Predicting Student Performance

November 27, 2017

Steven Brotz, Michael Li, Junno Tseng, Andrew Wang

Dataset Overview

- Performance data of high school students of two Portuguese schools
  - Two subjects: math (395) and portuguese (649)
- Data collected via school reports and questionnaires in 2008
- 30 attributes for each student
  - Demographic, social and school related features, activities, etc
- Three grades
  - First period (G1), second period (G2), and final grade (G3)
- Imputation of missing values using a weighted KNN approach

Understanding the problem

Motivation

- Schools often funded based on performance
- Enable schools to determine optimal allocation of limited resources
- Early identification of potential students at risk

Predicting Performance

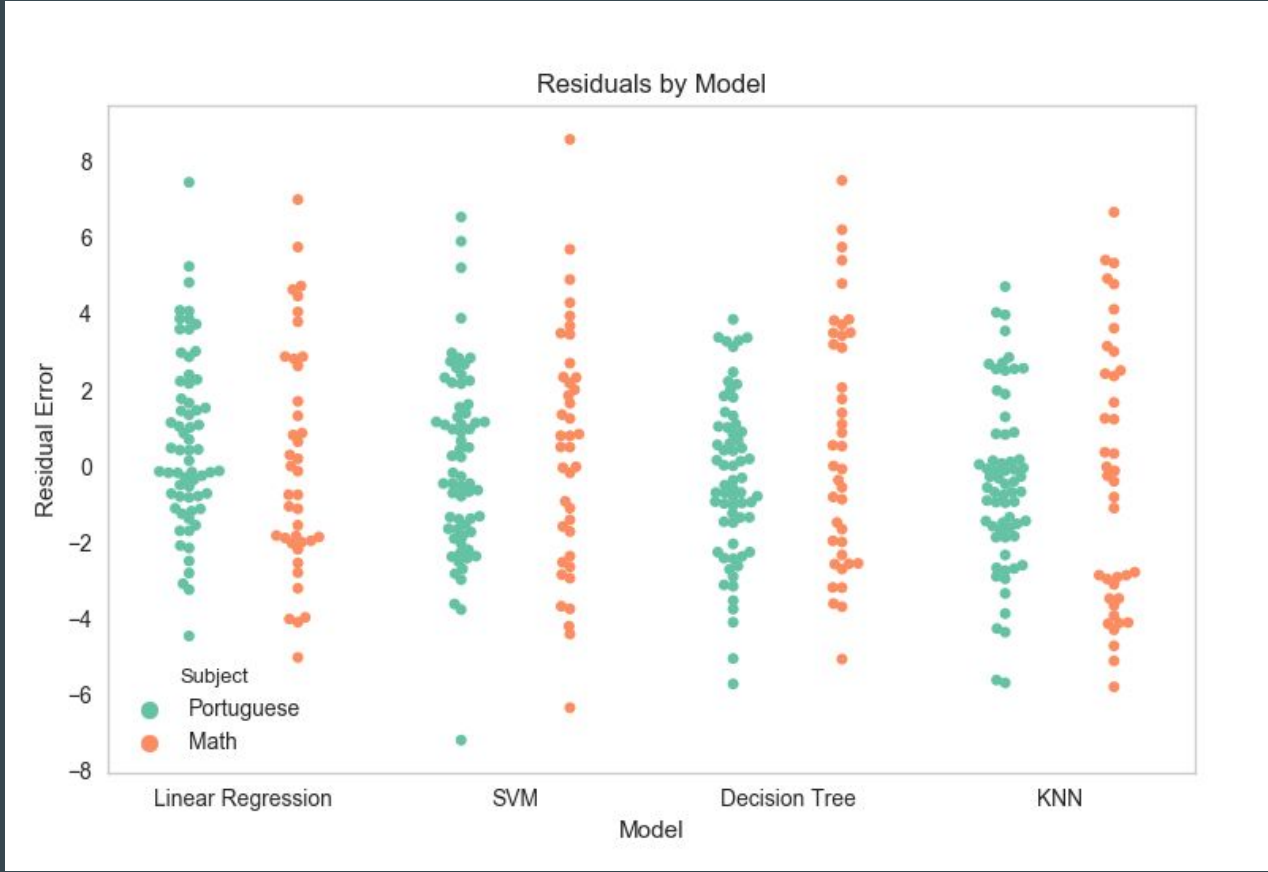
- Which factors most determine student performance?
- Can we predict performance over time?

