

The Effects Of Student Habit Behaviors on Exam Performance



A Report by Andrew En-maari Yahaya

June 14, 2025

Contents

1	Introduction	3
2	Objective	3
3	Dataset Overview	3
4	Exploratory Data Analysis (EDA)	4
4.1	Correlation Analysis	4
4.2	Visualization Insights	4
5	Methodology	4
5.1	Data Preprocessing	4
5.2	Model Selection	4
5.3	Evaluation Metrics	4
6	Results	5
7	Conclusion	5
8	Future Work	5
9	Appendix	6
A	Data Dictionary	6
B	Relationship between variables	6
C	Module Evaluation Function	8

Introduction

Understanding what drives academic performance is a key concern in the field of education. This study investigates how various student behaviors such as study time, sleep duration, internet usage, and extracurricular involvement correlate with exam performance. By leveraging a real-world dataset and applying data science methodologies, this report aims to identify behavioral patterns that significantly influence student results. The insights derived can support educators in developing more effective strategies to enhance learning and academic success.

Objective

The main objectives of this study are:

1. To explore and visualize behavioral trends among students.
2. To determine which behaviors have the strongest correlation with exam scores.
3. To build a predictive model capable of estimating exam scores based on observed behaviors.

Dataset Overview

The dataset used in this study contains behavioral and demographic data from students. Key features include:

1. Student Id
2. Study Time (hours)
3. Age
4. Gender
5. Social Media Hours
6. Part Time Job
7. Diet Quality
8. Exercise Frequency
9. Mental Health Rating
10. Netflix Hours
11. Sleep Duration (hours)
12. Internet Quality
13. Attendance
14. Parental Education Level
15. Participation in Extracurricular Activities
16. Exam Score (target variable)

The data was cleaned and preprocessed to handle missing values, normalize numerical features, and encode categorical variables while taking out columns that have no effects on the study or otherwise known as noise.

Exploratory Data Analysis (EDA)

4.1 Correlation Analysis

A strong positive correlation was observed between study time and exam scores. Excessive social media usage, netflix and frequent absences showed a negative or very weak correlation with performance. Students with regular sleep habits (7–9 hours or more) performed better on average (over 70%) than those with irregular or inadequate sleep.

4.2 Visualization Insights

Box plots revealed higher exam scores among students involved in extracurricular activities and also female students.

A heatmap of feature correlations highlighted study time, as a key performance predictor. In studying the effects of sleep hours on mental health and exam score, a pair plot was used. It was discovered that there was a lot of spread between the points of each feature compared to the other indicating that one feature alone cannot predict the exam score. This is shown in figure B

Methodology

5.1 Data Preprocessing

The only column with missing values was the parental education level which was filled out using the mode.

One-hot encoding was applied for categorical variables like parental education and activity involvement.

5.2 Model Selection

Multiple models were tested, including:

1. Random Forest
2. XGBoost
3. Linear Regression
4. Ridge Regression
5. Lasso Regression

5.3 Evaluation Metrics

Models were evaluated using:

1. R^2 Score
2. Mean Absolute Error (MAE)
3. Root Mean Squared Error (RMSE)

Results

The Random Forest Regressor and XGBoost performed best with an R^2 score of 0.9986883893072496 and 0.9984883628337156 respectively, indicating strong predictive capability. While the Linear Regression performed with an R^2 of 1.0 which suggest overfitting. The Lasso and Ridge regressions also performed very well with R^2 scores of 0.9965140898959688 and 0.9999792792893103 respectively.

Key features contributing to exam score predictions:

1. Study Time
2. Social Media
3. Attendance
4. Sleep Duration

These results suggest that behavioral factors play a significant role in academic outcomes and can be leveraged to forecast performance.

Conclusion

This study confirms that student behaviors such as consistent study routines, adequate sleep, and classroom attendance significantly impact academic success. By identifying these correlations, educators and parents can encourage habits that foster improved performance. Predictive modeling also shows promise in early intervention strategies to assist students at risk of underperforming.

