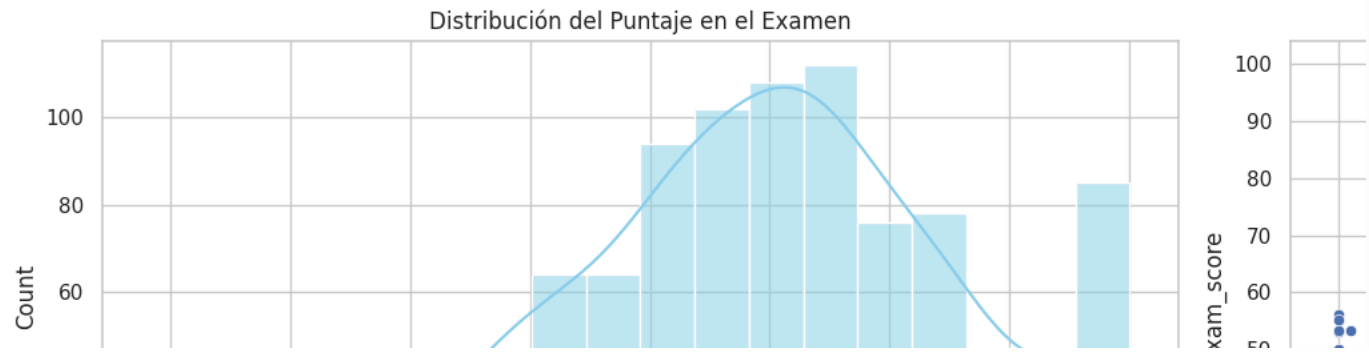
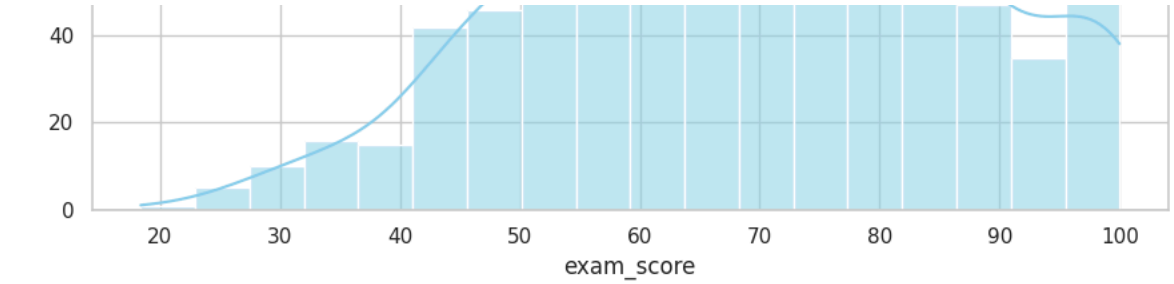
```
from google.colab import files
uploaded = files.upload()

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("student_habits_performance.csv")
df.head()
print(df)

sns.set(style="whitegrid")
#I use the code we work yesterday
# Crear múltiples subplots
fig, axs = plt.subplots(5, 2, figsize=(18, 24))
axs = axs.flatten()

# 1. Distribución del puntaje
sns.histplot(df['exam_score'], kde=True, ax=axs[0], color="skyblue")
axs[0].set_title("Distribución del Puntaje en el Examen")

# 2. Horas de estudio vs puntaje
sns.scatterplot(data=df, x='study_hours_per_day', y='exam_score', ax=axs[1])
axs[1].set_title("Horas de Estudio vs Puntaje en el Examen")

# 3. Salud mental vs puntaje
sns.boxplot(data=df, x='mental_health_rating', y='exam_score', ax=axs[2])
axs[2].set_title("Salud Mental vs Puntaje en el Examen")

# 4. Educación de los padres
sns.boxplot(data=df, x='parental_education_level', y='exam_score', ax=axs[3])
axs[3].set_title("Nivel Educativo de los Padres vs Puntaje")

# 5. Calidad del internet
sns.boxplot(data=df, x='internet_quality', y='exam_score', ax=axs[4])
axs[4].set_title("Calidad del Internet vs Puntaje")

# 6. Dieta y puntaje
sns.boxplot(data=df, x='diet_quality', y='exam_score', ax=axs[5])
axs[5].set_title("Calidad de la Dieta vs Puntaje")

# 7. Participación extracurricular
sns.boxplot(data=df, x='extracurricular_participation', y='exam_score', ax=axs[6])
axs[6].set_title("Actividades Extracurriculares vs Puntaje")

# 8. Redes sociales vs puntaje
sns.scatterplot(data=df, x='social_media_hours', y='exam_score', ax=axs[7])
axs[7].set_title("Horas en Redes Sociales vs Puntaje")

# 9. Trabajo medio tiempo
sns.boxplot(data=df, x='part_time_job', y='exam_score', ax=axs[8])
axs[8].set_title("Trabajo de Medio Tiempo vs Puntaje")

# 10. Sueño vs puntaje
sns.regplot(data=df, x='sleep_hours', y='exam_score', ax=axs[9], scatter_kws={'alpha':0.5})
axs[9].set_title("Horas de Sueño vs Puntaje")

plt.tight_layout()
plt.show()

# Heatmap de correlación entre variables numéricas
plt.figure(figsize=(12, 8))
corr = df.corr(numeric_only=True)
sns.heatmap(corr, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Heatmap de Correlaciones entre Variables Numéricas")
plt.show()