Understanding alcohol

- Drinking age of Portugal: 18
- How does drinking affect grades? Is this effect significant?
- Weekday vs. weekend drinking?

Primary objective:
Predict academic performance so schools can give personalized learning experiences to all students.

Model Comparison for G1 grades



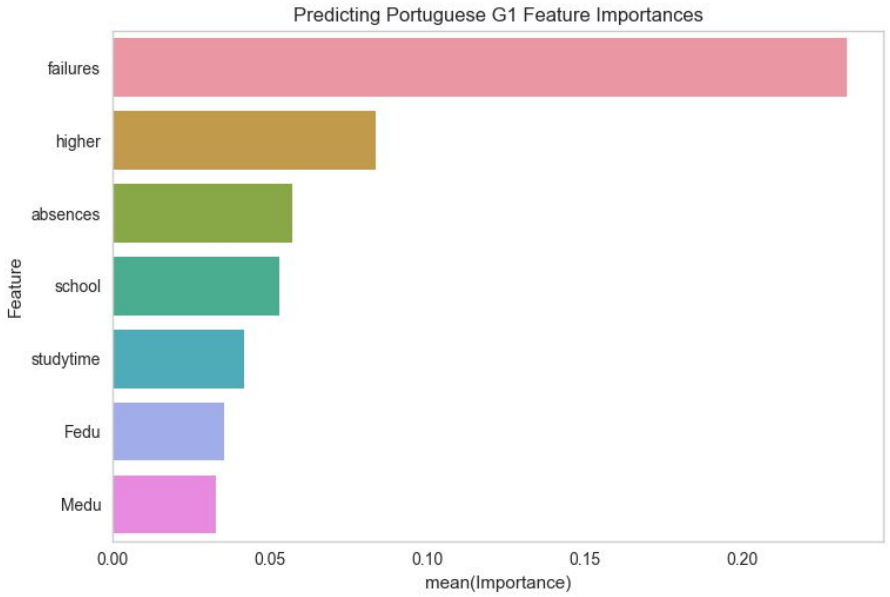
Portuguese

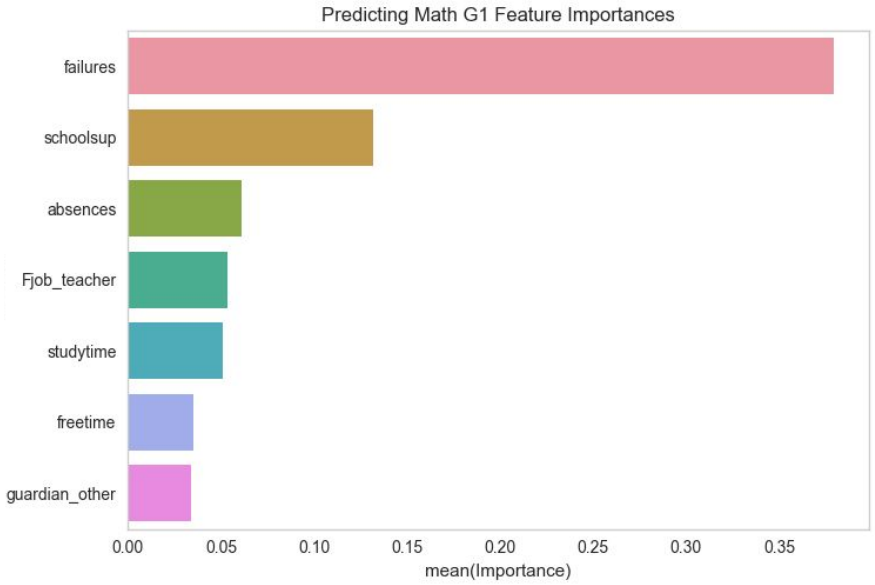
Model	R^2	*Mean squared error*	*Mean absolute error*
Naive (mean/median)	0.00	7.35	2.18
KNN	0.05	6.53	2.05
SVM	0.11	5.72	1.88
Linear regression	0.16	5.40	1.82
Regression Trees	0.21	5.26	1.80

