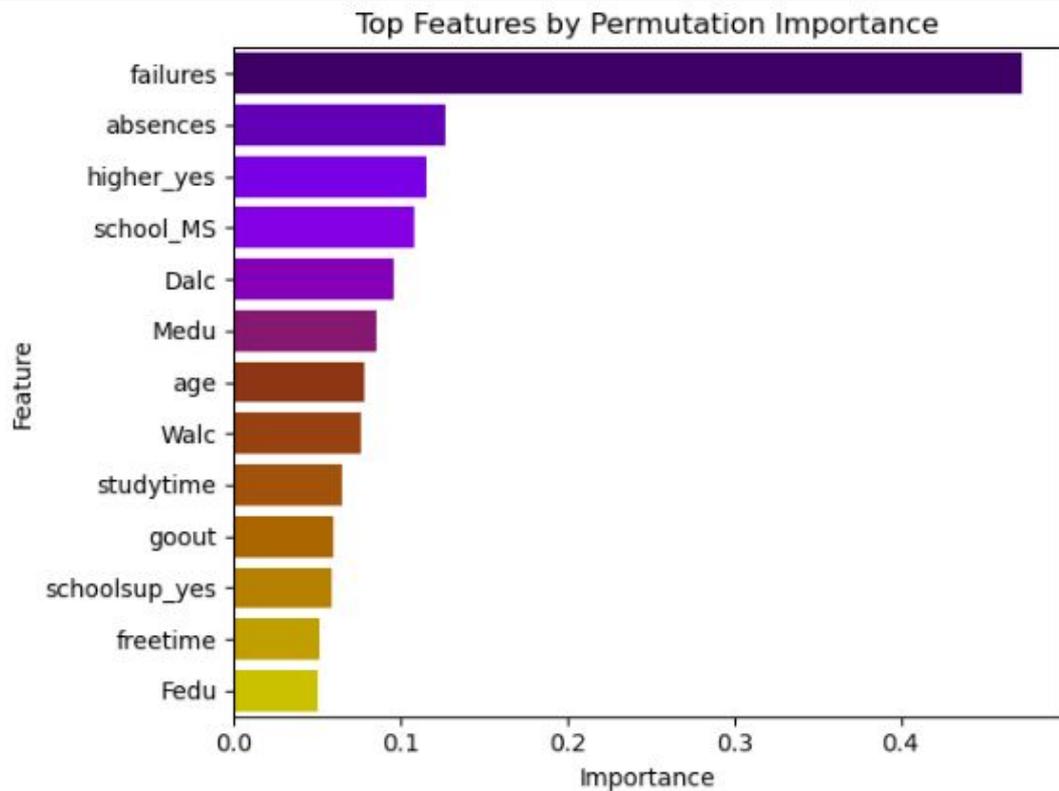
Feature	Importance
failures	0.47
absences	0.13
higher_yes	0.12
school_MS	0.11
Dalc	0.10
Medu	0.09
age	0.08
Walc	0.08
studytime	0.06
goout	0.06
schoolsup_yes	0.06
freetime	0.05
Fedu	0.05

RMSE on Training Data: 1.07
RMSE on Test Data: 2.46
Accuracy on Training Data: 0.92
Accuracy on Test Data: 0.83

- Measures how much a feature impacts the model's predictions by randomly shuffling its values and observing the change in performance
- Larger changes indicate greater importance