```

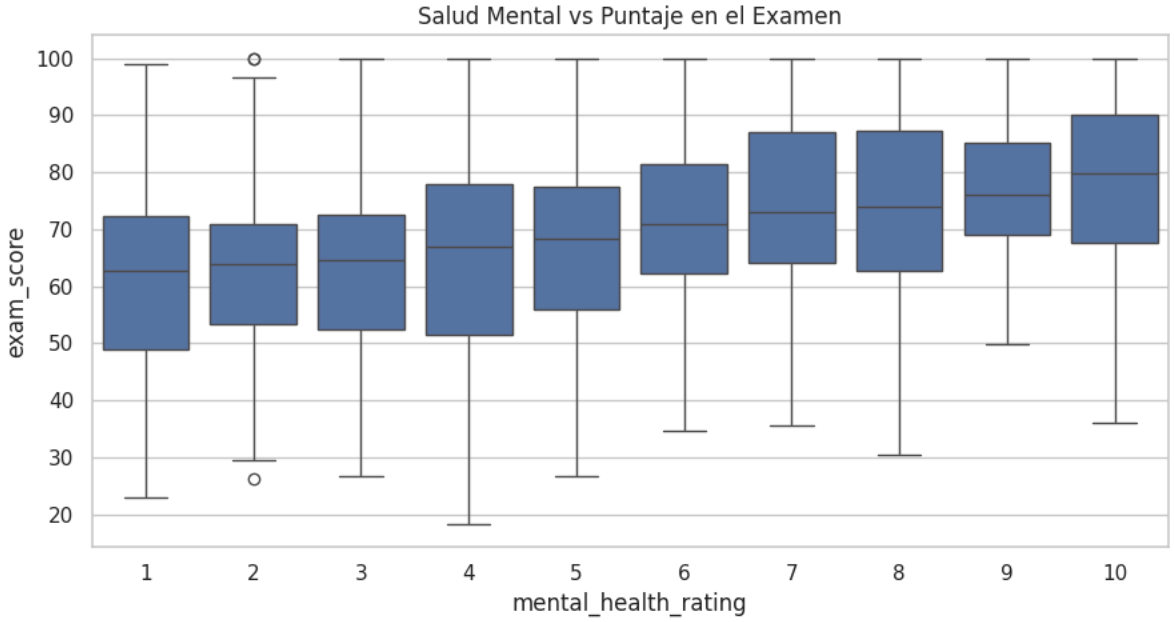
Elegir archivos student_hab…rmance.csv

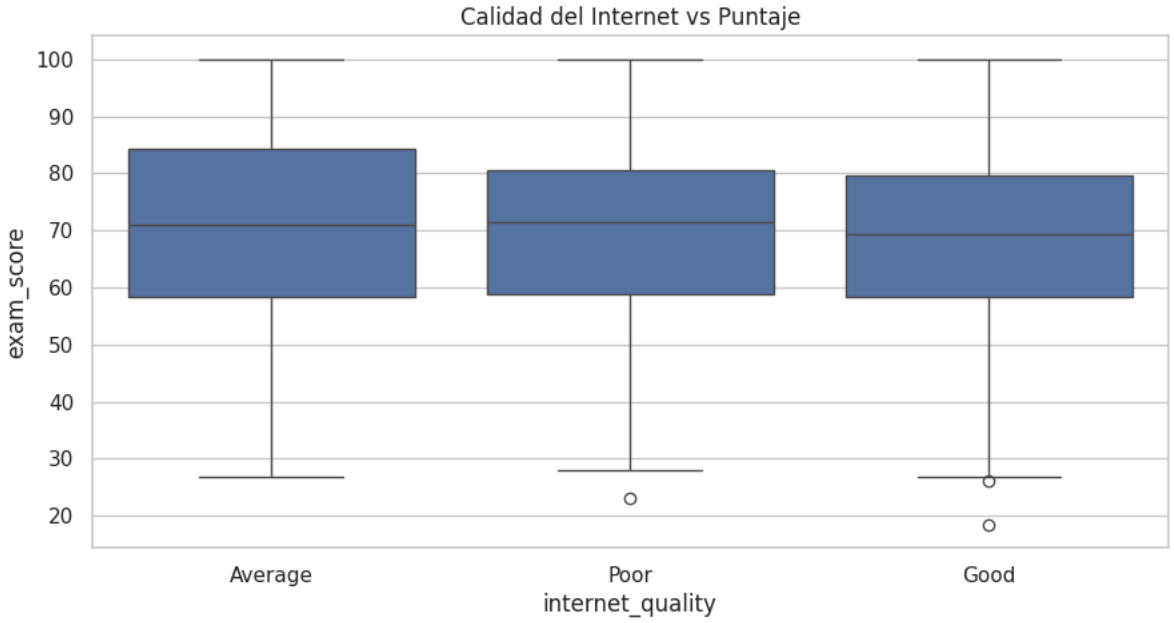
- **student_habits_performance.csv**(text/csv) - 73663 bytes, last modified: 6/5/2025 - 100% done

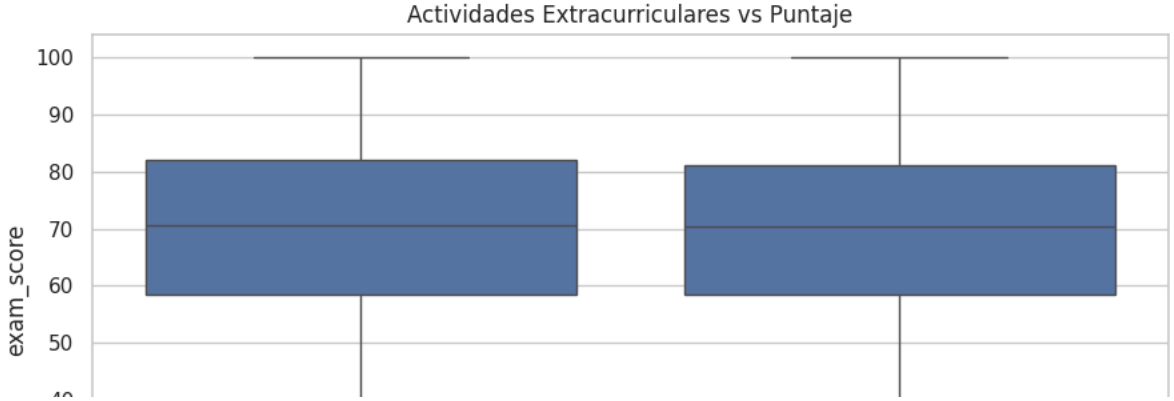
```
Saving student_habits_performance.csv to student_habits_performance (2).csv
    student_id  age  gender  study_hours_per_day  social_media_hours  \
0        S1000   23  Female                  0.0                 1.2
1        S1001   20  Female                  6.9                 2.8
2        S1002   21    Male                  1.4                 3.1
3        S1003   23  Female                  1.0                 3.9
4        S1004   19  Female                  5.0                 4.4
..         ...  ...     ...                  ...                 ...
995      S1995   21  Female                  2.6                 0.5
996      S1996   17  Female                  2.9                 1.0
997      S1997   20    Male                  3.0                 2.6
998      S1998   24    Male                  5.4                 4.1
999      S1999   19  Female                  4.3                 2.9

     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
..             ...           ...                    ...          ...
995            1.6            No                   77.0          7.5
996            2.4           Yes                   86.0          6.8
997            1.3            No                   61.9          6.5
998            1.1           Yes                  100.0          7.6
999            1.9            No                   89.4          7.1

    diet_quality  exercise_frequency parental_education_level  \
0           Fair                   6                  Master
1           Good                   6             High School
2           Poor                   1             High School
3           Poor                   4                  Master
4           Fair                   3                  Master
..           ...                 ...                     ...
995         Fair                   2             High School
996         Poor                   1             High School
997         Good                   5                Bachelor
998         Fair                   0                Bachelor
999         Good                   2                Bachelor

    internet_quality  mental_health_rating extracurricular_participation  \
0            Average                     8                           Yes
1            Average                     8                            No
2               Poor                     1                            No
3               Good                     1                           Yes
4               Good                     1                            No
..               ...                   ...                           ...
995             Good                     6                           Yes
996          Average                     6                           Yes
997             Good                     9                           Yes
998          Average                     1                            No
999          Average                     8                            No

     exam_score
0          56.2
1         100.0
2          34.3
3          26.8
4          66.4
..          ...
995        76.1
996        65.9
997        64.4
998        69.7
999        74.9

[1000 rows x 16 columns]
```

