LOVELY PROFESSIONAL UNIVERSITY

Course: Data Exploration and Preparation

Course Code: CAP482
CA 2
Dated: - 7/April/2025

Submitted By

Name: Aditya Kumar Yadav
Roll No: 42
Reg: 12307896
Section: De225, Group: 2

Name: Deepak Kuamr Rajbhar
Roll No: 5
Reg: 12307910
Section: De225, Group: 1

Submitted to

Ms. Ranjit Kaur Walia
UID: 28632
Assistant Professor
SCA, LPU

Lovely Faculty of Technology & Sciences
School of Computer Applications
Lovely Professional University
Punjab

Title:- ***Analyzing Student Depression Using R***

Project Overview

This project studies depression in students using a dataset. Many students feel mental pressure because of studies, personal life, and social expectations. The aim is to find the main reasons behind student depression, see common patterns, and understand how depression levels change among different students. This can help in knowing the real problems students face and suggest ways to improve their mental health.

Dataset Used

The dataset has student details like age, gender, academic performance, habits, sleep, and stress levels. It was downloaded from Kaggle: Student Depression Dataset. The columns "Work Pressure" and "Job Satisfaction" have been removed since these students do not have jobs.

Objectives

1. Understand the dataset (columns, missing values, data types).
2. Do basic analysis using filtering, grouping, and summarizing.
3. Find useful patterns in student depression and stress factors.
4. Gain insights to help in mental health support for students.

Levels of Data Analysis

Level 1: Basic Exploration

Key Insights:

1. The dataset contained several missing values, which could affect the accuracy of the analysis. Removing them helped improve data reliability.

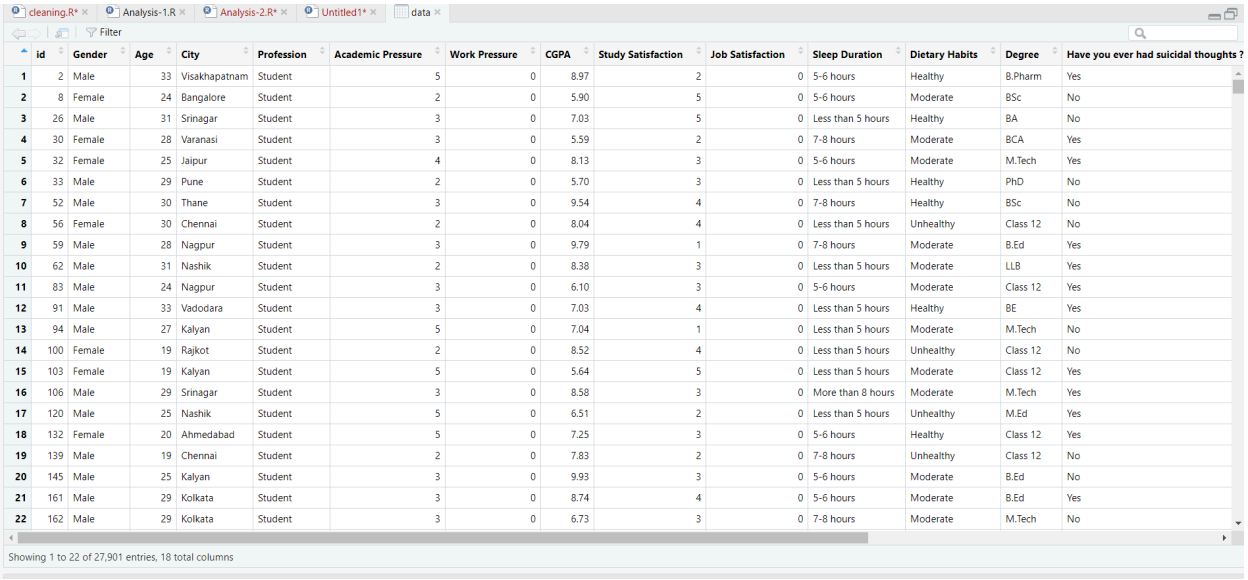
2. Basic statistical exploration provided an overview of the dataset, helping us understand which variables might be significant.

3. Calculating the percentage of students with depression gives us a strong metric that will be useful for deeper analysis and predictions.

```
# Load required libraries
library(readr)
library(dplyr)
library(tidyr)

#0: Load data set
data <- read_csv("C:/Users/Aditya Yadav/Downloads/student depression.csv")
View(data)
```

