

# **2704 Final Report**

**Dr Jong-Kyou Kim**

**April 23rd, 2025**

**Jackson McIntyre, Nicolas Serrano and Taryn Cail**

## Introduction

This study explores the factors influencing students academic performance, with a particular focus on the correlation between study habits and GPA. We particularly looked at the relationship between study hours per day and sleep hours per day and their effect on students' GPA.

## Background

Students' academic success (GPA) is linked to various features such as the amount of studying they do, their sleep schedules, eating habits and exercise patterns. But what feature affects their performance the most? We believe it is studying habits. Despite the multitude of influences, studying has long been considered the foundation of academic success. The thought process behind this is simple: the more time a student spends studying, the better they will understand the material and the better they will perform on assessments. However, data is necessary to confirm this hypothesis.

## Hypothesis

In this study, the Null Hypothesis is that there is no correlation between students' amount of studying and their academic performance. On the other hand, the Alternative Hypothesis is that the higher the amount of studying, the higher the student's academic performance will be. We wish to prove the Alternative Hypothesis, that the more studying a student performs the better they will achieve academically.

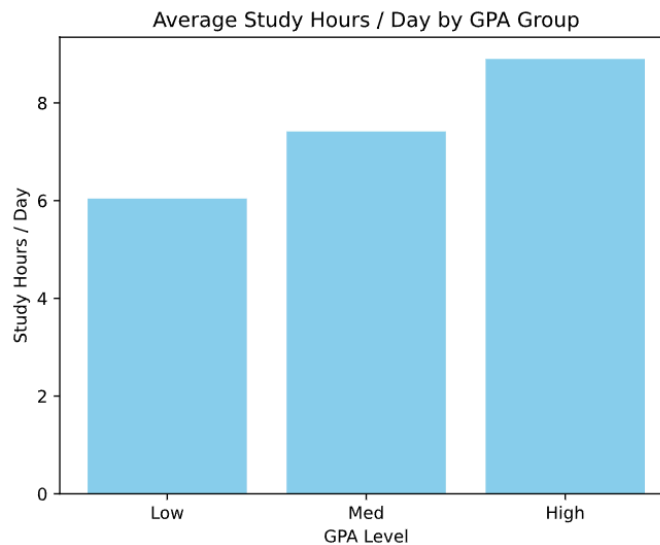
## The Dataset

The dataset used in this study was sourced from Kaggle, an open-source data website used for data-analytics and machine learning.

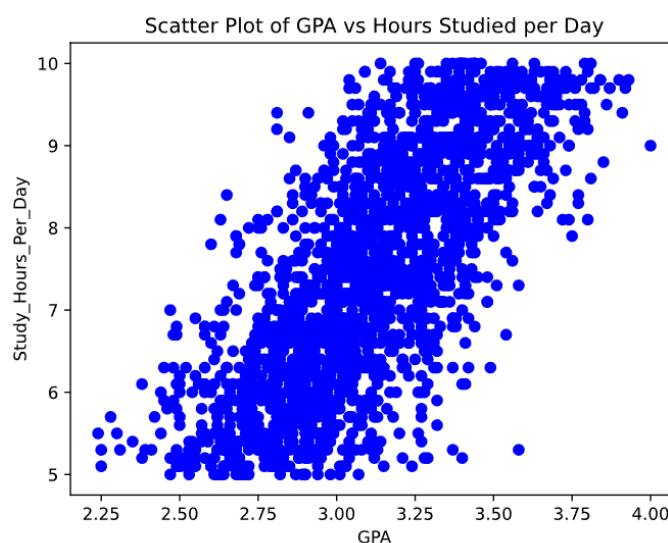
Student_ID	Study_Hours_Per_Day	Extracurricular_Hours_Per_Day	Sleep_Hours_Per_Day	Social_Hours_Per_Day	Physical_Activity_Hours_Per_Day	GPA	Stress_Level
1	6.9	3.8	8.7	2.8	1.8	2.99	Moderate
2	5.3	3.5	8	4.2	3	2.75	Low
3	5.1	3.9	9.2	1.2	4.6	2.67	Low
4	6.5	2.1	7.2	1.7	6.5	2.88	Moderate
5	8.1	0.6	6.5	2.2	6.6	3.51	High
6	6	2.1	8	0.3	7.6	2.85	Moderate
7	8	0.7	5.3	5.7	4.3	3.08	High
8	8.4	1.8	5.6	3	5.2	3.2	High
9	5.2	3.6	6.3	4	4.9	2.82	Low

