UNIVERSITI MALAYSIA PAHANG
AL-SULTAN ABDULLAH

BSD2223 DATA SCIENCE PROGRAMMING II
2023/2024 SEMESTER II

PROJECT TITLE :
COVID-19 AND ITS IMPACT ON STUDENT'S MENTAL HEALTH

PREPARED FOR :
DR. ROSLINAZAIRIMAH BINTI ZAKARIA

MATRIC ID	NAME	SECTION
SD22066	SHAHIRA BINTI MOHAIDEEN MEERA	02G
SD22041	TANUSHALANI A/P MUNIANDY	02G
SD22021	AMIRAH YASMIN BINTI ZAILANEE	02G
SD22058	SHAMIR BIN SHAHRIN	02G

ABSTRACT

The study titled "COVID-19 and Its Impact on Students' Mental Health" aims to analyze the psychological effects of the COVID-19 pandemic on students, identify the challenges they faced during this period and provide insights for future strategies to support mental well-being. Utilizing a comprehensive dataset from Kaggle, which includes survey responses on students' mental health during the pandemic, the research explores various stressors such as social isolation, academic pressure and uncertainty about the future. The analysis reveals significant increases in anxiety, depression and stress levels among students, highlighting the urgent need for accessible mental health resources and robust support systems. The study's findings offer valuable recommendations for educational institutions and policymakers to develop effective mental health interventions and resilience-building strategies. In conclusion, by understanding the specific factors that contributed to mental health issues during the pandemic, stakeholders can better prepare for future crises, ensuring a healthier, more supportive educational environment that promotes students' academic and personal growth even in challenging times.

TABLE OF CONTENTS

1.0 INTRODUCTION

The global impact of the COVID 19 pandemic has presented challenges affecting aspects of everyday life worldwide. One of the sectors that has been notably affected is education, with students experiencing a range of disruptions. The sudden switch to learning along with the toll of extended social distancing has created a unique and urgent situation.

The emergence of the pandemic resulted in the shutdown of schools and universities globally requiring students and teachers to adjust to methods of learning. Traditional face to face classes were replaced by platforms presenting personal obstacles. Many students struggled to adapt to learning due to resources like reliable internet connection and suitable devices for learning. Moreover the absence of a structured classroom setting often led to reduced motivation and participation levels among students making it challenging for them to stay on top of their studies.

In addition to challenges, the pandemic has had an impact on students' mental health. The abrupt transition to prolonged isolation has contributed to heightened levels of stress, anxiety, depression and other mental health concerns, among students. The changes in their habits, lack of connections and the unknown future have led to increased tension and emotional strain. These mental health obstacles have impacted not students' school achievements but their general state of being**.**

2.0 OBJECTIVES

Given the purpose of the project, the following study objectives can be outlined:

1. To analyze the impact of COVID-19 on students' mental health**.**
2. To identify the challenges faced by students during the pandemic.
3. To provide insights for future strategies.

3.0 PROJECT DESCRIPTION

This project investigates the impact of the COVID-19 pandemic on students' mental health, particularly focusing on students from various educational institutions within the Delhi National Capital Region (NCR). Utilizing a dataset comprising 1182 entries, each capturing a range of variables related to students' academic and personal lives during the pandemic, the primary objectives are to analyze the effects on mental health, identify the challenges faced, and provide insights for future strategies. The significance of this study lies in understanding the significant impact of global crises on students' mental health and well-being, aiding educators, policymakers, and mental health professionals in supporting students more effectively.

The project reveals important insights through thorough data analysis, such as the finding that students outside of Delhi spend more time taking online classes than their peers in the capital. This disparity could be caused by variations in the availability of physical resources, internet connectivity, and alternate teaching methods. In order to find trends and possible areas for helpful interventions, the project also looks at how time spent on different activities, such as sleep, social media, and fitness, corresponds with students' mental health.

Through the creative combination of in-depth data analysis and useful suggestions, this project provides a strong framework for comprehending and resolving the difficulties faced by students during pandemics. Prospective avenues for future research may comprise formulating a mental health support system grounded in the acquired insights, carrying out longitudinal investigations to monitor enduring consequences, and broadening the dataset to encompass a more heterogeneous student body. The project's ultimate goal is to provide stakeholders with the information and resources they need to help students in an efficient manner during times of global crisis, assuring their academic performance and mental health.

4.0 DATA DESCRIPTION

This dataset examines the impact of the COVID-19 pandemic with a sample size of 1182 students of different age groups from different educational institutions in Delhi National Capital Region (NCR). It covers various aspects of education and well-being, with data collected from the onset of the pandemic until recent years. The dataset contains 1182 rows with 19 columns.

No	Title	Explanation	Type
1	ID	Unique identifier for each student in the dataset	Quantitative
2	Region of residence	Indicates whether the student resides in Delhi-NCR or outside Delhi-NCR	Qualitative
3	Age	Age range of students from 7 to 59 years old.	Quantitative
4	Time spent on Online Class (Hours)	Range from 0.5 to 10 hours spent on attending online classes	Quantitative
5	Rating of Online Class experience	Options include Average, Excellent, Good, Poor, and Very Poor.	Qualitative
6	Medium for online class	Indicates the device used for attending online classes (e.g., Gadget, Laptop/Desktop, Smartphone, Tablet).	Qualitative
7	Time spent on self study (Hours)	Range from 0.5 to 18 hours spent on self-study	Quantitative

8	Time spent on fitness (Hours)	Range from 0.5 to 5 hours spent on fitness activities.	Quantitative
9	Time spent on sleep (Hours)	Range from 4 to 15 hours spent on sleep	Quantitative
10	Time spent on social media (Hours)	Range from 0.5 to 10 hours spent on social media	Quantitative
11	Preferred social media platform	Options include various social media platforms such as Facebook,Instagram and etc..	Qualitative
12	Time spent on TV (Hours)	Range from 0.1 to 15 hours spent on watching TV	Quantitative
13	Number of meals per day	Range from 1 to 8 times meals consumed per day	Quantitative
14	Change in weight	Indicates whether weight decreased, increased, or remained constant during the pandemic.	Qualitative
15	Health issue during lockdown	Indicates whether the student experienced health issues during the lockdown (Yes or No).	Qualitative
16	Stress busters	Various activities students engage in to reduce stress, with multiple options provided.	Qualitative
17	Time utilized	Indicates whether time during the pandemic was utilized effectively (Yes or No).	Qualitative

18	Do you feel closer to family, friends, or relatives?	Indicates whether students feel closer to family, friends, or relatives during the pandemic (Yes or No).	Qualitative
19	What do you miss the most?	Various activities or experiences students miss the most during the pandemic, with multiple options provided	Qualitative

5.0 DATA PREPARATION

As in the image below, the dataset we chose had missing values in the column of Rating of Online Class experience,Medium for online class and Prefered social media platform.

```
[44]: print("Missing values before handling:")
      print(df.isnull().sum())

Missing values before handling:
ID                                                                  0
Region of residence                                                 0
Age of Subject                                                      0
Time spent on Online Class                                          0
Rating of Online Class experience                                  24
Medium for online class                                            51
Time spent on self study                                            0
Time spent on fitness                                               0
Time spent on sleep                                                 0
Time spent on social media                                          0
Prefered social media platform                                     17
Time spent on TV                                                    0
Number of meals per day                                             0
Change in your weight                                               0
Health issue during lockdown                                        0
Stress busters                                                      0
Time utilized                                                       0
Do you find yourself more connected with your family, close friends , relatives  ?    0
What you miss the most                                              0
dtype: int64
```

In our data preprocessing step, we identified non-numeric columns such as 'Rating of Online Class experience', 'Medium for online class', and 'Preferred social media platform' using the select_dtypes() function.We then filled missing values in each of these non-numeric columns by replacing them with the mode, which represents the most frequent choice made by respondents.

```
[51]: numeric_cols = df.select_dtypes(include=[np.number]).columns
      numeric_cols

[51]: Index(['Age of Subject', 'Time spent on Online Class',
             'Time spent on self study', 'Time spent on fitness',
             'Time spent on sleep', 'Time spent on social media',
             'Number of meals per day'],
            dtype='object')

[52]: non_numeric_cols = df.select_dtypes(exclude=[np.number]).columns
      non_numeric_cols

[52]: Index(['ID', 'Region of residence', 'Rating of Online Class experience',
             'Medium for online class', 'Prefered social media platform',
             'Time spent on TV', 'Change in your weight',
             'Health issue during lockdown', 'Stress busters', 'Time utilized',
             'Do you find yourself more connected with your family, close friends , relatives  ?',
             'What you miss the most'],
            dtype='object')
```

```
[71]: df[numeric_cols]= df[numeric_cols].fillna(df[numeric_cols].mean())
      df[numeric_cols]
```

[71]:

	Age of Subject	Time spent on Online Class	Time spent on self study	Time spent on fitness	Time spent on sleep	Time spent on social media	Number of meals per day
0	21	2.0	4.0	0.0	7.0	3.0	4
1	21	0.0	0.0	2.0	10.0	3.0	3
2	20	7.0	3.0	0.0	6.0	2.0	3
3	20	3.0	2.0	1.0	6.0	5.0	3
4	21	3.0	3.0	1.0	8.0	3.0	4
...	...	...	...	...	...	...	...
1177	12	3.0	4.0	1.0	8.0	1.0	3
1178	14	6.0	4.0	1.0	9.0	1.0	4
1179	13	4.0	0.0	0.5	8.0	3.0	4
1180	14	5.0	3.5	1.0	8.0	0.5	4
1181	13	5.0	2.0	0.5	7.0	1.0	3