Output



cleaning.R* | Analysis-1.R | Analysis-2.R* | Untitled1* | data

Filter

	id	Gender	Age	City	Profession	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Sleep Duration	Dietary Habits	Degree	Have you ever had suicidal thoughts ?
1	2	Male	33	Visakhapatnam	Student	5	0	8.97	2	0	5-6 hours	Healthy	B.Pharm	Yes
2	8	Female	24	Bangalore	Student	2	0	5.90	5	0	5-6 hours	Moderate	BSc	No
3	26	Male	31	Srinagar	Student	3	0	7.03	5	0	Less than 5 hours	Healthy	BA	No
4	30	Female	28	Varanasi	Student	3	0	5.59	2	0	7-8 hours	Moderate	BCA	Yes
5	32	Female	25	Jaipur	Student	4	0	8.13	3	0	5-6 hours	Moderate	M.Tech	Yes
6	33	Male	29	Pune	Student	2	0	5.70	3	0	Less than 5 hours	Healthy	PhD	No
7	52	Male	30	Thane	Student	3	0	9.54	4	0	7-8 hours	Healthy	BSc	No
8	56	Female	30	Chennai	Student	2	0	8.04	4	0	Less than 5 hours	Unhealthy	Class 12	No
9	59	Male	28	Nagpur	Student	3	0	9.79	1	0	7-8 hours	Moderate	B.Ed	Yes
10	62	Male	31	Nashik	Student	2	0	8.38	3	0	Less than 5 hours	Moderate	LLB	Yes
11	83	Male	24	Nagpur	Student	3	0	6.10	3	0	5-6 hours	Moderate	Class 12	Yes
12	91	Male	33	Vadodara	Student	3	0	7.03	4	0	Less than 5 hours	Healthy	BE	Yes
13	94	Male	27	Kalyan	Student	5	0	7.04	1	0	Less than 5 hours	Moderate	M.Tech	No
14	100	Female	19	Rajkot	Student	2	0	8.52	4	0	Less than 5 hours	Unhealthy	Class 12	No
15	103	Female	19	Kalyan	Student	5	0	5.64	5	0	Less than 5 hours	Moderate	Class 12	Yes
16	106	Male	29	Srinagar	Student	3	0	8.58	3	0	More than 8 hours	Moderate	M.Tech	Yes
17	120	Male	25	Nashik	Student	5	0	6.51	2	0	Less than 5 hours	Unhealthy	M.Ed	Yes
18	132	Female	20	Ahmedabad	Student	5	0	7.25	3	0	5-6 hours	Healthy	Class 12	Yes
19	139	Male	19	Chennai	Student	2	0	7.83	2	0	7-8 hours	Unhealthy	Class 12	No
20	145	Male	25	Kalyan	Student	3	0	9.93	3	0	5-6 hours	Moderate	B.Ed	No
21	161	Male	29	Kolkata	Student	3	0	8.74	4	0	5-6 hours	Moderate	B.Ed	Yes
22	162	Male	29	Kolkata	Student	3	0	6.73	3	0	7-8 hours	Moderate	M.Tech	No

Showing 1 to 22 of 27,901 entries, 18 total columns

1. What are the column names and data types?

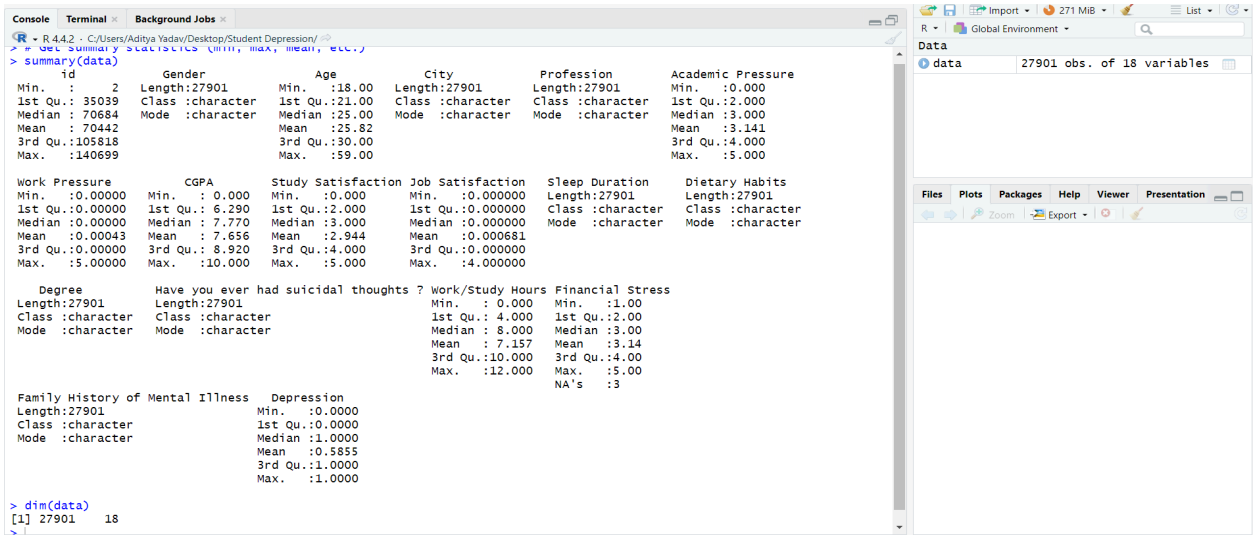
#1: Data understanding

```
# Check structure of dataset (data types of each column)
str(data)

# Get summary statistics (min, max, mean, etc.)
summary(data)

# Get data set dimensions (total rows and columns)
dim(data)
```