# LASSO Regression Combining School and Subject

LASSO-Selected Features
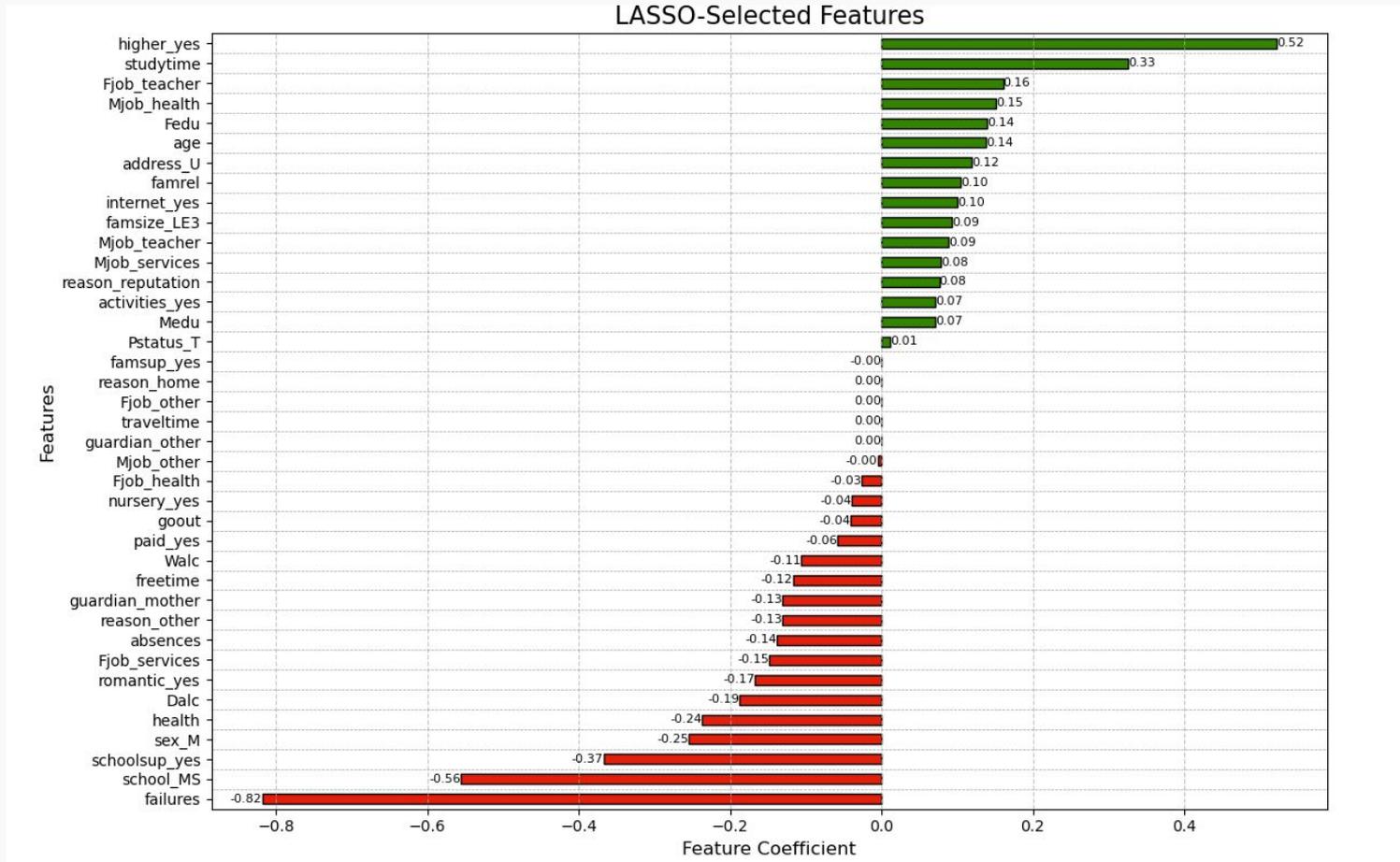
Features Affecting Final Math Grade

Math - GP School (LASSO)

Math - MS School (LASSO)
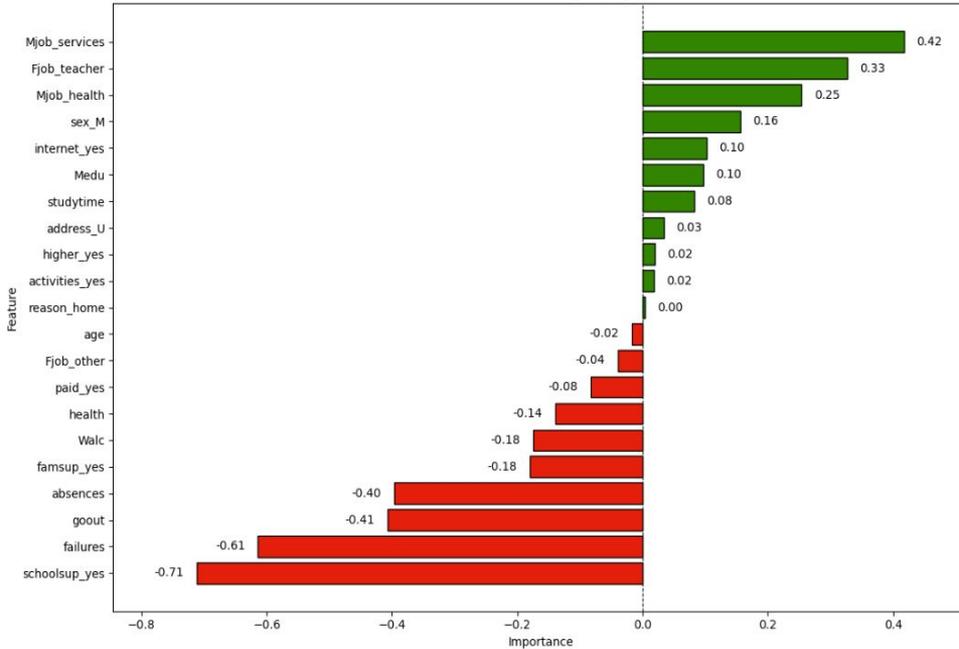
Results for Math-GP:
RMSE (Train): 2.67
RMSE (Test): 2.96
Accuracy (Train): 0.77
Accuracy (Test): 0.79

