

Egypt University of Informatics

Computer and Information Systems

Data Analysis Course

The Analysis of Student Habits vs Academic Performance.

…

Omar Ahmed Waleed 23-10132

Seif Alaa 23-201363

Mohamed Kamel 24-200150

# Introduction

The connection between student lifestyle habits and academic performance has long been a subject of debate in education. This topic holds significant relevance for students, parents, and educators alike, as understanding these patterns could lead to better learning strategies and improved outcomes. Through data-driven analysis, this report explores key behavioral factors such as study habits, screen time, and sleep to uncover their real impact on exam performance. Beyond mere statistics, we aim to reveal compelling insights and thought-provoking dilemmas that challenge common assumptions about academic success.

# Research Question

# How do daily habits like studying, screen time, sleep, and mental health affect students’ exam scores, and which ones have the biggest or most surprising impact?

# Hypothesis

*Null Hypotheses (H₀):*

1. There is no significant correlation between daily study hours and exam scores.

2. Attendance percentage has no effect on exam performance.

3. Screen time does not influence exam scores.

4. Sleep duration has no impact on academic performance.

5. Parental education level does not affect student performance.

6. Exercise frequency is unrelated to exam scores.

7. Study-sleep balance has no effect on academic performance.

8. Mental health status does not correlate with exam scores.

9. Diet quality has no impact on academic performance.

10. Internet quality does not influence exam outcomes.

*Alternative Hypotheses (H₁):*

1. Increased daily study hours significantly improve exam scores.

2. Higher attendance percentage leads to significantly higher exam scores.

3. Greater screen time significantly reduces exam performance.

4. Students with 7–9 hours of sleep score significantly higher.

5. Students with parents holding advanced degrees outperform others.

6. Regular exercise correlates with moderately higher exam scores.

7. A balanced study-sleep ratio leads to higher scores.

8. Better mental health ratings are linked to moderately higher scores.

9. Students with "Good" diets score significantly higher than those with "Poor" diets.

10. Students with reliable internet access achieve significantly higher scores.

*---*

# Population of Interest:

# The dataset used in this analysis was obtained from **Kaggle**, an online platform for sharing datasets and data science projects. It contains **1,000 student records**, each representing an individual student with detailed information on their **daily habits, lifestyle choices, and final exam scores**.

# Sampling Method:

# The dataset’s sampling method is not explicitly documented, which limits our ability to assess its representativeness. However, given the variables and sample size (n=1,000 ), we assume the data may have been collected through convenience sampling from a specific educational institution or region, because the of the IDs used.

# Bias Identification: We avoided using leading or emotionally loaded language in the data collection process, ensuring that questions did not push participants toward certain answers (e.g., not saying “Do you waste time on Netflix?” but rather “How many hours per day do you spend on Netflix?”).

based on 1,000 student entries, we ensured that our analysis treated all entries objectively and did not focus on or exclude any specific subgroup unfairly.

# Survey Questions/Collected Data/Dataset:

For this project, we used a dataset titled **“Student Habits vs Academic Performance ”**, which was sourced from Kaggle. It contains **1,000 individual student records**, each representing a unique entry of lifestyle habits and academic outcomes. The dataset is clean, structured, and suitable for educational and analytical purposes.Number of samples used

| **Variable Name** | **Description** |
| --- | --- |
| student\_id | Unique ID assigned to each student (not used in analysis) |
| study\_hours\_per\_day | Average number of hours the student studies each day |
| sleep\_hours\_per\_day | Average hours of sleep per day |
| attendance\_percentage | Percentage of classes attended by the student |
| screen\_time\_social\_media | Daily hours spent on social media |
| screen\_time\_netflix | Daily hours spent watching Netflix |
| mental\_health\_rating | Self-reported mental health status (scale: 1 to 10) |
| diet\_quality | Categorical variable (e.g., Poor, Average, Good) |
| exercise\_frequency\_per\_week | Number of days per week the student exercises |
| parent\_education\_level | Highest education level of parents (e.g., High School, Bachelor’s, Master’s, PhD) |
| internet\_quality | Quality of internet access (e.g., Poor, Average, Good) |
| exam\_score | Final exam score (0 to 100) |
| study\_sleep\_ratio | Derived variable: ratio between study hours and sleep hours |
| total\_screen\_time | Derived variable: social media + Netflix screen time combined |

# Analysis:

The relationship between **student lifestyle habits** and **academic performance** (measured by exam scores) was explored using the following methods:

* **Scatterplots**:  
  Visualized the trends between variables like **study hours, sleep duration, screen time**, and **exam scores**, highlighting both positive and negative associations.
* **Linear Regression**:  
  Demonstrated statistically significant positive relationships between **study hours**, **attendance**, and **exam scores**, while identifying a strong negative impact of **screen time** on academic performance.
* **Boxplots**:  
  Used to assess variation in exam scores across **categories such as diet quality, internet access, and parental education level**, making it easy to spot group-based differences.
* **ANOVA Tests**:  
  Conducted to test for statistically significant differences in exam scores between groups defined by **diet quality**, **internet speed**, and **parental education**.
* **T-tests**:  
  Compared groups with high and low mental health ratings, physical activity, and sleep levels, confirming statistically significant differences in performance.

# Hypothesis Testing Steps

# **Defined** null (H₀) and alternative (H₁) hypotheses for each habit-performance relationship.

# **Grouped** students based on habits (e.g. high vs. low study time, sleep, screen usage).

# **Calculated** average exam scores for each group.

# **Performed** t-tests and ANOVA where applicable.

# **Interpreted** p-values (p < 0.05) to test significance.

# Conclusion

# This analysis reveals that exam scores are most strongly increased by consistent study hours, high attendance, and reliable internet access, while decreased significantly by excessive screen time. Factors like parental education, diet quality, and a balanced study-sleep ratio also contribute positively, though less prominently. Surprisingly, sleep, mental health, and exercise show only weak correlations with performance, suggesting their roles are supportive but not decisive. Notably, variables such as part-time job status, age, and extracurricular participation had no statistically significant impact on exam scores indicating these factors neither enhance nor hinder academic outcomes in this dataset. While these findings highlight actionable strategies prioritizing focused study, minimizing distractions, and addressing resource gaps they also underscore the complexity of academic success, where disciplined habits and environmental advantages intertwine. Generalizability is limited by potential sampling biases and self-reported data, but the insights offer a roadmap for students, educators, and policymakers to target high-impact levers for improvement