Output



```
> summary(data)
       id            Gender               Age             City           Profession       Academic Pressure
 Min.   :     2   Length:27901       Min.   :18.00   Length:27901       Length:27901       Min.   :0.000
 1st Qu.: 35039   Class :character   1st Qu.:21.00   Class :character   Class :character   1st Qu.:2.000
 Median : 70684   Mode  :character   Median :25.00   Mode  :character   Mode  :character   Median :3.000
 Mean   : 70442                      Mean   :25.82                                         Mean   :3.141
 3rd Qu.:105818                      3rd Qu.:30.00                                         3rd Qu.:4.000
 Max.   :140699                      Max.   :59.00                                         Max.   :5.000

 Work Pressure          CGPA        Study Satisfaction Job Satisfaction   Sleep Duration     Dietary Habits
 Min.   :0.00000   Min.   : 0.000   Min.   :0.000      Min.   :0.000000   Length:27901       Length:27901
 1st Qu.:0.00000   1st Qu.: 6.290   1st Qu.:2.000      1st Qu.:0.000000   Class :character   Class :character
 Median :0.00000   Median : 7.770   Median :3.000      Median :0.000000   Mode  :character   Mode  :character
 Mean   :0.00043   Mean   : 7.656   Mean   :2.944      Mean   :0.000681
 3rd Qu.:0.00000   3rd Qu.: 8.920   3rd Qu.:4.000      3rd Qu.:0.000000
 Max.   :5.00000   Max.   :10.000   Max.   :5.000      Max.   :4.000000

    Degree          Have you ever had suicidal thoughts ? Work/Study Hours Financial Stress
 Length:27901       Length:27901                          Min.   : 0.000   Min.   :1.00
 Class :character   Class :character                      1st Qu.: 4.000   1st Qu.:2.00
 Mode  :character   Mode  :character                      Median : 8.000   Median :3.00
                                                         Mean   : 7.157   Mean   :3.14
                                                         3rd Qu.:10.000   3rd Qu.:4.00
                                                         Max.   :12.000   Max.   :5.00
                                                                          NA's   :3

 Family History of Mental Illness   Depression
 Length:27901                       Min.   :0.0000
 Class :character                   1st Qu.:0.0000
 Mode  :character                   Median :1.0000
                                    Mean   :0.5855
                                    3rd Qu.:1.0000
                                    Max.   :1.0000

> dim(data)
[1] 27901    18
```

Interpretation of Data Understanding :

1. **Dataset Size:** The dataset contains 27,901 rows and 18 columns, meaning we have data from 27,901 students with 18 features related to their academic and mental health conditions.

2. **Column Types:**
   - Numerical Data: Includes Age, Academic Pressure, CGPA, Study Satisfaction, Work/Study Hours, Financial Stress, and Depression (where 1 = Depressed, 0 = Not Depressed).
   - Categorical Data: Includes Gender, City, Profession, Sleep Duration, Dietary Habits, Degree, and Family History of Mental Illness.

3. **Columns to Remove:Work Pressure and Job Satisfaction:** These columns are irrelevant since students generally don't have jobs.

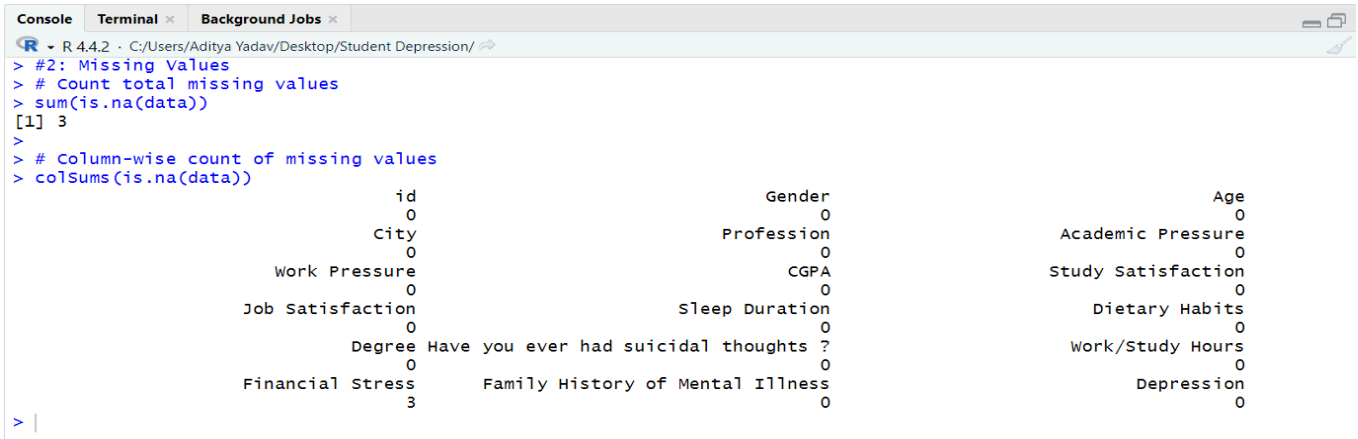
2. Are there any missing values?

