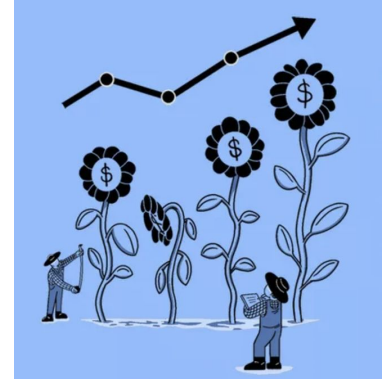
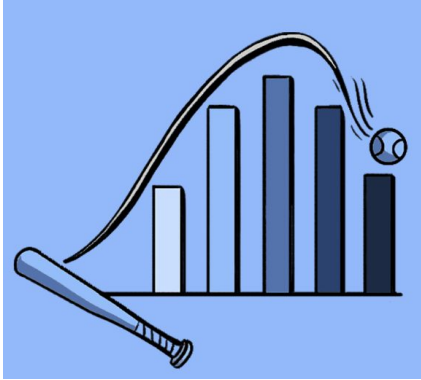




# Student Health Data Analysis

Group 10 - Project 1



Cetin, Ugur - Foster, Chandler - Khatri, Kriti - Patel, Aayushi - Ruiz, Willian

# Introduction and Data Overview

## About the Dataset / Student Health Data

This dataset focuses on assessing health risks in high-stress environments, such as academic and entrepreneurial settings, using physiological, psychological, and academic data from college students. It supports machine learning applications for predicting occupational health risks like stress and physical strain.

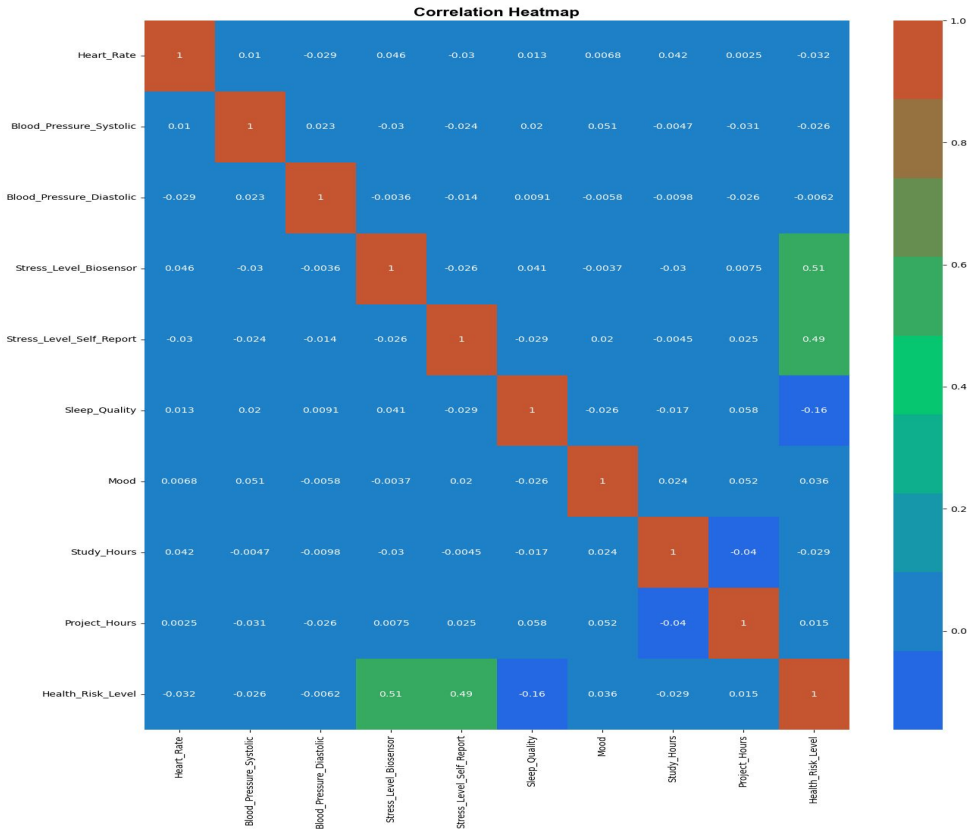
### Key Features:

- Demographics:
  - a. Study Group → 1000 people
  - b. Age → between 18 to 24
  - c. Gender → 519 male, 481 female
- Physiological Data:
  - a. Heart rate
  - b. Blood pressure
  - c. Biosensor-based stress levels → 1 to 10 scale
- Psychological Data:
  - a. Self-reported stress → 1 to 10 scale
  - b. Mood states → Happy - Neutral - Stressed
- Academic Activity: Hours spent on academic and entrepreneurial tasks.
- Lifestyle Factors:
  - a. Daily physical activity → High - Moderate - Low
  - b. Sleep quality → Good - Moderate - Poor
- Health Risk Label: Low, moderate, or high risk derived from combined metrics.