The dataset contains 2000 rows and 8 columns, including data on various student metrics such as student ID, study hours, extracurricular hours, sleep hours, social hours, physical activity hours per day as well as GPA levels and stress levels. This dataset offers a diverse overview of students' daily schedules and their corresponding academic performance, allowing for both correlation and predictive analysis.

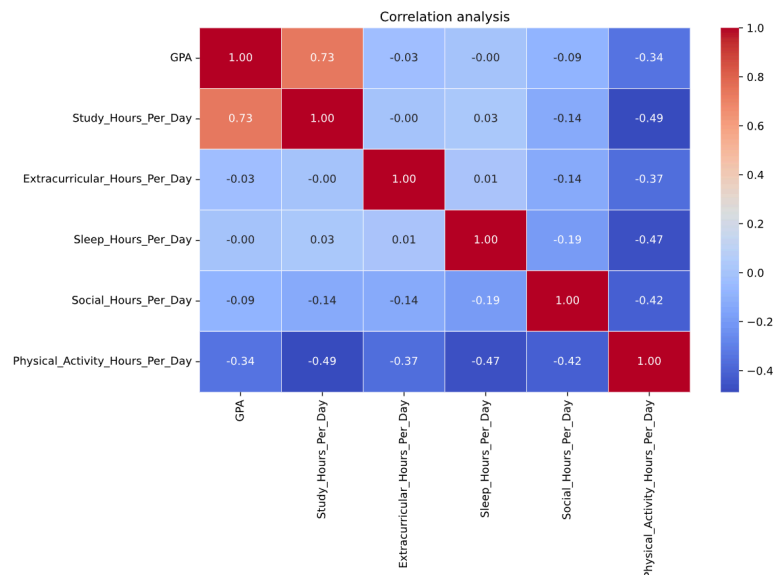
## Graphs



This bar graph further illustrates the relationship between a higher GPA and the amount of hours studied per day. This is evident from the larger bar in the 'High' GPA section compared to the 'Low' GPA section, indicating that students with a higher GPA score tend to spend more time studying. The more hours studied per day, the higher the GPA range they are placed into. This visually reinforces our hypothesis and highlights a clear trend between study hours and academic success.



The Scatter Plot further demonstrates a positive linear relationship between Study Hours per day and GPA. Each point represents a student, and as the study hours increase so does the GPA. While there is some variance, the overall pattern indicates that students who spend time studying may see better academic performance compared to their counterparts.



The heatmap showcases the correlation matrix for all numeric variables in the dataset. The closer the correlation value is to 1, the stronger the relationship. For instance, the correlation between GPA levels and Study Hours per Day is strong, with a value of 0.73. This suggests a positive relationship where an increase to one variable leads to an increase in the other. In contrast, the other variables do not exhibit a strong positive or negative correlation.

## P-Value Test and Pearson Correlation Test

We performed two statistical tests to decide whether we are able to accept or reject our hypothesis, these were the P-Value test and the Pearson Correlation Test. For the P-Value Test, a P-Value less than 0.05 indicates that the correlation is statistically significant. For the Pearson Correlation Test, the coefficient ( $r$ ) indicates a positive or negative relationship between the two variables. Specifically an ( $r$ ) value greater than 0.7 suggests a strong positive linear relationship.