#2: Missing Values

```
# Count total missing values
sum(is.na(data))

# Column-wise count of missing values
colSums(is.na(data))
```

Output

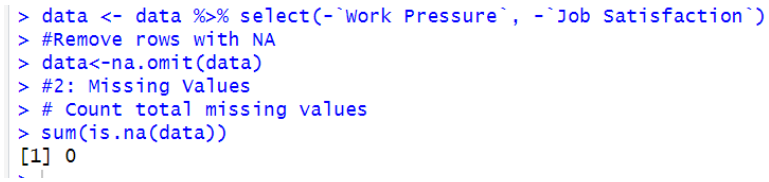


3. Are there any unnecessary columns ?

#3: Clean Data

```
# Remove unnecessary columns
data <- data %>% select(-`Work Pressure`, -`Job Satisfaction`)
#Remove rows with NA
data<-na.omit(data)
```

Output



Interpretation for Missing Values and Data Cleaning:

- Only 3 missing values found in the Financial Stress column.

Data Cleaning Steps Taken:

- Removed "Work Pressure" & "Job Satisfaction" as they are irrelevant for student depression analysis.
- Deleted rows with missing values (NA) to ensure clean data.

Final Dataset:

- No missing values.
- Refined for further analysis .

4. What is the percentage of students with depression?

```
# 4: Calculate the percentage of students with depression
percentage_depressed <- mean(data$Depression) * 100
print(paste("Percentage of students with depression:", round(percentage_depressed, 2), "%"))
```

Output

```
> percentage_depressed <- mean(data$Depression) * 100
> print(paste("Percentage of students with depression:", round(percentage_depressed, 2), "%"))
[1] "Percentage of students with depression: 58.55 %"
```

Interpretation of Depression Percentage

1. **High Depression Rate:**
   - 58.55% of students in the dataset show signs of depression, indicating a major concern.
2. **Significance:**
   - More than half of the students are struggling with depression, which could be linked to factors like academic pressure, financial stress, or personal issues.
3. **Next Steps:**
   - Further analysis is needed to identify key factors contributing to this high depression rate.
   - Possible interventions can be explored for mental health support.

Level 2: Identifying Patterns

Key Insights from Level 2:

1. **Academic Pressure vs. Financial Stress:**
   - One of these factors plays a bigger role in student depression.
   - Solutions can be provided based on which factor is dominant.

2. **Depressed Students List:**
   - Helps analyze common traits in depressed students.
   - Can be used for targeted support programs.

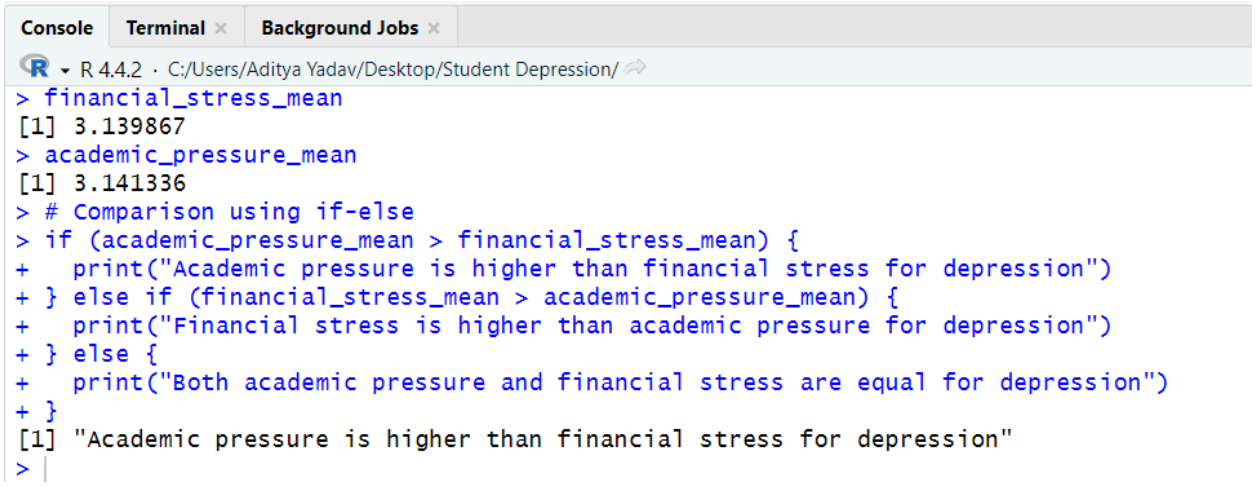
3. **Low CGPA & Depression Connection:**
   - If many students with low CGPA are depressed, universities should address academic-related stress and provide additional support.

5. Find the most common stress factors among students.

```
# Mean calculation with NA handling
financial_stress_mean <- mean(data$`Financial Stress`, na.rm = TRUE)
academic_pressure_mean <- mean(data$`Academic Pressure`, na.rm = TRUE)

# Comparison using if-else
if (academic_pressure_mean > financial_stress_mean) {
  print("Academic pressure is higher than financial stress for depression")
} else if (financial_stress_mean > academic_pressure_mean) {
  print("Financial stress is higher than academic pressure for depression")
} else {
  print("Both academic pressure and financial stress are equal for depression")
}
```

Output



Interpretation :

- Academic pressure is the primary stressor among students, slightly exceeding financial stress in its contribution to depression.