Math

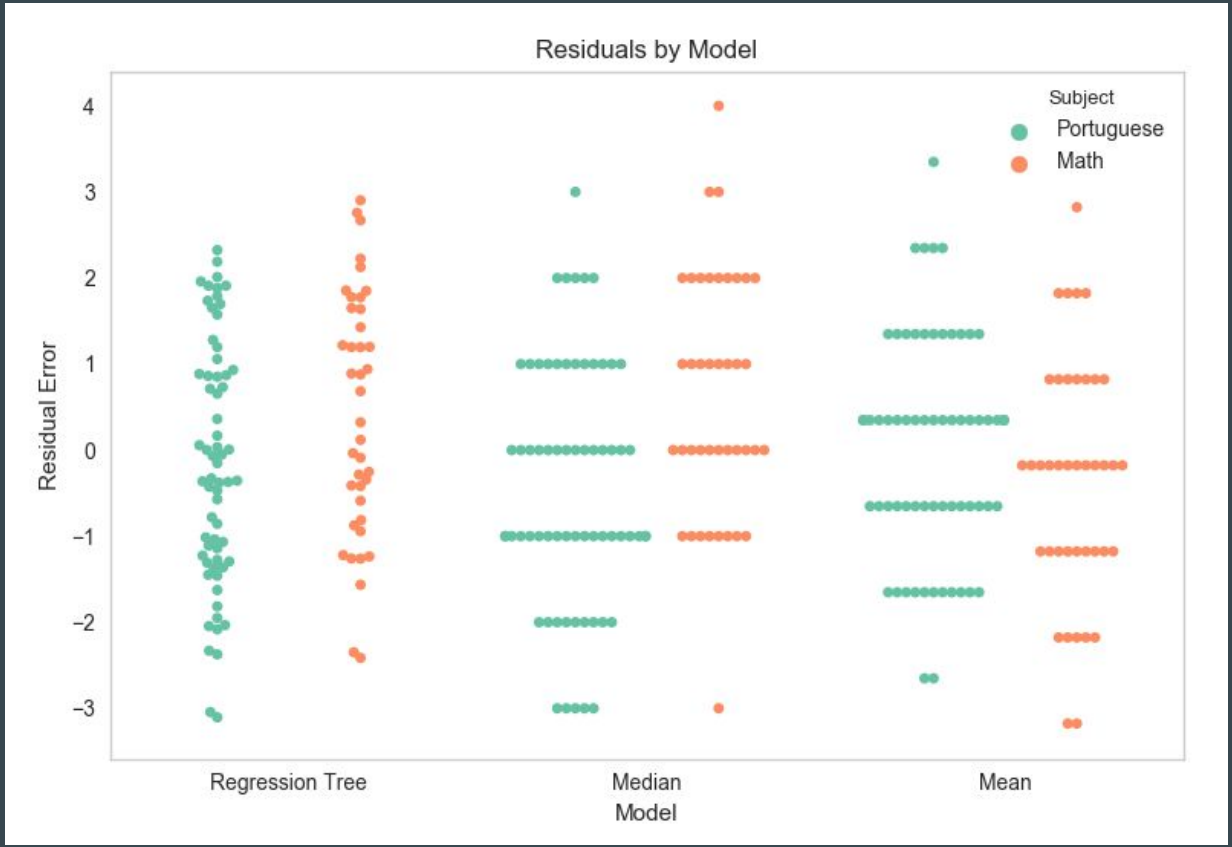
Naive (mean/median)	0.00	10.99	2.75
KNN	0.00	10.76	2.74
SVM	0.01	10.41	2.64
Linear regression	0.10	9.70	2.53
Regression Trees	0.12	9.42	2.53

