

"Analyzing the Impact of Lifestyle and Study Habits on Academic Performance"

Team :VS

Date: 24 June 2025

Problem Statement

- How do lifestyle factors like sleep, study time, diet, and screen use affect exam scores?

Objective: Explore and model these effects using statistical techniques.

Dataset Overview

- Brief explanation of features: demographics, habits, academic indicators
- Target variable: exam_score

	student_id	age	gender	study_hours_per_day	social_media_hours	\
0	S1000	23	Female		0.0	1.2
1	S1001	20	Female		6.9	2.8
2	S1002	21	Male		1.4	3.1
3	S1003	23	Female		1.0	3.9
4	S1004	19	Female		5.0	4.4

	netflix_hours	part_time_job	attendance_percentage	sleep_hours	\
0	1.1	No	85.0	8.0	
1	2.3	No	97.3	4.6	
2	1.3	No	94.8	8.0	
3	1.0	No	71.0	9.2	
4	0.5	No	90.9	4.9	

	diet_quality	exercise_frequency	parental_education_level	internet_quality	\
0	Fair		6	Master	Average
1	Good		6	High School	Average
2	Poor		1	High School	Poor
3	Poor		4	Master	Good
4	Fair		3	Master	Good

	mental_health_rating	extracurricular_participation	exam_score
0	8	Yes	56.2
1	8	No	100.0
2	1	No	34.3
3	1	Yes	26.8
4	1	No	66.4

Literature Review

- Sánchez-Hernando, B., Gasch-Gallén, Á., Antón-Solanas, I., Gea-Caballero, V., Juárez-Vela, R., Gállego-Diéguéz, J., Carboneres-Tafaner, M. I., Echániz-Serrano, E., Lasso-Olayo, L., & Santolalla-Arnedo, I. (2022). A comparative study of life skills, lifestyle habits and academic performance in health promoting and non-health promoting schools in the Autonomous Community of Aragon, Spain. PeerJ, 10, e13041. <https://doi.org/10.7717/peerj.13041>
- Kang, J., & Chen, S. (2009). Effects of an irregular bedtime schedule on sleep quality, daytime sleepiness, and fatigue among university students in Taiwan. BMC Public Health, 9(1). <https://doi.org/10.1186/1471-2458-9-248>
- Golombek, D. A., & Cardinali, D. P. (2008). Mind, brain, education, and biological timing. Mind Brain and Education, 2(1), 1–6. <https://doi.org/10.1111/j.1751-228x.2008.00022.x>



TYPE OF PROBLEM & STATISTICAL MODELS

PROBLEM TYPE:

- Regression: to predict marks from numbers like sleep hours.
- Group Comparison: to compare marks between groups like good vs poor diet.

STATISTICAL MODELS:

- Linear Regression: To predict exam_score from numerical predictors.
- T-Test / ANOVA: To compare group performance (e.g., breakfast habits, diet quality).
- Correlation Analysis: To assess linear relationships (e.g., study hours vs score).

LINEAR REGRESSION

- Linear regression helps us guess marks using things like how much the student studies or sleeps.
- We used this method to check if more sleep or better mental health means better marks.
- It draws a line to show the relationship between habits and exam scores.

Data SET Used

← → ⌂ kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance

≡ **kaggle**

Create

Home

Competitions

Datasets

Models

Code

Discussions

Search

JAYANTA NATH · UPDATED 2 MONTHS AGO

▲ 686

Student Habits vs Academic Performance

Great for regression, classification, EDA, visualization, and even ML practice.

Data Card Code (92) Discussion (3) Suggestions (0)

A screenshot of a web browser showing a Kaggle dataset page. The URL in the address bar is kaggle.com/datasets/jayaantanaath/student-habits-vs-academic-performance. The page title is "Student Habits vs Academic Performance". On the left, there's a sidebar with navigation links: Create, Home, Competitions, Datasets (which is highlighted), Models, Code, and Discussions. The main content area shows the dataset details: it was created by JAYANTA NATH and updated 2 months ago, with 686 versions. Below the title is a description: "Great for regression, classification, EDA, visualization, and even ML practice." At the bottom, there are four buttons: Data Card, Code (92), Discussion (3), and Suggestions (0). The "Data Card" button is underlined, indicating it's the active tab.