Kaggle Link → <https://www.kaggle.com/datasets/ziva07/student-health-data>

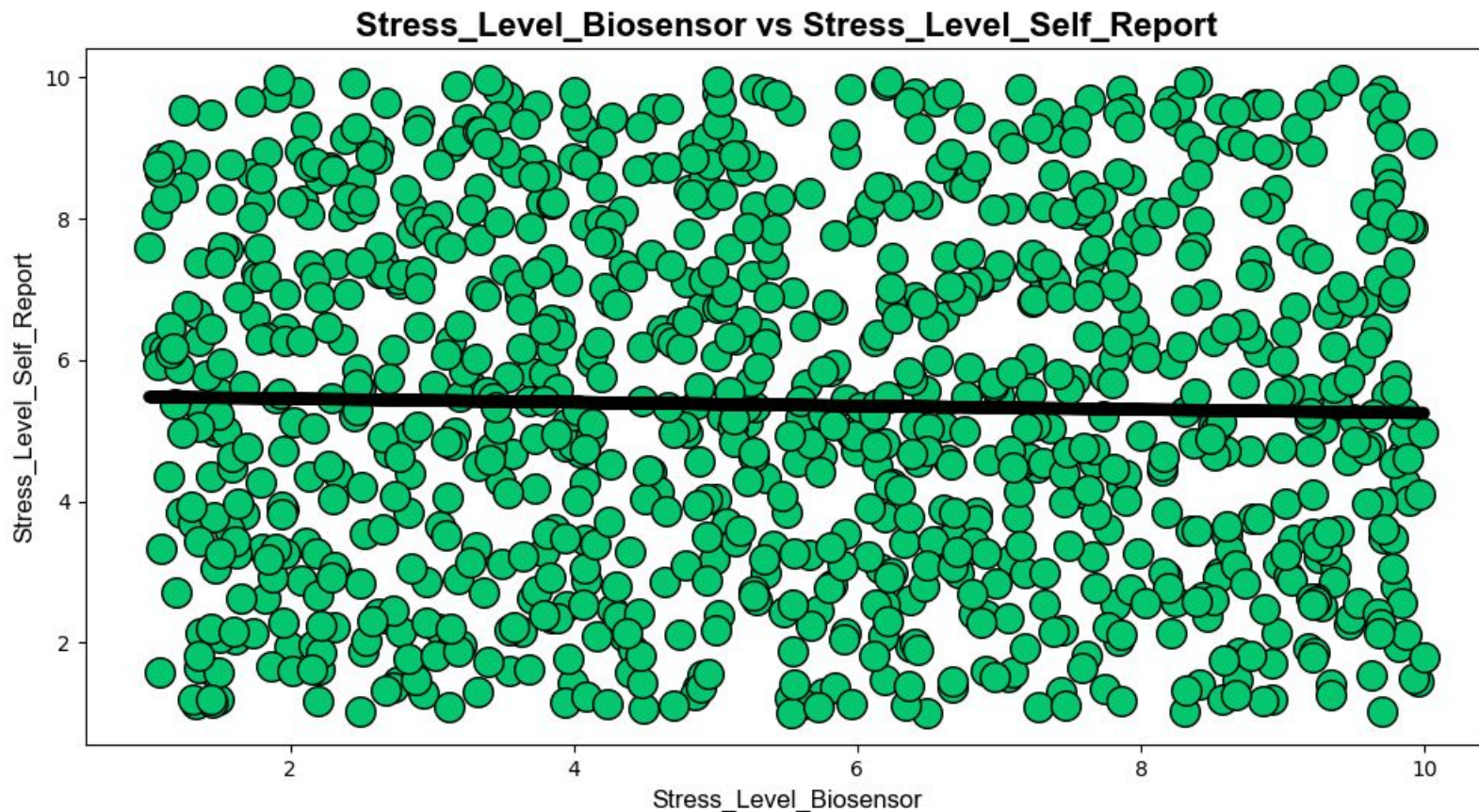


# Question 1: What is the relationship between Stress and Health Risk?

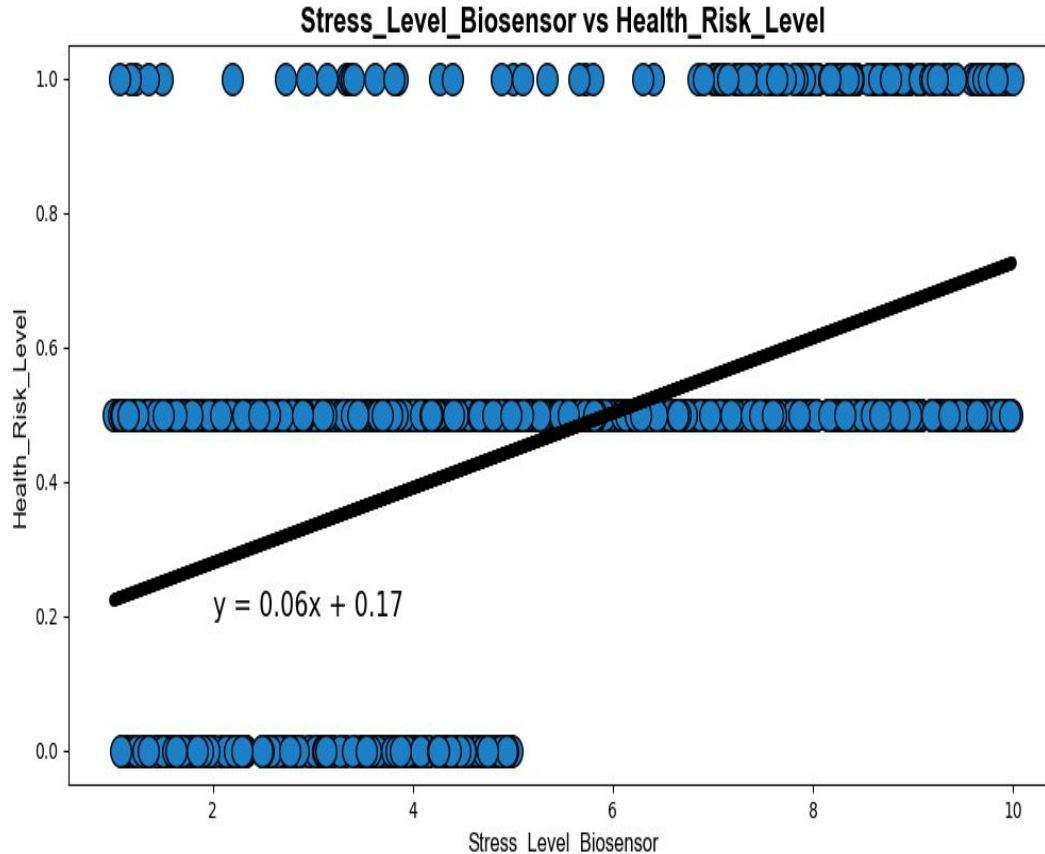


An initial look at potential stress factors or indicators shows that no singular stress factor or indicator correlates strongly with Health Risk. The only strong correlation this data suggests is between stress level and health risk level. Self reported stress levels and biosensor stress levels were very close, which means that self-reports were relatively accurate. This data suggests that higher stress leads to an increased Health Risk Level.

Subjects with higher biosensor stress levels had a slight tendency to self report lower stress levels.



It is clear that increased stress levels correlate with increased health risk.