Predicting G1 Feature Importance



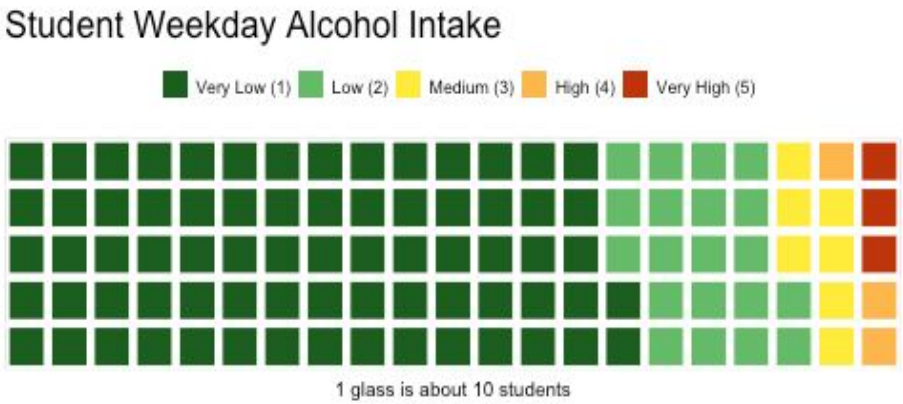


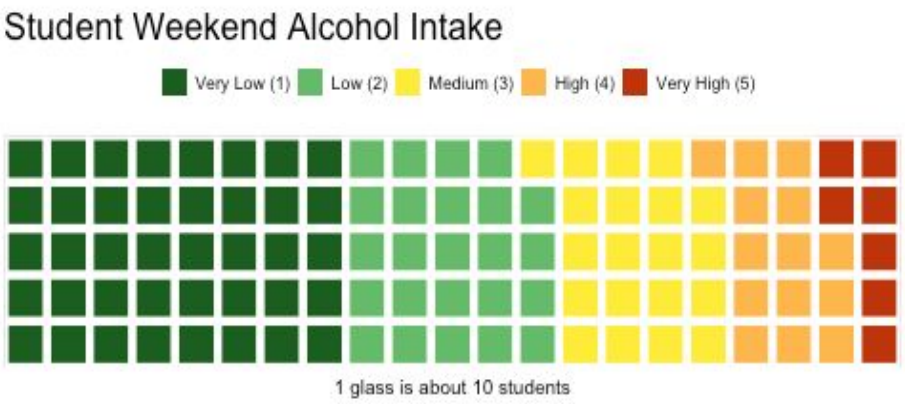
Predicting G1 - G3: Ineffective



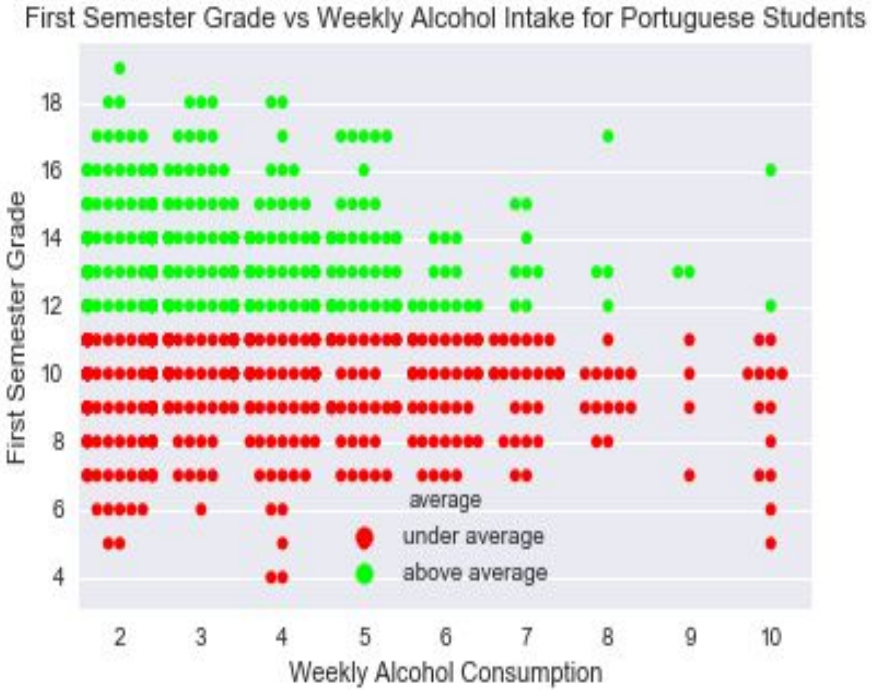
Secondary objective: Drinking vs. student performance