1182 rows × 7 columns

```
[72]: for col in non_numeric_cols :
          df[col].fillna(df[col].mode()[0], inplace=True)
```

After completing the preprocessing steps, we addressed missing values in several columns of our dataset. This is the result we obtained after handling the missing values.

```
[57]: print("Missing values after handling:")
      df.isnull().sum()
```

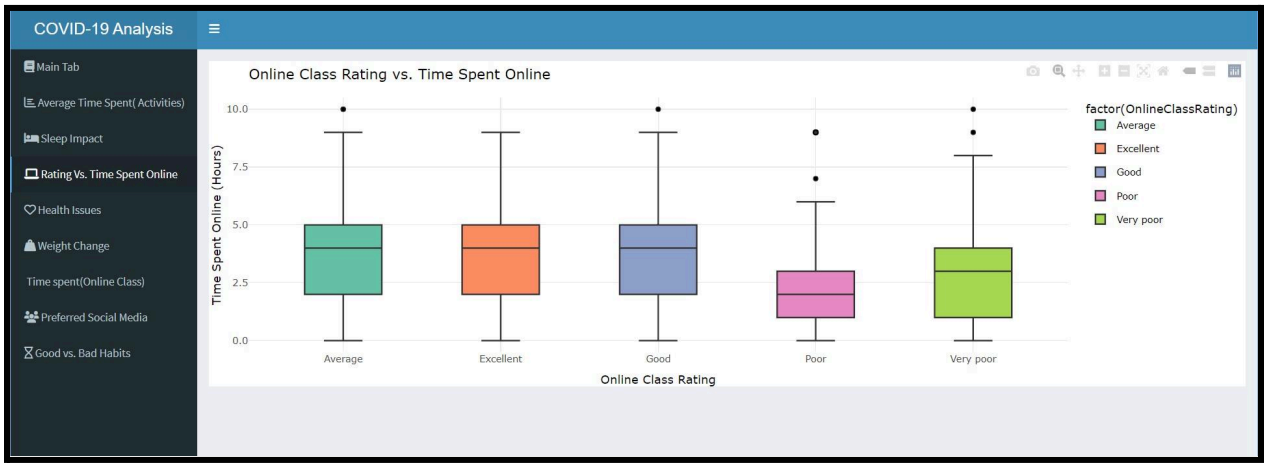
```
Missing values after handling:
```

[57]:

```
ID                                                                  0
Region of residence                                                 0
Age of Subject                                                      0
Time spent on Online Class                                          0
Rating of Online Class experience                                   0
Medium for online class                                             0
Time spent on self study                                            0
Time spent on fitness                                               0
Time spent on sleep                                                 0
Time spent on social media                                          0
Prefered social media platform                                      0
Time spent on TV                                                    0
Number of meals per day                                             0
Change in your weight                                               0
Health issue during lockdown                                        0
Stress busters                                                      0
Time utilized                                                       0
Do you find yourself more connected with your family, close friends , relatives  ?    0
What you miss the most                                              0
dtype: int64
```

6.0 DATA ANALYSIS, RESULTS AND DISCUSSION

Boxplot: Online Class Rating vs Time Spent Online



This boxplot visualization provides a detailed overview of the relationship between the time spent online in hours and students' ratings of their online classes. The ratings are categorized as "Average," "Excellent," "Good," "Poor," and "Very poor." For students who rated their online classes as "Average," the distribution shows a maximum online time of 10 hours, with an outlier at this upper limit. The interquartile range (IQR) for this group is from 2 hours (Q1) to 5 hours (Q3), with a median of 4 hours, indicating that most students in this category spend between 2 to 5 hours online.

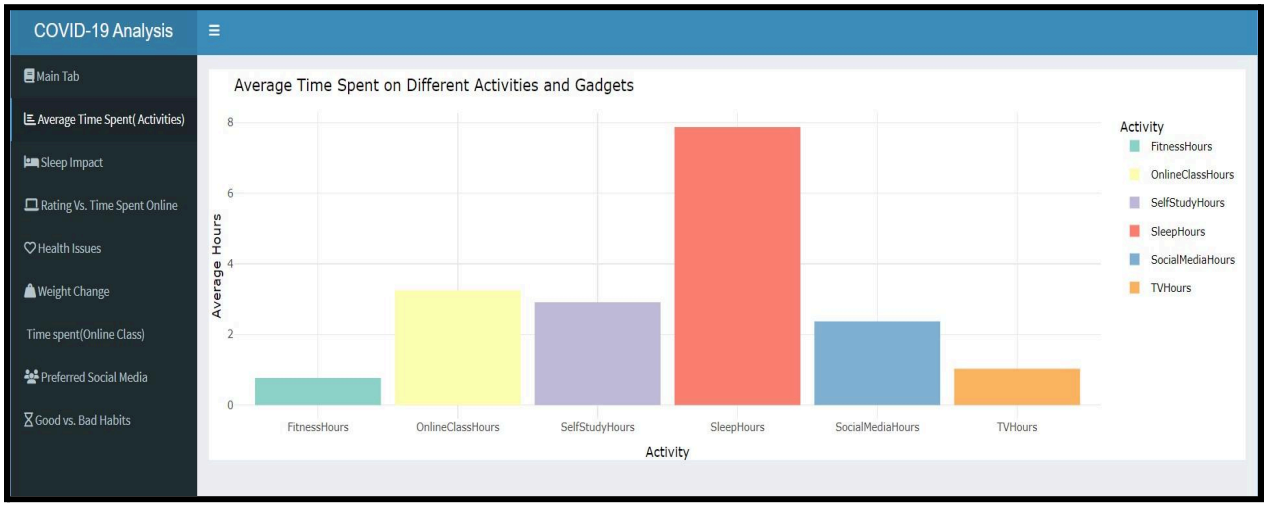
For the "Excellent" rating category, the maximum online time is 9 hours, with no outliers. The IQR remains consistent with the "Average" group, ranging from 2 hours (Q1) to 5 hours (Q3), and the median is also at 4 hours. This consistency in the distribution suggests that students who are very satisfied with their online classes tend to spend a moderate amount of time engaged in these activities, similar to those who rate their experience as average.

Students who rated their online classes as "Good" also show a maximum time of 10 hours, with outliers at this upper limit. The IQR for this group mirrors the previous categories,

with Q1 at 2 hours, Q3 at 5 hours, and a median of 4 hours. The similarity in time distribution across the "Average," "Excellent," and "Good" ratings suggests that a moderate amount of time spent online is a common factor among students who perceive their online classes positively, whether they are satisfied or highly satisfied.

In conclusion, students who rated their classes as "Poor" or "Very poor" spent less time online. The "Poor" category shows a maximum time of 9 hours with significant outliers at 7 and 9 hours. The IQR is narrower, from 1 hour (Q1) to 3 hours (Q3), with a median of 2 hours. The "Very poor" category also has outliers at 9 and 10 hours, with an IQR from 1 hour (Q1) to 4 hours (Q3), and a median of 3 hours. These distributions indicate that students who are dissatisfied with their online classes tend to spend less time engaged with them, although there are a few exceptions who spend significantly more time online. This variability in time commitment among the dissatisfied groups may reflect struggles with the content or engagement issues, leading to lower satisfaction ratings.

Bar Chart: Average Time Spent on Different Activities and Gadgets



The bar chart provided illustrates the average time spent on different activities and gadgets during the COVID-19 pandemic. The x-axis represents various activities, namely FitnessHours, OnlineClassHours, SelfStudyHours, SleepHours, SocialMediaHours, and TVHours, while the y-axis measures the average hours spent on each activity. Each activity is represented by a distinct color, as indicated in the legend on the right.

The chart reveals that SleepHours, represented in red, dominate the average time spent, reaching close to 8 hours. This suggests that individuals have allocated a significant portion of their day to sleep, potentially influenced by the flexibility of remote work or study schedules during the pandemic.

