CSCI 4343 DATA SCIENCE

SEMESTER 1, 2025/2026

SECTION 2

Assignment 3:

Analysis of Student Lifestyle Patterns: Correlating Daily Habits with Stress Levels and Academic Performance

LECTURER: Asst. Prof. Dr. Raini Hassan

PREPARED BY:

NAME	MATRIC NUMBER
IMRAN HAKIM BIN MOHD NORHISHAM	2317783
AHMAD FAHEEM BIN SHARIFUDIN	2311771
MUHAMMAD AZIM BIN MUSTAFA KHALIL	2313415
MOHAMMAD HAKIM BIN MAHYUDIN	2313673
AHMAD AZZAM BIN ABDUL JAMIL	2313099

TABLE OF CONTENTS

1. INTRODUCTION

University students today face increasing pressure to balance academic responsibilities with personal health and social activities. This balancing act often leads to varying levels of stress, which can significantly impact both their mental well-being and academic performance. Understanding the relationship between daily lifestyle habits such as study hours, sleep duration, and physical activity and stress is crucial for promoting a healthier student environment. This project aims to analyze a dataset of student lifestyle patterns to uncover insights into how these daily activities correlate with stress levels and Grade Point Average (GPA), providing a descriptive overview of the factors contributing to student well-being.

A) Research Question

"How does daily lifestyle habits, specifically sleep schedule, study duration, and physical activity, vary among students with different stress levels and academic performance?"

B) Dataset Selection

- Dataset Name: Student Lifestyle and Stress Analysis Dataset
- Source: 📚 Student Stress & Performance Insights 🌟
- Description: The dataset contains 2,000 records representing university students. It includes 8 columns capturing various lifestyle and academic attributes.
- **Key Variables:**

    - **Numerical: (6 variables)**

      Study_Hours_Per_Day, Sleep_Hours_Per_Day, Physical_Activity_Hours_Per_Day, Social_Hours_Per_Day, Extracurricular_Hours_Per_Day, GPA.

    - **Categorical: (1 variable)** Stress_Level (Low, Moderate, High)

2. RESEARCH OBJECTIVES

1. To compare the average daily lifestyle habits (specifically sleep duration, study hours, and physical activity) among students with Low, Moderate, and High stress levels.
2. To analyse the relationship between Academic Performance (GPA) and Stress Levels to determine if higher academic achievement is associated with increased stress.

3. DATA PREPROCESSING

A) Feature Selection (Remove column ID)

The data set was not found to have any major issues such as outliers, missing or null values,
and duplicates. The feature selection process that we performed was to remove column ID from the original dataset as it's not needed for our research.

```
> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

> df <- df %>%
+ select(-Student_ID)
> colnames(df)
[1] "Study_Hours_Per_Day"            "Extracurricular_Hours_Per_Day"
[3] "Sleep_Hours_Per_Day"            "Social_Hours_Per_Day"
[5] "Physical_Activity_Hours_Per_Day" "GPA"
[7] "Stress_Level"
```

	Study_Hours_Per_Day	Extracurricular_Hours_…¹	Sleep_Hours_Per_Day	Social_Hours_Per_Day	Physical_Activity_Ho…²	GPA	Stress_Level
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
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

B) Renaming Columns

Then the next preprocessing is done by renaming all the columns to remove insignificant symbols such as underscore and normalize them into "Daily ..(variable)… (Hour)" and make it easier for next step of data wrangling process.

- Study_Hours_Per_Day => Daily Revision (Hour)
- Sleep_Hours_Per_Day => Daily Sleep Duration (Hour)
- Physical_Activity_Hours_Per_Day => Daily Exercise (Hour)
- Social_Hours_Per_Day => Daily Social Activity (Hour)
- Extracurricular_Hours_Per_Day => Daily Cocu Activity (Hour)
- GPA
- Stress_Level => Stress Level

```
> df <- df %>%
+ select(-Student_ID)
> colnames(df)
[1] "Study_Hours_Per_Day"            "Extracurricular_Hours_Per_Day"
[3] "Sleep_Hours_Per_Day"            "Social_Hours_Per_Day"
[5] "Physical_Activity_Hours_Per_Day" "GPA"
[7] "Stress_Level"
>
> df <- df %>%
+   rename(
+     `Daily Revision (Hour)` = Study_Hours_Per_Day,
+     `Daily Sleep Duration (Hour)` = Sleep_Hours_Per_Day,
+     `Daily Exercise (Hour)` = Physical_Activity_Hours_Per_Day,
+     `Daily Social Activity (Hour)` = Social_Hours_Per_Day,
+     `Daily Cocu Activity (Hour)` = Extracurricular_Hours_Per_Day,
+     `Stress Level` = Stress_Level
+   )
> colnames(df)
[1] "Daily Revision (Hour)"        "Daily Cocu Activity (Hour)"
[3] "Daily Sleep Duration (Hour)"  "Daily Social Activity (Hour)"
[5] "Daily Exercise (Hour)"        "GPA"
[7] "Stress Level"
```

	`Daily Revision (Hour)`	`Daily Cocu Activity (Hour)`	Daily Sleep Duration…	Daily Social Activit…	Daily Exercise (Hour…	GPA
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	6.9	3.8	8.7	2.8	1.8	2.99
2	5.3	3.5	8	4.2	3	2.75
3	5.1	3.9	9.2	1.2	4.6	2.67
4	6.5	2.1	7.2	1.7	6.5	2.88
5	8.1	0.6	6.5	2.2	6.6	3.51
6	6	2.1	8	0.3	7.6	2.85

