

Thursday, March 26th, 2026

Ante-Join

# Student Performance Based on Various Factors

---

Ethan Huang, Hugo Hill, Anis Ahmed, Leo Bai

## Why student performance?

- We are an all-freshmen group adjusting to college academics and finding strategies to maintain good grades
- We wanted to explore what factors can most or least affect our performance on exams

### RESEARCH QUESTION:

Which factors are most strongly associated with higher student exam performance?

# THE DATASET

## - Dataset & Variables:



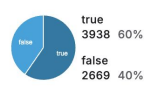
Kaggle; focused on 6 key variables

## - Outcome Measure:

Exam scores put in two groups: high ( $>70$ ) and low ( $\leq 70$ )

## - Data Preparation:

Cleaned and standardized variables, removed missing categorical data, and reorganized categories.

# Hours_Studied	# Attendance	△ Parental_Involve...	△ Access_to_Resou...	✓ Extracurricular_A...
Daily study hours	Class attendance percentageClass attendance percentage	Parents' academic support	Learning resource availability	Participation in activities
		Medium 51%	Medium 50%	
1	60	High 29%	High 30%	true 3938 60%
44	100	Other (1337) 20%	Other (1313) 20%	false 2669 40%
23	84	Low	High	No
19	64	Low	Medium	No
24	98	Medium	Medium	Yes
29	89	Low	Medium	Yes
19	92	Medium	Medium	Yes
19	88	Medium	Medium	Yes
29	84	Medium	Low	Yes
25	78	Low	High	Yes
17	94	Medium	High	No
23	98	Medium	Medium	Yes
17	80	Low	High	No
17	97	Medium	High	Yes

## Numerical Variables

**Hours Studied** → Time spent studying (1–44 hrs/week)

**Class Attendance** → % of classes attended (60–100%)

**Hours of Sleep** → Avg. sleep per night (4–10 hrs)

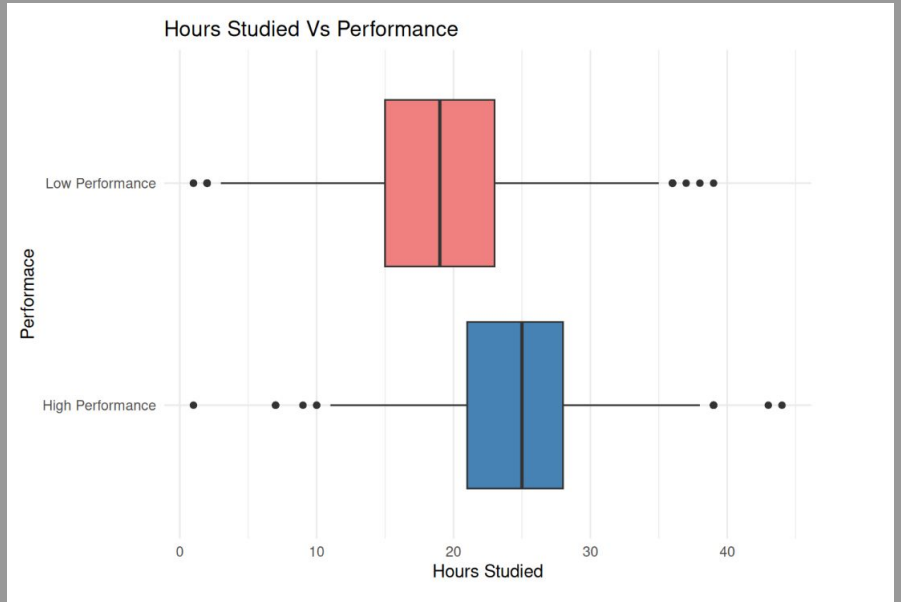
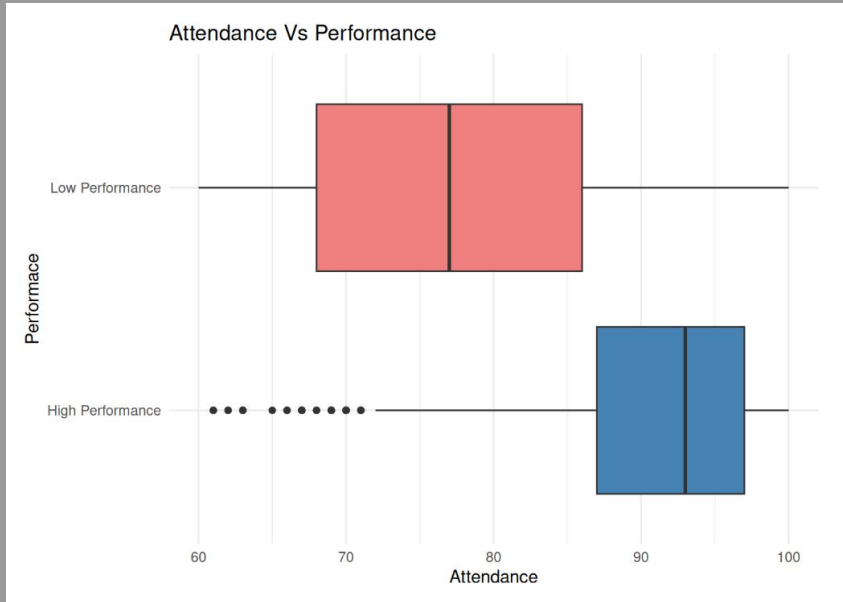
## Categorical Variables