Data Preprocessing

- Removed duplicate records
- Handled missing values using forward fill
- Fix Wrong Data Types
- Rename columns
- Save cleaned file
- Prepared clean dataset for analysis

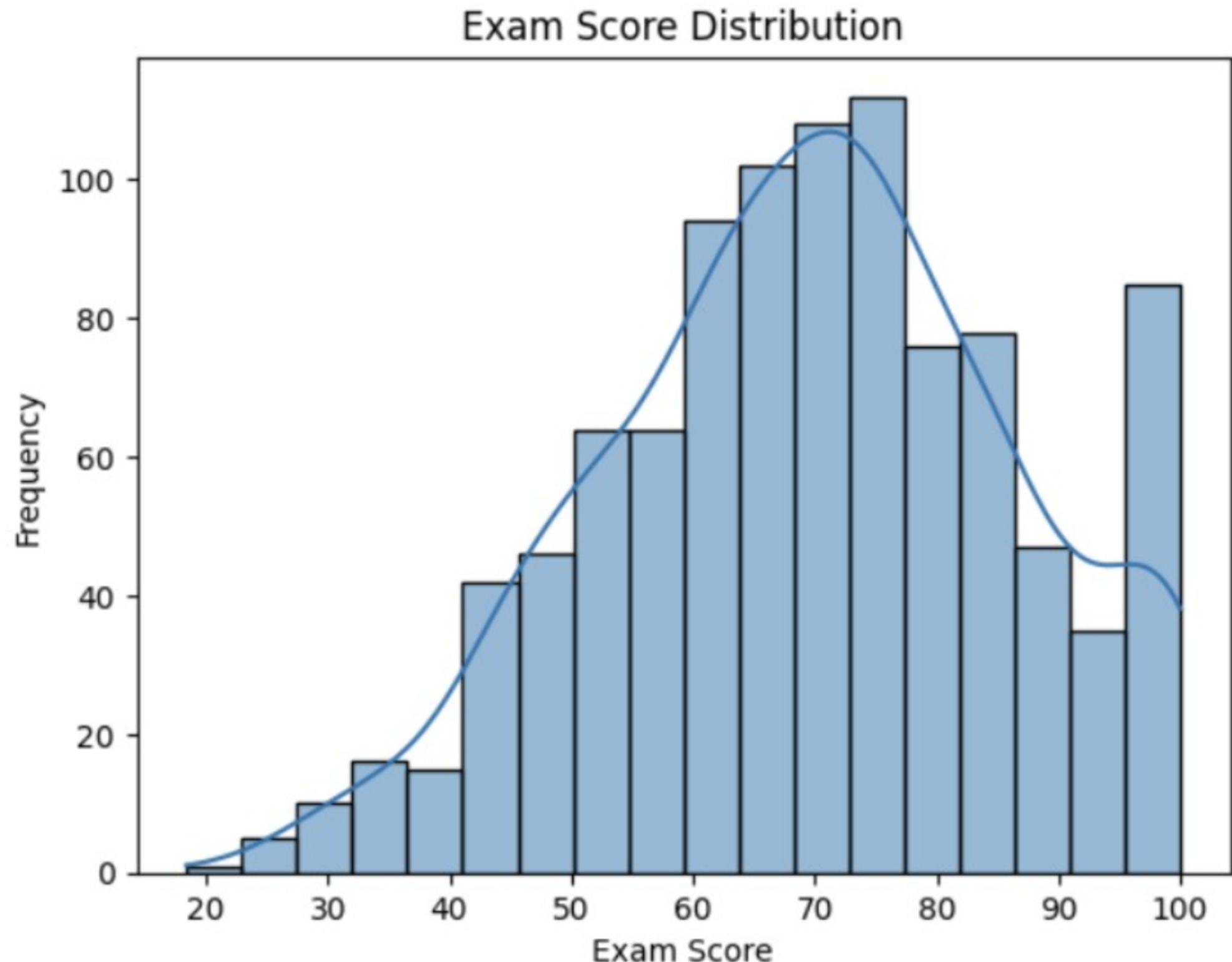
```
# Drop duplicates
df.drop_duplicates(inplace=True)

# Handle missing values
df.fillna(method='ffill', inplace=True)
```

