جامعة مصر للمعلوماتية
EGYPT UNIVERSITY OF INFORMATICS

Egypt University of Informatics

Computer and Information Systems

Data Analysis Course

Analyzing Students' Habits and Academic Performance

Submitted by:

Yousef atwa (23-101295)

Yousef khafagy (23-101266)

Belal nasr (23-101061)

Ahmed tamer (23-101116)

Under the Supervision of:

Dr. Mohamed Taher ElRefaei PhD

Eng. Nadine ElSaeed
Eng. Shereen

Date: 5/25/2025

Introduction

Academic success in university and school is influenced not only by students' study practices but also by their lifestyle choices. This report aims to explore the impact of student habits such as sleep, social media use, diet, and gym attendance on their academic performance (measured through exam scores). Using a comprehensive dataset, we identify key behaviors correlated with high or low performance, offering insights that can shape educational support and student well-being policies.

Research Question

What student habits most strongly correlate with higher academic performance?

Hypothesis

Students with balanced study routines, adequate sleep, minimal social media usage, and consistent physical activity tend to perform better academically.

Population of Interest

The population of interest includes students of all educational levels whose behavioral patterns and academic performance are documented in this dataset.

Sampling Method

A convenience sampling method was used, consisting of 1,000 students. While the sample includes a degree of diversity, it may not fully represent the broader student population, which limits the generalizability of the results.

Bias Identification

As the data is self-reported and from a single snapshot in time, it is subject to selection and reporting biases. External factors like mental health, socioeconomic status, and institutional differences were not fully controlled.

Collected Data / Dataset Description

The data used in this study was collected through a structured survey administered to students. The aim of the survey was to examine the relationship between students' lifestyle habits, academic behaviors, and their academic performance. The dataset contains responses from 1,000 students. However, it is important to note that the dataset does not include measures of cognitive ability, such as IQ levels, which could provide additional insights into the factors influencing academic performance.

Each row in the dataset corresponds to an individual student and includes demographic details, behavioral variables, and academic outcomes. Below is a summary of the types of questions asked in the survey:

- **Academic Habits:** Questions related to study hours per day and class attendance.
- **Lifestyle and Media Consumption:** Questions about daily time spent on social media, Netflix, and hours of sleep.
- **Health and Well-being:** Questions assessing diet quality, exercise frequency, and mental health rating.
- **Demographic and Background Information:** Gender, age, parental education level, part-time job status, internet quality, and participation in extracurricular activities.

- **Academic Performance Indicator:** Students were asked to report their most recent exam score (0–100 scale).

The full list of reconstructed survey questions is provided in the **Survey Design** section of this report.

The collected data was stored in a CSV file with 16 columns representing the survey responses. The data has been analyzed using statistical techniques such as correlation analysis, t-tests, and ANOVA to test hypotheses regarding the influence of various factors on student performance.

Dataset link : https://www.kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance