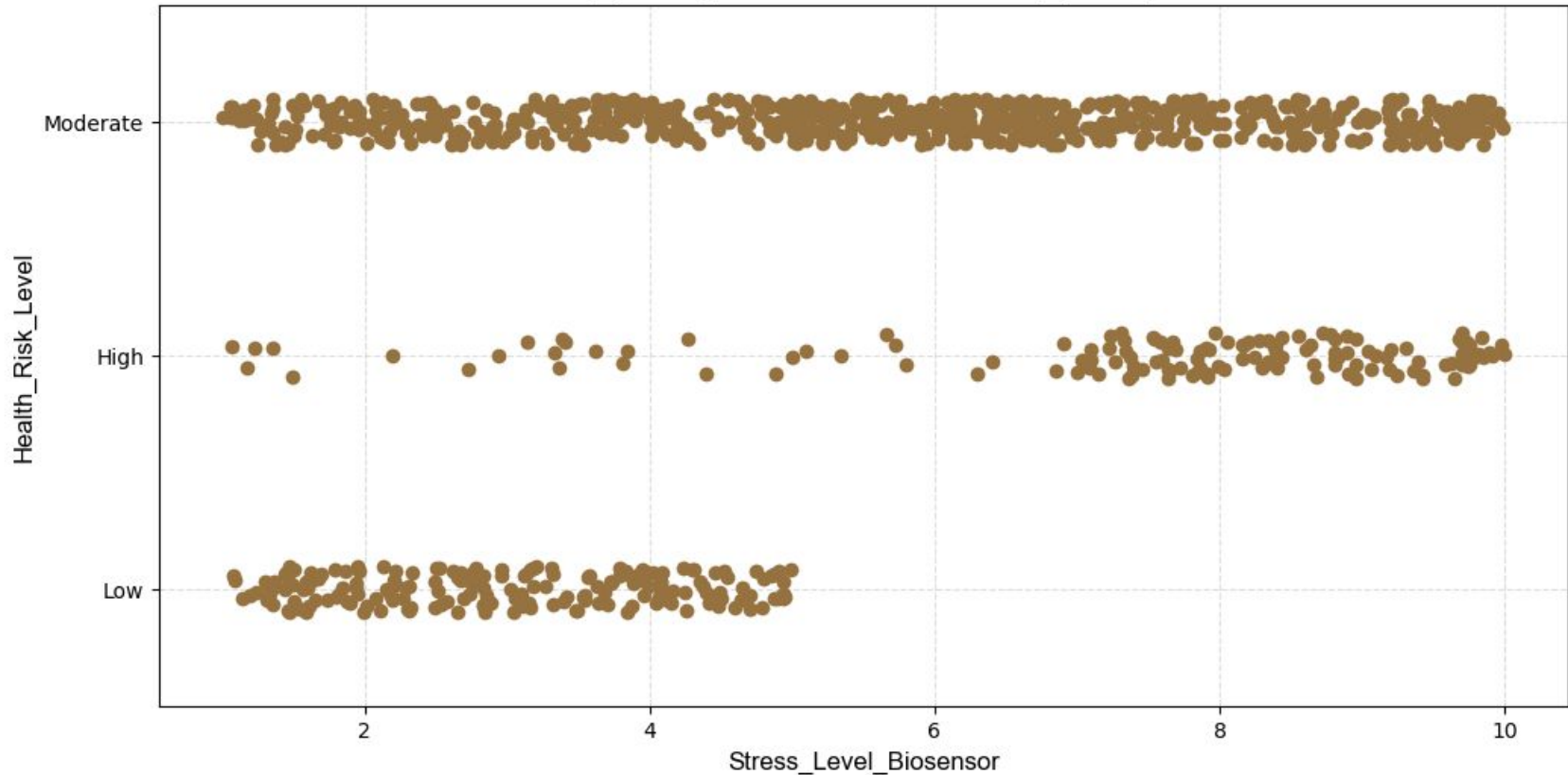


In order to create regression plots involving categorical data, values had to be mapped to a quantitative equivalent. Here, 0 represents a low health risk, while 1 represents a high risk. Moderate health risks are represented as 0.5. This is why the data is congregated along three points on the y axis.

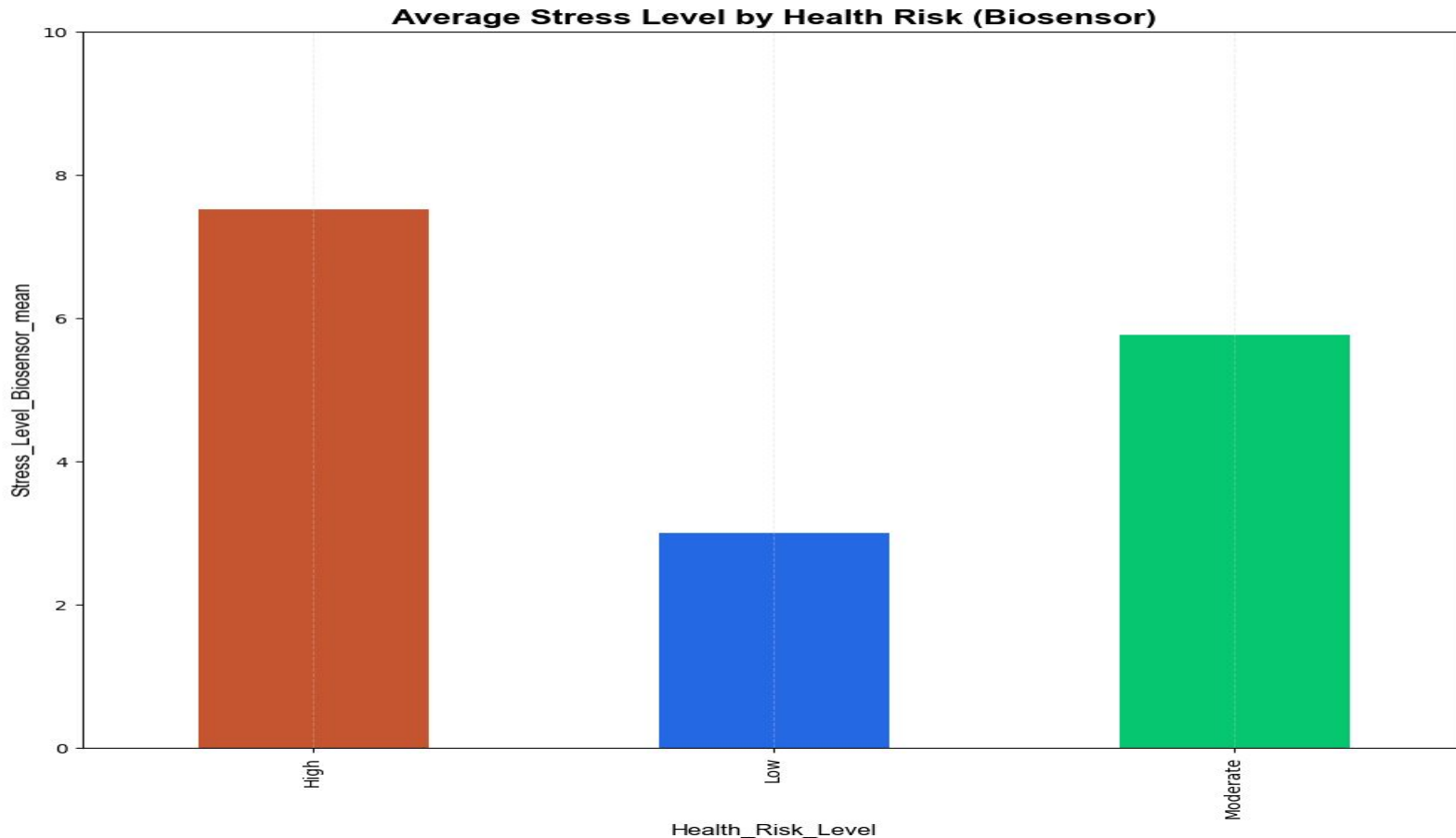


Whereas this visualization represents the range of stress levels measured in each health risk category.