Future Work

Incorporate psychological variables such as motivation or stress levels.

Expand the dataset across different institutions and age groups.

Develop an interactive dashboard for real-time behavioral analysis and academic forecasting.

Appendix

Data Dictionary

- **student_id** – Unique identifier for each student
- **age** – Age of the student
- **gender** – Gender of the student (e.g., Male/Female/Other)
- **study_hours_per_day** – Average number of hours spent studying per day
- **social_media_hours** – Daily hours spent on social media platforms
- **netflix_hours** – Daily hours spent watching Netflix
- **part_time_job** – Whether the student holds a part-time job (Yes/No)
- **attendance_percentage** – Class attendance rate as a percentage
- **sleep_hours** – Average number of hours of sleep per night
- **diet_quality** – Self-reported diet quality (e.g., Poor, Average, Good)
- **exercise_frequency** – Number of days per week the student exercises
- **parental_education_level** – Highest education level achieved by either parent
- **internet_quality** – Quality of internet connection (e.g., Poor, Average, Excellent)
- **mental_health_rating** – Mental health status (scale 1–10)
- **extracurricular_participation** – Participation in extracurricular activities (Yes/No)
- **exam_score** – Final exam score (numeric)

Relationship between variables

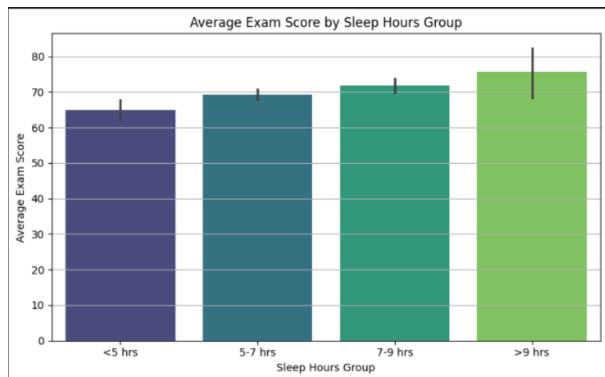
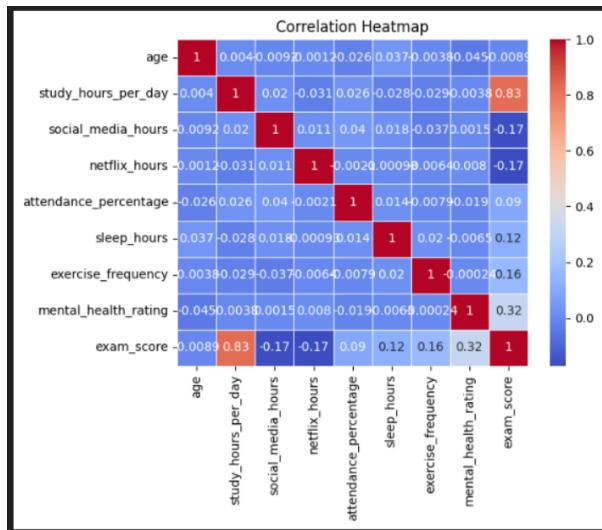
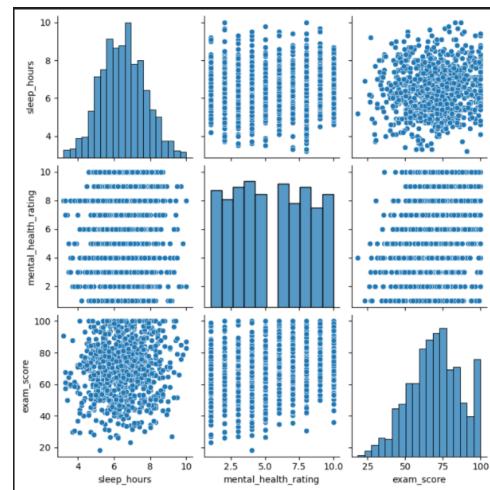