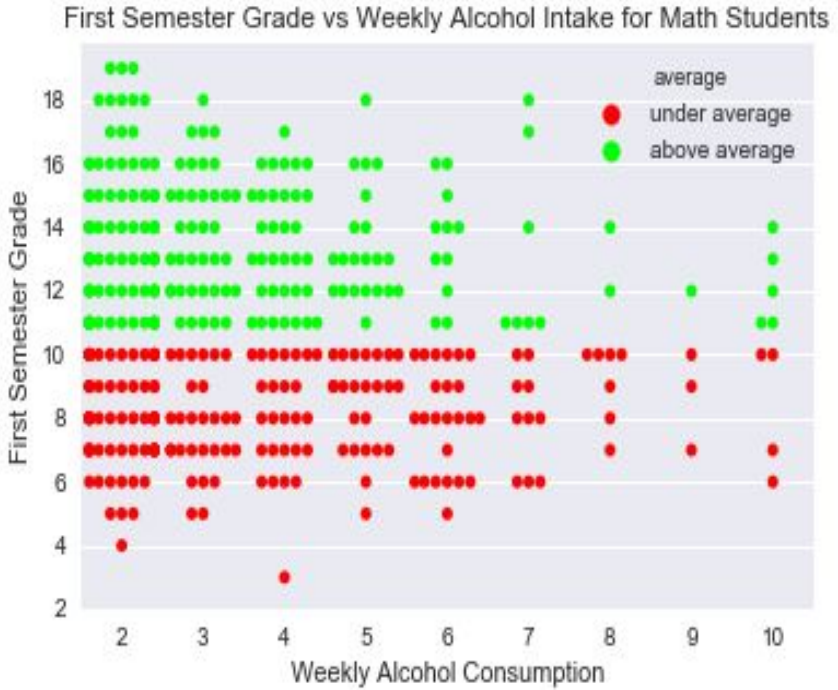
Distribution of Drinking Levels



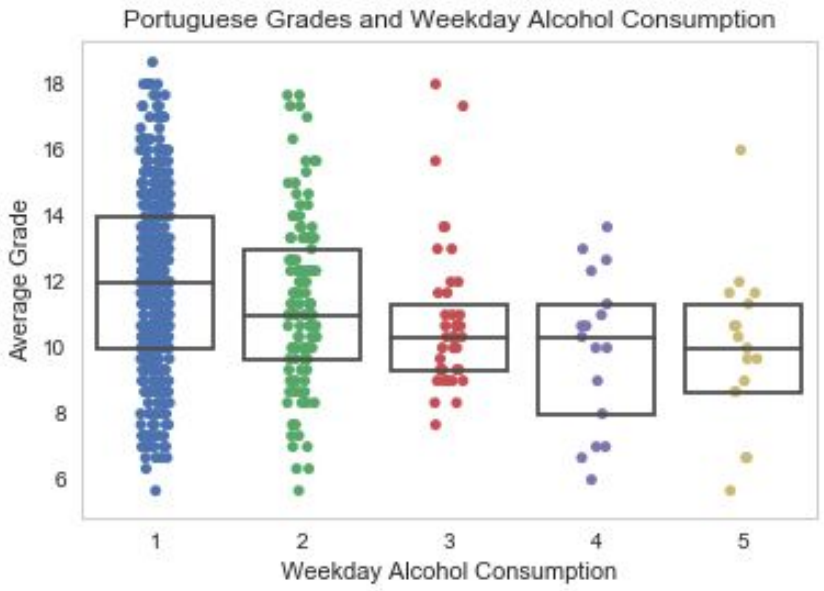


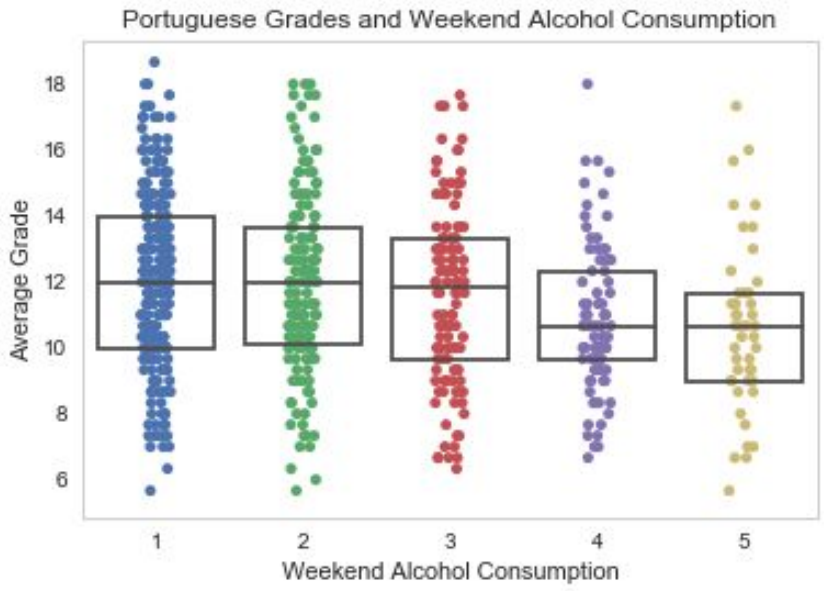
G1 Distribution by Weekly Alcohol Intake



