

Student Success: Factors and Insights

Huiting Wu, Eli Singh, Evan Nagesh

2025-10-10

Introduction

Since academic success is something that three of us are completely unfamiliar with, we wanted to learn about some of the most important factors that contribute to it. We sought to understand the interplay between the variables that impact final exam scores and discover which variables had the strongest association with exam scores. For categorical variables, we looked at the frequency of high-performing students in each category. For quantitative variables, we looked at averages and scatter plots to assess which quantitative variables had the strongest relationship with exam scores.

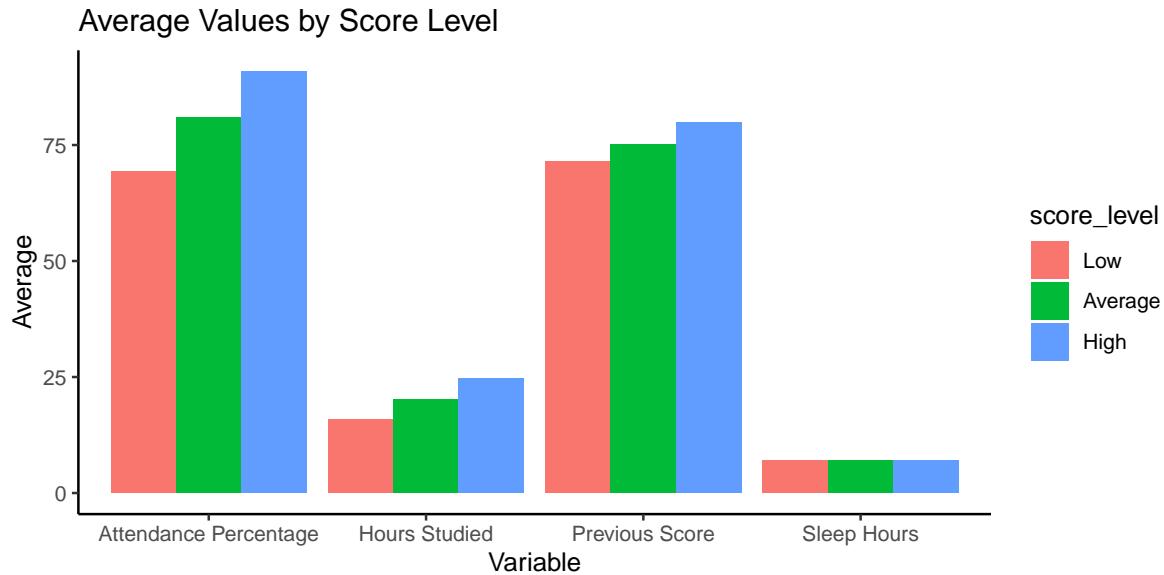
Data Description

The dataset was obtained from Kaggle. It was obtained from a simulated student performance survey designed to explore the relationships between study habits, motivation, and environmental factors on academic outcomes. The dimensions of our dataset are 6607 rows by 20 columns. Variables of interest were Previous_Scores, Exam_Score, Attendance, Hours_Studied, Sleep Hours, Tutoring Sessions, and Physical Activity. Exam_Score was our primary variable of interest which we treated as the outcome variable. The other variables were all categorical variables, as noted in the introduction. Previous_Scores was the result of previous academic assessments with a maximum grade of 100. Attendance was a percentage of class sessions attended. Tutoring sessions recorded the number of weekly tutoring sessions attended by each student. Physical Activity measured the number of hours per week of physical activity that were conducted by each student. Hours_Studied represents the number of hours per week a student dedicated to studying. Sleep Hours recorded the number of hours of sleep a student was getting per night. Parental_Involvement indicates the level of parental engagement in the student's education and has three levels: High, Medium, and Low. Peer_Influence represents the degree of influence peers have on the student, categorized as Positive, Neutral, or Negative. Family_Income reflects the income level of the student's family or father and is classified as

High, Medium, or Low. Motivation_Level describes the student's level of motivation toward academic work, also categorized as High, Medium, or Low.