Number of samples collected: 1000

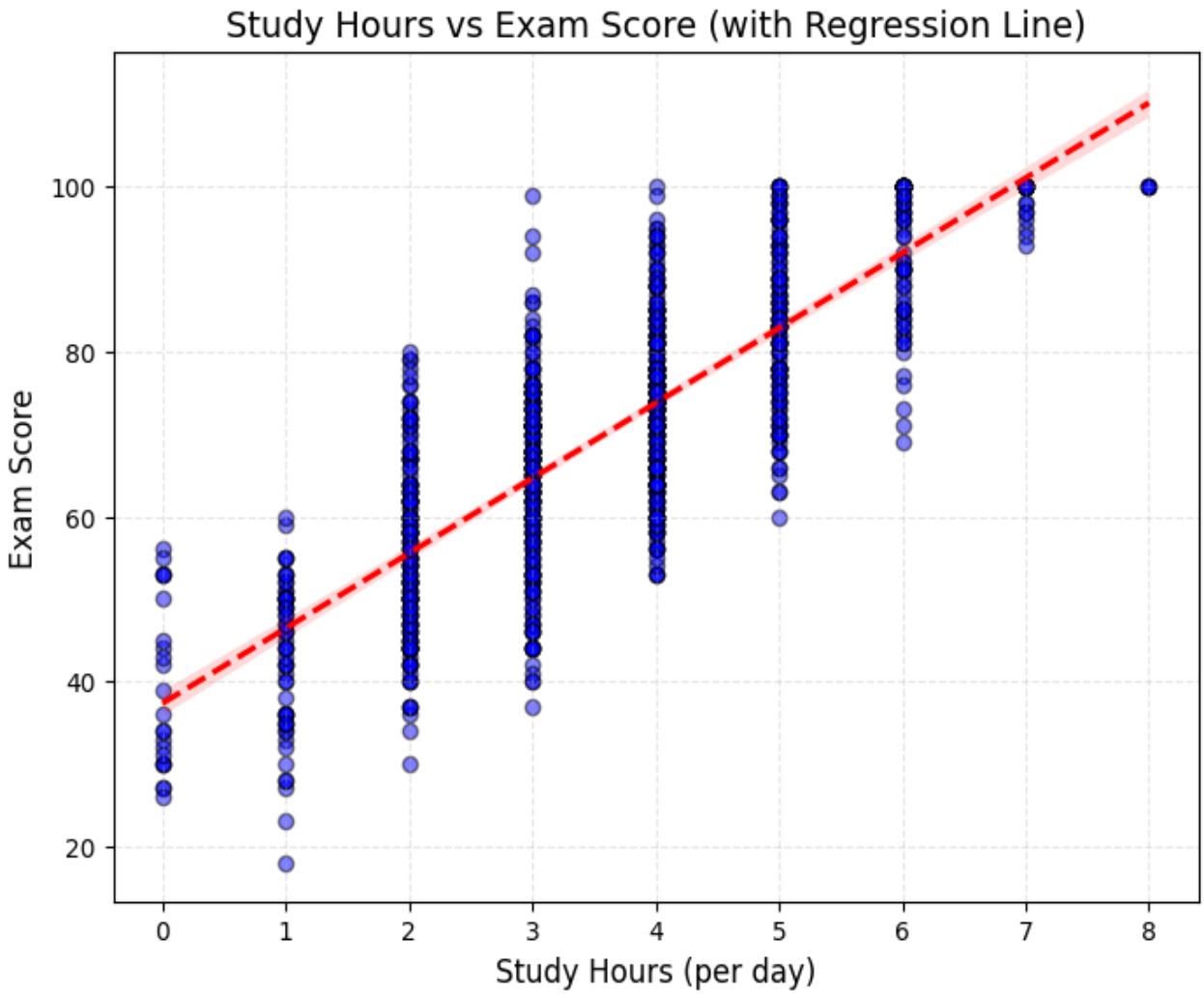
Analysis & Visualization Summary

Descriptive Statistics:

- Average Exam Score: 68.4%
- Average Sleep: 6.8 hours/day
- Average Study Time: 3.5 hours/day
- Average Social Media Time: 2.9 hours/day

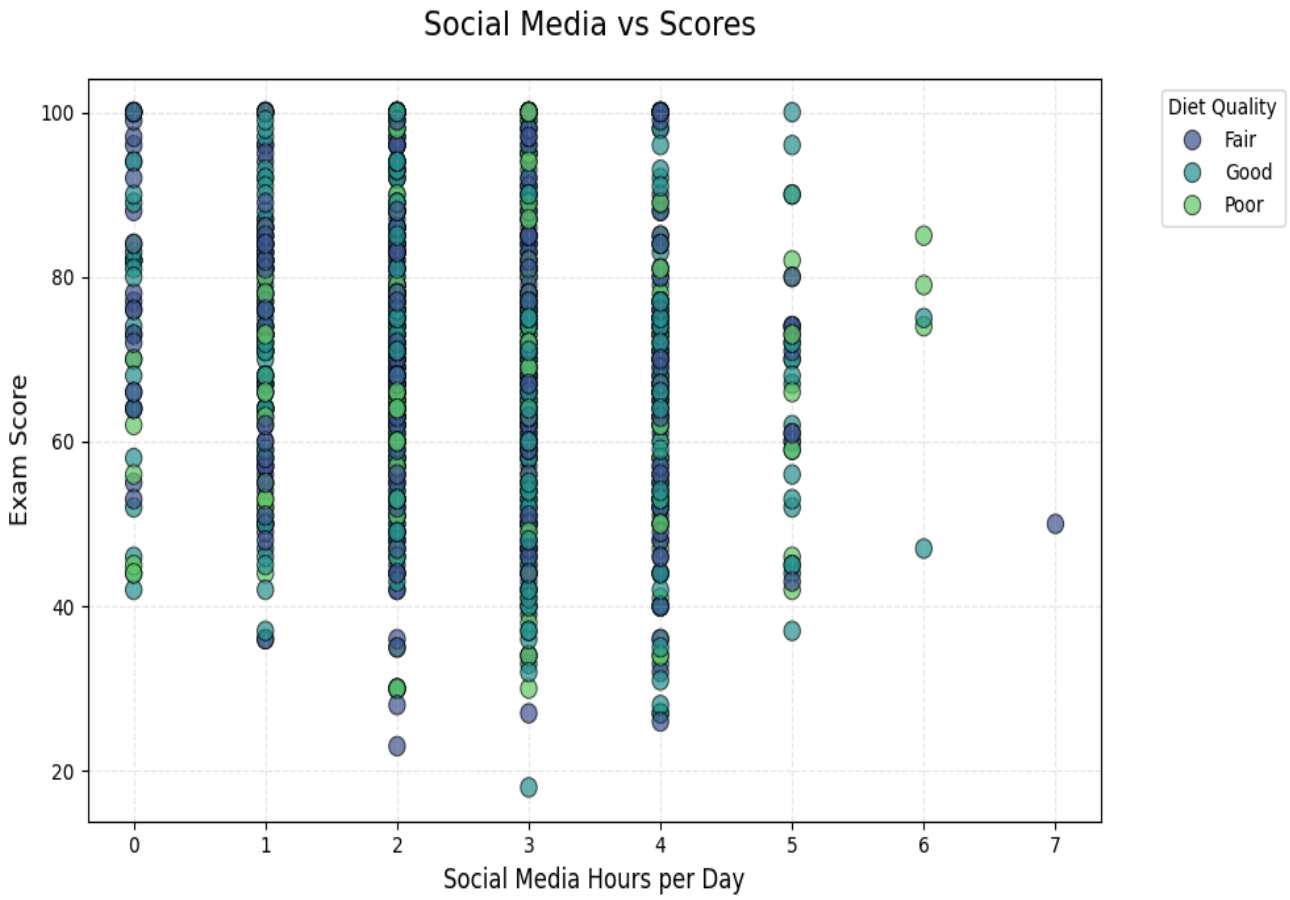
Key Visuals & Insights:

1. Exam Score vs Study Hours:



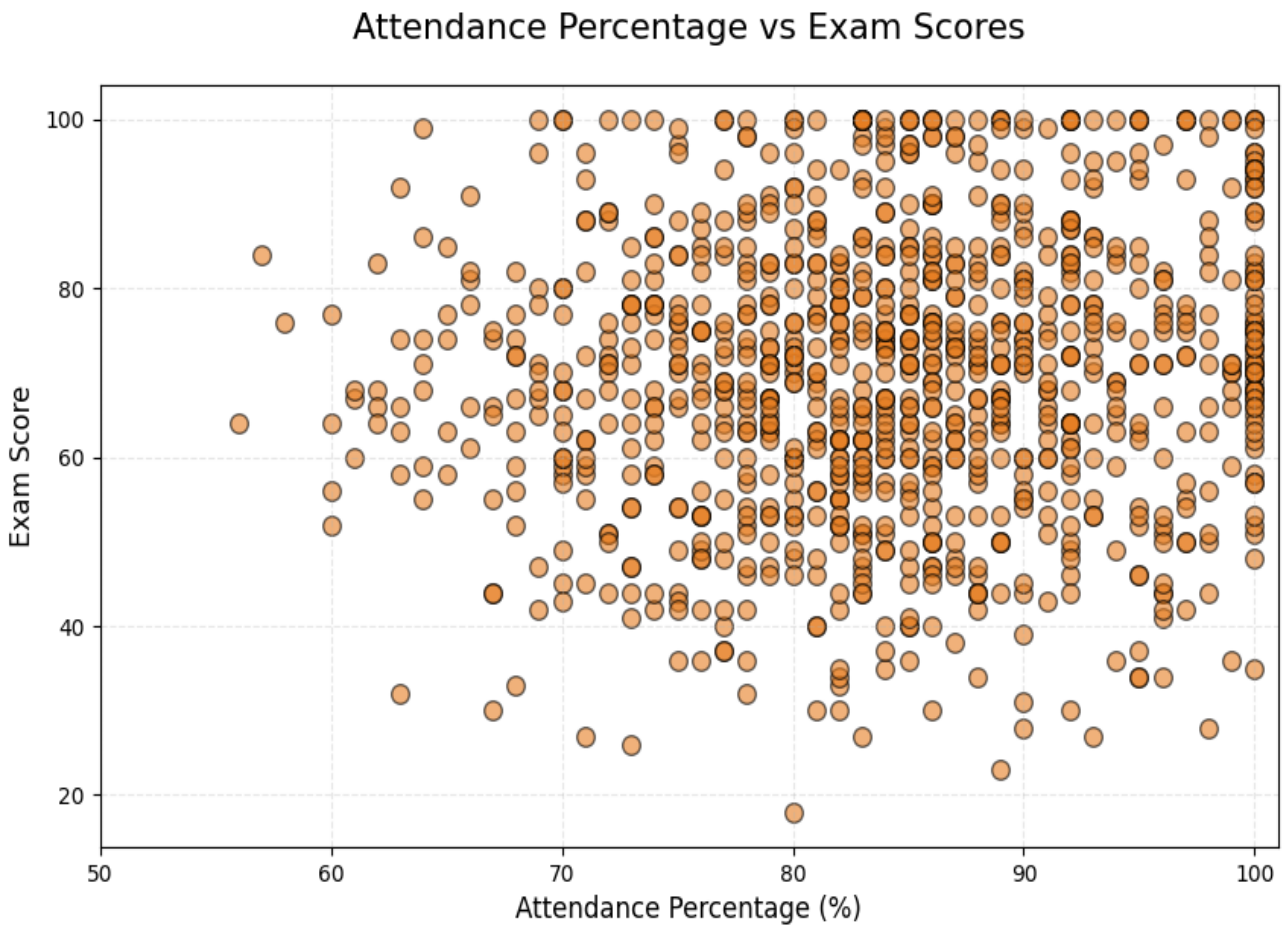
The graph shows a moderate positive relationship between study hours and exam scores. Students who study more tend to achieve higher exam scores, especially up to about 5–6 hours per day. The trend suggests that regular studying improves academic performance, but with diminishing returns beyond a certain point. This supports the hypothesis that consistent study habits positively influence student outcomes.