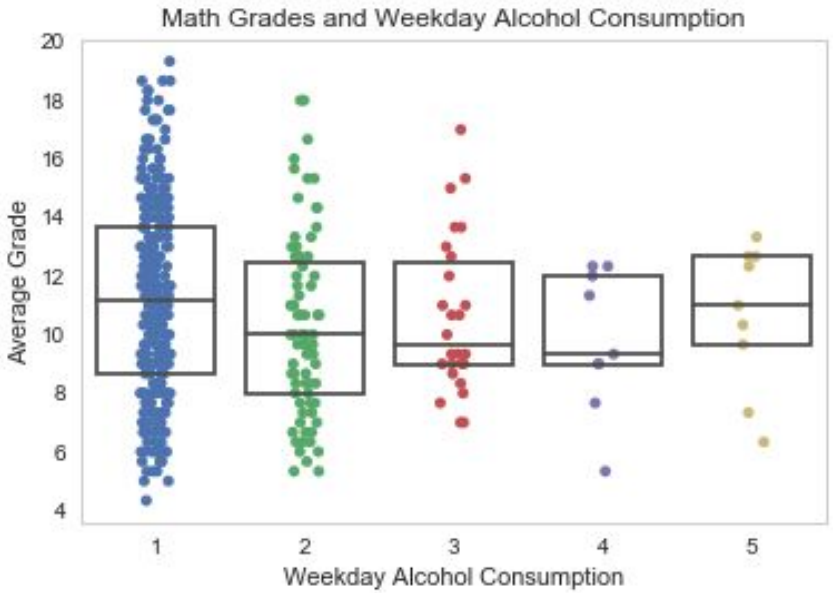


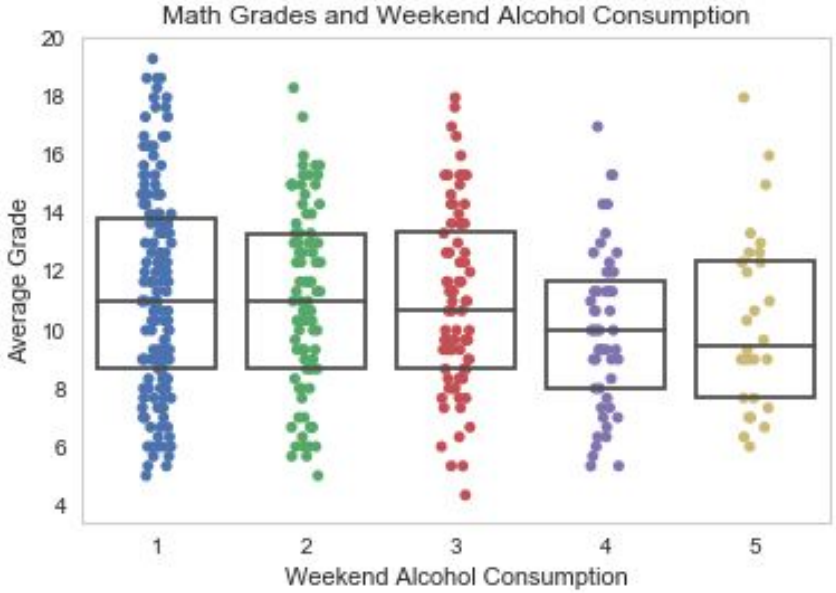
Portuguese : Drinking and Average Grades





Math: Drinking and Average Grades





Chi-Square Test of Independence of G1 vs Alcohol Intake

Null hypothesis: G1 and alcohol intake are independent
0.05 significance level

Model	*P-value*	***Independent***
Weekday Intake	0.01225	**No**
Weekend Intake	0.2391	**Yes**
Total Intake	0.2059	**Yes**

Key Takeaways

Predicting performance

- MAE =1 .80: within 10% of true grade
- Previous class failures are significant
- Tutoring / additional support more important in math than portuguese
- Mother and father's education more important for language than math
- Predicting improvement/ decline is very difficult

Understanding alcohol's academic impact

- More weekly drinking, the more likely to score below average
- Weekday drinking matters, weekend not so much
- Very low -> low drinking for weekday is a big deal