Stress\_Level\_Biosensor vs Health\_Risk\_Level



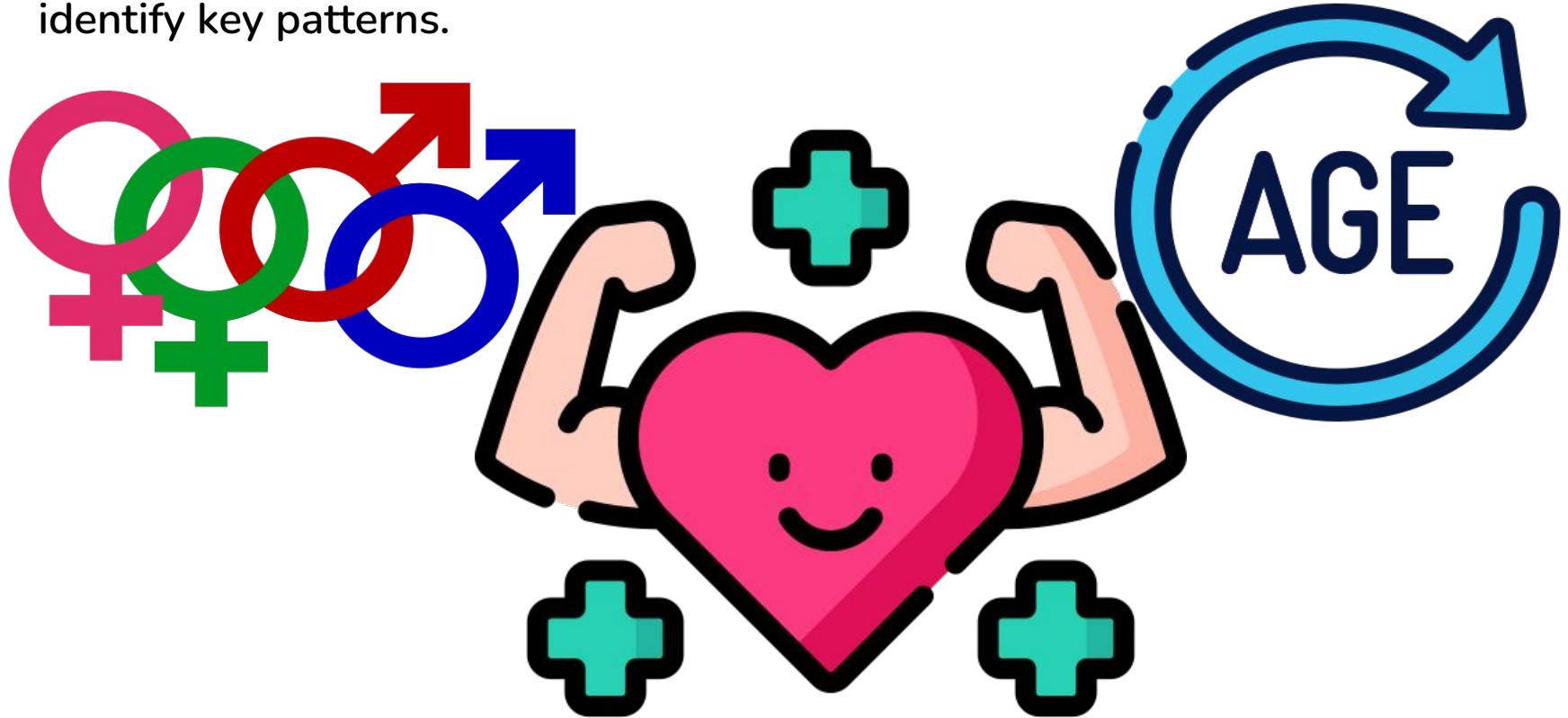
We can visualize the increase in health risk as stress level rises.



# Understanding Age, Gender, and Health Levels

**Risk**

Question 2: how age and gender relate to health risk levels and identify key patterns.





# Age vs. Health Risk Levels (Box Plot)

## Observations:

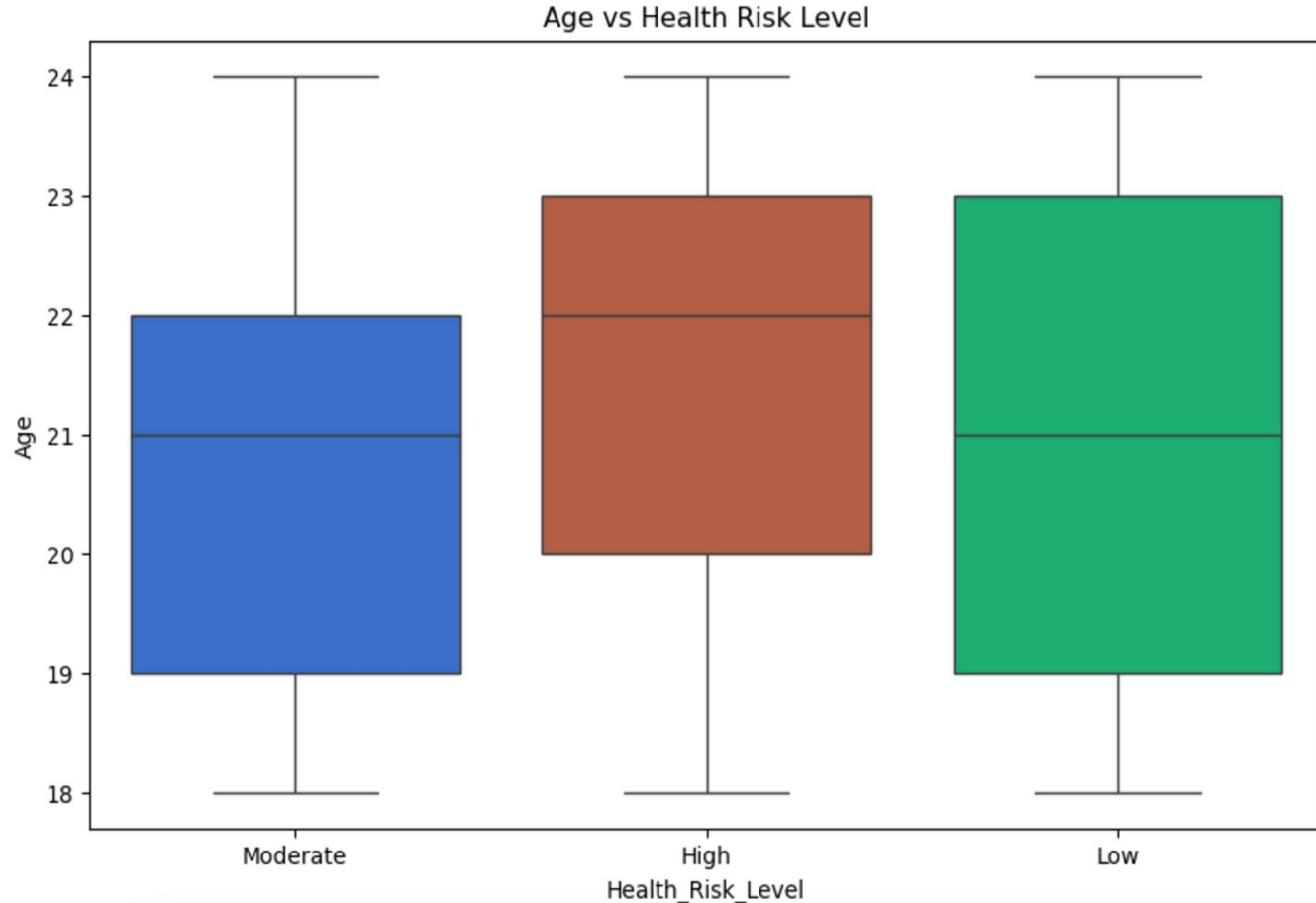
Median age is consistent at around 21 for all groups.