4. DATA WRANGLING

Using the data that has been processed in the previous phase, we will manipulate the data for better analysis. The data will be separated into 5 dataframes with each is one of the factors contributing to various GPA status by stress level.

```
> #data Wrangling
> # 1. Study (Revision) vs GPA vs Stress
> StudyvsGPAbyStress <- df[c("Daily Revision (Hour)", "GPA", "Stress Level")]
> # 2. Sleep vs GPA vs Stress
> SleepvsGPAbyStress <- df[c("Daily Sleep Duration (Hour)", "GPA", "Stress Level")]
> # 3. Exercise vs GPA vs Stress
> ExercisevsGPAbyStress <- df[c("Daily Exercise (Hour)", "GPA", "Stress Level")]
> # 4. Social Activity vs GPA vs Stress
> SocialvsGPAbyStress <- df[c("Daily Social Activity (Hour)", "GPA", "Stress Level")]
> # 5. Co-curriculum (Cocu) vs GPA vs Stress
> CocuvsGPAbyStress <- df[c("Daily Cocu Activity (Hour)", "GPA", "Stress Level")]
>
```

```
> head(StudyvsGPAbyStress)
# A tibble: 6 × 3
  `Daily Revision (Hour)`   GPA `Stress Level`
                    <dbl> <dbl> <chr>
1                     6.9  2.99 Moderate
2                     5.3  2.75 Low
3                     5.1  2.67 Low
4                     6.5  2.88 Moderate
5                     8.1  3.51 High
6                     6    2.85 Moderate
> head(SleepvsGPAbyStress)
# A tibble: 6 × 3
  `Daily Sleep Duration (Hour)`   GPA `Stress Level`
                          <dbl> <dbl> <chr>
1                           8.7  2.99 Moderate
2                           8    2.75 Low
3                           9.2  2.67 Low
4                           7.2  2.88 Moderate
5                           6.5  3.51 High
6                           8    2.85 Moderate
> head(ExercisevsGPAbyStress)
# A tibble: 6 × 3
  `Daily Exercise (Hour)`   GPA `Stress Level`
                    <dbl> <dbl> <chr>
1                     1.8  2.99 Moderate
2                     3    2.75 Low
3                     4.6  2.67 Low
4                     6.5  2.88 Moderate
5                     6.6  3.51 High
6                     7.6  2.85 Moderate
> head(SocialvsGPAbyStress)
# A tibble: 6 × 3
  `Daily Social Activity (Hour)`   GPA `Stress Level`
                          <dbl> <dbl> <chr>
1                           2.8  2.99 Moderate
2                           4.2  2.75 Low
3                           1.2  2.67 Low
4                           1.7  2.88 Moderate
5                           2.2  3.51 High
6                           0.3  2.85 Moderate
> head(CocuvsGPAbyStress)
# A tibble: 6 × 3
  `Daily Cocu Activity (Hour)`   GPA `Stress Level`
                        <dbl> <dbl> <chr>
1                         3.8  2.99 Moderate
2                         3.5  2.75 Low
3                         3.9  2.67 Low
4                         2.1  2.88 Moderate
5                         0.6  3.51 High
6                         2.1  2.85 Moderate
>
```

A) Daily Revision (Hour) vs GPA by Stress Levels

From the table, we can group the students into categories to see the pattern easier. By that, we create a new column called “Study Intensity”. Students that have below 6 hours of daily revision will be classed as “Low Intensity”. Students that have a range from 6 until 8 study hours of daily revision will be classed as “Medium Intensity”. Students that have more than 8 hours of daily revision will be classed as “High Intensity”.

```
> library(dplyr)
> StudyIntensity <- StudyvsGPAbystress %>%
+   mutate(`Study Intensity` = case_when(
+     `Daily Revision (Hour)` < 6 ~ "Low Intensity",
+       `Daily Revision (Hour)` >= 6 & `Daily Revision (Hour)` <= 8 ~
"Medium Intensity",
+     `Daily Revision (Hour)` > 8 ~ "High Intensity"
+   ))
> head(StudyIntensity)
```

	Daily Revision (Hour)	GPA	Stress Level	Study Intensity
1	6.9	2.99	Moderate	Medium Intensity
2	5.3	2.75	Low	Low Intensity
3	5.1	2.67	Low	Low Intensity
4	6.5	2.88	Moderate	Medium Intensity
5	8.1	3.51	High	High Intensity
6	6.0	2.85	Moderate	Medium Intensity

Then, the number of students and the average GPA will be grouped by ‘Study Intensity’ and ‘Stress level’ to compare performance among different categories of ‘Study Intensity’ and ‘Stress level’.

```
> Intensity_Analysis <- StudyIntensity %>%
+   group_by(`Study Intensity`, `Stress Level`) %>%
+   summarise(
+     Average_GPA = mean(GPA),
+     Student_Count = n()
+   ) %>%
+   arrange(desc(Average_GPA))
> print(Intensity_Analysis)
```

```
# A tibble: 5 × 4
# Groups:   Study Intensity [3]
  `Study Intensity` `Stress Level` Average_GPA Student_Count
  <chr>             <chr>                <dbl>         <int>
1 High Intensity    High                  3.36           766
2 Medium Intensity  High                  3.05           197
3 Medium Intensity  Moderate              3.02           674
4 Low Intensity     Low                   2.82           297
5 Low Intensity     High                  2.80            66
```

B) Daily Sleep Duration (Hour) vs GPA by Stress Levels

From the original table, we can group the students into multiple categories to assess the pattern. By performing aggregation on the dataset into more general classes, we can identify the sleep habits of students and their correlation with stress level and GPA. This new table classifies students who have below or above 8 hours sleep recommendation, including their stress level and GPA status, and a new column of the total of students for each class. However, the percentage shown in this table is only exclusive to the same stress level and sleep status.

```
> library(dplyr)
> SleepvsGPAbystress <- df %>%
 + mutate(
 +   Stress_Level = factor(Stress_Level, levels = c("High", "Moderate",
"Low"))
 + ),
+ Daily_Sleep_Duration(Hour) = ifelse(`Daily Sleep Duration (Hour)` <
8, "< 8 hours", ">= 8 hours") ,
#GPA Status Classification :
  Excellent >= 3.5, Good = <3.5 && >= 3.0,  Average <3.0
 + GPA_Status = case_when(
 + GPA >= 3.5 ~ "Excellent",
 + GPA >= 3.0 & GPA < 3.5 ~ "Good",
 + GPA < 3.0 ~ "Average"
 +  ) %>%
 + group_by(Daily_Sleep_Duration(Hour), Stress_Level) %>%
 + summarise(
 + Total_Student = n(),
 +   Excellent = paste0(round(sum(GPA_Status == "Excellent") / n() *
100, 2), "%"),
```

```
 +  Good = paste0(round(sum(GPA_Status == "Good") / n() * 100, 2),
"%"),
 +   Average  =  paste0(round(sum(GPA_Status== "Average") / n() *
100, 2), "%"),
 + .groups = "drop"
 + )
>print(SleepvsGPAbystress)
```