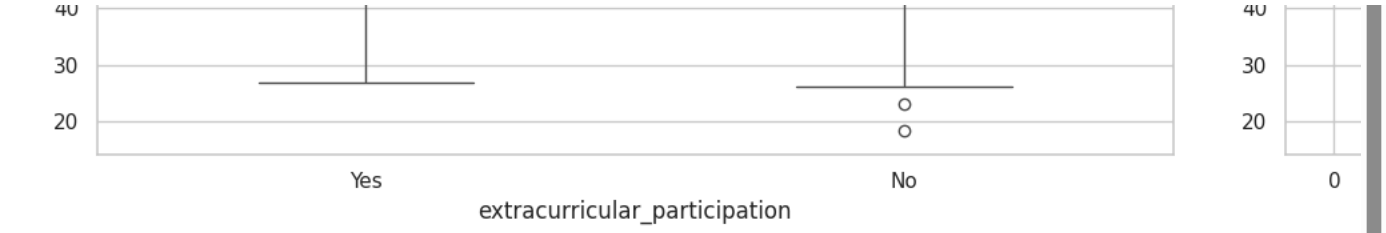


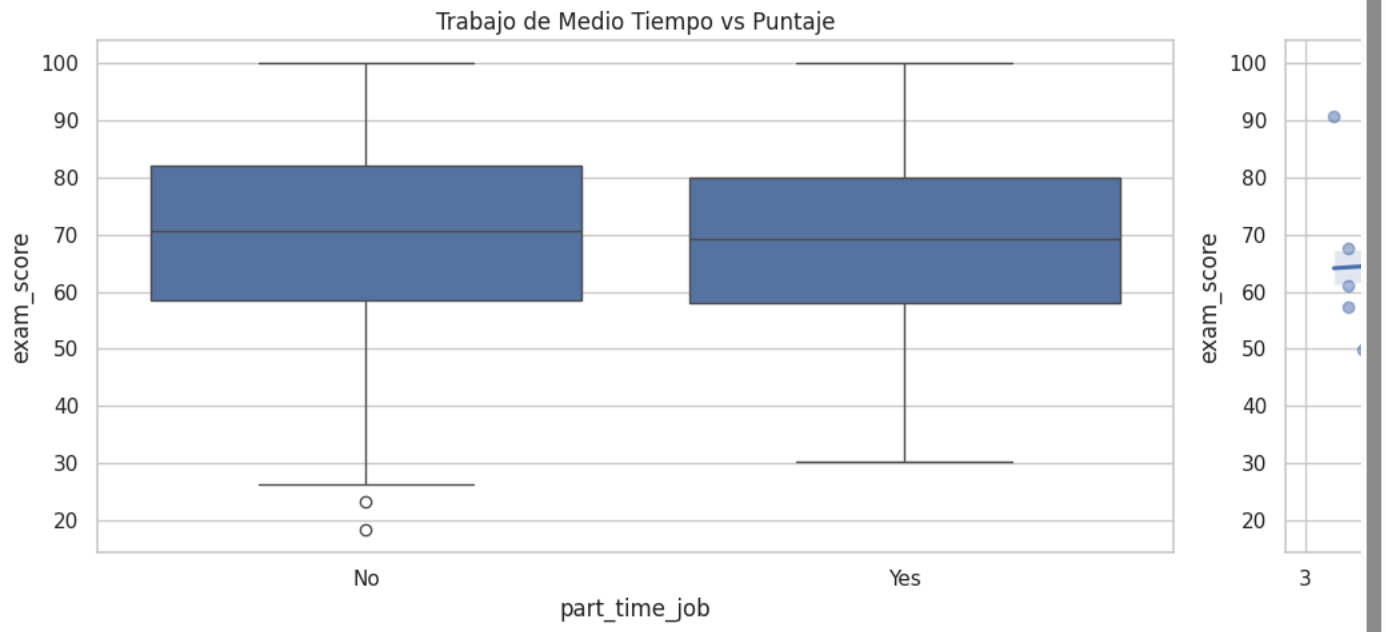

exam_score

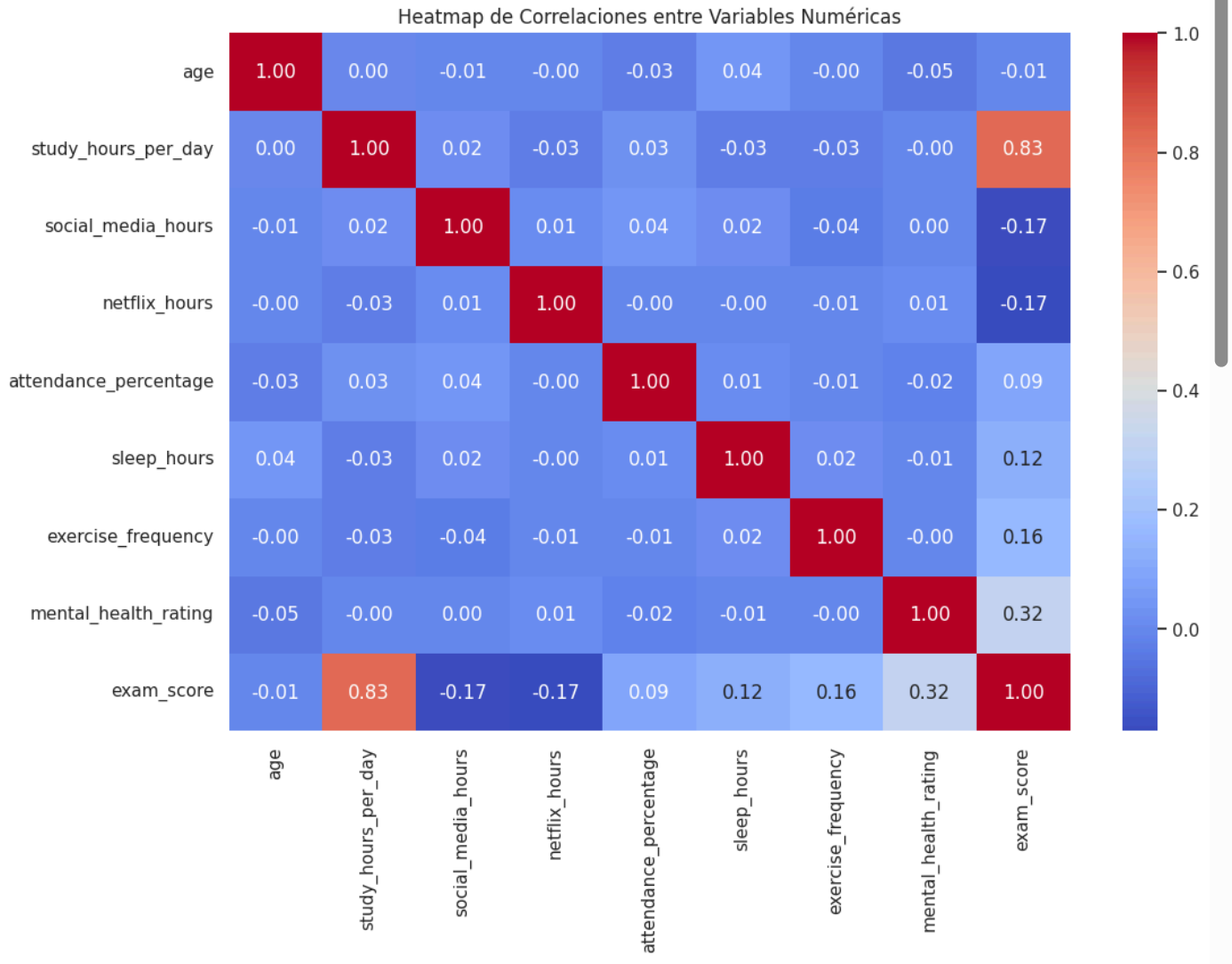

Salud Mental vs Puntaje en el Examen
exam_score
mental_health_rating


Calidad del Internet vs Puntaje
exam_score
Average
Poor
Good
internet_quality


Actividades Extracurriculares vs Puntaje
exam_score


40
30
20
Yes
No
extracurricular_participation
40
30
20
0


Trabajo de Medio Tiempo vs Puntaje
exam_score
100
90
80
70
60
50
40
30
20
No
Yes
part_time_job
exam_score
100
90
80
70
60
50
40
30
20
3


Heatmap de Correlaciones entre Variables Numéricas
age 1.00 0.00 -0.01 -0.00 -0.03 0.04 -0.00 -0.05 -0.01
study_hours_per_day 0.00 1.00 0.02 -0.03 0.03 -0.03 -0.03 -0.00 0.83
social_media_hours -0.01 0.02 1.00 0.01 0.04 0.02 -0.04 0.00 -0.17