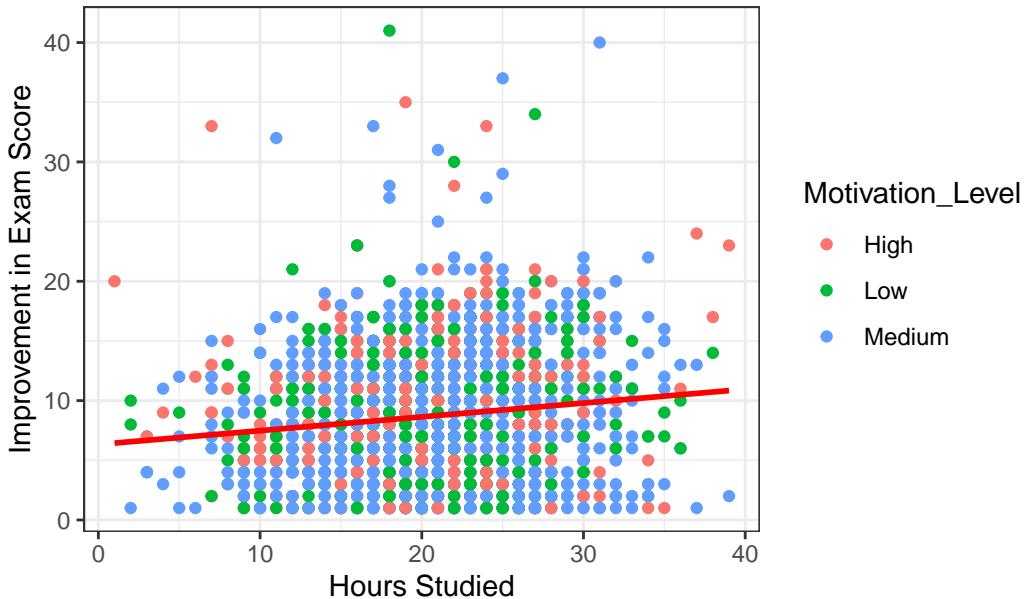
Results

To examine the relationship between the quantitative variables in our dataset and the final exam score, we decided to group scores into three categories: low, average and high. The thresholds for these three categories were selected by looking at the distribution of exam scores and selecting values that would ensure a good amount of data fell into each category. Ultimately, we selected 71 or higher for the level "high", 65 through 70 for the level "average", and 64 or below for the level "low". This selection method guaranteed at least 1000 data points in each category. From there, we examined the averages for some of our most important quantitative variables for each category. Our results are shown below.



The high scorers had the highest average attendance percentage followed by the average scorers. High scorers also had the highest average number of hours studied and the highest average previous score. In contrast, low scorers had the lowest average numbers of average hours studied and the lowest average previous score. The average number of hours slept was pretty consistent across the different score levels.

Hours Studied vs. Improvement in Exam Score



This scatterplot shows the relationship between hours studied and improvement in exam score among the students who performed better on the final exam than the previous exams. The color of the data points corresponds to the students' self-reported motivation level. Overall, there is a slight positive correlation between hours studied and score improvement, indicated by the trendline. On average, students who study more tend to have a larger improvement in their exam scores, though the effect is relatively weak. The spread of the data points shows that there is a large amount of variation in this sample, and several students had large gains in their exam scores even with relatively fewer hours spent studying. Motivation level does not appear to have a strong or consistent effect, with all three groups spread pretty evenly throughout the plot.

For further study, we were interested in seeing which factors led to the biggest improvements from previous score to final score. To that end, we split our dataset into students that improved their score and students saw their scores decrease.

Table 1: Average of Numerical Factors for Students Who Improved vs. Not Improved

Factor	Average_Improve	Average_Not_Improve
Attendance	82	79
Sleep_Hours	7	7
Hours_Studied	20	20
Tutoring_Sessions	2	1
Physical_Activity	3	3

Table 2: Proportion(Factor: Improved vs. Not Improved)

Factor	Level_Improve	Proportion_Improve	Level_Not_Improve	Proportion_Not_Improve
Parental_Involvement	Medium	50.51	Medium	51.06
Peer_Influence	Positive	42.34	Neutral	39.58
Family_Income	Medium	39.77	Low	40.83

These tables present the most common levels of key categorical factors for students who improved and those who did not. Across both groups, medium parental involvement is the most frequent, though its proportion is slightly higher among non-improved students. Improved students are more likely to experience positive peer influence and come from middle-income families, whereas non-improved students are more often influenced by neutral peers and come from low-income families. While several factors are similar between the groups, differences in peer influence and family income suggest that social and economic contexts contribute to academic improvement. Additionally, students who improved show slightly higher attendance (82% vs. 79%) and participate in more tutoring sessions compared to those who did not improve, indicating that consistent class attendance and engagement in tutoring are associated with better academic outcomes.