✓ 0.0s

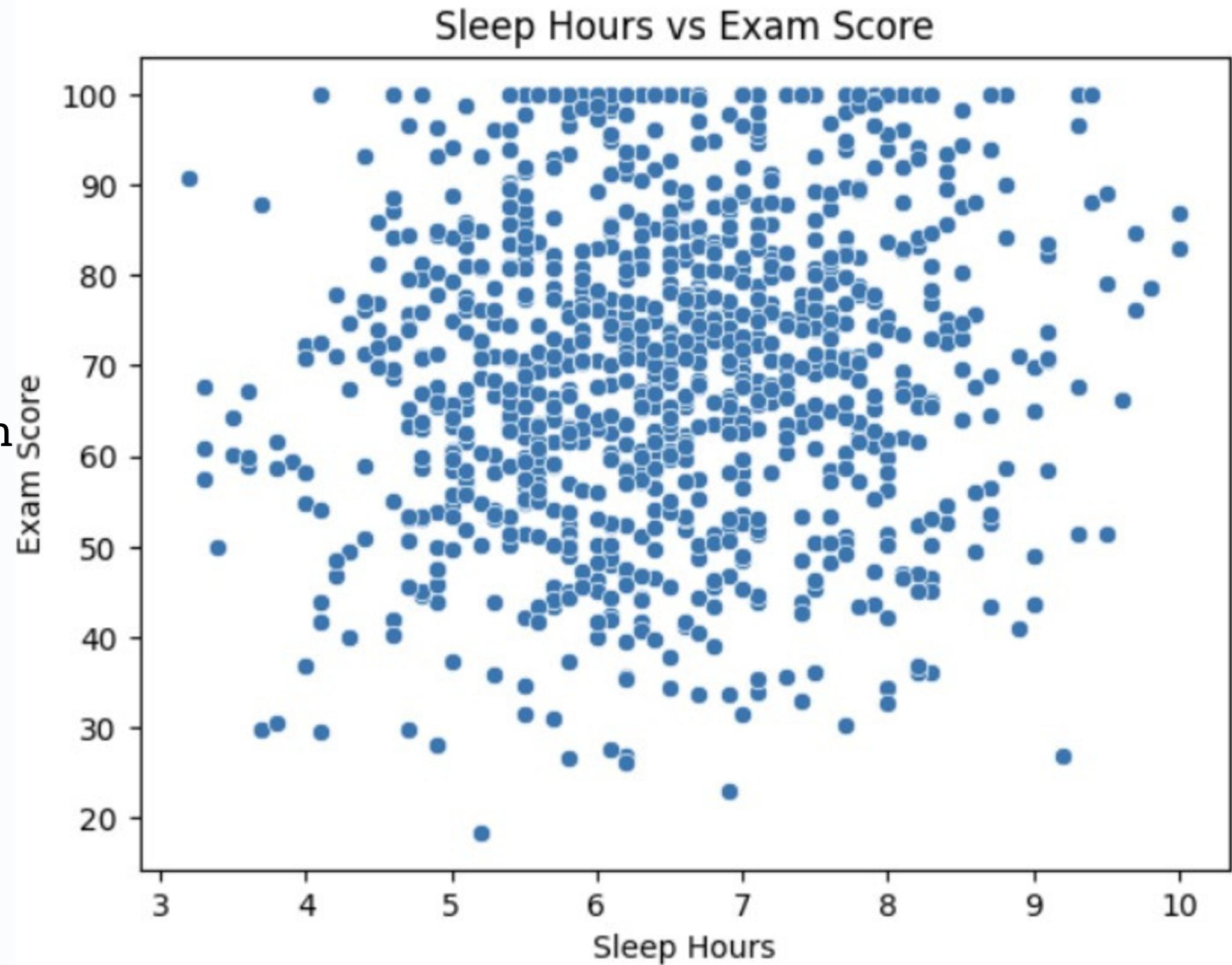
Exam Score Distribution

- Most students scored between 70 and 75.
- This shows a bell-shaped (normal) distribution.
- It means many students performed around the average.