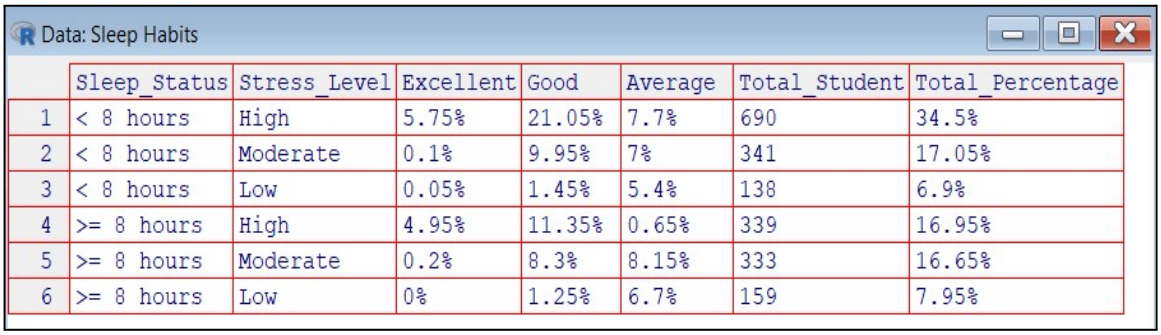
	Stress Level	Sleep_Status	GPA_Status	Total_Students	Percentage
1	High	Sleep < 8 hours	Good	421	61.01%
2	High	Sleep < 8 hours	Average	154	22.32%
3	High	Sleep < 8 hours	Excellent	115	16.67%
4	High	Sleep >= 8 hours	Good	227	66.96%
5	High	Sleep >= 8 hours	Excellent	99	29.2%
6	High	Sleep >= 8 hours	Average	13	3.83%
7	Low	Sleep < 8 hours	Average	108	78.26%
8	Low	Sleep < 8 hours	Good	29	21.01%
9	Low	Sleep < 8 hours	Excellent	1	0.72%
10	Low	Sleep >= 8 hours	Average	134	84.28%
11	Low	Sleep >= 8 hours	Good	25	15.72%
12	Moderate	Sleep < 8 hours	Good	199	58.36%
13	Moderate	Sleep < 8 hours	Average	140	41.06%
14	Moderate	Sleep < 8 hours	Excellent	2	0.59%
15	Moderate	Sleep >= 8 hours	Good	166	49.85%
16	Moderate	Sleep >= 8 hours	Average	163	48.95%
17	Moderate	Sleep >= 8 hours	Excellent	4	1.2%

Thus, further wrangling needs to be performed by creating extra columns to represent GPA status. The GPA status is now divided into three separate columns.The percentage of students in each class is modified to represent the percentage from the total of 2000 students in a new dataframe called Sleep Habits. By performing this process, the data is more feasible to get a deeper insight of the association between sleep habits and GPA status by stress level for the next step of data analysis.

```
#To find total students = 2000 data
> totalStudents<- nrow(df)
>
> `Sleep Habits` <- df %>%
+   mutate(
+     Stress_Level = factor(Stress_Level, levels = c("High", "Moderate",
"Low")),
+     Sleep_Status = ifelse(Sleep_Hours_Per_Day < 8, "< 8 hours",
">= 8 hours"),
+     GPA_Status = case_when(
+       GPA >= 3.5 ~ "Excellent",
+       GPA >= 3.0 & GPA < 3.5 ~ "Good",
+       GPA < 3.0 ~ "Average"
+     )
+   ) %>%
+   group_by(Sleep_Status, Stress_Level) %>%
+ #Separating the GPA into 3 columns to represent the grade in
percentage from 2000  students
+   summarise(
+     Exc_Val = sum(GPA_Status == "Excellent") / totalStudents * 100,
+     Good_Val = sum(GPA_Status == "Good") / totalStudents * 100,
+     Ave_Val = sum(GPA_Status == "Average") / totalStudents * 100,
+
```

#Find the total students and their percentage according to each GPA status

```
+     Total_Student = n(),
+     Excellent = paste0(round(Exc_Val, 2), "%"),
+     Good = paste0(round(Good_Val, 2), "%"),
+     Average = paste0(round(Ave_Val, 2), "%"),
+     Total_Percentage = paste0(round(Exc_Val + Good_Val +
Ave_Val, 2), "%"),
+     .groups = "drop"
+   ) %>%
+select(Sleep_Status, Stress_Level, Excellent, Good, Average,
Total_Student, Total_Percentage) %>%
+   arrange(Sleep_Status, Stress_Level)
> View(`Sleep Habits`)
```



Data: Sleep Habits

	Sleep_Status	Stress_Level	Excellent	Good	Average	Total_Student	Total_Percentage
1	< 8 hours	High	5.75%	21.05%	7.7%	690	34.5%
2	< 8 hours	Moderate	0.1%	9.95%	7%	341	17.05%
3	< 8 hours	Low	0.05%	1.45%	5.4%	138	6.9%
4	>= 8 hours	High	4.95%	11.35%	0.65%	339	16.95%
5	>= 8 hours	Moderate	0.2%	8.3%	8.15%	333	16.65%
6	>= 8 hours	Low	0%	1.25%	6.7%	159	7.95%

C) Social (Hour) vs GPA by Stress Levels

From the table, we can identify the social level or social category of the students through grouping them into different categories. By that, we can clearly see the correlation between Social Hours Per Day with their GPA performances, whether it really does affect or not. Thus, we created a new column that categorizes student social level into three categories: Low Social (Social Hours < 2), Moderate Social (2 < Social Hours < 4), and High Social (Social Hours > 4).

```
> SocialvsGPA <- SocialvsGPAbystress %>%
+   mutate(Social_Category = case_when(
+     ‘Daily Social Activity (Hour)’ < 2 ~ "Low Social",
+     ‘Daily Social Activity (Hour)’ >= 2 & Daily Social Activity (Hour) <=
4 ~ "Moderate Social",
+     ‘Daily Social Activity (Hour)’ > 4 ~ "High Social"))
> head(SocialvsGPA)
```

	Daily Social Activity (Hour)	GPA	Stress Level	Social_Category
1	2.8	2.99	Moderate	Moderate Social
2	4.2	2.75	Low	High Social
3	1.2	2.67	Low	Low Social
4	1.7	2.88	Moderate	Low Social
5	2.2	3.51	High	Moderate Social
6	0.3	2.85	Moderate	Low Social

Further wrangling needs to be performed by grouping students based on their social category and also their stress level . With that , a new column was also created which are Average_GPA and Jumlah_Student to see how many students with different levels of social and stress level perform in their GPA by averaging all of the total students in that group. Thus, we are able to gain much more clear insight , which group , and how many students achieve the highest average GPA compared to the others group.

```
> Grouped_Social <- SocialvsGPA %>%
+   group_by(Social_Category, Stress_Level) %>%
+   summarise(
+     Average_GPA = mean(GPA, na.rm = TRUE),
+     Jumlah_Student = n(),
+     .groups = "drop"
+   ) %>%
+   arrange(
+     match(Social_Category, c("Low Social", "Moderate Social", "High
Social")),
+     Stress_Level
+   )
> print(Grouped_Social)
```

```
# A tibble: 9 × 4
  Social_Category Stress_Level Average_GPA Jumlah_Student
  <chr>           <chr>               <dbl>          <int>
1 Low Social      High                 3.29            415
2 Low Social      Low                  2.79            106
3 Low Social      Moderate             3.03            256
4 Moderate Social High                 3.27            362
5 Moderate Social Low                  2.80            101
6 Moderate Social Moderate             3.04            238
7 High Social     High                 3.20            252
8 High Social     Low                  2.86             90
9 High Social     Moderate             3.00            180
```