Based on our results:

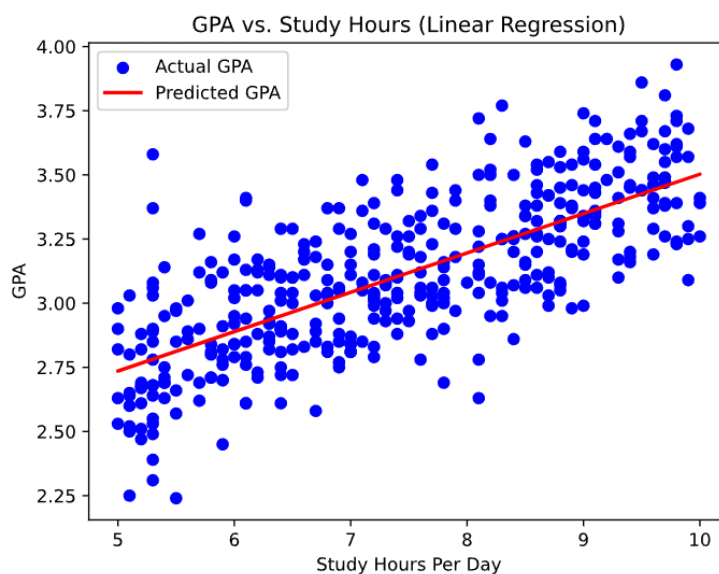
- P-Value =  $\sim 0$
- $r = 0.7345$

We can confidently reject the null hypothesis and conclude that an increase in study hours is associated with a higher GPA.

## Predictive Analytics

GPA level, which reflects the outcome we want to predict, serves as the response variable in this analysis. The amount of time spent studying each day is the independent variable, and is the input which may influence or affect the GPA outcome.

### Simple Linear Regression Model



A simple linear regression model was trained on 80% of the project's dataset (~1600 data points) to predict GPA based on study hours, with the remaining 20% used for testing (~400 data points). The model's slope and intercept provide insight into the expected increase in GPA for each additional hour of studying.

### Custom Input Analysis for Simple Linear Regression

Hours Studied Per Day (Input, Independent Variable)	Predicted GPA (Output, Dependent Variable)
0	1.97
2	2.28
4	2.58
6	2.89
8	3.20

10	3.50
----	------

Testing with sample study hours such as 2, 4, 6, etc. showed a steady increase in predicted GPA values, suggesting a positive relationship between study time and GPA level.

## Linear Discriminant Analysis (LDA)

To classify GPA levels (High, Medium, Low) based on study hours, an LDA model was trained. Similar to the Simple Linear Regression Model, 80% of the dataset was used for training and the other 20% used for testing. This provided a probabilistic prediction of GPA category based on input features. The following classification report was generated from the Linear Discriminant Analysis model after training:

Accuracy: 0.67%

	Precision	Recall	F1-Score	Support
High	0.66	0.59	0.62	94
Low	0.65	0.51	0.57	88
Med	0.68	0.77	0.72	218
Accuracy			0.67	400
Macro Average	0.66	0.62	0.64	400
Weighted Average	0.67	0.67	0.67	400

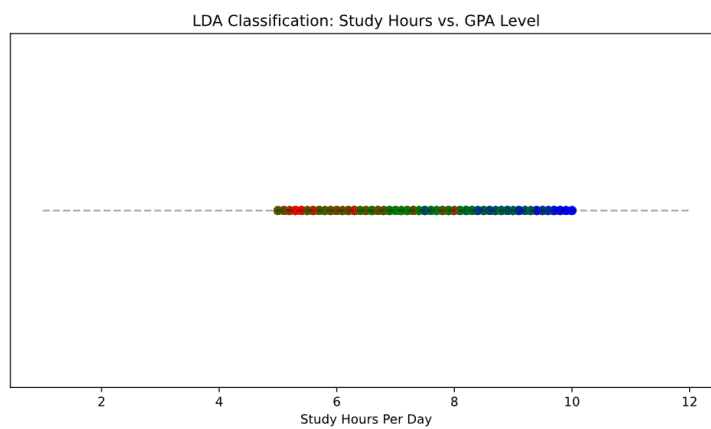
## Custom Input Analysis for Linear Discriminant Analysis

Hours Studied Per Day (Input, Independent Variable)	GPA Level (Output, Dependent Variable)
0	Low
2	Low
4	Low
6	Med
8	Med