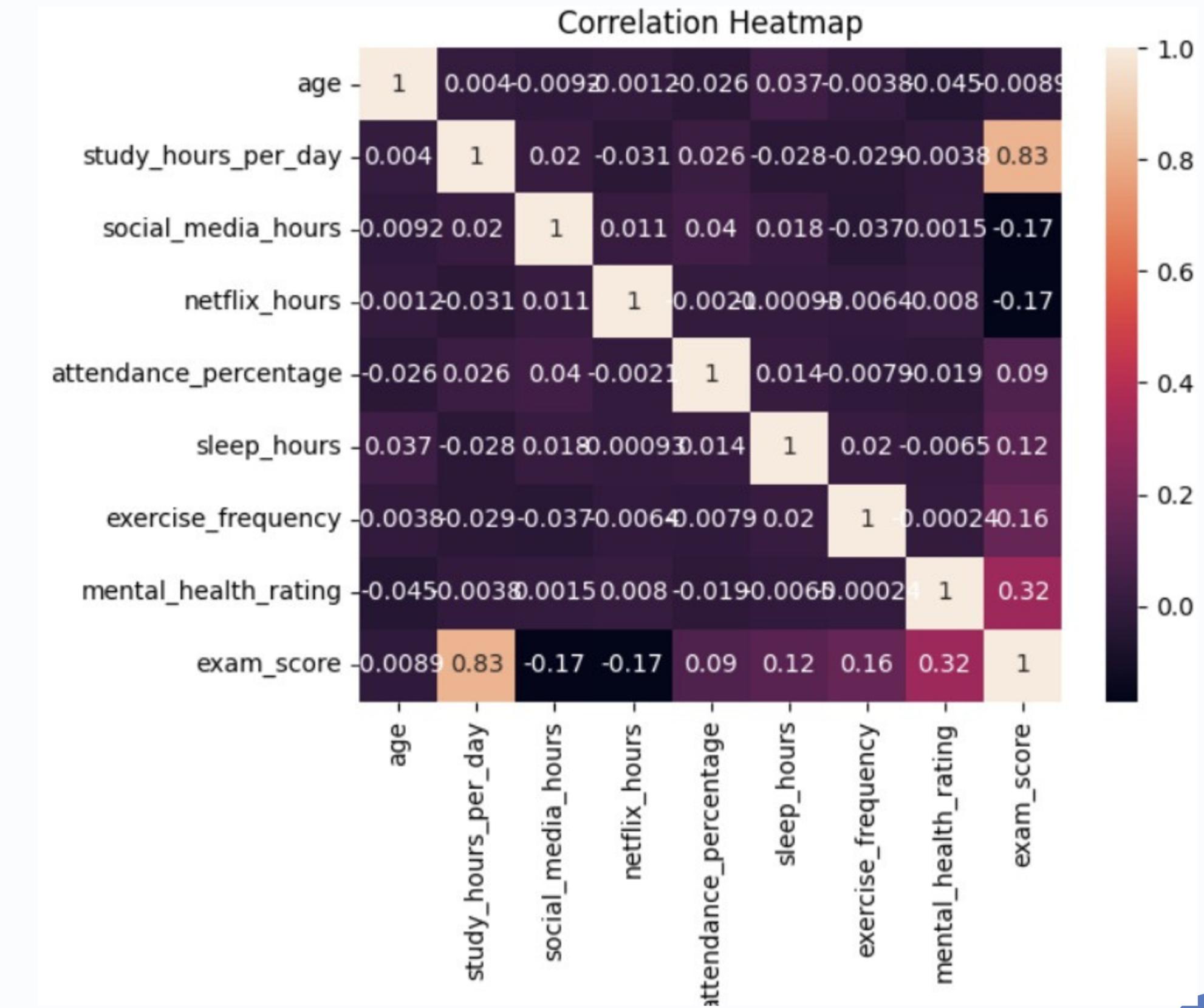
Sleep vs Exam Score – "Does Sleeping More Help?"

- Scatter plot shows weak or no direct correlation between sleep hours and performance