D) Daily Exercise (Hour) vs GPA by Stress Levels

To analyze the impact of physical health on academic performance, the continuous variable "Daily Exercise (Hour)" was first segmented into distinct categories. This was done to identify non-linear patterns that raw hours might obscure. Students were classified into three groups: "Sedentary" (less than 3 hour of daily exercise), "Active" (3 to 6 hours), and "Highly Active" (more than 6 hours)

```
> library(dplyr)
> ActivityLevel <- ExercisevsGPAbystress %>%
+ mutate(`Activity Level` = case_when(
+ `Daily Exercise (Hour)` < 3 ~ "Sedentary",
+ `Daily Exercise (Hour)` >= 3 & `Daily Exercise (Hour)` <= 6 ~ "Active",
+ `Daily Exercise (Hour)` >6 ~ "Highly Active"
+ ))
> head(ActivityLevel)
```

	Daily Exercise (Hour)	GPA	Stress Level	Activity Level
1	1.8	2.99	Moderate	Sedentary
2	3.0	2.75	Low	Active
3	4.6	2.67	Low	Active
4	6.5	2.88	Moderate	Highly Active
5	6.6	3.51	High	Highly Active
6	7.6	2.85	Moderate	Highly Active

Following categorization, the data was grouped by both 'Activity Level' and 'Stress Level'. For each subgroup, the average GPA was calculated to serve as the primary performance metric, while the student count was recorded to ensure sample size validity. This aggregation allows for a direct comparison of how different combinations of physical activity and stress influence academic results.

```
> Exercise_Analysis <- ActivityLevel %>%
+ group_by(`Activity Level`, `Stress Level`) %>%
+ summarise(
+ Average_GPA = mean(GPA),
+ Student_Count = n(),
+ .groups = 'drop'
+ ) %>%
```

```
+ arrange(desc(Average_GPA))
> print(Exercise_Analysis)
```

```
# A tibble: 9 × 4
  `Activity Level` `Stress Level` Average_GPA Student_Count
  <chr>            <chr>                <dbl>         <int>
1 Sedentary        High                  3.37           418
2 Active           High                  3.27           388
3 Highly Active    High                  3.05           223
4 Active           Moderate              3.03           307
5 Highly Active    Moderate              3.02           162
6 Sedentary        Moderate              3.02           205
7 Sedentary        Low                   2.86            41
8 Highly Active    Low                   2.81           127
9 Active           Low                   2.81           129
```

E) Daily Co-curricular Activity vs GPA by Stress Levels

To investigate the relationship between extracurricular involvement and academic success, the variable "Daily Cocu Activity (Hour)" was segmented into three distinct intensity levels. This classification was necessary to handle the variance in student schedules. Students involved for less than 1 hour daily were labeled "Low Intensity", those dedicating between 1 and 3 hours were labeled "Medium Intensity", and those exceeding 3 hours were classified as "High Intensity"

- **Low Intensity:** < 1 hour (Students with minimal involvement).
- **Medium Intensity:** 1 to 3 hours (Students with moderate/balanced involvement).
- **High Intensity**: > 3 hours (Students with heavy involvement).

```
> # 2. Create 'Cocu Intensity' categories
> CocuIntensity <- CocuvsGPAbyStress %>%
+  mutate(`Cocu Intensity` = case_when(
+    `Daily Cocu Activity (Hour)` < 1 ~ "Low Intensity",
+    `Daily Cocu Activity (Hour)` >= 1 & `Daily Cocu Activity (Hour)` <=
3 ~ "Medium Intensity",
+    `Daily Cocu Activity (Hour)` > 3 ~ "High Intensity"
+  ))
```

`Daily Cocu Activity (Hour)`	GPA	`Stress Level`	`Cocu Intensity`
<dbl>	<dbl>	<chr>	<fct>
3.8	2.99	Moderate	High Intensity
3.5	2.75	Low	High Intensity
3.9	2.67	Low	High Intensity
2.1	2.88	Moderate	Medium Intensity
0.6	3.51	High	Low Intensity
2.1	2.85	Moderate	Medium Intensity

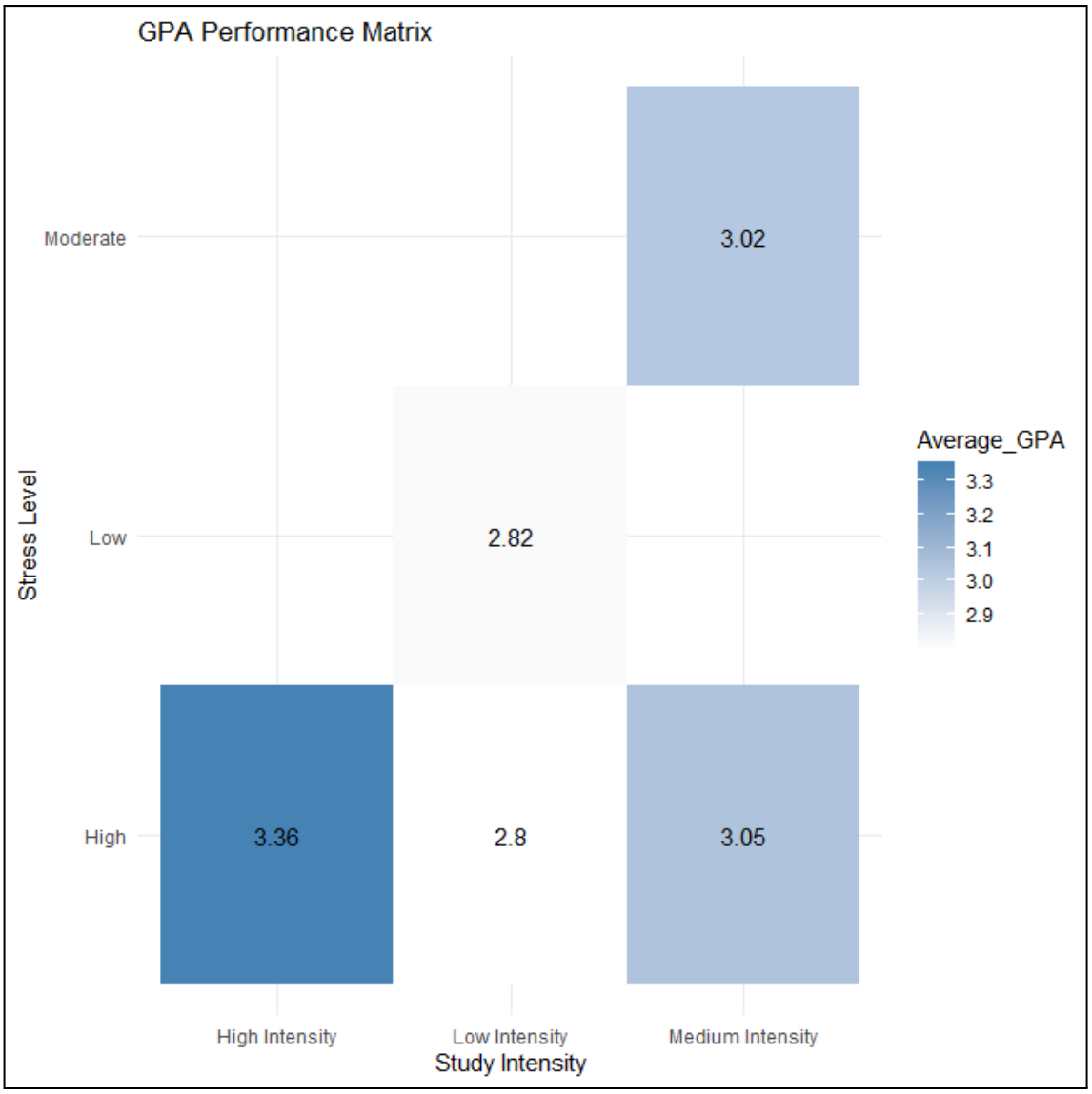
The dataset was then grouped by both 'Cocu Intensity' and 'Stress Level'. For each of these subgroups, the Average GPA was calculated to assess performance, and the student counts was recorded to ensure that the sample size for each category was sufficient for valid comparison.

```
> # Group by Intensity and Stress Level to find Average GPA
> Cocu_Analysis <- CocuIntensity %>%
+  group_by(`Cocu Intensity`, `Stress Level`) %>%
+  summarise(
+      Average_GPA = mean(GPA, na.rm = TRUE), # na.rm = TRUE handles missing values safely
+    Student_Count = n(),
+    .groups = 'drop'
+  ) %>%
+  arrange(desc(Average_GPA))
> print(Cocu_Analysis)
```