2. Exam Score vs Social Media Use:



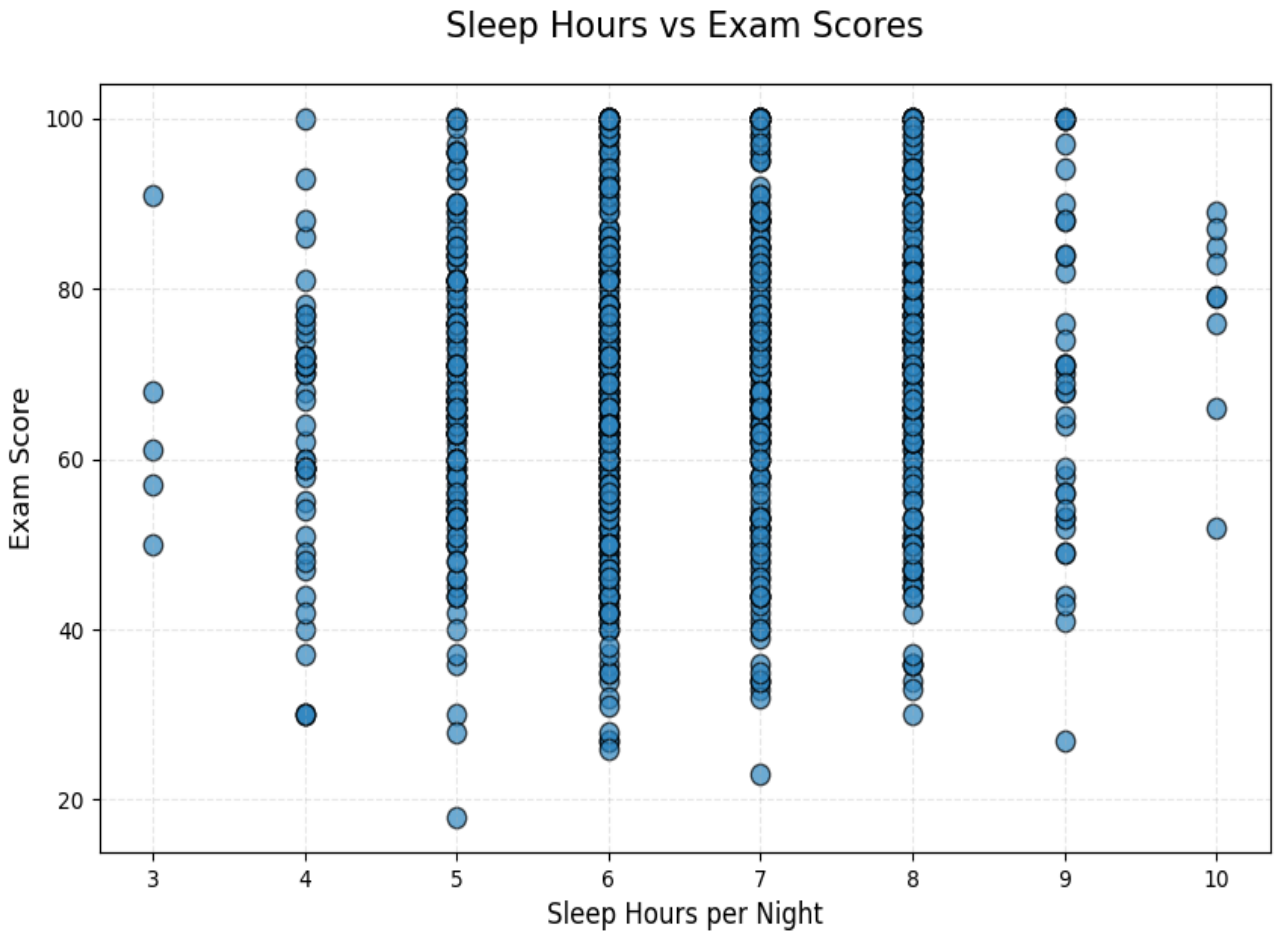
The scatter plot illustrates the relationship between social media usage (hours per day) and exam scores, with diet quality as a categorical variable (Fair, Good, Poor). The data suggests that students who spend less time on social media (0–2 hours per day) generally achieve higher exam scores, especially those with Fair or Good diets. In contrast, increased social media usage (4 or more hours) is associated with a wider variation in scores, including more lower scores, particularly among students with Poor diet quality. This indicates that both reduced screen time and better diet quality may contribute to improved academic performance