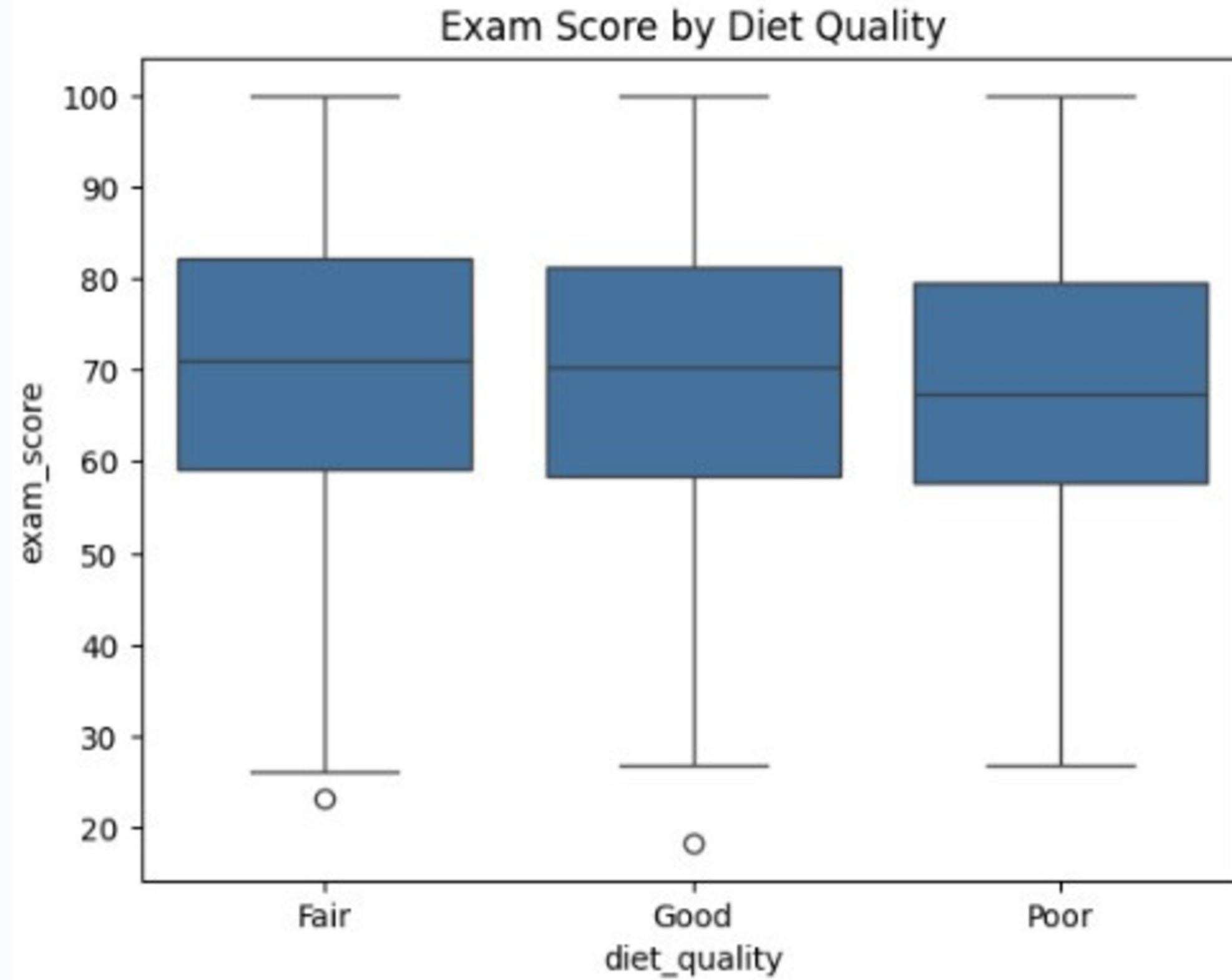
Correlation Heatmap

- `study_hours_per_day` shows strongest correlation with exam score ($r = 0.83$)
- Moderate positive with mental health & exercise