	Cocu Intensity	Stress Level	Average_GPA	Student_Count
1	Low Intensity	High	3.272581	248
2	Medium Intensity	High	3.261599	538
3	High Intensity	High	3.251975	243
4	Low Intensity	Moderate	3.042566	152
5	Medium Intensity	Moderate	3.025553	371
6	High Intensity	Moderate	3.005232	151
7	High Intensity	Low	2.825270	74
8	Low Intensity	Low	2.820000	75
9	Medium Intensity	Low	2.811081	148

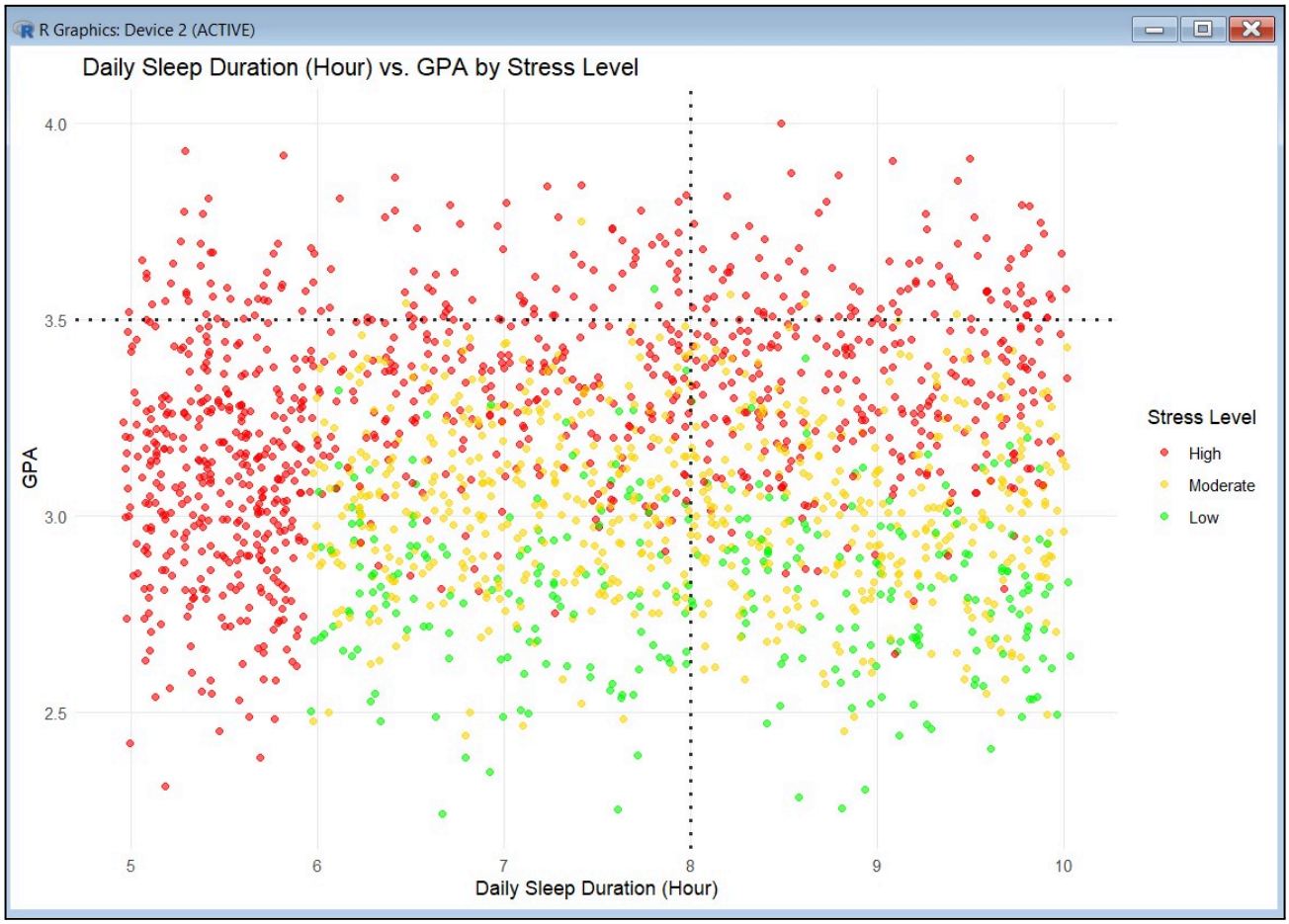
5. DATA ANALYSIS

A) Daily Revision (Hour) vs GPA by Stress Levels



After the data has been processed and manipulated from the table, we manage to plot the data into a graph to be analyzed. Heatmap graph has been used for this particular table for data analysis. According to the figure below, the higher the study intensity of a student, the higher GPA a student will achieve on average. Alongside the strong correlation, stress level of a student will also increase with the students' study intensity.