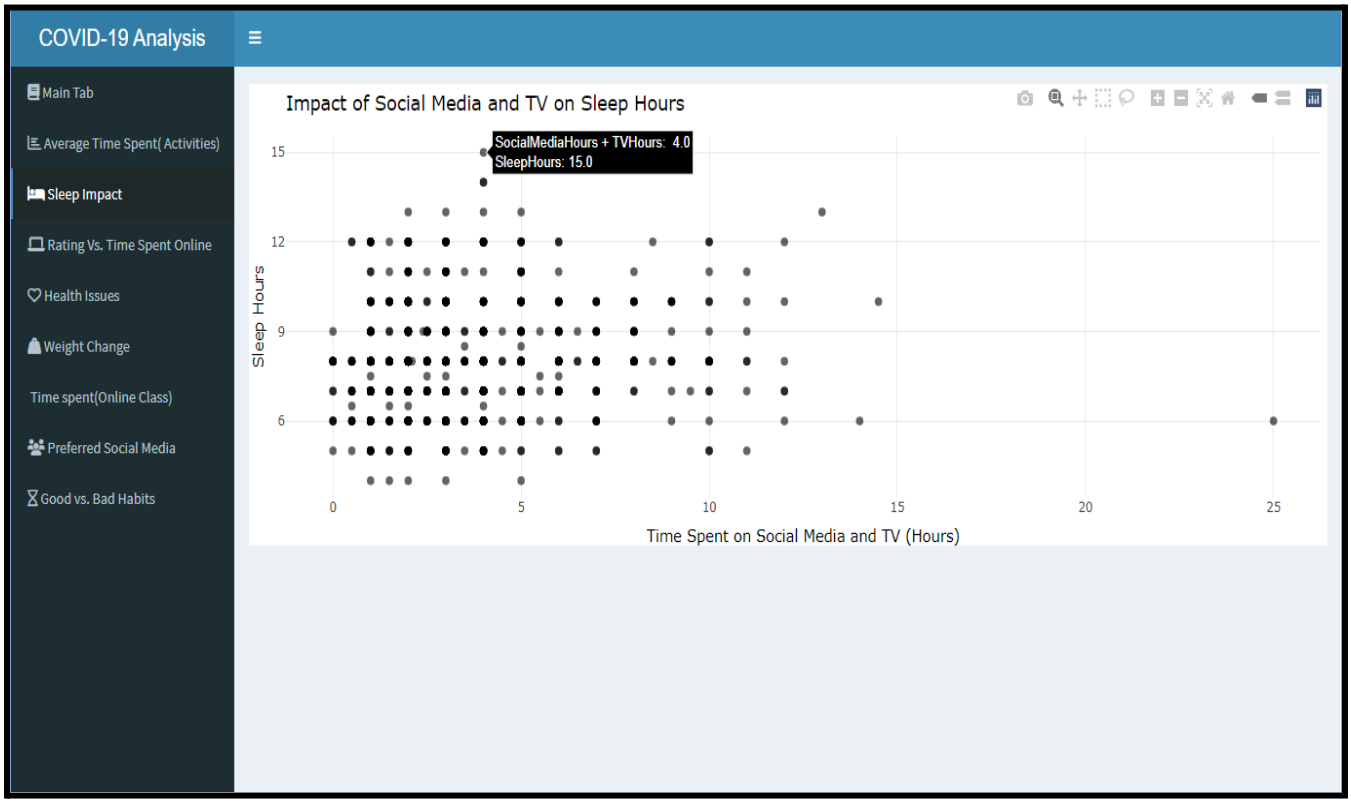
OnlineClassHours, depicted in yellow, and SelfStudyHours, shown in purple, also consume considerable time, with each averaging around 4 hours. This highlights the substantial time commitment students and professionals have dedicated to online education and self-directed learning during the pandemic. The transition to online learning has evidently required a substantial daily investment of time.

SocialMediaHours, illustrated in blue, and TVHours, in orange, are moderate in comparison, averaging around 2 hours each. These activities represent the time spent on digital

entertainment and social interaction, indicating that while these activities are significant, they do not overshadow educational and sleep commitments.

Lastly, FitnessHours, represented in light blue, account for the least amount of time, averaging around 1 hour. This suggests that despite the recognized importance of physical activity, individuals might have struggled to prioritize fitness during the pandemic, potentially due to restrictions on outdoor activities or access to gyms. Overall, the chart provides a comprehensive view of how daily routines have been reshaped, with a noticeable emphasis on sleep and online education.

Scatter Plot: Impact of Social Media and TV on Sleep Hours



The scatter plot presented demonstrates the relationship between the amount of time students spend on social media and television, and their corresponding sleep hours. At first glance, it becomes apparent that the majority of data points are densely packed between 0 to 5

hours of screen time, regardless of the time spent on sleep. This clustering suggests that a significant number of students spend minimal time on social media and television, with a relatively broad range of sleep hours, predominantly from 6 to 12 hours. This indicates that while many students might be engaging in these activities for limited periods, their sleep duration varies significantly, hinting at other factors possibly influencing their sleep patterns.

As we delve deeper into the data, a more nuanced picture emerges regarding the impact of increased screen time on sleep. When examining the range of 6 to 10 hours spent on social media and television, a noticeable trend appears: sleep hours generally tend to decrease. Most students in this bracket receive between 6 to 9 hours of sleep, reflecting a modest reduction in sleep with increased screen exposure. This section of the plot, however, contains fewer data points, suggesting that not many students spend extensive amounts of time on social media and television. It implies that, for those who do, the increased screen time could be linked to reduced sleep duration.

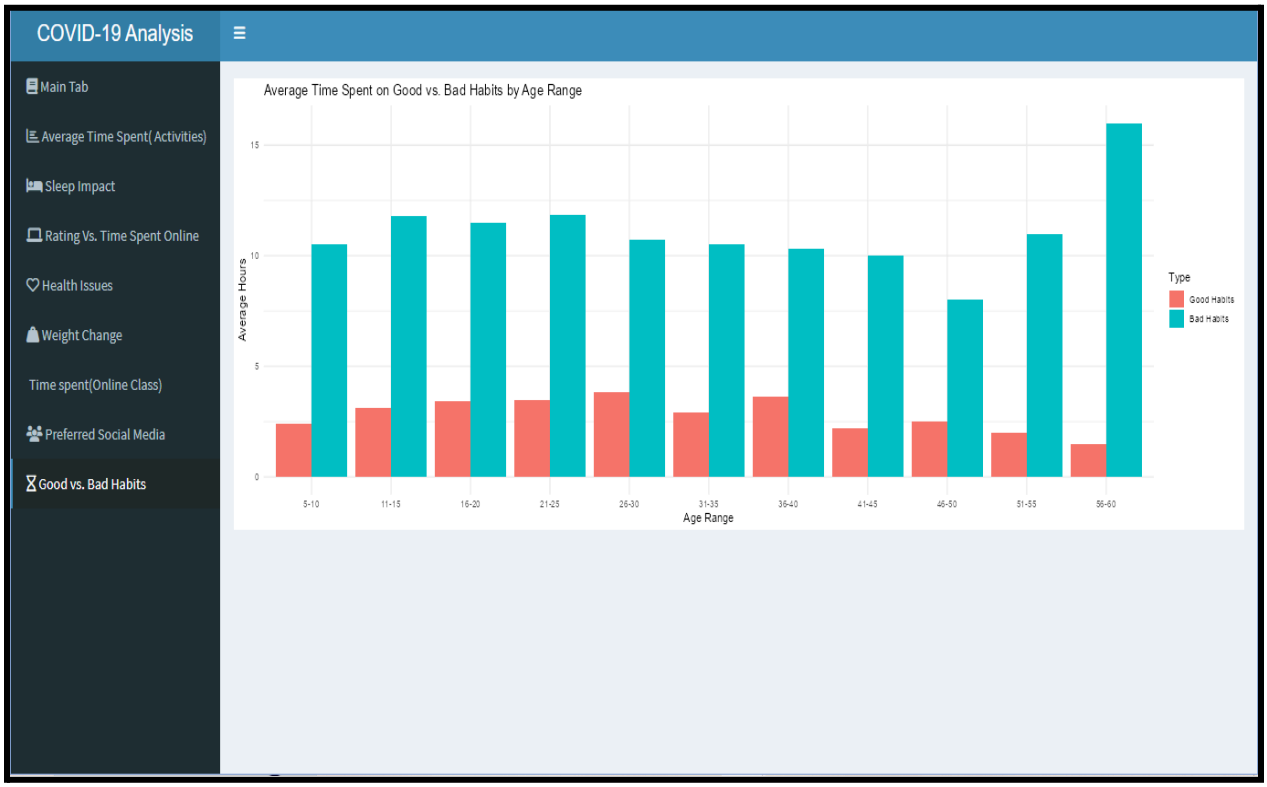
Further into the analysis, the sparse distribution of data points among students who spend more than 10 hours on social media and TV becomes evident. While some students in this category manage to secure a higher amount of sleep, ranging from 10 to 12 hours, these instances are rare. The scarcity of data points in this range indicates that it is uncommon for students to spend such large amounts of time on these activities. Those who do, tend to sleep less, although some outliers exist where students manage to maintain significant sleep hours despite high screen time.

Overall, the plot suggests a general inverse relationship between the amount of time spent on social media and television and the hours of sleep students get. Although the reduction in sleep with increased screen time is apparent, it is not starkly dramatic, indicating that the impact of social media and TV on sleep is moderate. Many students seem capable of balancing their screen time with adequate sleep. However, the scattered points where sleep drops to 4-6 hours, even with minimal screen exposure, point to the presence of additional factors that might be affecting their sleep quality.

In conclusion, the scatter plot reveals that while higher engagement in social media and television might be associated with reduced sleep hours, the effect is moderate and varies

significantly among individuals. The key takeaway is the importance of balancing digital consumption with sufficient sleep, a balance that seems achievable for many students. Nevertheless, there remains a portion of the student population for whom excessive screen time could be detrimental to maintaining healthy sleep patterns, highlighting the need for mindful media consumption.

Grouped Bar Chart: Average Time Spent on Good Vs. Bad Habit by Age Range



The grouped bar chart depicting the average time spent on good versus bad habits across various age ranges reveals significant insights into the lifestyle patterns of different age groups during the COVID-19 pandemic. The analysis indicates that individuals in the 56-60 age range dedicate the most time to good habits, averaging 16 hours, which is the highest across all age

groups. Conversely, this age group spends the least time on bad habits, around 1.5 hours, suggesting a healthier balance between productive and less productive activities.