Age range overlaps across all categories, from 18 to about 24.

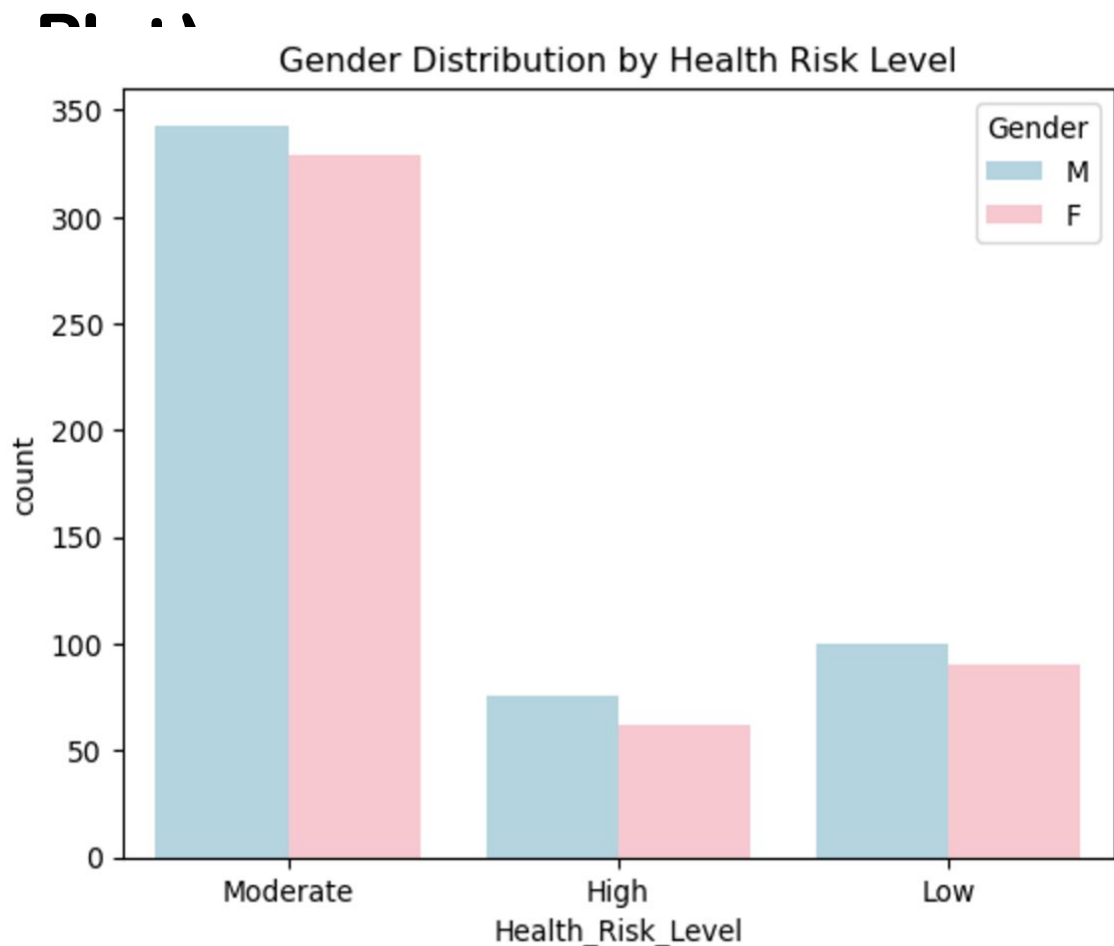
Slightly more variation in the Moderate and Low groups compared to High risk.

## Conclusion:

Age doesn't seem to play a significant role in differentiating between health risk levels. Other factors may have more influence.



# Gender Distribution by Health Risk Levels (Bar



## Observations:

The Moderate risk level has the most people, with nearly equal males and females.

High risk has the smallest group, slightly more males than females.

Low risk has fewer people than Moderate but more than High, again with a near gender balance.

## Conclusion:

Gender distribution is fairly balanced across all groups, meaning gender doesn't appear to strongly affect health risk levels.

# Combined Effect of Age and Gender (Violin plot)

## Observations:

Neither age nor gender alone (or together) strongly predicts health risk levels.

Females in the Low risk group show slightly more variation in age compared to males.

## Conclusion:

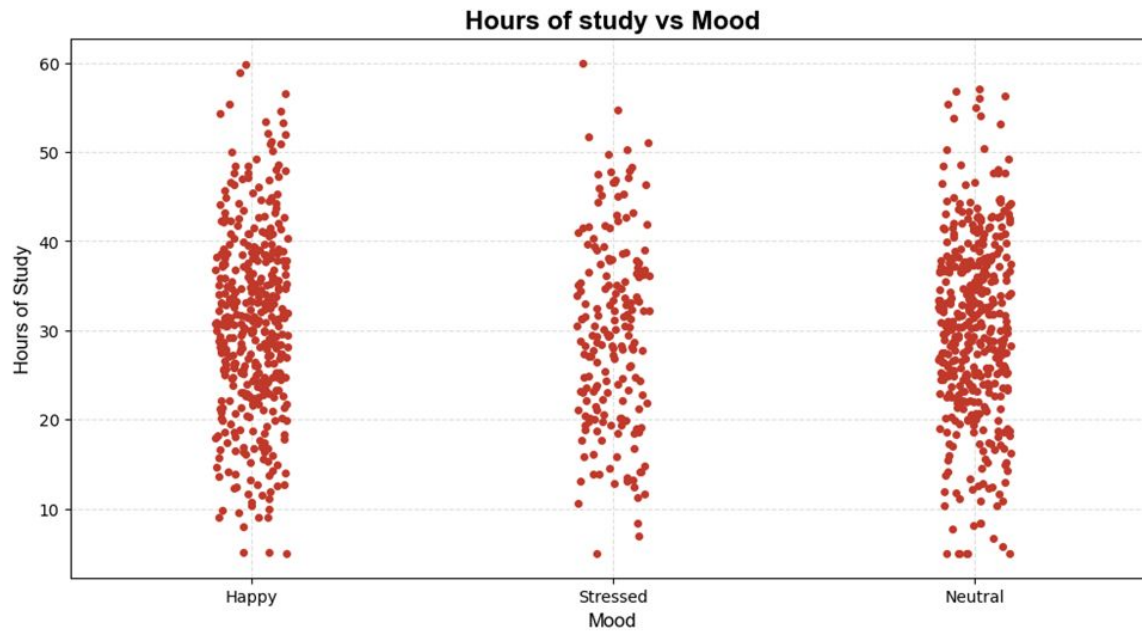
While age and gender provide useful baseline information, they don't seem to explain health risk levels on their own. Further analysis with more variables could help uncover deeper patterns and improve health risk assessments.