B) Daily Sleep Duration (Hour) vs GPA by Stress Levels



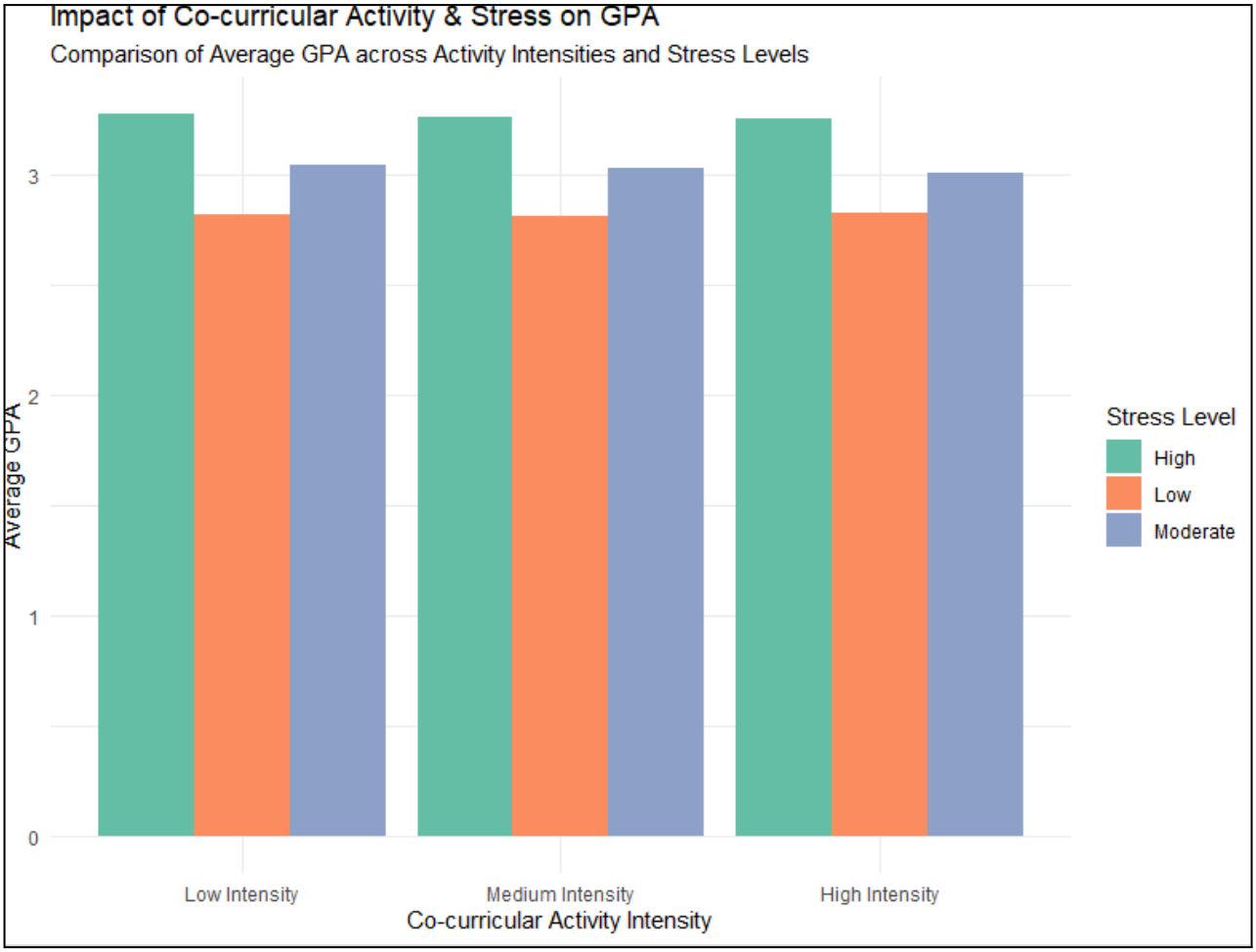
After the data has been processed and manipulated from the table, we manage to plot the data into a graph to be analyzed. Based on the scatter plot Daily Sleep Duration (Hour) vs GPA by Stress above, we could get a few insights on how sleep habits affected both GPA and stress level of 2000 students There are two main intercept lines at X-intercept (Daily Sleep Duration(Hour)) and Y-intercept (GPA) to provide a clear insight on students performance based on their sleep habits supported by color labeling of the stress level intensity.

The graph shows that the distribution for most of the excellent students with above 3.5 GPA are having quality sleep habits on a daily basis (above 8 hours). However, most of them also have the highest level of intensity of stress and a few on moderate levels, and no lowest level of stress is recorded . On the other hand, the excellent GPA obtained by a low stress level student is recorded only one student with below daily 8 hours sleep. From both insights, we could conclude that there is a correlation between sleep habits and GPA by stress level intensity.

To support the claim, there are more students with low stress levels for above 8 hours sleep, but traded with GPA on the lower end. The data suggests that students' GPA is driven more by the stress levels than by the sleep duration, but the sleep duration is the key factor in controlling stress.The trends from the graph also suggests that if the students want to achieve an excellent grade, they need more than 8 hours of sleep and are more prone to get stress with the highest intensity level.