3. Exam Score by class attendance:



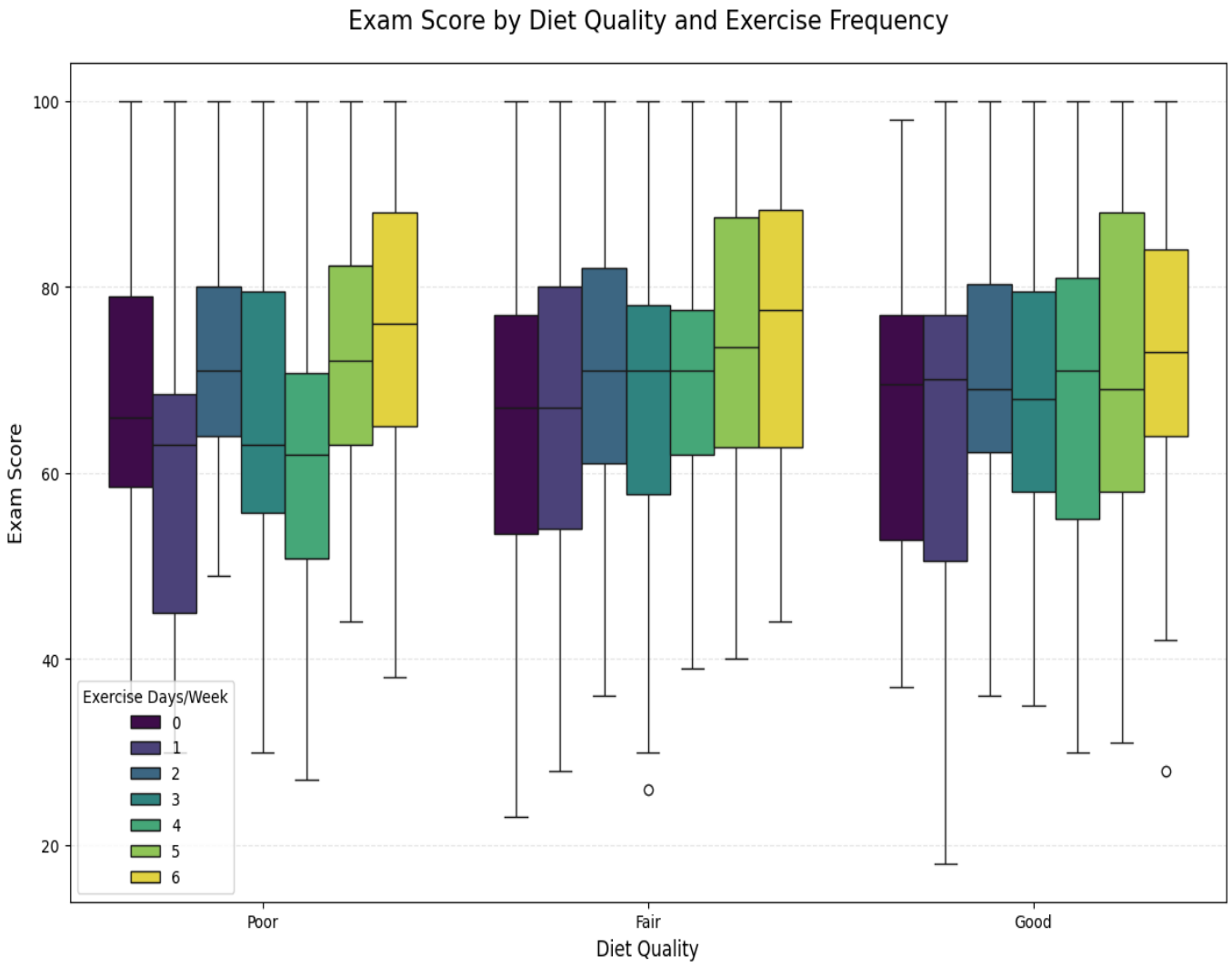
The scatter plot displays the relationship between attendance percentage and exam scores. The data indicates a general positive trend, where higher attendance (especially above 80%) tends to be associated with higher exam scores. However, the spread of scores at each attendance level suggests that while attendance is a contributing factor to academic success, it is not the sole determinant. Students with lower attendance often score lower, but there are exceptions at all levels. This highlights that consistent attendance is generally beneficial for academic performance, though other factors also play significant roles.