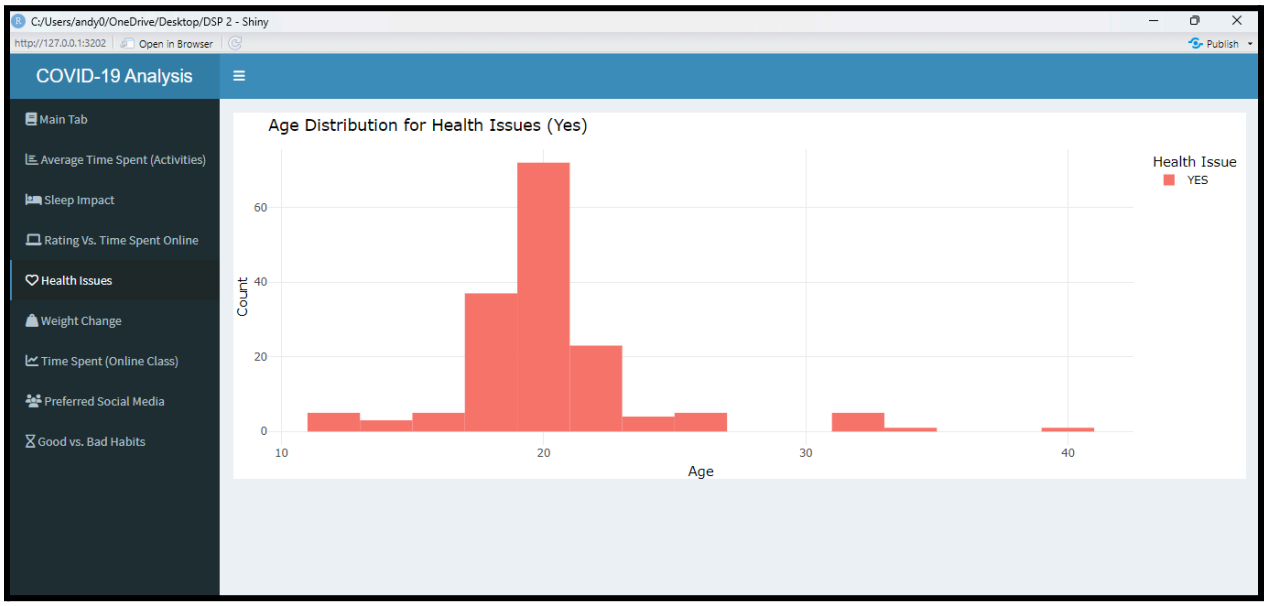
In contrast, the 26-30 age range exhibits the highest average time spent on bad habits, with approximately 3.8 hours. This group also spends a considerable amount of time on good habits, about 11.3 hours, but their engagement in bad habits is notably higher than in other age brackets. This pattern might reflect the challenges faced by individuals in their late twenties to early thirties, balancing career pressures and personal well-being.

The younger age groups, particularly those aged 5-10 and 11-15, demonstrate a substantial commitment to good habits, averaging around 10 to 11.7 hours. Their time spent on bad habits is relatively low, around 2 to 3.2 hours, indicating a generally healthy lifestyle balance. These younger individuals have possibly benefited from structured routines, whether through school or parental guidance, that emphasize positive activities.

For the age groups 31-35 and 41-45, there is a noticeable higher proportion of time allocated to bad habits, averaging around 3.2 to 3.7 hours, compared to other age ranges. Their time spent on good habits, though still significant, is lower than that of the younger and older groups, averaging around 9.8 to 10 hours. This suggests that individuals in these age groups may face more distractions or stressors that lead to increased engagement in less productive activities.

Overall, the data reveals a clear trend where older adults, particularly those in the 56-60 age range, prioritize good habits significantly more than younger adults and middle-aged individuals. Younger age groups maintain a healthy balance between good and bad habits, while those in their late twenties to mid-thirties appear to struggle more with maintaining this balance. This analysis highlights the varied impacts of the COVID-19 pandemic on different age groups, shedding light on their mental health and lifestyle adaptations.

Histogram: Age Distribution for Health Issues



The histogram illustrating the age distribution of health issues among students during the COVID-19 pandemic. This visualization provides significant insights into how different age groups were affected. The x-axis represents the age of students, segmented into bins, while the y-axis shows the count of students within each age group. The bars are divided into two colours to indicate whether students reported experiencing health issues during the lockdown (Yes). In a look, the histogram highlights specific trends and patterns by displaying clusters of health conditions across different age groups.

For younger students, for example those aged between 7 and 15, the incidence of health issues appears lower. Bars representing health issues 'Yes' are minimal, suggesting fewer reported health concerns within this cohort. This might indicate a fewer stressors impacting health in younger students. In contrast, students aged 16 to 25 show a higher incidence of health issues, with a noticeable increase in the height of bars for 'Yes'. This age group likely faced significant stressors related to educational pressures, social isolation, and uncertainty about the future, contributing to a higher prevalence of reported health issues.

For students aged 26 and above, there is a slight decline in the number of reported health issues. However, there are still significant bars indicating 'Yes' for health issues, suggesting that while the intensity might have reduced compared to the mid-age group, a substantial number of older students still experienced health-related challenges during the pandemic. The highest concentration of health issues among students aged 16-25 highlights a particularly vulnerable group during the pandemic, suggesting a need for targeted mental health support and resources for this demographic. The relatively lower incidence of health issues among younger students suggests better coping mechanisms or fewer external stressors affecting this group, though ongoing support remains essential.

Overall, the visualization suggests a bell-shaped curve with the mid-age groups being the most affected by health issues during the pandemic. Understanding these patterns can help in designing age-specific interventions to better support student well-being.

Grouped Bar Chart: Average Change in Weight by Health Issues

The bar plot comparing the average change in weight among students, grouped by whether they experienced health issues during the pandemic, reveals interesting insights into the relationship between mental health and physical health outcomes. The x-axis differentiates between students with ('Yes') and without ('No') health issues during the lockdown, while the y-axis indicates the average weight change. Bars are divided by colours to reflect this differentiation, providing a clear visual comparison.

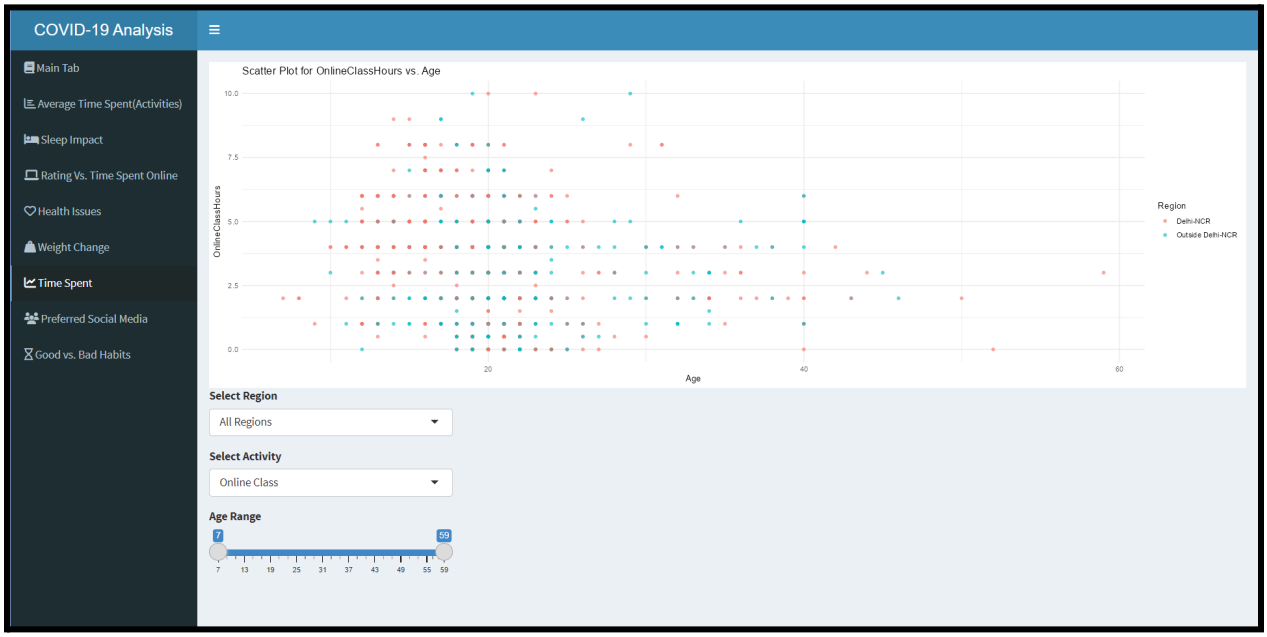
Interestingly, both charts show similar outputs, highlighting that health issues and weight change are closely interlinked. The bars for students who reported health issues 'Yes' indicate a notable weight change, suggesting that these students experienced significant physical health impacts during the pandemic. This could be due to stress-related eating habits, decreased physical activity, and other factors associated with mental health challenges.

Conversely, the bars representing students without health issues ('No') show a more stable weight pattern, with minimal average weight change. This stability suggests that students without health issues managed to maintain their physical health better, possibly due to more consistent routines and healthier lifestyle choices.

The similarity between the two visualizations underscores the strong correlation between mental health issues and weight changes. Students who faced health issues during the pandemic were more likely to experience significant weight fluctuations, highlighting the need for integrated health strategies that address both mental and physical well-being. Providing resources for stress management, promoting physical activity, and ensuring access to nutritious food can help mitigate the negative impacts of health issues on weight.

Overall, these visualizations emphasize the importance of supporting both mental and physical health among students, especially during challenging times like the COVID-19 pandemic. By understanding and addressing the interconnected nature of these aspects, we can better support students' overall well-being.