### Question 3:

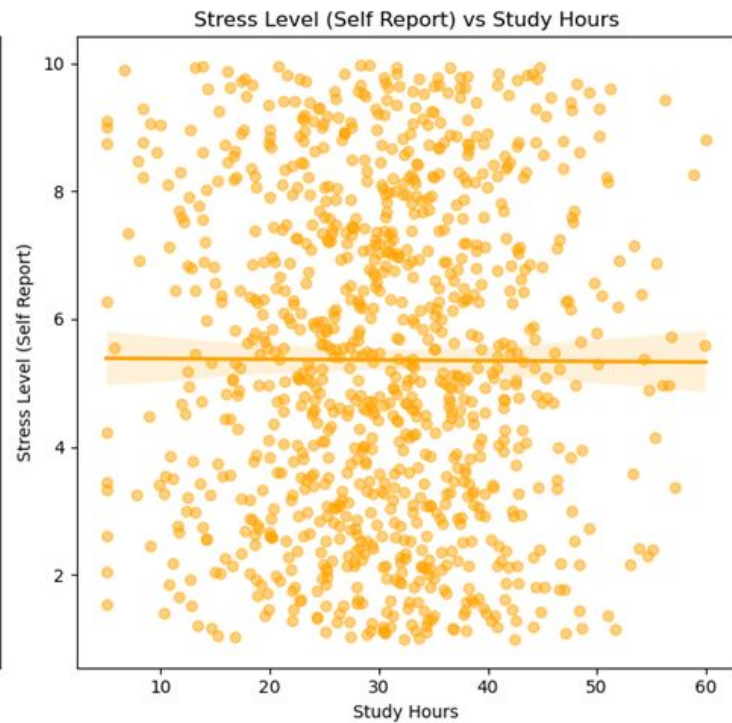
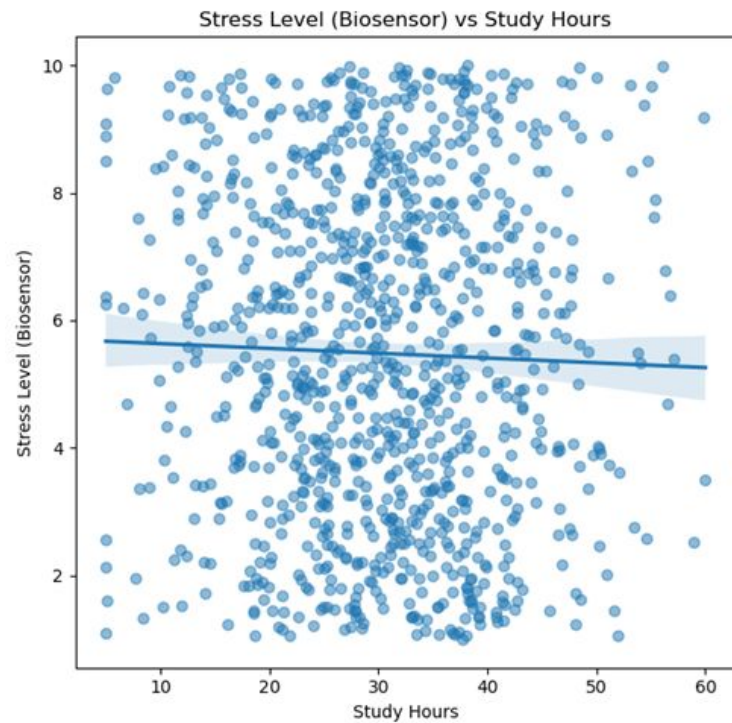
**Does the amount of time a student spend studying affect student's stress and/or their health risk?**





- **Mood and Study Hours:**
  - There is no strong or clear differentiation in study hours between the moods. Students with all moods (Happy, Stressed, Neutral) appear to study for a similar range of hours.
  - Students who study excessively (above 50 hours) seem distributed across all mood categories, suggesting other factors (like stress tolerance, work-life balance, or coping strategies) may influence mood.





1. Stress Level (Biosensor) vs. Study Hours:
  - The regression line is almost flat, indicating no strong relationship between study hours and biosensor-measured stress levels.
2. Stress Level (Self Report) vs. Study Hours:
  - Similarly, the regression line shows little to no slope, indicating a weak or nonexistent relationship between study hours and self-reported stress levels.