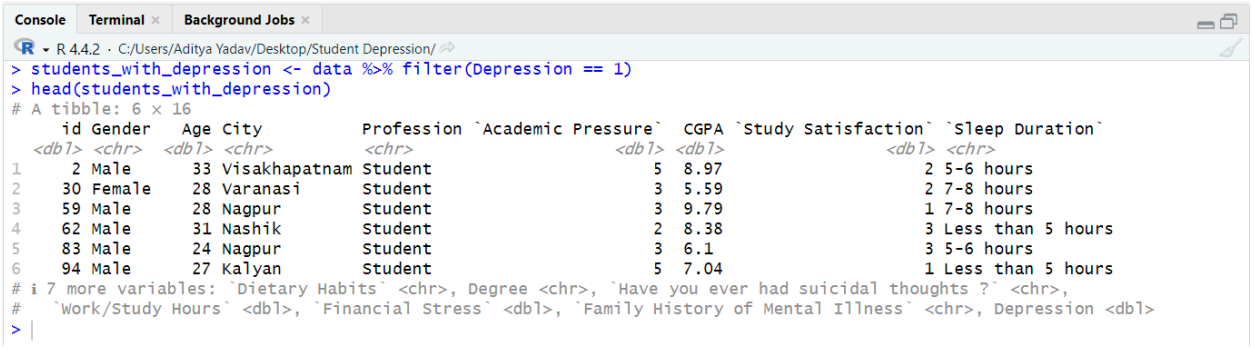
- The small difference (3.141 vs. 3.139) suggests that both factors are almost equally significant, indicating a need for balanced mental health support addressing both academic and financial concerns.

- Universities and institutions should focus on reducing academic pressure while also offering financial aid and counseling services to help students manage stress effectively.

6. List students who have depression (Depression = 1).

```
students_with_depression <- data %>% filter(Depression == 1)
head(students_with_depression)
```

Output



```
Console   Terminal   Background Jobs
R 4.4.2 · C:/Users/Aditya Yadav/Desktop/Student Depression/
> students_with_depression <- data %>% filter(Depression == 1)
> head(students_with_depression)
# A tibble: 6 × 16
     id Gender   Age City          Profession `Academic Pressure`  CGPA `Study Satisfaction` `Sleep Duration`
  <dbl> <chr>  <dbl> <chr>         <chr>                    <dbl> <dbl>                <dbl> <chr>
1     2 Male      33 Visakhapatnam Student                      5  8.97                    2 5-6 hours
2    30 Female    28 Varanasi      Student                      3  5.59                    2 7-8 hours
3    59 Male      28 Nagpur        Student                      3  9.79                    1 7-8 hours
4    62 Male      31 Nashik        Student                      2  8.38                    3 Less than 5 hours
5    83 Male      24 Nagpur        Student                      3  6.1                     3 5-6 hours
6    94 Male      27 Kalyan        Student                      5  7.04                    1 Less than 5 hours
# i 7 more variables: `Dietary Habits` <chr>, Degree <chr>, `Have you ever had suicidal thoughts ?` <chr>,
#   `Work/Study Hours` <dbl>, `Financial Stress` <dbl>, `Family History of Mental Illness` <chr>, Depression <dbl>
>
```

Interpretation

- The dataset contains students diagnosed with depression, showing **varied academic pressure, CGPA, and study satisfaction levels**.

- **Sleep duration is inconsistent**, with some students getting less than 5 hours of sleep, which may contribute to mental health issues.

7. Identify students with low academic performance and depression.

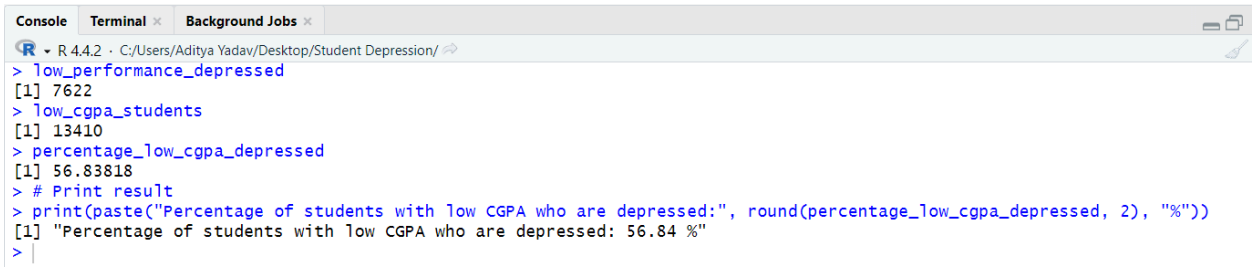
```
# Total depressed students with CGPA below average
low_performance_depressed <- data %>%
  filter(CGPA < mean(CGPA) & Depression == 1)%>%
  nrow()

# Total students with CGPA below average
low_cgpa_students <- data %>%
  filter(CGPA < mean(CGPA)) %>%
  nrow()

# Calculate percentage
percentage_low_cgpa_depressed <- (low_performance_depressed / low_cgpa_students) * 100

# Print result
print(paste("Percentage of students with low CGPA who are depressed:",
round(percentage_low_cgpa_depressed, 2), "%"))
```

Output



Interpretation

- 56.84% of students with below-average CGPA are depressed, indicating a strong link between academic performance and mental health.
- Students struggling with grades may experience higher academic pressure, stress, or self-doubt, leading to depression.
- This insight suggests the need for mental health support, counseling, or academic assistance for students with lower CGPAs.