Scatter Plot: Time Spent on Different Activities based on Age and Region

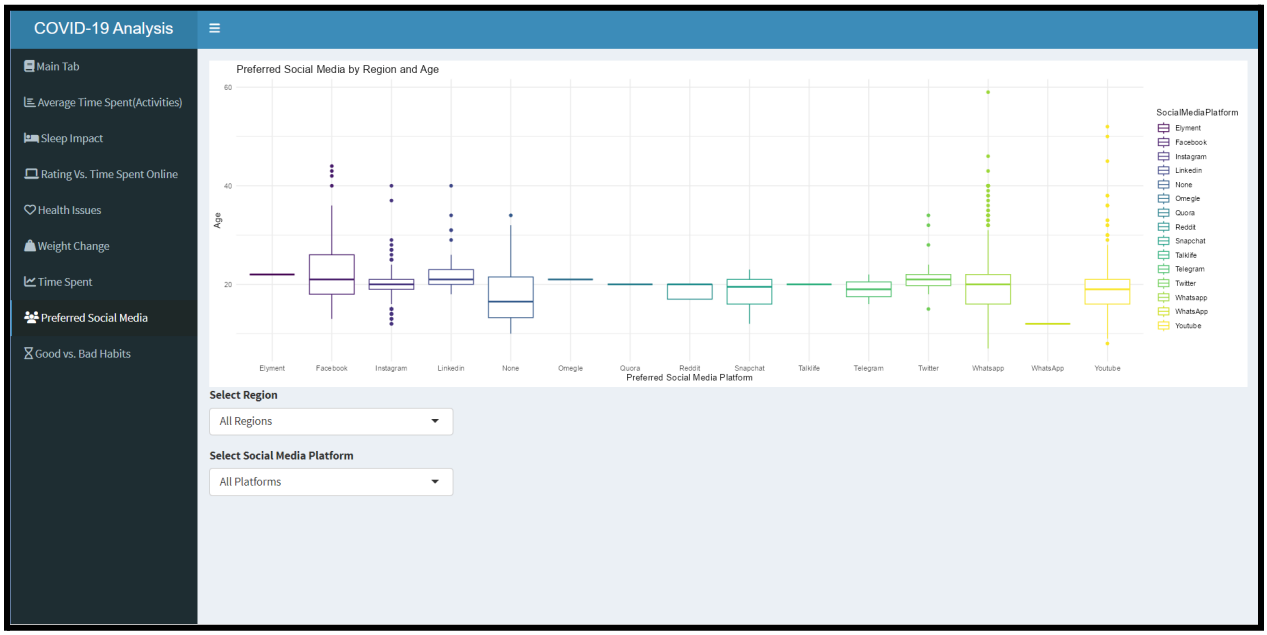


This scatter plot shows that the older a person becomes, the less time they spend on virtual education. The age of those being taught through online classes ranges from 7 to 59 years. Ideally, younger children spend more hours in these forums as opposed to their parents. This trend is worldwide without any exemptions. People from all parts of the world exhibit similar behaviors in terms of digital studying platforms. However, this does not mean that there are no differences at all between various regions around the globe and their involvement levels with e-learning systems. The major difference comes when we start comparing students' hours of attendance based on their geographical areas; there seems to be almost no similarity among them although generalization cannot be made at this point since statistically insignificant data were found during research. The scatter plot illustrates the relationship between the number of hours spent in online classes and the age of individuals with data points color-coded by area (Delhi-NCR-red, Outside Delhi-NCR-blue). Participants' age ranges from 7 to 59 years; most hours are clustered between 0 and 10 with very few outliers beyond this range.

For kids and teens (7-19 years old), there is an upward trend on average time spent per week while attending virtual lessons whose peak lies within 5 hours – showing higher participation levels for these groups respectively. However, among young adults (20-30) we

observe that as they grow older there are fewer hours taken during each session indicating declining interest towards this mode of study which may be attributed to a number of factors such as work commitments or lack thereof where necessary qualifications have been met already and therefore one finds no need for furthering their education through such means in case they were using them partly because they couldn't access traditional schools while others had different reasons altogether maybe academic excellence which rendered physical classrooms unnecessary. Moving on rapidly into adulthood (31-59), graphs show us considerable reduction both average but also more spread out thus highlighting limited number engaged fulltime employment possibly due inability secure formal placement within competitive job market that would provide better incomes enabling them meet financial obligations involving higher learning.

Box Plot: : Preferred Social Media Based on Age and Region



The box plot shows the social media platforms that are most popular among different age groups and regions. The plot uses different colors to represent various platforms while also using box plots to show how user age is distributed across these platforms. From the diagrams, it can be observed that each social platform has its own unique user base when it comes to age. With

these box plots, one can easily see the median age, quartiles as well as any outliers for each site. A few people seem to know about Elyment which means it is not very popular perhaps the reason being that majority its users are about 30 years old only. The age range of Facebook users is large spanning from around mid-30s all way upto 60 years with significant numbers being recorded at mid-30s while some go beyond 60. On the other hand Instagram has more younger people using it, with many of them having an average age of about mid-20s although there few elderly outliers too. Typically professionals in their early thirties dominate LinkedIn user base though there are older individuals also present among them. Some individuals do not like any particular social media platform and they can be of any age. Moreover, Omegle, Quora, Reddit, Snapchat, Talklife and Telegram are predominantly used by younger users generally while that of Twitter cuts across various ages including those between 20-40 years old but WhatsApp takes the lead towards older adults specifically having a median age in the late thirties. YouTube has been accredited with being loved by people from all walks of life where there seems to more concentration around 30 years. These findings suggest that there are no significant differences in regional preference towards social networks since both regions were combined in this study; however attention was paid more on how ages are shared among different areas than anything else.

7.0 CONCLUSION

Findings from COVID-19 impact analysis on the daily activities and psychological states of students are therefore immense in contributing quality information. The key findings help to summarize how students allocated their time for activities during the pandemic; extending online class hours and self-study hours were seen across most age groups and regions.

Especially, which records a definite relation between the number of sleep hours and general well-being, such that students conforming to stable sleep patterns indicated better mental health and higher satisfaction from online classes. The study also pinpointed a large variance in online class ratings; hence, there have been excellent satisfaction discrepancies between students. The efficacy of online classes was found to be good and appreciated by age categories between the ages of 15 and 25 years.

Health issues, and that including mental health concerns, were widely reported during the pandemic. The mental burdens, predominantly stress and anxiety, were of concern, further ensuring the need to factor in psychological support and mental health resources among students. Furthermore, all the age groups showed a typical trend in the analysis of social media preferences; finding, therefore, was that WhatsApp, Instagram, and Facebook were the leading entries. Usage depended on age and geography, eliciting information on how adolescents remained connected and received liabilities during COVID-19.

These findings will help stakeholders, such as educators, policymakers, and mental health professionals, to design interventions and initiate explicit support systems.

8.0 LIMITATIONS OF THE STUDY

Although these analyses report many rich insights, there are various limitations to be considered. First, the analysis is based on one specific dataset, which does not represent in totality the range of experiences of students during the pandemic. The dataset is limited by geography and time, and hence, the generalization of findings may also be limited.

Second, there is the potential for sample bias as regards the representativeness of the entire student community. The sampling biases might have contributions from factors such as the methods of data collection and the availability and selection of participants. The study also lacks an exploration of a defined list of variables, including but not limited to family dynamics, socioeconomic status, and access to technology.

Third, reliance on self-reported data could introduce biases, as students might under-report or over-report their activities and feelings. This can taint the truth of findings drawn from the data. Furthermore, the analysis uses certain algorithms and statistical methods for analysis. The efficiency and accuracy of the algorithms vary, and a repetitive analysis of the model with different datasets and methods might enhance the robustness of the results.

The final labor is that the study does not consider longitudinal changes and trends. It therefore fails to offer deeper insights into the long-term impact of the pandemic on the mental quality life and activities of students. In conclusion, although the present study contributes significantly to unraveling the impact of COVID-19 in students' mental health and their daily activities, addressing these limitations through further research is mandatory. Therefore, an expansion of the dataset, inclusion of more variables, and use of proper validation methods will increase the reliability and applicability of the findings.

9.0 APPENDIX

DATASET

https://drive.google.com/drive/folders/1jMPihixGYA93NEvMg763s3iKviMJX5z8?usp=drive_link