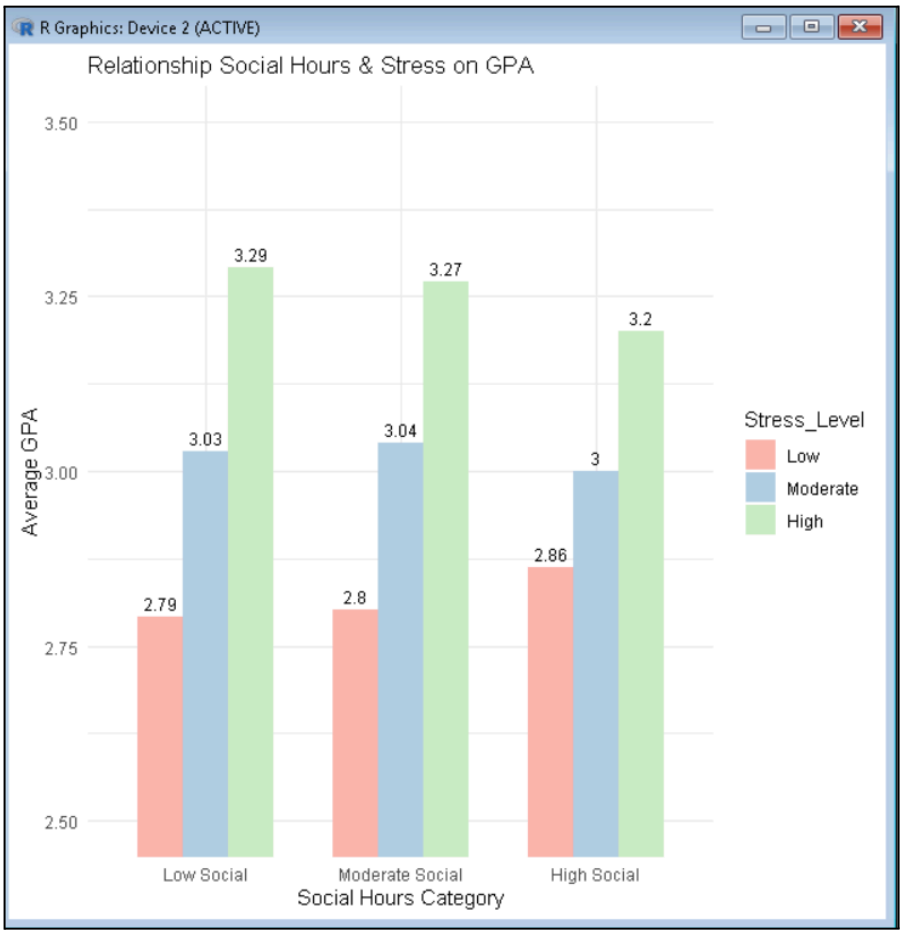
C) Daily Co-curricular Activity vs GPA by Stress Levels

We use a grouped bar chart for this visualization because it allows us to see if students with "High Intensity" co-curricular activities maintain good GPAs even under high stress.



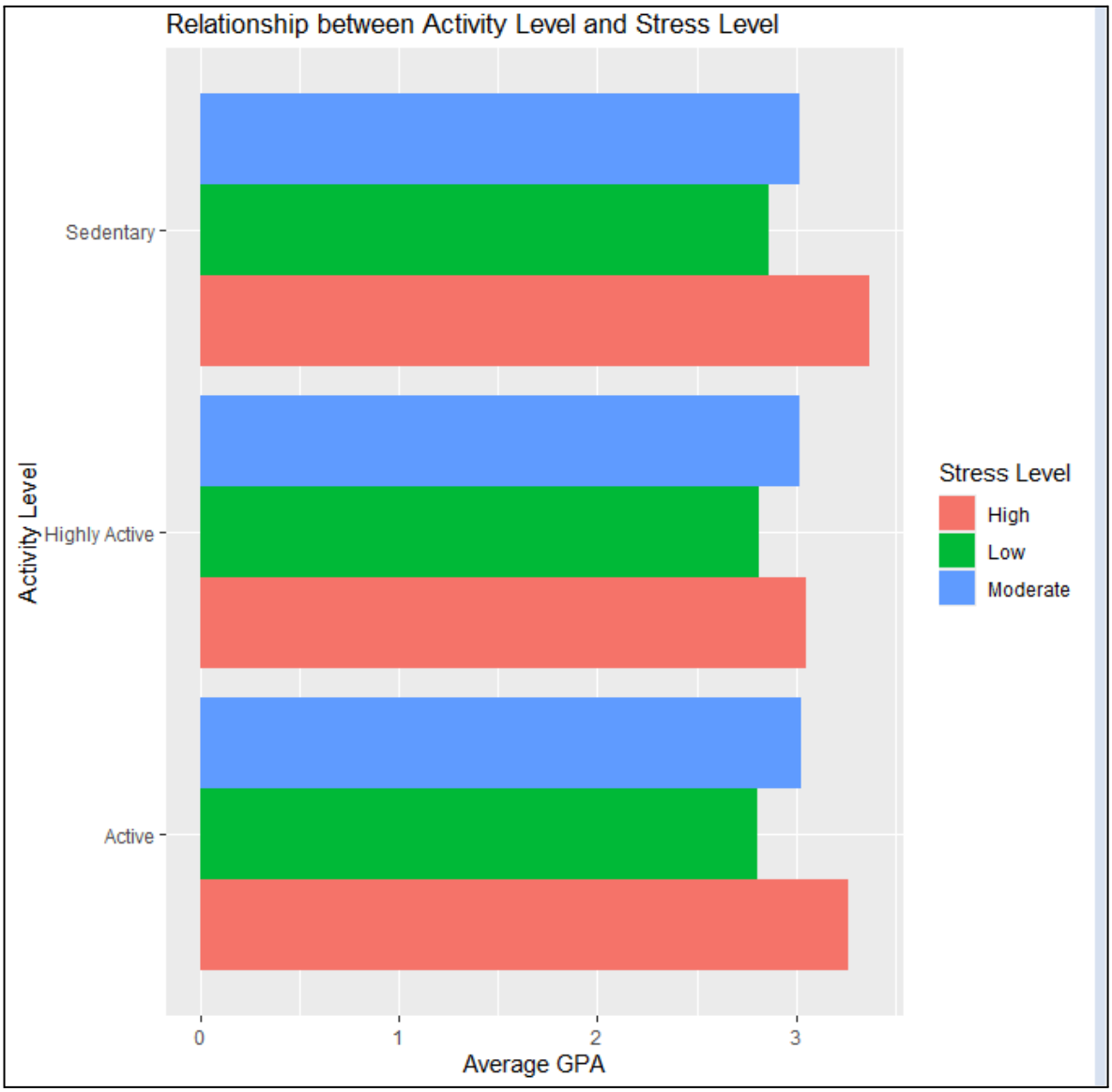
A horizontal comparison of the activity groups reveals a remarkable stability in Average GPA, as the performance metrics for High, Moderate, and Low stress categories remain nearly identical across Low, Medium, and High activity intensities. This consistency indicates that the volume of co-curricular engagement does not significantly negatively impact academic performance, as students with high participation levels maintain grades comparable to their less active peers. Consequently, these findings effectively reject the assumption that dedicating time to co-curricular activities inevitably compromises academic scores, suggesting that students are able to successfully balance heavy extra-curricular commitments without affecting their GPA.