**Parental Involvement** → Low / Medium / High support

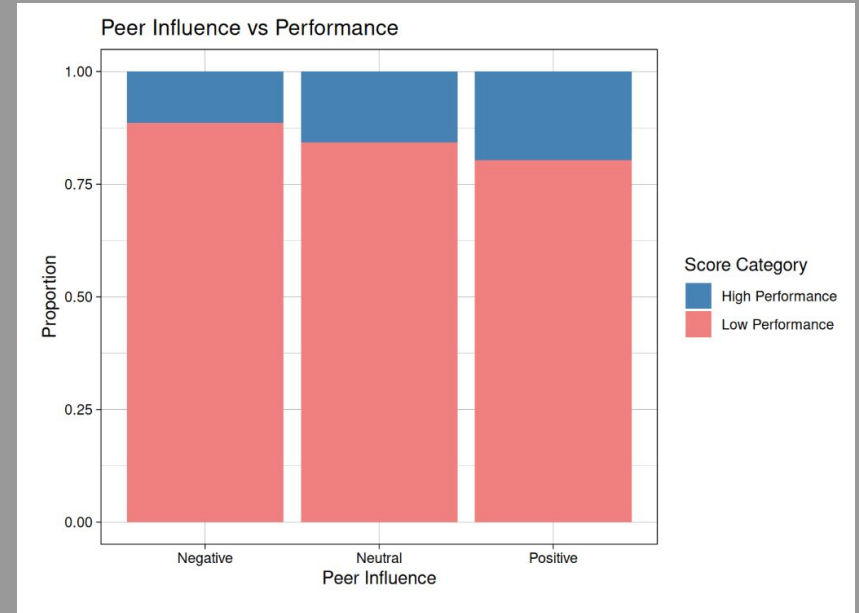
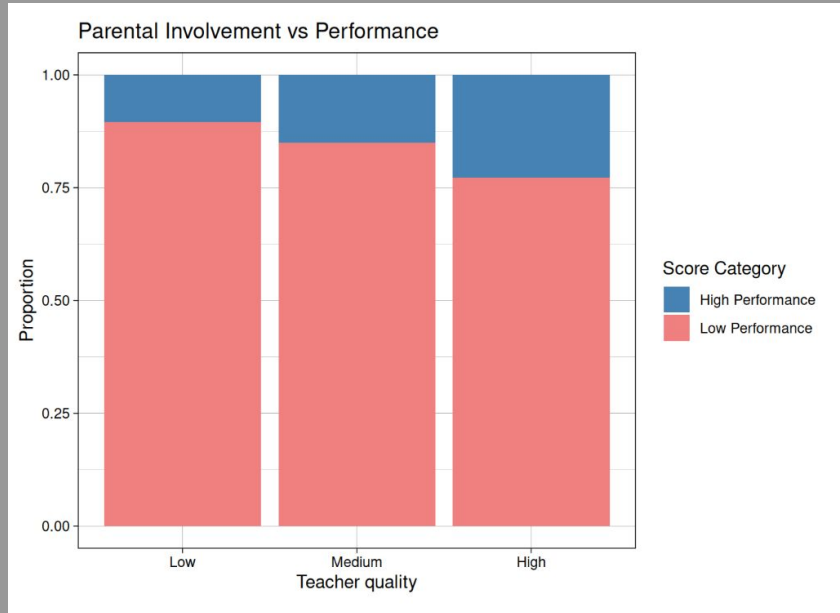
**Teacher Quality** → Low / Medium / High effectiveness

**Peer Influence** → Negative / Neutral / Positive influence

From the numerical variables we focused on, attendance and hours studied had the largest impact:



From the categorical variables we focused on, parental involvement and peer influence had the largest impact:



# Overall Findings

## Numerical Variables:

- **Attendance** → highest adjusted  $R^2 = 0.3375$ 
  - High-performance students: 91% attendance
  - Low-performance students: 78% attendance

## Categorical Variables:

- **Parental Involvement** → Largest difference
  - High parental involvement: 23% high-performing students
  - Low parental involvement: 10% high-performing students

# THINGS TO CONSIDER

## Socioeconomic Status

Socioeconomic status can influence multiple variables at once, such as access to better teachers, more parental support, and healthier sleep routines, thereby inflating their apparent individual impacts

## Self-Reported Measurement Issues

Since we are not given much information on the background of the data, we are not sure of how reliable it is. The methods, sample, and design of the survey were not listed, so the accuracy of results could be confounded

## Overlap in Results

Strong students potentially choose better peers, attend class more regularly, and study more, making it hard to examine the effects of each individual variable alone. A lot of variables simultaneously play a role in impacting student performance.

## Biased Perceptions

Students may have biased perceptions of their teachers or parents based on the fact that they are being “forced” to do work, particularly if they are younger students

THANK YOU