R SOURCE CODE

```
library(shiny)
library(shinydashboard)
library(tidyverse)
library(plotly)
library(anytime)
library(ggplot2)
library(dplyr)
library(tidyr)
library(janitor)
library(shinythemes)
library(readr)
library(DT)
library(viridis)

# Load and clean the data
data <- read_csv("C:\\Users\\shami\\OneDrive\\Documents\\DSP 2\\COVID-19 and Its Impact On Students Mental Health.csv") %>%
  clean_names() %>%
  rename(Region = region_of_residence)

# Rename columns for easier referencing
colnames(data) <- c("ID", "Region", "Age", "OnlineClassHours", "OnlineClassRating",
                    "OnlineClassMedium", "SelfStudyHours", "FitnessHours", "SleepHours",
                    "SocialMediaHours", "SocialMediaPlatform", "TVHours",
                    "MealsPerDay", "WeightChange", "HealthIssue", "StressBuster",
```

```
                "TimeUtilized", "FeelCloser", "MissMost")

# Create age range categories
data <- data %>%
  mutate(AgeRange = cut(Age, breaks = c(5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60),
                        labels = c("5-10", "11-15", "16-20", "21-25", "26-30", "31-35", "36-40", "41-45",
"46-50", "51-55", "56-60")))

# Aggregate the data for good and bad things
data <- data %>%
  mutate(GoodThings = SelfStudyHours + FitnessHours + SleepHours,
         BadThings = SocialMediaHours + TVHours)

# UI for the Shiny app
ui <- dashboardPage(
  dashboardHeader(title = "COVID-19 Analysis"),
  dashboardSidebar(
    sidebarMenu(
      menuItem("Main Tab", tabName = "cover", icon = icon("book")),
      menuItem("Average Time Spent (Activities)", tabName = "time_activities", icon =
icon("chart-bar")),
      menuItem("Sleep Impact", tabName = "sleep_impact", icon = icon("bed")),
      menuItem("Rating Vs. Time Spent Online", tabName = "online_class", icon =
icon("laptop")),
      menuItem("Health Issues", tabName = "health_issues", icon = icon("heart")),
      menuItem("Weight Change", tabName = "weight_change", icon = icon("weight-hanging")),
      menuItem("Time Spent (Online Class)", tabName = "scatter_plot", icon = icon("chart-line")),
# Changed icon to chart-line
      menuItem("Preferred Social Media", tabName = "social_media", icon = icon("users")),
      menuItem("Good vs. Bad Habits", tabName = "good_bad", icon = icon("hourglass"))
    )
```

```
  ),
  dashboardBody(
    tabItems(
      tabItem(
        tabName = "cover",
        h1("Welcome to COVID-19 Analysis Dashboard"),
        p("This dashboard provides insights into the impact of COVID-19 on students' mental
health."),
        tags$img(src = "Image for dashboard.jpg", height = 560, width = 1200),
        p("Click on the sidebar menu to explore different sections of the analysis.")
      ),
      tabItem(
        tabName = "time_activities",
        plotlyOutput("plot_time_activities")
      ),
      tabItem(
        tabName = "sleep_impact",
        plotlyOutput("plot_sleep_impact")
      ),
      tabItem(
        tabName = "online_class",
        plotlyOutput("plot_online_class")
      ),
      tabItem(
        tabName = "health_issues",
        plotlyOutput("plot_health_issues")
      ),
      tabItem(
        tabName = "weight_change",
        plotlyOutput("plot_weight_change")
      ),
```

```
      tabItem(
        tabName = "scatter_plot",
        plotOutput("plot_scatter"),
        selectInput("region", "Select Region", choices = c("All Regions", unique(data$Region)),
selected = "All Regions"),
        selectInput("activity", "Select Activity", choices = c("Online Class" = "OnlineClassHours",
"Self Study" = "SelfStudyHours", "Fitness" = "FitnessHours", "Sleep" = "SleepHours", "Social
Media" = "SocialMediaHours", "TV" = "TVHours"), selected = "OnlineClassHours"),
        sliderInput("slider", "Age Range", min = min(data$Age, na.rm = TRUE), max =
max(data$Age, na.rm = TRUE), value = c(min(data$Age, na.rm = TRUE), max(data$Age, na.rm
= TRUE)))
      ),
      tabItem(
        tabName = "social_media",
        plotOutput("plot_social_media"),
        selectInput("region_social", "Select Region", choices = c("All Regions",
unique(data$Region)), selected = "All Regions"),
        selectInput("social_media", "Select Social Media", choices = c("All",
unique(data$SocialMediaPlatform)), selected = "All")
      ),
      tabItem(
        tabName = "good_bad",
        plotOutput("plot_good_bad")
      )
    )
  )
)

# Server logic
server <- function(input, output) {
  output$plot_time_activities <- renderPlotly({
```

```
    data_long <- data %>%
      select(OnlineClassHours, SelfStudyHours, FitnessHours, SleepHours, SocialMediaHours,
TVHours) %>%
      pivot_longer(cols = everything(), names_to = "Activity", values_to = "Hours")

    p <- ggplot(data_long, aes(x = Activity, y = Hours, fill = Activity)) +
      geom_bar(stat = "summary", fun = "mean") +
      labs(title = "Average Time Spent on Different Activities and Gadgets", x = "Activity", y =
"Average Hours") +
      theme_minimal() +
      scale_fill_brewer(palette = "Set3")

    ggplotly(p)
  })

  output$plot_sleep_impact <- renderPlotly({
    p <- ggplot(data, aes(x = SocialMediaHours + TVHours, y = SleepHours)) +
      geom_point(alpha = 0.6) +
      labs(title = "Impact of Social Media and TV on Sleep Hours", x = "Time Spent on Social
Media and TV (Hours)", y = "Sleep Hours") +
      theme_minimal()

    ggplotly(p)
  })

  output$plot_online_class <- renderPlotly({
    p <- ggplot(data, aes(x = factor(OnlineClassRating), y = OnlineClassHours, fill =
factor(OnlineClassRating))) +
      geom_boxplot() +
      labs(title = "Online Class Rating vs. Time Spent Online", x = "Online Class Rating", y =
"Time Spent Online (Hours)") +
```

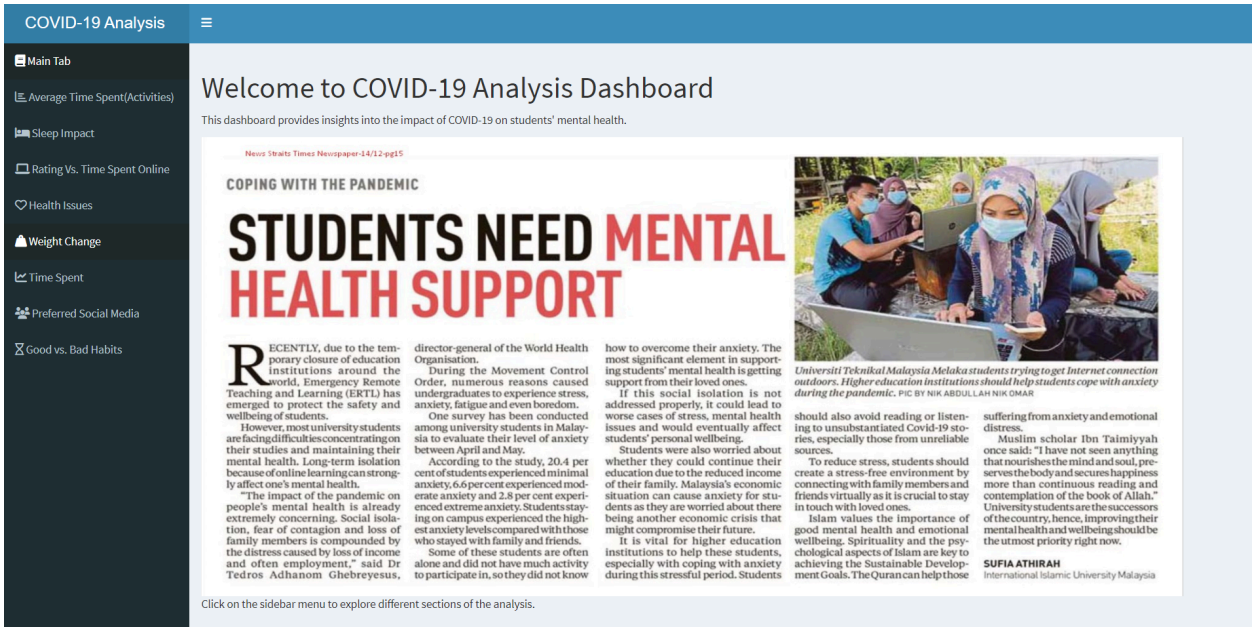
```
    theme_minimal() +
    scale_fill_brewer(palette = "Set2")

  ggplotly(p)
 })

 output$plot_health_issues <- renderPlotly({
  filtered_data <- data %>% filter(HealthIssue == "YES")

  p <- ggplot(filtered_data, aes(x = Age, fill = HealthIssue)) +
    geom_histogram(binwidth = 2, position = "dodge") +
    labs(title = "Age Distribution for Health Issues (Yes)", x = "Age", y = "Count", fill = "Health
Issue") +
    theme_minimal()

  ggplotly(p)
 })

 output$plot_weight_change <- renderPlotly({
  p <- ggplot(data, aes(x = HealthIssue, y = WeightChange, fill = HealthIssue)) +
    geom_bar(stat = "summary", fun = "mean", position = "dodge") +
    labs(title = "Average Change in Weight by Health Issues", x = "Health Issue", y = "Average
Weight Change") +
    theme_minimal() +
    scale_fill_brewer(palette = "Pastel1")

  ggplotly(p)
 })

 output$plot_scatter <- renderPlot({
```

```
   scatter_plot(data = data, activity = input$activity, min_age = input$slider[1], max_age =
input$slider[2], region = input$region)
  })

  # Reactive expression to filter data based on selected region and social media platform for
social media plot
  filtered_data_social <- reactive({
    if (input$region_social == "All Regions" & input$social_media == "All") {
      data
    } else if (input$region_social == "All Regions") {
      data %>% filter(SocialMediaPlatform == input$social_media)
    } else if (input$social_media == "All") {
      data %>% filter(Region == input$region_social)
    } else {
      data %>% filter(Region == input$region_social, SocialMediaPlatform ==
input$social_media)
    }
  })

  output$plot_social_media <- renderPlot({
    filtered_data <- filtered_data_social()
    ggplot(filtered_data, aes(x = SocialMediaPlatform, y = Age, color = SocialMediaPlatform)) +
      geom_boxplot(outlier.colour = "red", outlier.shape = 16, outlier.size = 2, lwd = 0.5) +
      labs(
        title = "Preferred Social Media by Region and Age",
        x = "Social Media Platform",
        y = "Age"
      ) +
      theme_minimal() +
      scale_color_viridis(discrete = TRUE)
  })
```