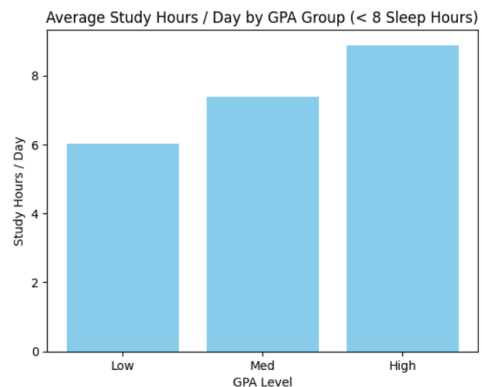
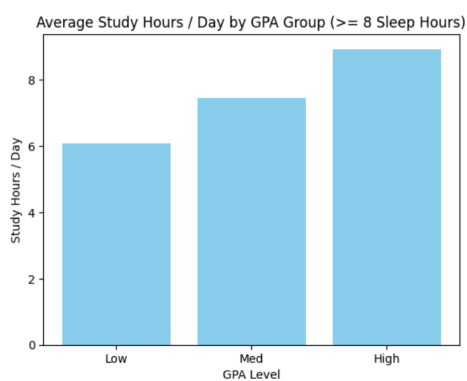
Testing with sample study hours such as 2, 4, 6, etc. showed a steady increase in predicted GPA levels, suggesting a positive relationship between study time and GPA level.

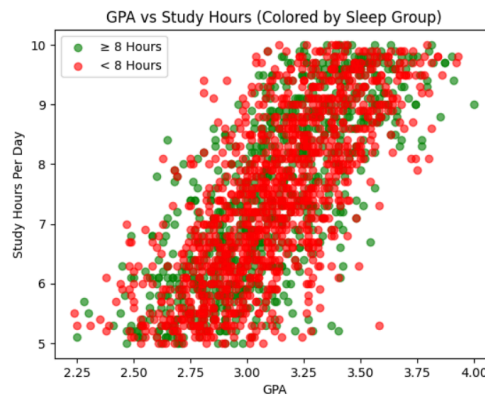
## GPA vs Study Hours based on Sleep Hours

In the following graph we can see that the heavy overlap shows that study hours per day may not be the sole determinant for GPA levels.



Therefore, returning to our heatmap we decided to investigate whether sleep, which had the largest negative correlation in the heatmap, had an effect on students GPA's.





After conducting the secondary analysis it was revealed that whether students consistently received more than 8 hours of sleep or less than 8 hours of sleep it ultimately did not affect overall academic success.

## Discussion

The use of visual tools such as bar graphs, scatter plots and heatmaps proved effective in visualizing trends and identifying patterns and relationships in the dataset. The heatmap provided valuable insights into the correlation between variables. Meanwhile the Simple Linear Regression model and Linear Discriminant Analysis model helped illustrate patterns of change and growth, making the relationship between study hours and GPA more interpretable.

One challenge encountered was interpreting the p-value, as it produced unexpected results and ultimately equaled zero. This suggests that one or more variables may have had minimal variation and a low standard deviation, potentially affecting the reliability of the statistical analysis.

## Analysis and Implication

These findings imply that institutions should emphasize not just class attendance or curriculum quality, but also encourage structured and consistent study habits. Academic support programs might consider promoting time management tools, study groups, or workshops on productive study techniques.

## Further Research

To enhance the study's accuracy and reliability, future work should focus on expanding the dataset by collecting more diverse and extensive data. This would help improve statistical robustness and provide a more comprehensive understanding of the relationship between study hours and GPA.



## **Conclusion**

A correlation coefficient ( $r$ ) greater than 0.7 indicates a strong positive relationship between study hours per day and GPA. Additionally, a p-value below 0.05 signifies that the correlation is statistically significant. Our analysis yielded an ( $r$ ) value equal to 0.7345 and our p-value near 0, providing strong evidence to confidently reject the null hypothesis and accept the alternative hypothesis. Therefore, we can confidently conclude that there is a statistically significant positive correlation between study hours per day and GPA level, suggesting that students who spend more hours studying are more likely to achieve higher GPAs.

## References

Kumar, Sumit. "Student Lifestyle Dataset." Kaggle, November 11, 2024.  
<https://www.kaggle.com/datasets/steve1215rogg/student-lifestyle-dataset>.

GitHub Link: <https://github.com/Taryn-Cail/Python-Group-Project-2704>