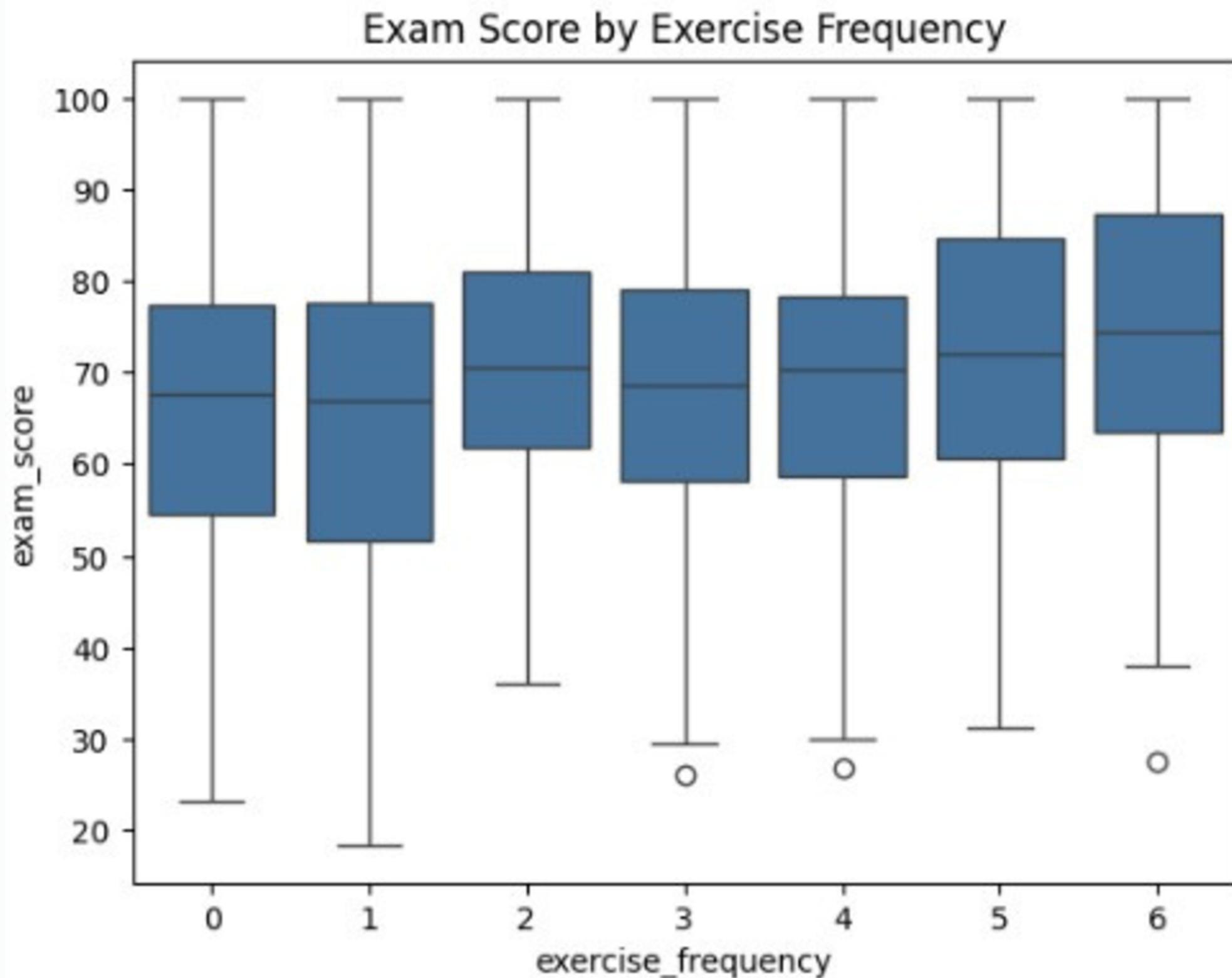
Lifestyle Factors

- Diet quality shows minimal difference



Lifestyle Factors

- Boxplots: Better exercise frequency shows higher scores



Regression Analysis

- OLS regression confirms: every 1 extra study hour increases score by ~9.49
 - $R^2 = 0.681$ (68% variance explained)



Hypothesis Testing

Hypotheses:

- H1: Study time improves scores
- H2: Sleep affects performance
- H3: Better diet = better scores

Tests Performed:

- Linear Regression
- T-test (Good vs Poor Diet)
- ANOVA (Sleep groups)

Week 6 - Insights and Reflection

Insights:

- Study time = strongest predictor
- Sleep pattern also significant
- Diet → less impact

Reflections:

- Behavior-based interventions can boost performance





Week 7 - Streamlit Application

Developed a local Python app using Streamlit:

- User inputs habits (sleep, study hours, etc.)
- Predicts performance category
- Visualizes output

Main Files:

- Academic_Prediction.py- This is the main file. It runs the app.
- retrain_model.py- This file trains or updates the prediction model.
- model/model.pkl- This is the saved machine learning model file.

Key Findings & Conclusion

- Most influential: Study hours > Mental health > Exercise
- Negative impact: Excess Netflix/social media
- Students who study more tend to score better.
- Good mental health and regular exercise help improve results.
- Diet and gender didn't make much difference.



APPLICATION DEMO

THANK
YOU!