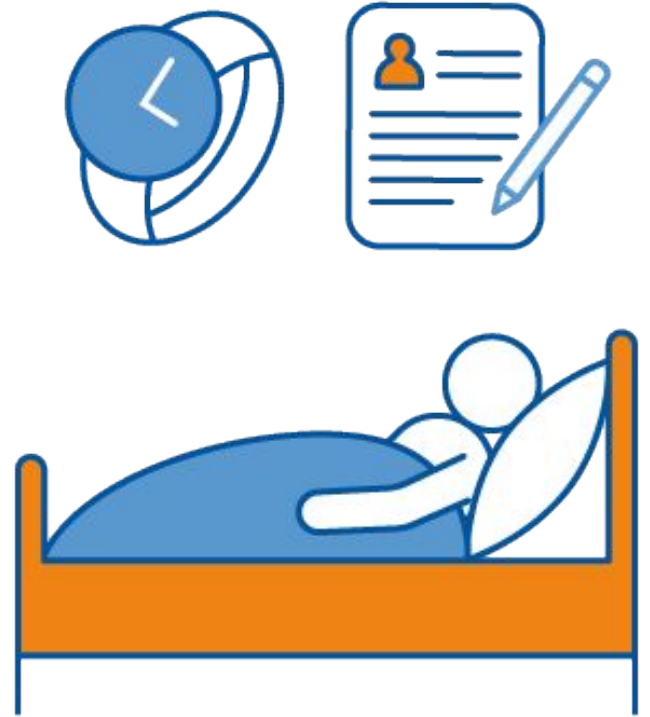


### Study Hours vs. Health Risk:

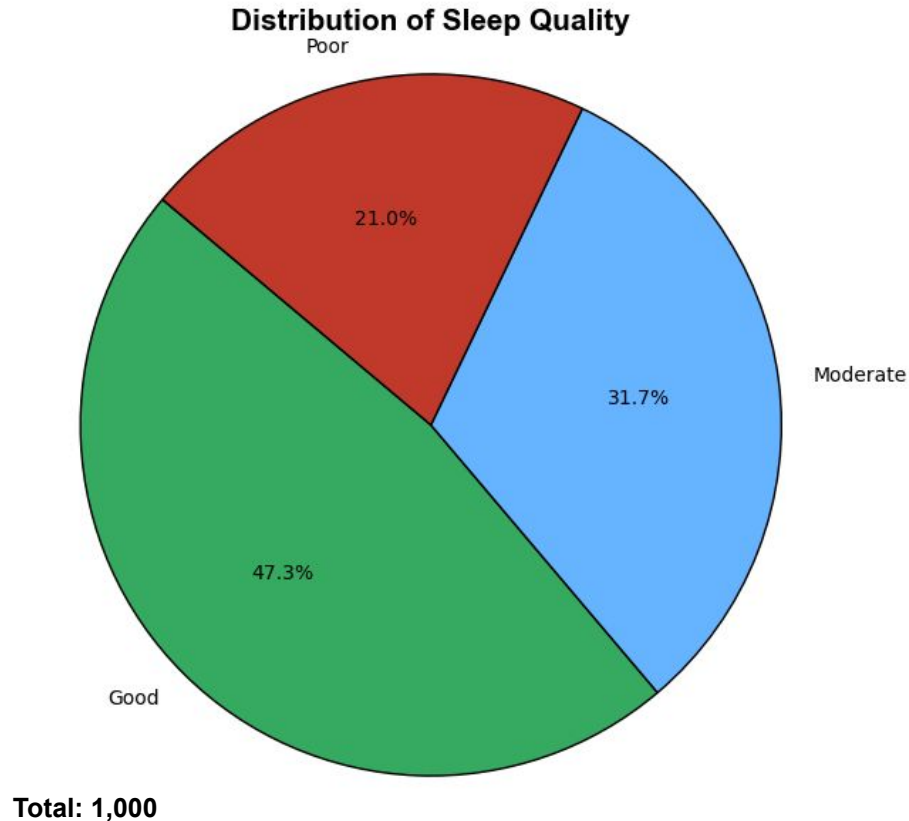
- In this strip plot, There doesn't appear to be a clear or strong relationship between hours of study and health risk levels.
- Students with moderate, high, or low health risks study for a wide range of hours, suggesting that factors other than study hours might play a larger role in determining health risk.

## Question 4:

Relationship between  
Sleep Quality, Physical  
Activity, Mood and Health  
Risk Level



# Distribution of Sleep Quality among Students

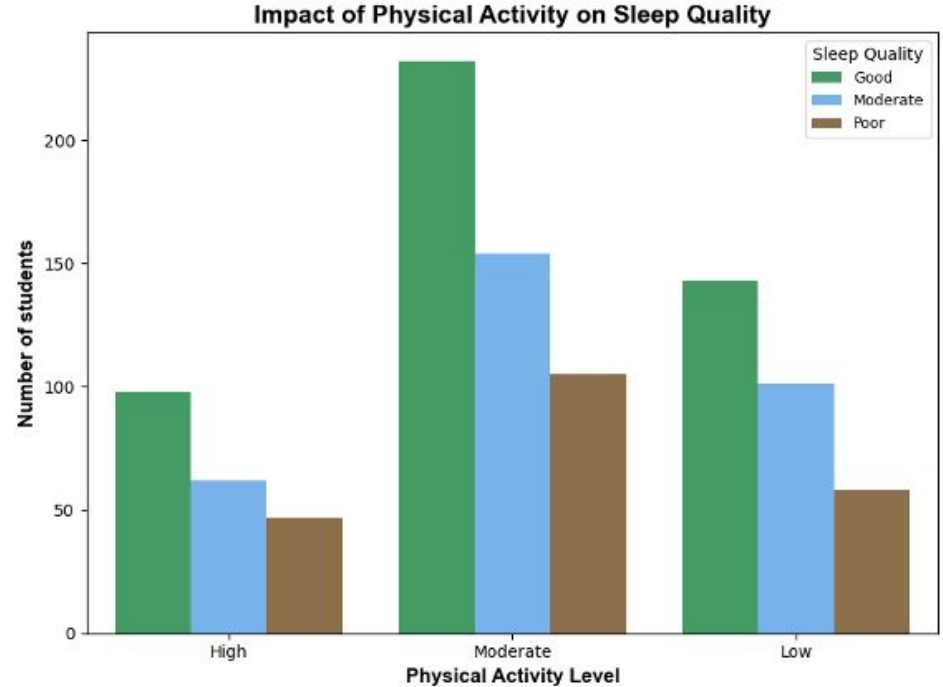


- ☐ Impact of physical activity on sleep quality
- ☐ Sleep quality on measured stress level and Health risk