*Low test accuracy in MS possibly a result of smaller sample size*
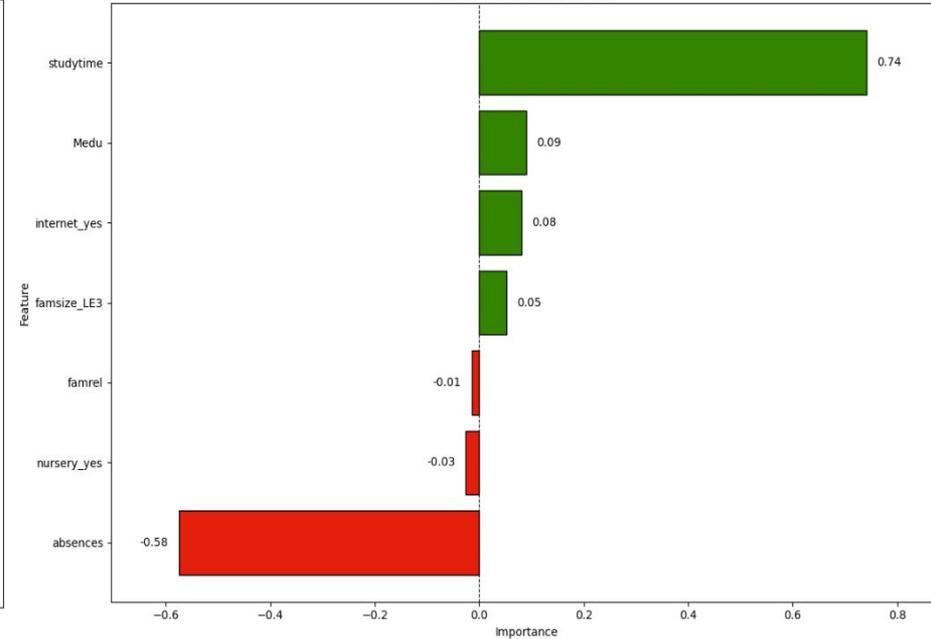
Results for Math-MS:
RMSE (Train): 2.54
RMSE (Test): 2.48
Accuracy (Train): 0.82
Accuracy (Test): 0.56

Features Affecting Final Portuguese Grade

Portuguese - GP School (LASSO)

Feature	Importance
higher_yes	0.51
age	0.32
studytime	0.22
activities_yes	0.19
Mjob_health	0.17
Fjob_teacher	0.17
reason_reputation	0.16
Medu	0.12
address_U	0.08
internet_yes	0.07
reason_home	0.07
Mjob_services	0.06
Mjob_teacher	0.02
freetime	-0.01
Fjob_other	-0.05
guardian_other	-0.05
Mjob_other	-0.07
Pstatus_T	-0.07
paid_yes	-0.08
traveltime	-0.11
Dalc	-0.11
romantic_yes	-0.13
goout	-0.23
Fjob_services	-0.23
health	-0.24
absences	-0.29
sex_M	-0.31
schoolsup_yes	-0.50
failures	-0.55

Portuguese - MS School (LASSO)

Feature	Importance
Fjob_teacher	0.34
higher_yes	0.33
Medu	0.28
Mjob_teacher	0.24
famrel	0.15
traveltime	0.10
studytime	0.10
Fedu	0.10
reason_reputation	0.09
Mjob_other	0.07
age	0.03
Fjob_health	-0.06
reason_other	-0.08
Mjob_health	-0.10
guardian_mother	-0.16
famsup_yes	-0.25
Walc	-0.31
absences	-0.41
failures	-0.70