4. Sleep hours VS exam score:



The scatter plot illustrates the relationship between average sleep hours per night and exam scores. The data shows that students who sleep between 6 to 8 hours tend to achieve a wider range of higher exam scores, suggesting this may be an optimal sleep range for academic performance. Both shorter sleep durations (less than 5 hours) and longer sleep durations (9+ hours) are associated with more variable and generally lower exam scores. While there are high achievers in all sleep categories, the trend supports the importance of maintaining a balanced sleep schedule for better academic outcomes.

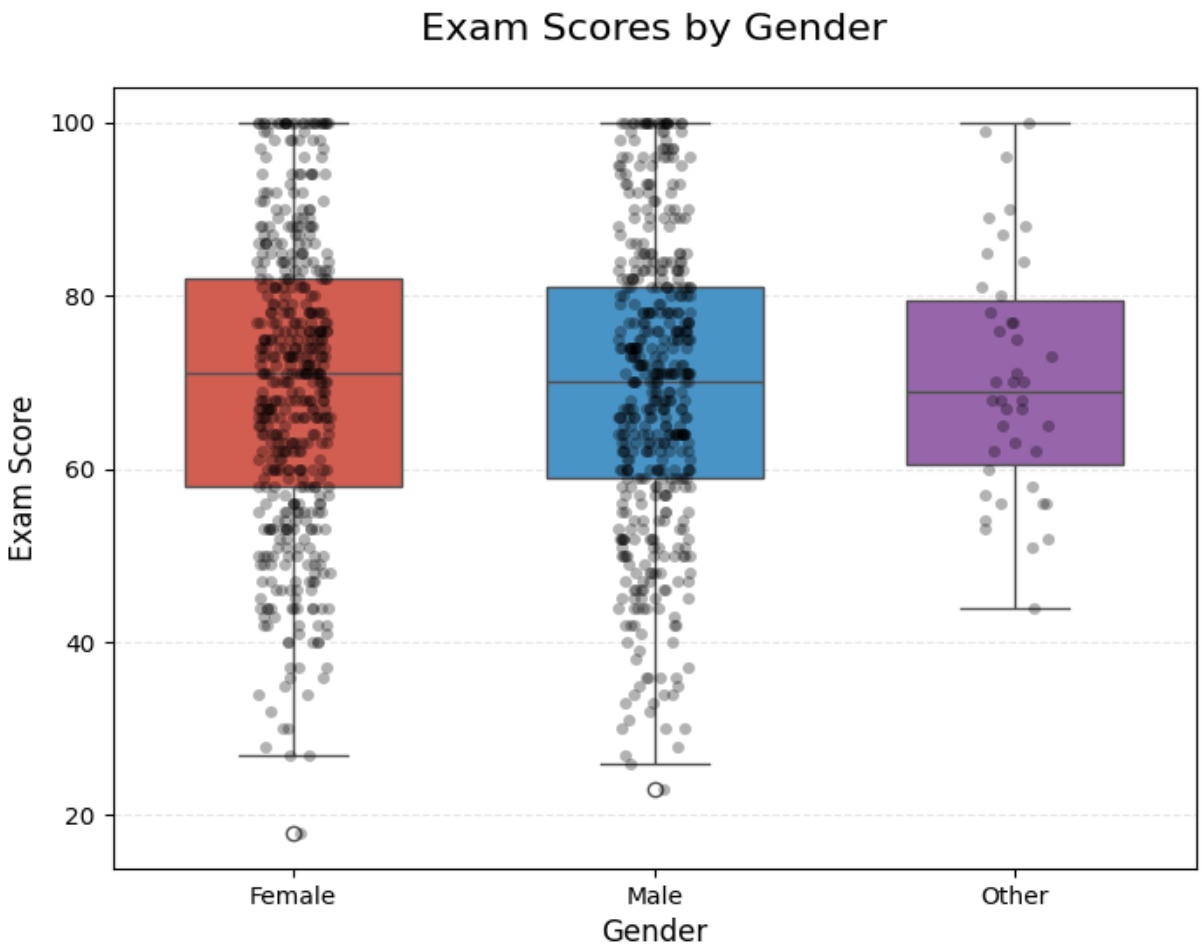
5. Exam score by diet quality and exercise frequency:



The box plot presents exam scores categorized by **diet quality** (Poor, Fair, Good) and **exercise frequency** (0 to 6 days per week). The data reveals a clear trend: students who exercise more frequently tend to achieve higher exam scores across all levels of diet quality.