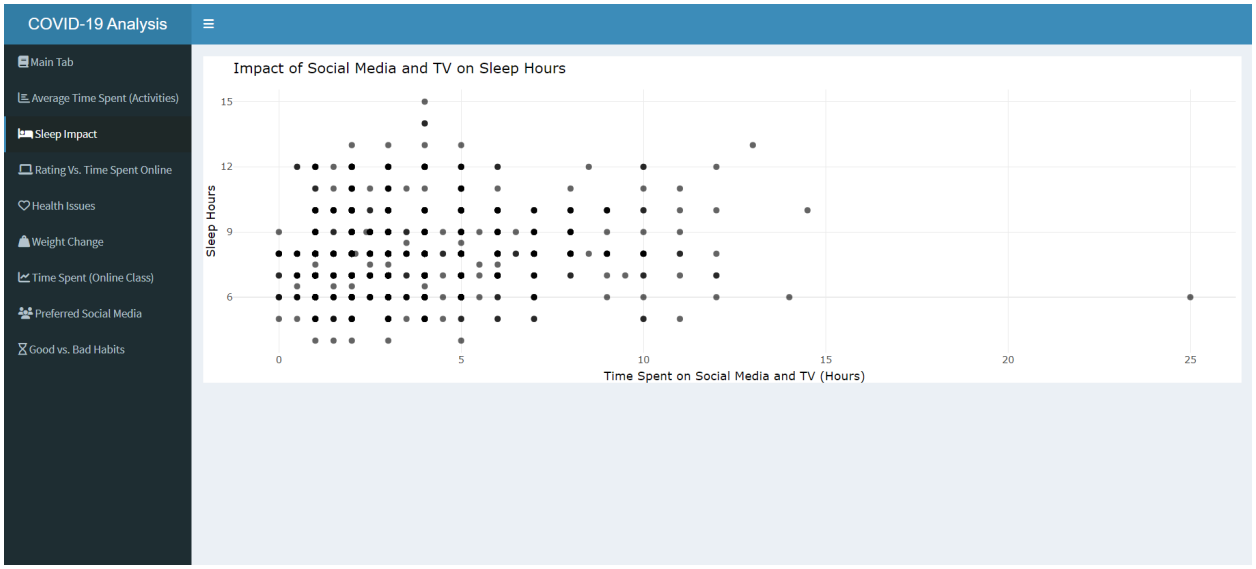
```
  output$plot_good_bad <- renderPlot({
    good_bad_data <- data %>%
      group_by(AgeRange) %>%
      summarize(
        MeanGood = mean(GoodThings, na.rm = TRUE),
        MeanBad = mean(BadThings, na.rm = TRUE)
      ) %>%
      pivot_longer(cols = c("MeanGood", "MeanBad"), names_to = "Type", values_to =
"MeanHours")

    ggplot(good_bad_data, aes(x = AgeRange, y = MeanHours, fill = Type)) +
      geom_bar(stat = "identity", position = "dodge") +
      labs(
        title = "Comparison of Good vs. Bad Habits by Age Range",
        x = "Age Range",
        y = "Average Hours",
        fill = "Habit Type"
      ) +
      theme_minimal()
  })
}

# Function to create scatter plots based on user input
scatter_plot <- function(data, activity, min_age, max_age, region) {
  if (region == "All Regions") {
    filtered_data <- data %>% filter(Age >= min_age, Age <= max_age)
  } else {
    filtered_data <- data %>% filter(Age >= min_age, Age <= max_age, Region == region)
  }
```

```
  ggplot(filtered_data, aes_string(x = "Age", y = activity, color = "Region")) +
    geom_point(alpha = 0.6) +
    labs(
      title = paste("Time Spent on", gsub("([a-z])([A-Z])", "\\1 \\2", activity), "by Age"),
      x = "Age",
      y = "Hours"
    ) +
    theme_minimal() +
    scale_color_brewer(palette = "Dark2")
}

# Run the Shiny app
shinyApp(ui = ui, server = server)
```