netflix_hours -0.00 -0.03 0.01 1.00 -0.00 -0.00 -0.01 0.01 -0.17
attendance_percentage -0.03 0.03 0.04 -0.00 1.00 0.01 -0.01 -0.02 0.09
sleep_hours 0.04 -0.03 0.02 -0.00 0.01 1.00 0.02 -0.01 0.12
exercise_frequency -0.00 -0.03 -0.04 -0.01 -0.01 0.02 1.00 -0.00 0.16
mental_health_rating -0.05 -0.00 0.00 0.01 -0.02 -0.01 -0.00 1.00 0.32
exam_score -0.01 0.83 -0.17 -0.17 0.09 0.12 0.16 0.32 1.00
age
study_hours_per_day
social_media_hours
netflix_hours
attendance_percentage
sleep_hours
exercise_frequency
mental_health_rating
exam_score
1.0
0.8
0.6
0.4
0.2
0.0

Are there any variables that do not provide information?

Variables like student_id are identifiers and do not provide useful information for analysis

If you had to eliminate variables, which ones would you remove and why?

I would remove student_id because it's an identifier

Are there any variables with unusual data?

Some outliers may be seen in the exam_score histogram or in the scatterplot of study_hours_per_day

If you compare the variables, are they all in similar ranges?

No, variables like exam_score (0–100) and social_media_hours or sleep_hours have different scales.

Do you think this affects the data analysis? Can you find any similar groups? What are these groups?

Yes, normalization or standardization is needed when comparing across different scales.

The scatterplots and boxplots show some trends, like students with more study hours or better mental health generally scoring higher. These may form clusters, such as:

- High performers with good habits
- Low performers with poor health or habits
- Average students with mixed patterns

https://github.com/SvpremeLemus/Act