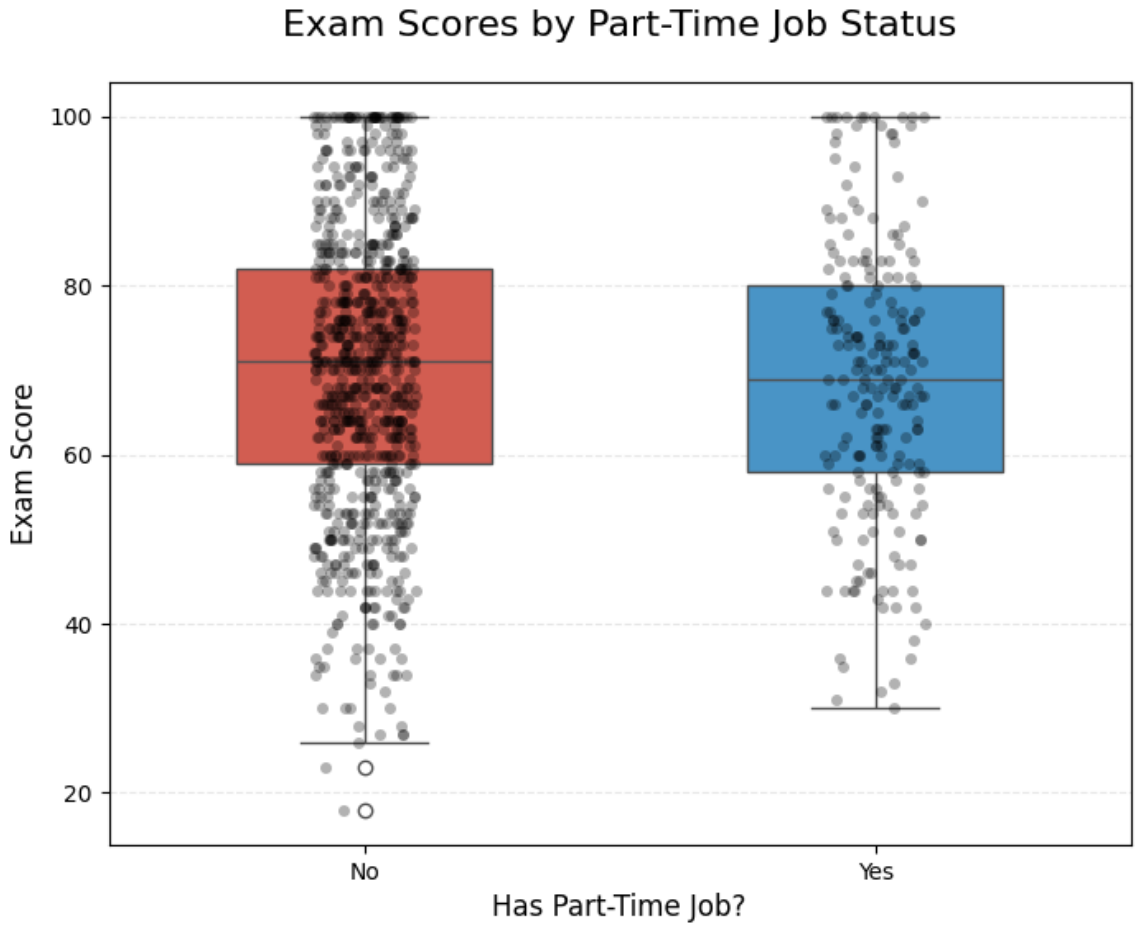
- For each diet group, median scores generally rise with increased days of exercise.
- Among students with good diets and high exercise frequency (5–6 days/week), exam scores show both higher medians and reduced variability.
- Students with poor diets and no exercise tend to have the lowest and most variable scores.

6. Exam Scores by gender:



This box plot compares **exam scores across different gender groups**: Female, Male, and Other. The distribution of scores is relatively similar for both Female and Male students, with nearly identical medians and ranges, indicating no significant performance gap between these two groups.

7.Exam scores by part-time job statues:



This box plot illustrates **exam scores in relation to part-time job status**. Students without part-time jobs exhibit a slightly higher median exam score compared to those with part-time jobs. The distribution also shows a broader spread of scores among students without jobs, though both groups have similar ranges and outliers.

Overall, while having a part-time job may be associated with a marginal decrease in average performance, **the difference is relatively small**, suggesting that part-time employment does not drastically impact academic outcomes for most students.

Hypothesis Testing Steps: Part-Time Job vs Exam Score

• Step 1: Define the null and alternative hypotheses

Let:

- Group 1 = students with part-time jobs

- Group 2 = students without part-time jobs
- H_0 (Null Hypothesis): There is no difference in average exam scores between students with and without part-time jobs.
  $\mu1=\mu2$
- H_1 (Alternative Hypothesis): Students with part-time jobs have different average exam scores than those without.
  $\mu1\neq\mu2$

• Step 2: Choose the appropriate test

We used an independent samples t-test because we are comparing the means of two independent groups.