D) Social (Hour) vs GPA by Stress Levels



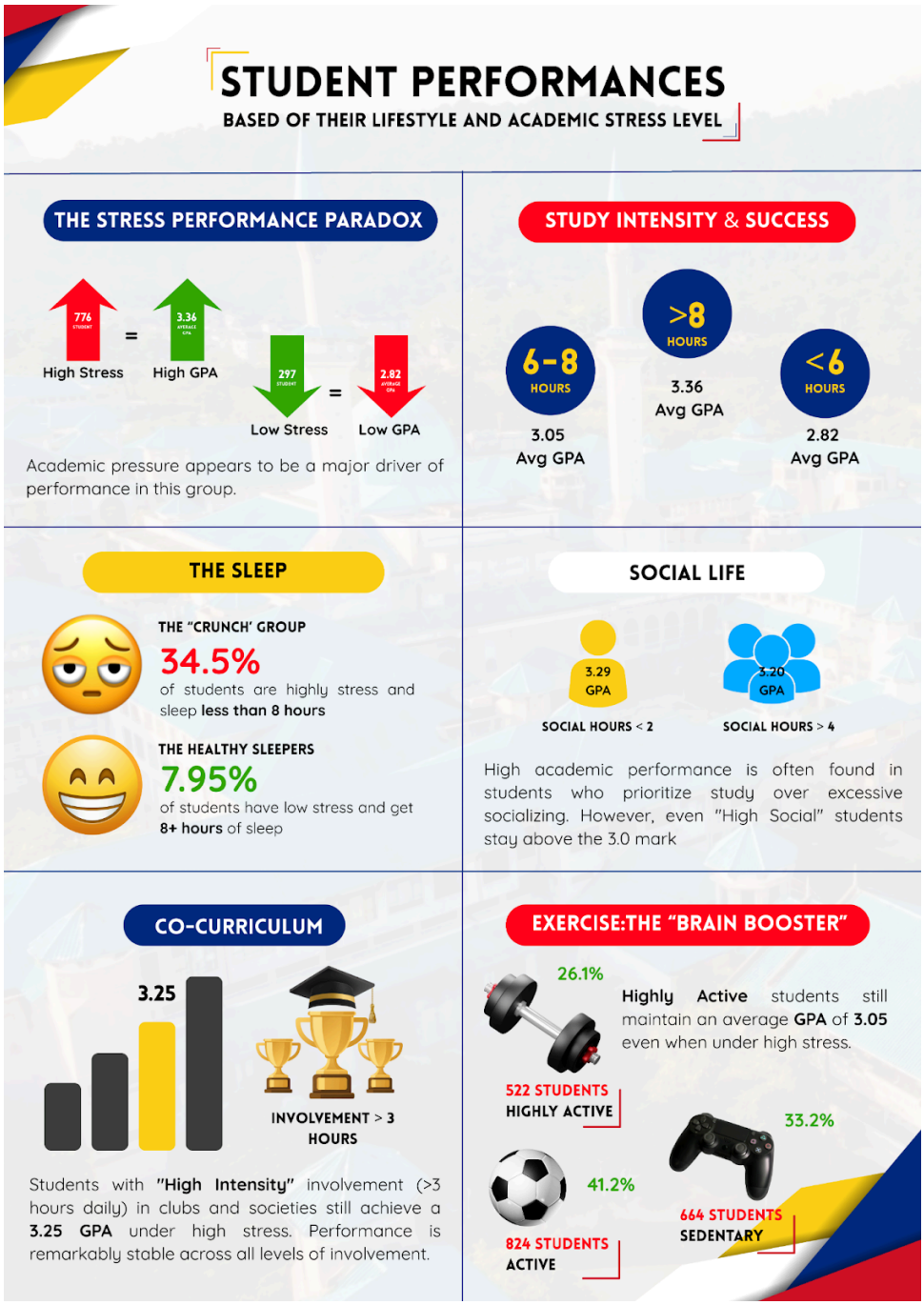
The bar chart analysis reveals that stress level is a primary determinant of academic performance, with the High stress group consistently achieving the highest average GPA (above 3,20) across all social hour categories.In contrast,the Low and Moderate Stress groups show significantly lower GPA averages,hovering between 2.79 and 3.04,suggesting that a lack of academic pressure may correlate with lower performance in this sample. Interestingly, while high stress is associated with superior academic results in the dataset, an inverse relationship is observed between social engagement and GPA within this group.Students in the High Stress/Low Social category peaked at a GPA of 3.29,which slightly declined to 3.20 for those in the High Social category. This suggest that while academic pressure may drive higher performance,excessive social hours still introduce.

E) Daily Exercise (Hour) vs GPA by Stress Levels



The horizontal bar chart shows that the average GPA is quite similar across all three activity levels. Whether students are active, highly active or sedentary, their GPA does not change very much. The same pattern across stress levels as well. Students with high, moderate or low stress all show the same GPA within each activity group. Overall, this suggests that being more physically active does not harm academic performance. Students who take part in more activities are still able to achieve grades that are comparable to those who are less active. In another words, being busy with other physical activity does not automatically lead to lower GPA and students seem able to manage both academics and activities at the same time.

6. DATA VISUALIZATION



Based on the data analysis, it is identified that academic pressure acts as a primary driver of success ,with High Stress students achieving a significantly higher average GPA of 3.36 compared to their low-stress peers.In terms of study habits, students investing more than 8 hours daily record the maximum academic performance, although this often comes at cost to wellbeing, as 34.5% of students fall into the “Crunch” group with high stress and insufficient sleep and also low social hours.This may cause , that 34.5% students spend most of their time learning and exploring things due to academic pressure in order to achieve an excellent pointer in the end of semester which sacrifice their sleep and social time.Despite these pressures, the data proves that maintaining a balanced lifestyle through exercise , social and co-curricular involvement (over 3 hours daily) allows students to sustain a stable GPA above the 3.0 mark. Ultimately , the infographic summarizes that while high intensity and stress are correlated with top-tier results, a holistic approach to lifestyle helps mitigate the negative impacts of academic strain.

7. WORK DISTRIBUTION

NO	TASKS	ASSIGNED TO
1.	● Introduction ● Research Question ● Data Analysis: Daily Co-curricular Activity vs GPA by Stress Levels ● Data Wrangling: Daily Co-curricular Activity vs GPA by Stress Levels	Hakim
2.	● Dataset ● Data Preprocessing ● Data Wrangling: Daily Sleep Duration (Hour) vs GPA by Stress Levels ● Data Analysis: Daily Sleep Duration (Hour) vs GPA by Stress Levels	Azzam
3.	● Data Wrangling: Daily Revision (Hour) vs GPA by Stress Levels ● Data Analysis: Daily Revision (Hour) vs GPA by Stress Levels	Imran
4.	● Research objective ● Data Analysis: Daily Exercise (Hour) vs GPA by Stress Levels ● Data Wrangling: Daily Exercise (Hour) vs GPA by Stress Levels	Faheem
5	● Data Visualisation : ● Data Analysis: Daily Social Activity (Hour) vs GPA by Stress Levels ● Data Wrangling: Daily Social Activity (Hour) vs GPA by Stress Levels	Azim

8. REFERENCES

Ishara, S. (2024). *Student stress & performance insights*. Kaggle. https://www.kaggle.com/code/sulaniishara/student-stress-performance-insights