Results for Portuguese-GP:
RMSE (Train): 1.97
RMSE (Test): 2.27
Accuracy (Train): 0.94
Accuracy (Test): 0.92
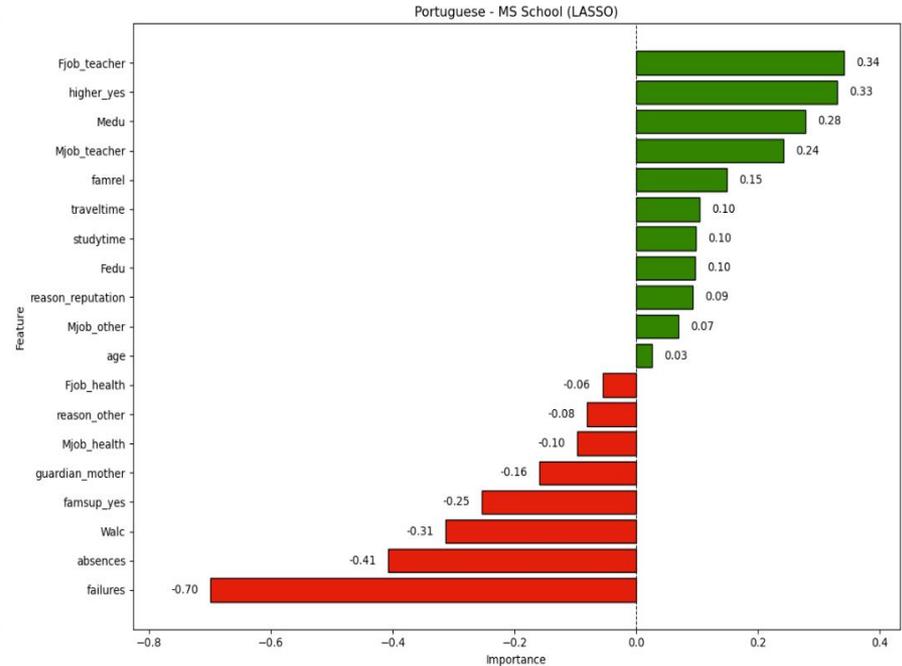
Results for Portuguese-MS:
RMSE (Train): 2.20
RMSE (Test): 2.38
Accuracy (Train): 0.78
Accuracy (Test): 0.79

# Decision Tree Visualization

Decision Tree: Pros and Cons

| Advantages	Disadvantages
— Does not require normalizing or scaling the data — Works with categorical and numerical data — Is easy to understand and interpret	— Can lead to overfitting — Is highly sensitive to slight changes in data — Struggles with highly nonlinear or complex datasets

Main Takeaways

- Failures and absences are consistently the most impactful factors across all schools and subjects
- When analyzed separately, additional features such as study time, family support, travel time/transportation, intention of pursuing higher education, and other external characteristics emerge as important influencers
- Indicates that school board strategies should be adapted to address the unique challenges/characteristics of each school and subject

Mouzinho da Silveira (MS) School

- Study time is lower on average
- Final grades are lower on average
- Lower predictability based on the data
- Only 11% of students are from MS for math data
- Only 35% of students are from MS for Portuguese data

Average Study Time by School

Average Final Exam Score by School
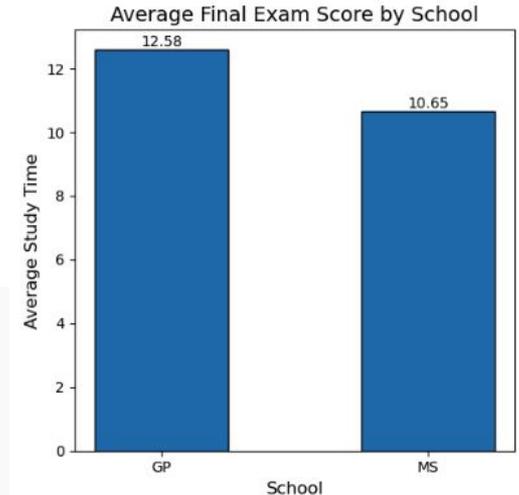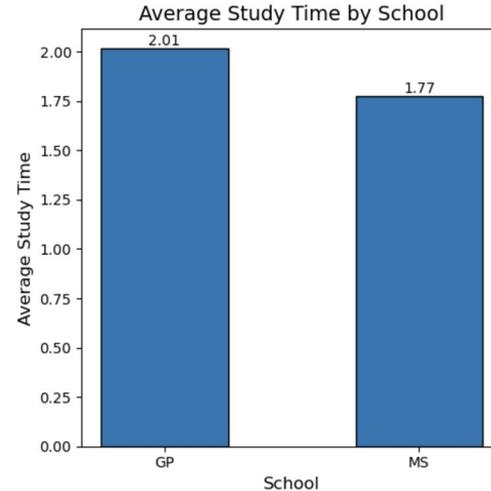
School Board Strategies to Help Student Performance

Failures
Implement programs early on to help students who have already struggled in the past

Higher Education
Promote college-prep programs and emphasize importance of higher education

Studytime
Promote after school programs or convenient tutoring sessions

Absences
Enforce stricter attendance policies and address issues pertaining the students

Parental Involvement
Engage parents through meetings about their child's progress and promote the value of education

Decreases Grades
Increases Grades

Future Topic Recommendations

- How do the Parent Jobs in terms of sectors and salary impact student performance?
- How would this analysis look chronologically? (i.e. absences go up/down over the three exams: G1, G2, G3)
- Is the overall lower level of MS school compared to GP school a result of the smaller sample size, or does it indicate an overall trend? Investigate potential factors contributing to this discrepancy.

Thank You