

Student Habits and Academic Success: A Behavioral Analysis

Goal: Understand how student lifestyle habits relate to academic performance

Research Question:

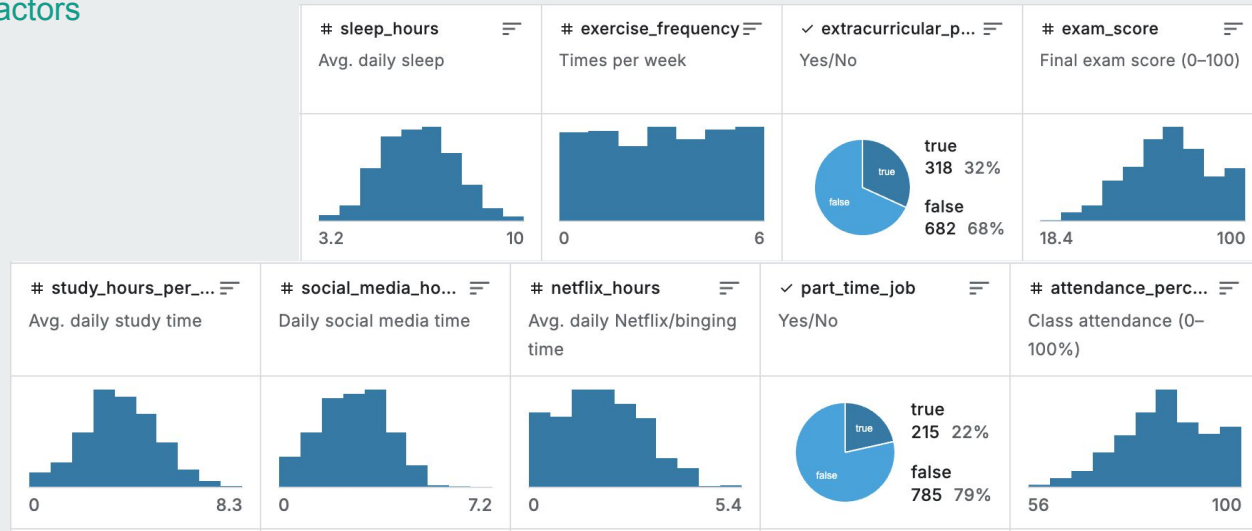
“How do study time, sleep, screen use, and other daily habits affect exam scores?”

Joshua Zhang | Yuyao Huang | Jiawen Yu



Dataset Information

- Kaggle Dataset: <https://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance>
- Sample Size: 1,000 students
- Structure: 16 columns
- Key Variables:
 - study_hours_per_day
 - social_media_hours / netflix_hours
 - sleep_hours / exercise_frequency
 - attendance_percentage
 - diet_quality / mental_health_rating
 - exam_score (target variable)
- Preprocessing: Removing student_id
Converting character columns to factors
Checking for missing values



Methods

We used multiple linear regression to model exam scores as a function of student lifestyle habits.

Started with all 15 behavioral and demographic variables.

Applied stepwise variable selection (via stepAIC) to remove insignificant predictors.

Final model includes 7 significant variables:

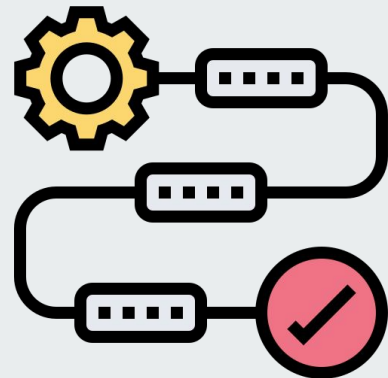
- study_hours_per_day
- sleep_hours
- attendance_percentage
- exercise_frequency
- mental_health_rating
- social_media_hours (–)
- netflix_hours (–)

Tools: R (lm, stepAIC from MASS, broom, ggplot2)

Coefficients:

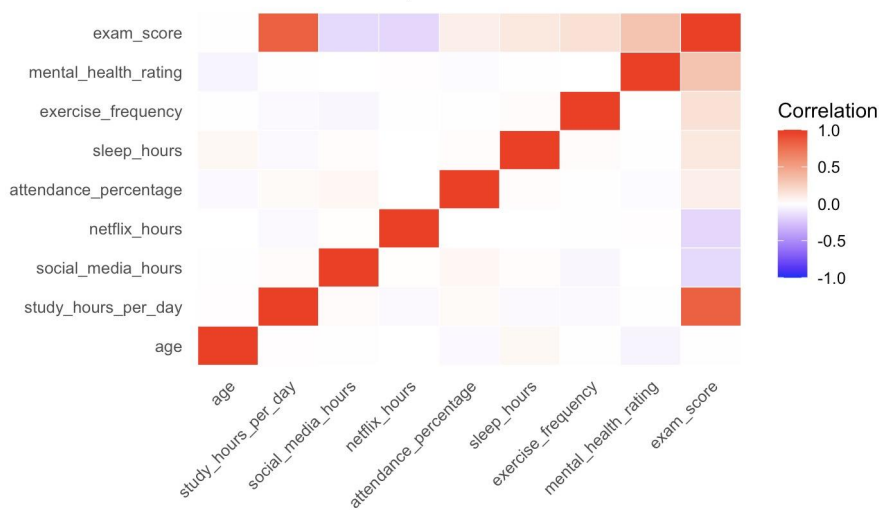
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.15722	1.89252	3.253	0.00118	**
study_hours_per_day	9.57456	0.11503	83.238	< 2e-16	***
social_media_hours	-2.61978	0.14413	-18.177	< 2e-16	***
netflix_hours	-2.27708	0.15697	-14.507	< 2e-16	***
attendance_percentage	0.14473	0.01797	8.054	2.28e-15	***
sleep_hours	2.00462	0.13764	14.564	< 2e-16	***
exercise_frequency	1.45187	0.08338	17.413	< 2e-16	***
mental_health_rating	1.94891	0.05924	32.897	< 2e-16	***

```
library(MASS)
step_model <- stepAIC(model, direction = "both")
summary(step_model)
```



Correlation heatmap

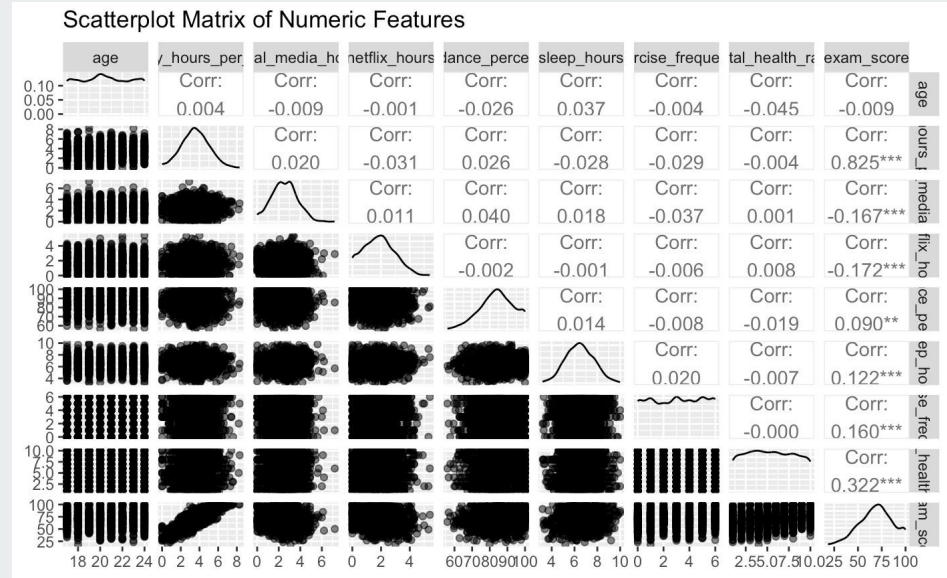
Correlation Heatmap of Numeric Predictors



- **Strong positive:**
 - Study hours ↔ Exam score
 - Attendance percentage ↔ Exam score
 - Sleep hours & Mental health ↔ Exam score
- **Strong negative:**
 - Social media hours ↔ Exam score
 - Netflix hours ↔ Exam score
- **Weak or negligible:**
 - Exercise frequency ↔ Exam score

Scatterplot Matrix Highlights

- **Exam vs. Study Hours:** Clear upward trend—more study, higher scores
- **Exam vs. Screen Time:** Downward slopes for social media & Netflix use
- **Exam vs. Well-Being:**
 - Sleep & Mental Health show mild positive association
 - Attendance slightly positive, exercise diffuse
- **Distributions:**
 - Study hours & exam scores skewed toward higher values
 - Screen-time variables more varied at lower exam scores



Regression Model Performance

```
lm(formula = exam_score ~ study_hours_per_day + social_media_hours +  
  netflix_hours + attendance_percentage + sleep_hours + exercise_frequency +  
  mental_health_rating, data = df)
```

Residuals:

Min	1Q	Median	3Q	Max
-21.9509	-3.3953	-0.0283	3.6680	15.9059

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.15722	1.89252	3.253	0.00118 **
study_hours_per_day	9.57456	0.11503	83.238	< 2e-16 ***
social_media_hours	-2.61978	0.14413	-18.177	< 2e-16 ***
netflix_hours	-2.27708	0.15697	-14.507	< 2e-16 ***
attendance_percentage	0.14473	0.01797	8.054	2.28e-15 ***
sleep_hours	2.00462	0.13764	14.564	< 2e-16 ***
exercise_frequency	1.45187	0.08338	17.413	< 2e-16 ***
mental_health_rating	1.94891	0.05924	32.897	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.331 on 992 degrees of freedom

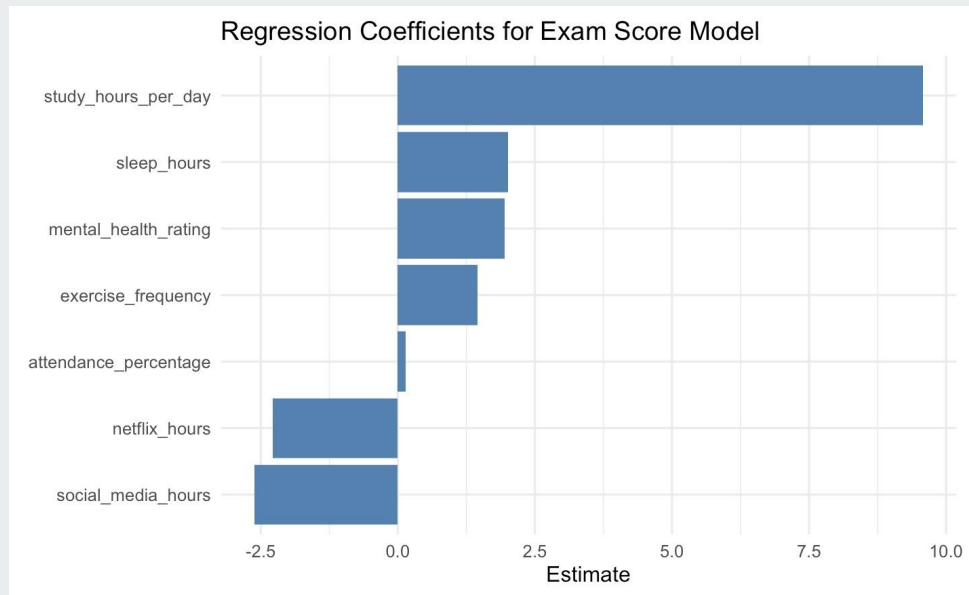
Multiple R-squared: 0.9011, Adjusted R-squared: 0.9004

F-statistic: 1291 on 7 and 992 DF, p-value: < 2.2e-16

- **Explained variance:** Adjusted $R^2 \approx 0.90$
- **All predictors highly significant** ($p < 0.001$)
- **Largest effects:**
 - +9.57 points per additional study hour
 - -2.62 points per additional social media hour
 - -2.28 points per additional Netflix hour
- **Positive contributors** (smaller):
 - +0.14 per % attendance
 - +2.00 per hour of sleep
 - +1.45 per exercise unit
 - +1.95 per mental health point

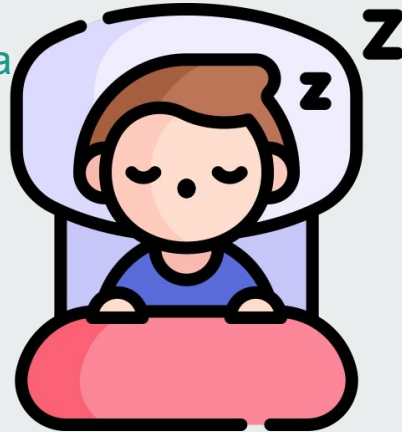
Predictor Importance Ranking

- **Top three:**
 - Study hours per day
 - Mental health rating
 - Sleep hours
- **Mid level:**
 - Exercise frequency
 - Attendance percentage
- **Negative drivers:**
 - Social media hours (most)
 - Netflix hours
- **Actionable insight:**
 - Prioritize study time, mental health & sleep
 - Mitigate excessive screen use for better exam outcomes



If We Were to Give Advice...

- Boost study time where possible — it has the largest return
- Aim for 7–8 hours of sleep for academic + mental benefit
- Stay active: exercise helps, even in small doses
- Prioritize mental health — it's strongly linked to outcomes
- Be mindful of screen time — especially social media
- Consistent attendance also important





Thank You