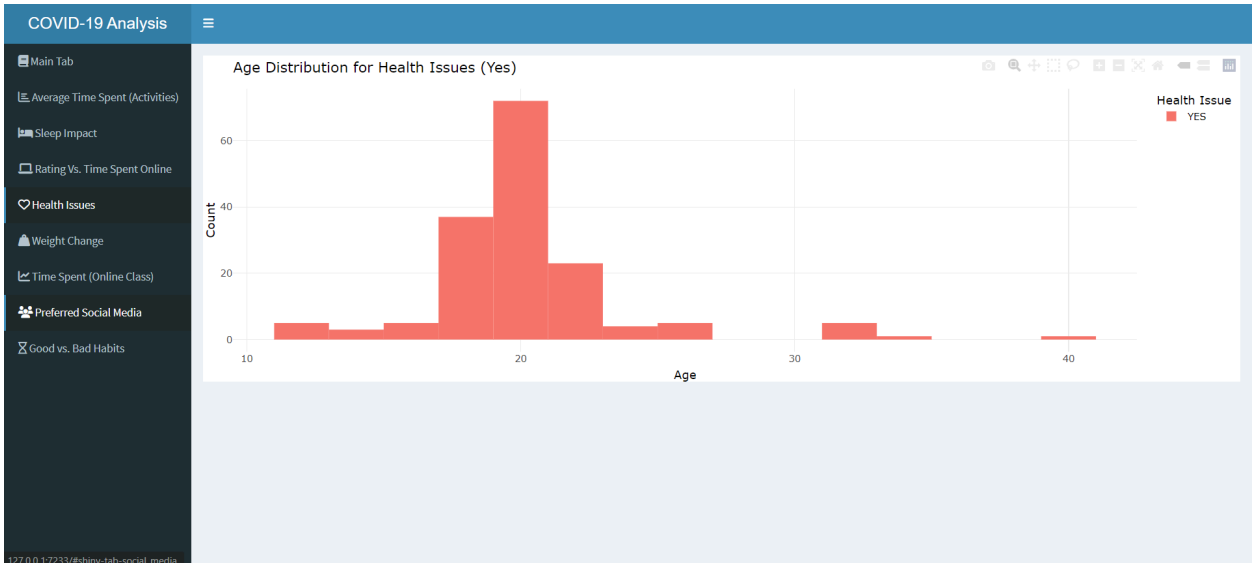
COMPLETE GUI

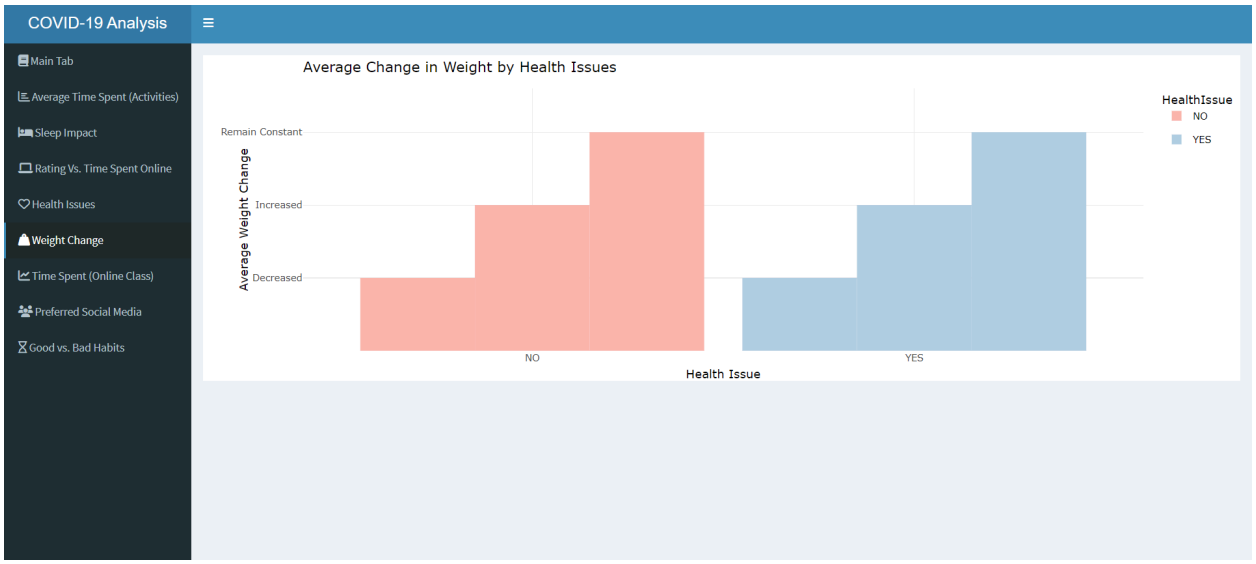


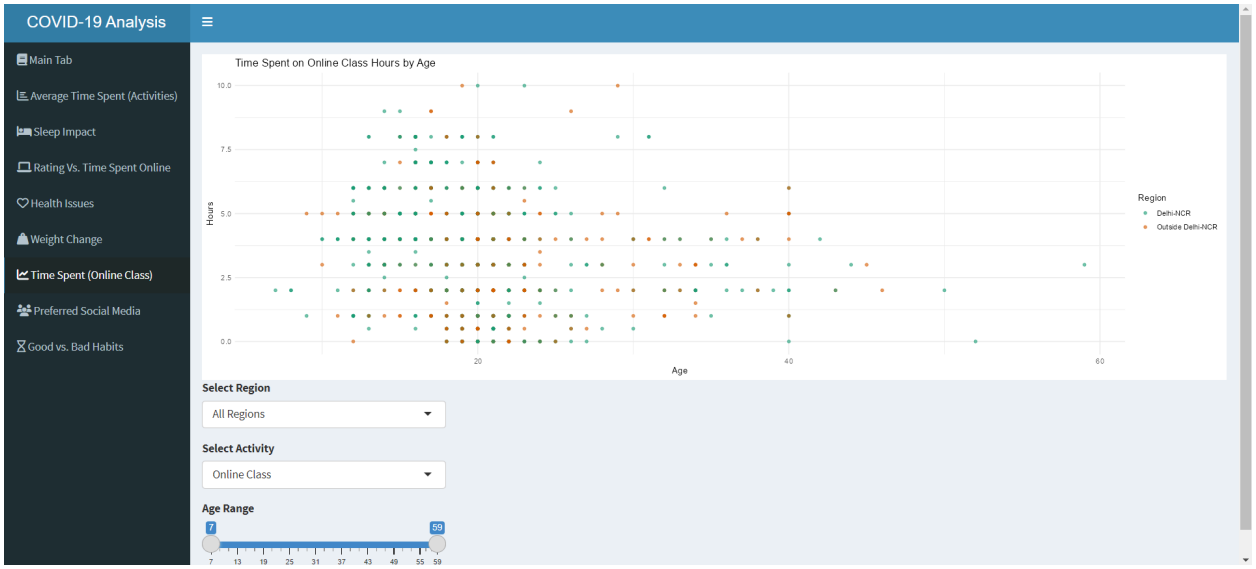

COVID-19 Analysis
Main Tab
Average Time Spent (Activities)
Sleep Impact
Rating Vs. Time Spent Online
Health Issues
Weight Change
Time Spent (Online Class)
Preferred Social Media
Good vs. Bad Habits
Average Time Spent on Different Activities and Gadgets
Average Hours
Activity
FitnessHours
OnlineClassHours
SelfStudyHours
SleepHours
SocialMediaHours
TVHours


COVID-19 Analysis
Main Tab
Average Time Spent (Activities)
Sleep Impact
Rating Vs. Time Spent Online
Health Issues
Weight Change
Time Spent (Online Class)
Preferred Social Media
Good vs. Bad Habits
Impact of Social Media and TV on Sleep Hours
Sleep Hours
Time Spent on Social Media and TV (Hours)


COVID-19 Analysis
Main Tab
Average Time Spent (Activities)
Sleep Impact
Rating Vs. Time Spent Online
Health Issues
Weight Change
Time Spent (Online Class)
Preferred Social Media
Good vs. Bad Habits
Online Class Rating vs. Time Spent Online
Time Spent Online (Hours)
10.0
7.5
5.0
2.5
0.0
Average
Excellent
Good
Poor
Very poor
Online Class Rating
factor(OnlineClassRating)
Average
Excellent
Good
Poor
Very poor


COVID-19 Analysis
Main Tab
Average Time Spent (Activities)
Sleep Impact
Rating Vs. Time Spent Online
Health Issues
Weight Change
Time Spent (Online Class)
Preferred Social Media
Good vs. Bad Habits
Age Distribution for Health Issues (Yes)
Count
60
40
20
0
10
20
30
40
Age
Health Issue
YES


COVID-19 Analysis
Main Tab
Average Time Spent (Activities)
Sleep Impact
Rating Vs. Time Spent Online
Health Issues
Weight Change
Time Spent (Online Class)
Preferred Social Media
Good vs. Bad Habits
Average Change in Weight by Health Issues
HealthIssue
NO
YES
Remain Constant
Increased
Decreased
Average Weight Change
NO
YES
Health Issue


COVID-19 Analysis
Main Tab
Average Time Spent (Activities)
Sleep Impact
Rating Vs. Time Spent Online
Health Issues
Weight Change
Time Spent (Online Class)
Preferred Social Media
Good vs. Bad Habits
Time Spent on Online Class Hours by Age
Hours
Age
Region
Delhi-NCR
Outside Delhi-NCR
Select Region
All Regions
Select Activity
Online Class
Age Range
7
59


COVID-19 Analysis
Main Tab
Average Time Spent (Activities)
Sleep Impact
Rating Vs. Time Spent Online
Health Issues
Weight Change
Time Spent (Online Class)
Preferred Social Media
Good vs. Bad Habits
Preferred Social Media by Region and Age
Age
Social Media Platform
Elyment
Facebook
Instagram
Linkedin
None
Omegle
Quora
Reddit
Snapchat
Talklife
Telegram
Twitter
Whatsapp
WhatsApp
Youtube
SocialMediaPlatform
Select Region
All Regions
Select Social Media
All


COVID-19 Analysis
Main Tab
Average Time Spent (Activities)
Sleep Impact
Rating Vs. Time Spent Online
Health Issues
Weight Change
Time Spent (Online Class)
Preferred Social Media
Good vs. Bad Habits
Comparison of Good vs. Bad Habits by Age Range
Average Hours
Age Range
5-10
11-15
16-20
21-25
26-30
31-35
36-40
41-45
46-50
51-55
56-60
Habit Type
MeanBad
MeanGood

GROUP PROJECT PLAN & APPROVAL

Note 1: Print this form and obtain your lecturer's approval before starting the project's tasks.

2. Attached this form with the final report.

SECTION NO.	02G	
GROUP NAME	PANDEMIC EXPERT	
GROUP MEMBERS (Leader's name is the first in the list)	**Id. No.**	**Name**
	SD22058	SHAMIR BIN SHAHRIN
	SD22066	SHAHIRA BINTI MOHAIDEEN MEERA
	SD22041	TANUSHALANI A/P MUNIANDY
	SD22021	AMIRAH YASMIN BINTI ZAILANEE
PROJECT TITLE	Covid-19 and Its Impact On Student Mental Health	
PROJECT DESCRIPTION	The project involves analyzing a case study on the impact of Covid-19 on students,aiming to understand the challenges they faced during the pandemic and provide insights for future strategies. The motivation lies in addressing the significant disruption caused by the pandemic in education and the importance of supporting students through evidence and interventions.	
DATA DESCRIPTION	14 variables 1000 ++ data Qualitative data - rating,medium online class,social media platform Quantitative data - time spent, self study, number of meals	
APPROVED BY (signature)	Associate Professor Dr. Roslinazairimah Zakaria	
DATE	3/4/2024	

Marking Sheet (For Lecturer's use only)

	DATA SCIENCE PROGRAMMING II (BSD2223)		**MARKS: 120 (30%)**
	GROUP LEADER: SHAMIR BIN SHAHRIN	**ID NO : SD22058**	
	GROUP PROJECT	**SECTION NO: 01G/02G/03G/04G** **DUE DATE: 19/6/2023**	/120

RUBRICS FOR CLO2/PLO2

CLO2: Analyse and summarise data using appropriate programming tools	PLO2: Cognitive Skills and Functional work skills with focus on Numeracy skills. **C4: Analysis**	/40	/10

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to **obtain appropriate data** for the project	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10
Ability to **summarise data numerically and graphically using R**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **analyse data using R and obtain data insights**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **provide conclusion** and recommendation from the project.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10

RUBRICS FOR CLO3/PLO3

CLO3: Develop programming codes to solve problems	PLO3: Functional work skills with focus on Practical, and Digital skills. **P4: Mechanism**	/40	/10

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to **construct R codes to summarise data numerically and graphically**.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	2	10
Ability to **construct R codes for data analysis**.	Unable to do the task.	Limited ability to do the task.	Able to do the task with errors.	Able to do the task with no errors but wrong answer.	Able to do the task with no errors.	Able to do the task efficiently and correctly with no errors.	2	10
Ability to **develop a dashboard (GUI) to present the data using Rshiny.**	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	4	20

RUBRICS FOR CLO4/PLO5

CLO4: Demonstrate verbal and written communication skills	PLO5:Functional work skills with focus on communication skills. **A3: Valuing**	/20	/5

	Criteria	Achievement Level						Weightage	Score
		0	1	2	3	4	5		
		Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Written	Ability to write the report findings coherently.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
	Ability to present the project report in the given format which include data description, data analysis, results and discussion.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
Communication	Ability to present the project proficiently by organizing and communicating the results in a clear, logical, and easy-to-follow manner.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5
	Ability to deliver the dashboard to summarise the project findings.	Unable to do the task.	Limited ability to do the task.	Reasonable ability to do the task.	Able to do the task with effort.	Good effort to do the task .	Able to do the task efficiently.	1	5

RUBRICS FOR CLO5/PLO8

CLO5: Relate entrepreneur skills in assigned task	PLO8:Entrepreneural skills **A4: Organising values**	/20	/5

Criteria	Achievement Level						Weightage	Score
	0	1	2	3	4	5		
	Incompetent	Inadequate	Emerging	Developing	Good	Excellent		
Ability to articulate the given/ chosen case study related to entrepreneurship.	Unable to articulate the given/ chosen case study.	Able to articulate the given/ chosen case study fairly weak.	Able to articulate the given/ chosen case study fairly well.	Able to articulate the given/ chosen case study well.	Able to articulate the given/ chosen case study reasonably well.	Able to articulate the given/ chosen case study excellently.	2	10
Ability to deliver entrepreneur ideas.	Unable to deliver any entrepreneur idea.	Delivery of idea is unclear, vague and not systematic.	Delivery of idea is less clear, vague and systematic.	Delivery of idea is moderately clear, vague and systematic.	Delivery of idea is clear and systematic.	Delivery of idea is very clear and systematic.	2	10