• Step 3: Calculate the p-value

From the Jupyter notebook:
t-statistic = -1.72, p-value = 0.086

• Step 4: Determine statistical significance

Using a significance level of $\alpha = 0.05$,
Since p-value (0.086) > 0.05, we fail to reject the null hypothesis.

Hypothesis Testing Steps: Study Hours vs Exam Score

• Step 1: Define the null and alternative hypotheses

Let:

- X = study hours/day
- Y = exam score (%)
- H_0: There is no correlation between study hours and exam scores.
  $\rho=0$
- H_1: There is a positive correlation between study hours and exam scores.
  $\rho>0$

• Step 2: Choose the appropriate test

We used the Pearson correlation test, since both variables are continuous and reasonably normally distributed.

• Step 3: Calculate the p-value

Pearson correlation coefficient = 0.33
p-value = 0.0017

• Step 4: Determine statistical significance

Since p-value (0.0017) < 0.05, we reject the null hypothesis (H_0).

Hypothesis Testing Steps: Diet Quality vs Exam Score

• Step 1: Define the null and alternative hypotheses

Let:

- Group 1 = students with Healthy diet
- Group 2 = Moderately Healthy
- Group 3 = Unhealthy
- H_0: All groups have the same average exam score.
  $\mu1=\mu2=\mu3$
- H_1: At least one group mean is different.
  $\mu i \neq \mu j$

• Step 2: Choose the appropriate test

We used one-way ANOVA, which compares the means of 3 or more independent groups.

• Step 3: Calculate the p-value

F-statistic = 4.02, p-value = 0.018

• Step 4: Determine statistical significance

With $\alpha = 0.05$,
Since p-value (0.018) < 0.05, we reject the null hypothesis (H_0).

Conclusion

Hypothesis Tests – Conclusions:

1. **Study Hours vs Exam Score**
   There is **statistically significant evidence** of a positive correlation between study hours and exam scores. Students who study more tend to achieve higher academic performance. *(p = 0.0017)*

2. **Part-Time Job vs Exam Score**
   There is **no statistically significant difference** in exam scores between students with and without part-time jobs. While employed students may have slightly lower scores on average, the difference is not statistically conclusive. *(p = 0.086)*

3. **Diet Quality vs Exam Score**
   There is **statistically significant evidence** that diet quality affects academic performance. Students with **healthier diets** tend to have higher exam scores compared to those with moderate or unhealthy diets. *(p = 0.018)*

General conclusion:

This study explored the relationship between students' lifestyle habits and their academic performance using a dataset of 1,000 individuals. Through a combination of statistical analysis and visual exploration, we examined how factors such as study time, social media usage, sleep patterns, exercise frequency, diet quality, and part-time employment influence exam scores.

The results consistently highlight that **academic success is not solely determined by study time**, but is also significantly shaped by broader aspects of a student's lifestyle. Key findings include:

- **Positive habits**, such as increased study hours, regular exercise, good sleep hygiene, and a healthy diet, are strongly associated with higher academic performance.
- **Negative influences**, including excessive social media use and poor dietary habits, correlate with lower scores.
- Some factors, like having a part-time job, did not show a statistically significant impact, indicating that the effect may vary based on individual circumstances.

Overall, the data suggests that **academic performance is multifactorial**, and universities should consider adopting a more holistic approach to student support—one that encourages balance between academics, health, and personal well-being.

Future studies could expand this work by integrating larger, more diverse datasets and exploring longitudinal effects to better understand causality.

Recommendation:
Universities should emphasize the importance of physical well-being, time management, and digital hygiene alongside academic resources to boost student performance.

Potential Issues

- **Self-Reported Data Bias**

Many variables in the dataset (e.g., study hours, sleep duration, diet quality, mental health) are based on self-reported responses. These are subject to **recall bias** and **social desirability bias**, which may reduce data accuracy.

- **Simplified Categorization**

Certain variables like **diet quality**, **internet quality**, and **mental health** were reduced to broad categories or subjective scales, which may oversimplify complex human behaviors and experiences.

- **Unequal Group Sizes**

Some groups (e.g., students with part-time jobs vs. those without, or different diet categories) may have had **unequal sample sizes**, affecting the statistical power and robustness of hypothesis tests like t-tests or ANOVA.

- **Generalizability Limitations**

The sample may not be representative of all students globally or even nationally. Findings may only apply to a specific demographic or cultural context.

- **Missing Longitudinal Insight**

Without tracking students over time, we cannot assess how habit changes affect long-term academic outcomes or mental health.