Level 3: Grouping & Summarization

Key Insights from Level 3:

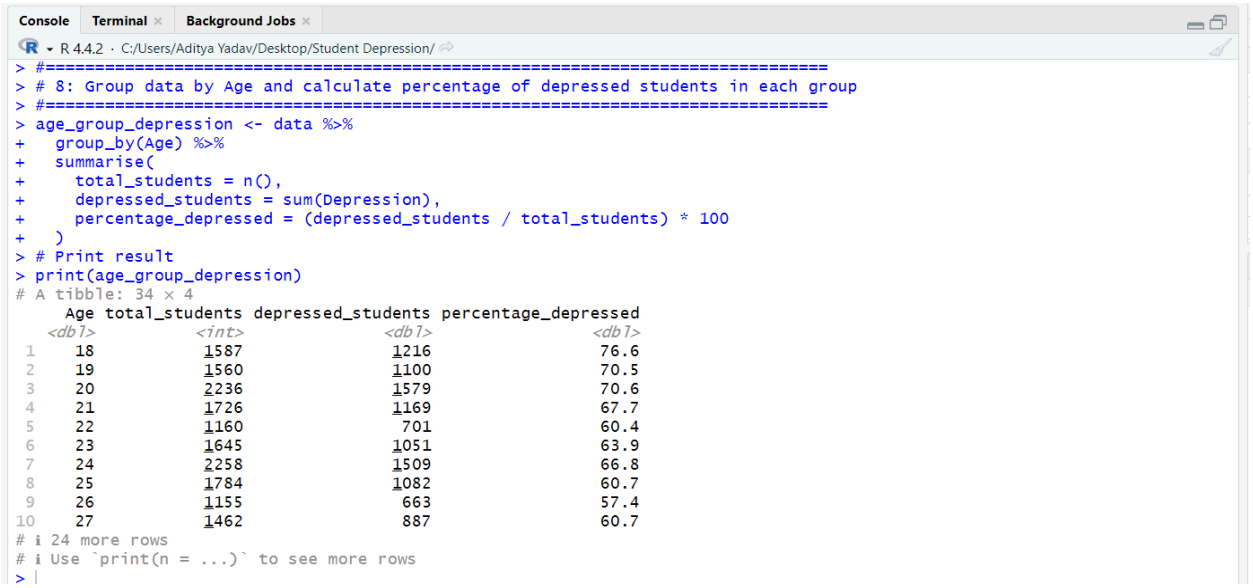
1. **Depression by Age Group:**
   Older students may experience more depression due to increasing pressure.
2. **CGPA & Depression relation:**
   Students with **low CGPA (0-4) are most affected by depression**.
   Academic stress plays a big role in students' mental health.

8. Find the percentage of students with depression by age group.

```
# 8: Group data by Age and calculate percentage of depressed students in each group
age_group_depression <- data %>%
  group_by(Age) %>%
  summarise(
    total_students = n(),
    depressed_students = sum(Depression),
    percentage_depressed = (depressed_students / total_students) * 100
  )

# Print result
print(age_group_depression)
```

Output



```
> #============================================================================
> # 8: Group data by Age and calculate percentage of depressed students in each group
> #============================================================================
> age_group_depression <- data %>%
+   group_by(Age) %>%
+   summarise(
+     total_students = n(),
+     depressed_students = sum(Depression),
+     percentage_depressed = (depressed_students / total_students) * 100
+   )
> # Print result
> print(age_group_depression)
# A tibble: 34 × 4
     Age total_students depressed_students percentage_depressed
   <dbl>          <int>              <dbl>                <dbl>
 1    18           1587               1216                 76.6
 2    19           1560               1100                 70.5
 3    20           2236               1579                 70.6
 4    21           1726               1169                 67.7
 5    22           1160                701                 60.4
 6    23           1645               1051                 63.9
 7    24           2258               1509                 66.8
 8    25           1784               1082                 60.7
 9    26           1155                663                 57.4
10    27           1462                887                 60.7
# i 24 more rows
# i Use `print(n = ...)` to see more rows
>
```

Interpretation :

- Teenagers (18-20 years) have the highest depression rates (above 70%), likely due to academic pressure, career uncertainty, and social stress.
- Depression decreases slightly with age, with students aged 26+ having lower depression rates (~57-60%), possibly due to better coping mechanisms or life stability.
- These insights highlight the need for early mental health support, stress management programs, and counseling for younger students.