# Physical Activity and Sleep Quality among Students

Physical Activity	Student Count
High	207
Moderate	491
Low	302

Sleep Quality	Student Count
Good	473
Moderate	317
Poor	210

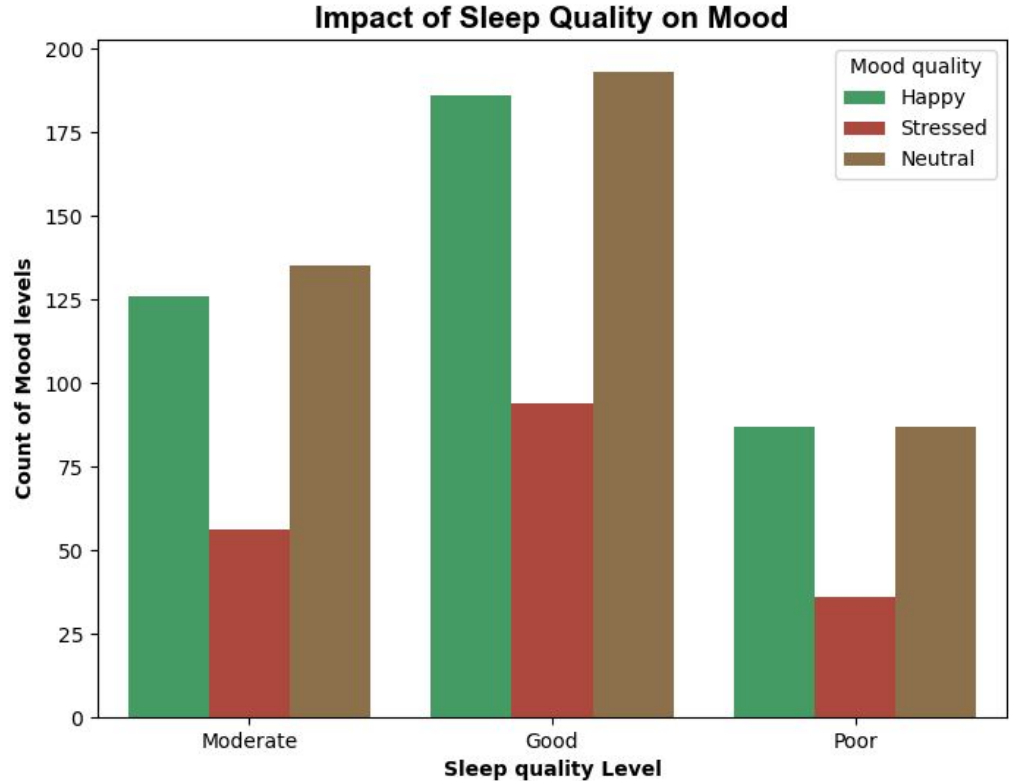


- Sleep quality is not impacted drastically by the level of physical activity

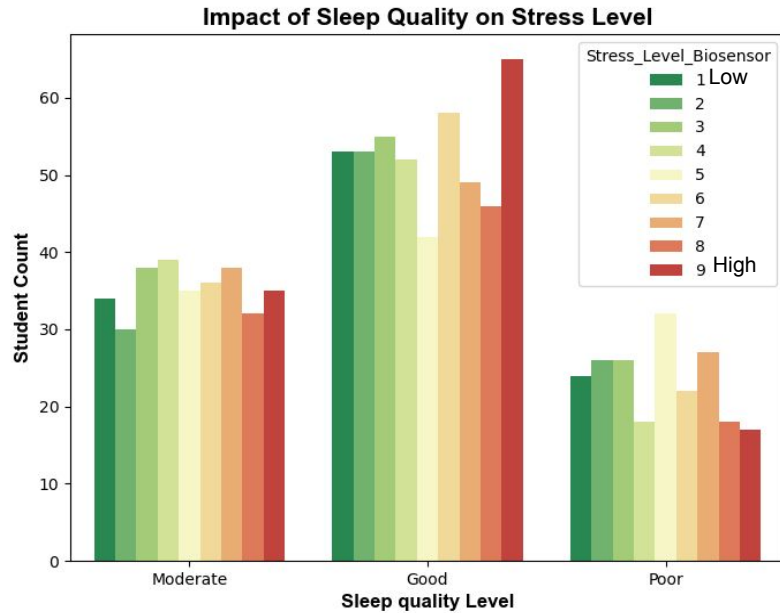
# Impact of Sleep Quality on Mood

Mood	Student Count
Happy	399
Neutral	415
Stressed	186

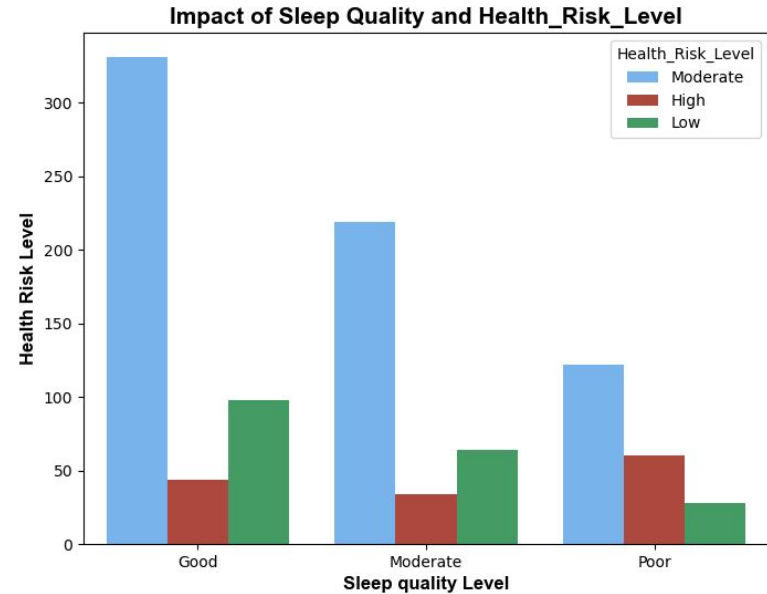
- Relatively, more students tend to have “Happy” or “Neutral” mood than “stressed” in all sleep quality group.
- Suggests that sleep quality is not impacting mood of the students



# Impact of Sleep Quality on Stress Level and Health Risk level



stress level distribution is similar among students in all sleep quality category



Health risk level is higher among students with poor sleep quality than among good or moderate sleep quality groups.



# Conclusion

- The charts show that stress does seem to directly lead to higher health risks.
- Age and gender doesn't affect health risk, but certain groups have a higher risk, especially in the moderate category.
- The amount of time students spend studying doesn't clearly affect their stress or health risks; stress levels stay about the same no matter how much they study.
- Sleep quality may impact on health risks. Students with poor sleep tend to have higher health risks, while those with good sleep quality have lower health risks

# Call to Action

The data clearly demonstrates that stress levels significantly impact health risks.

Consequently, students and their support networks should focus on developing strategies to reduce student stress.

However, the data provides limited insight into the factors that contribute to stress levels.

# Bias And Limitation

## 1. Bias:

- Only youth age group (18-24 years) is included
- Only Male or Female Gender is included

## 2. Limitation:

- Methodologies about survey whether interview or questionnaire is not clear
- Time-Frame of the data (we're assuming it's one week's data)
- Medical background of students
- No mention of geographical demography of students
- Unclear about nature of physical activity

# Future Work

- Incorporate larger sample population
- Incorporate more variables (like hobby, food, exercise, past medical history and such) in determining health risk level
- Role of domestic/school environment in influencing student mental and physical well being

# Works Cited

Ziya. *Student Health Data*. Kaggle, n.d.,  
<https://www.kaggle.com/datasets/ziya07/student-health-data/code>.  
Accessed 25 Nov. 2024.

Google.  
<https://www.google.com/search?client=firefox-b-1-d&q=how+to+change+matplotlib+color+palette>. Generative AI. Accessed 30 November, 2024.

Google.  
<https://www.google.com/search?client=firefox-b-1-d&q=seaborn+set+color+palette>. Generative AI. Accessed 30 Nov. 2024.

XPert Learning Assistant. Accessed 30 Nov. 2024.

# Q&A