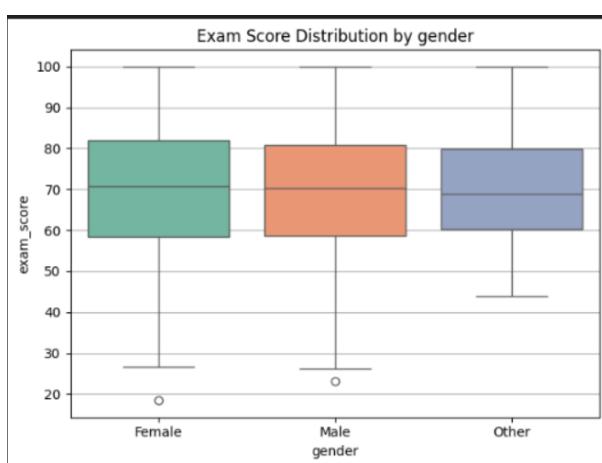
Figure 1: Bar plot of sleep hours against exam score.



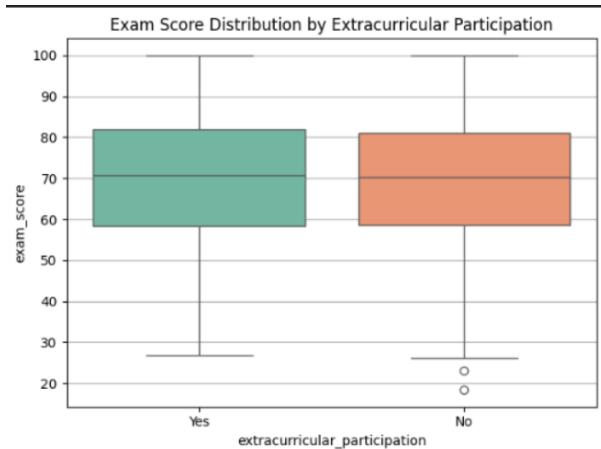
(a) Heatmap of Correlation of variables



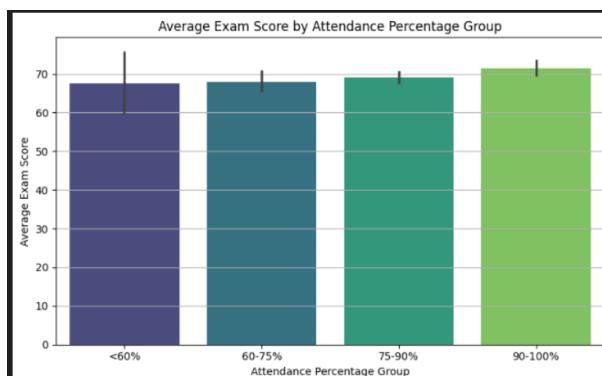
(b) Pair plot showing relationships between sleep hours, mental health rating, and exam score.



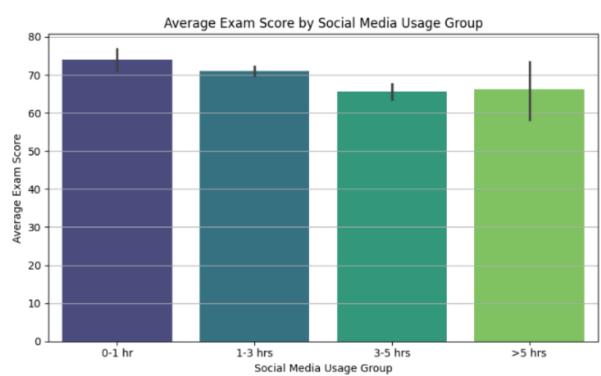
(a) Boxplot of Gender Against Exam Score



Boxplot of Extracurricular participation against Exam Score



Barplot of Attendance percentage against Exam Score



Barplot of hous spent on social media against Exam Score

Module Evaluation Function

```
def evaluate(model,y_test,y_pred):  
    print(f'Evaluation of {model}')  
    print(f'R2Score: ',r2_score(y_test,y_pred))  
    print("MAE:", mean_absolute_error(y_test, y_pred))
```