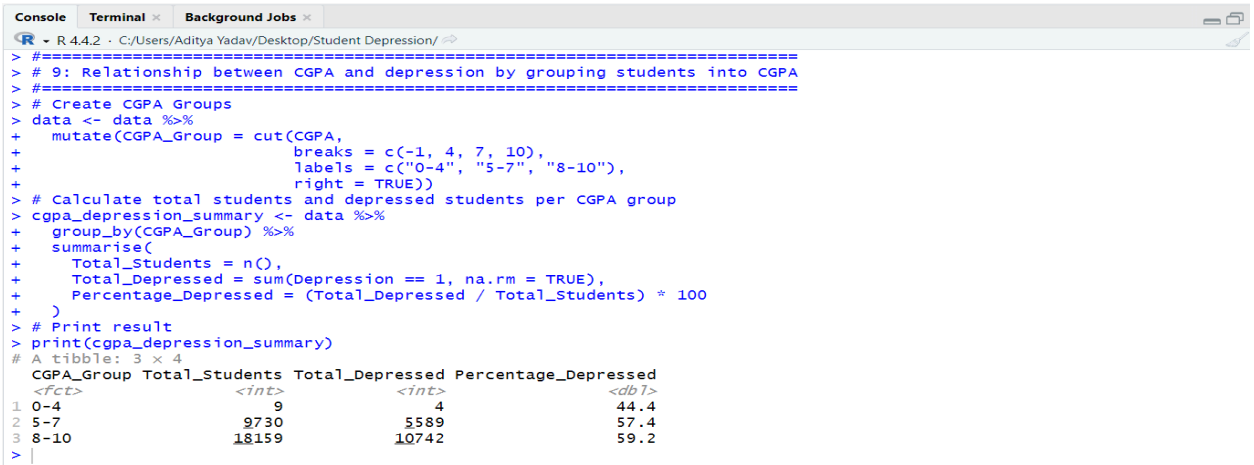
9. Analyze the relationship between CGPA and depression by grouping students into CGPA categories (0-4, 5-7, 8-10).

```
# Create CGPA Groups
  data <- data %>%
    mutate(CGPA_Group = cut(CGPA,
                            breaks = c(-1, 4, 7, 10),
                            labels = c("0-4", "5-7", "8-10"),
                            right = TRUE))

  # Calculate total students and depressed students per CGPA group
  cgpa_depression_summary <- data %>%
    group_by(CGPA_Group) %>%
    summarise(
      Total_Students = n(),
      Total_Depressed = sum(Depression == 1, na.rm = TRUE),
      Percentage_Depressed = (Total_Depressed / Total_Students) * 100
    )

  # Print result
  print(cgpa_depression_summary)
```

Output



```
> #=============================================================================
> # 9: Relationship between CGPA and depression by grouping students into CGPA
> #=============================================================================
> # Create CGPA Groups
> data <- data %>%
+   mutate(CGPA_Group = cut(CGPA,
+                           breaks = c(-1, 4, 7, 10),
+                           labels = c("0-4", "5-7", "8-10"),
+                           right = TRUE))
> # Calculate total students and depressed students per CGPA group
> cgpa_depression_summary <- data %>%
+   group_by(CGPA_Group) %>%
+   summarise(
+     Total_Students = n(),
+     Total_Depressed = sum(Depression == 1, na.rm = TRUE),
+     Percentage_Depressed = (Total_Depressed / Total_Students) * 100
+   )
> # Print result
> print(cgpa_depression_summary)
# A tibble: 3 × 4
  CGPA_Group Total_Students Total_Depressed Percentage_Depressed
  <fct>               <int>           <int>                <dbl>
1 0-4                     9               4                 44.4
2 5-7                  9730            5589                 57.4
3 8-10                18159           10742                 59.2
>
```

Interpretation:

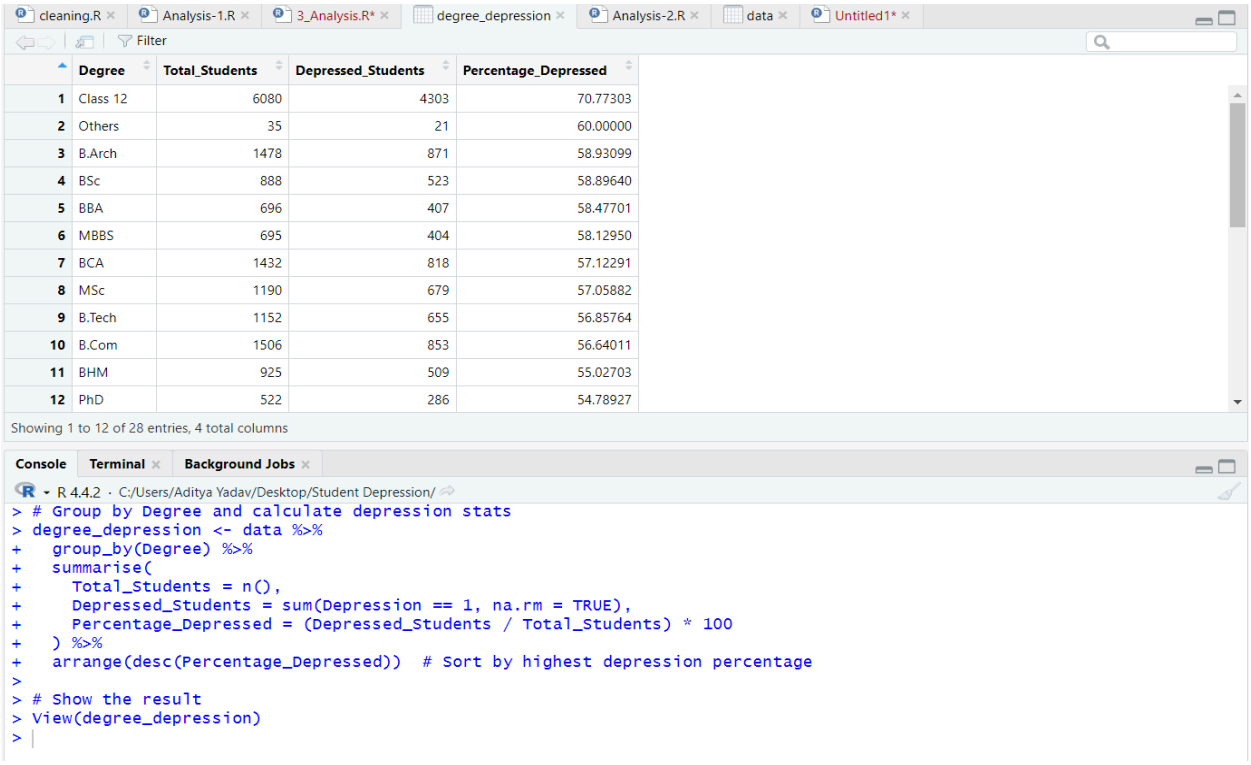
- **Low CGPA (0–4)**: 44.4% of students are depressed — relatively high, but this group has very few students (only 9), so the result might not be fully reliable.
- **Average CGPA (5–7)**: 57.4% of students are depressed — shows that a majority of average performers are experiencing depression.
- **High CGPA (8–10)**: 59.2% of students are depressed — surprisingly, even top performers have high depression levels, possibly due to pressure to maintain performance.

10. Which degree program has the highest number and percentage of depressed students?

```
# Group by Degree and calculate depression stats
degree_depression <- data %>%
  group_by(Degree) %>%
  summarise(
    Total_Students = n(),
    Depressed_Students = sum(Depression == 1, na.rm = TRUE),
    Percentage_Depressed = (Depressed_Students / Total_Students) * 100
  ) %>%
  arrange(desc(Percentage_Depressed))  # Sort by highest depression percentage

# Show the result
View(degree_depression)
```

Output



cleaning.R | Analysis-1.R | 3_Analysis.R* | degree_depression | Analysis-2.R | data | Untitled1*

Filter

	Degree	Total_Students	Depressed_Students	Percentage_Depressed
1	Class 12	6080	4303	70.77303
2	Others	35	21	60.00000
3	B.Arch	1478	871	58.93099
4	BSc	888	523	58.89640
5	BBA	696	407	58.47701
6	MBBS	695	404	58.12950
7	BCA	1432	818	57.12291
8	MSc	1190	679	57.05882
9	B.Tech	1152	655	56.85764
10	B.Com	1506	853	56.64011
11	BHM	925	509	55.02703
12	PhD	522	286	54.78927

Showing 1 to 12 of 28 entries, 4 total columns

Console | Terminal | Background Jobs

R 4.4.2 · C:/Users/Aditya Yadav/Desktop/Student Depression/

```
> # Group by Degree and calculate depression stats
> degree_depression <- data %>%
+   group_by(Degree) %>%
+   summarise(
+     Total_Students = n(),
+     Depressed_Students = sum(Depression == 1, na.rm = TRUE),
+     Percentage_Depressed = (Depressed_Students / Total_Students) * 100
+   ) %>%
+   arrange(desc(Percentage_Depressed))  # Sort by highest depression percentage
>
> # Show the result
> View(degree_depression)
>
```

Interpretation :

- **Class 12** students have the highest depression rate at **70.77%**, indicating that pre-university pressure and uncertainty may be a major factor.
- Other degrees like **B**.**Arch**, **BSc**, and **BBA** also show high depression rates (around **58–59%),** possibly due to academic pressure or career concerns.
- Advanced degrees like **M.Tech, M.Ed,** and **MHM** show relatively lower depression rates (around **50–52%**), suggesting that students at this stage may be better at handling academic stress or have more clarity in their goals.

Level 4: Ranking & Comparison

11. Rank students by academic performance and depression levels.

```
# 11: Rank students based on CGPA and Depression levels
ranked_data <- data %>%
  # Rank based on CGPA (higher CGPA = better rank)
  mutate(CGPA_Rank = dense_rank(desc(CGPA))) %>%
  # Arrange by Depression (1 first) and CGPA_Rank
  arrange(desc(Depression), CGPA_Rank) %>%
  # Assign final ranking
  mutate(Final_Rank = row_number())

# Select relevant columns
ranked_students <- ranked_data %>%
  select(id, CGPA, CGPA_Rank, Depression, Final_Rank)

# Print top 10 ranked students
print(head(ranked_students, 10))
```

Output:

```
Console   Terminal ×   Background Jobs ×
R 4.4.2 · C:/Users/Aditya Yadav/Desktop/Student Depression/
> ranked_data <- data %>%
+   # Rank based on CGPA (higher CGPA = better rank)
+   mutate(CGPA_Rank = dense_rank(desc(CGPA))) %>%
+   # Arrange by Depression (1 first) and CGPA_Rank
+   arrange(desc(Depression), CGPA_Rank) %>%
+   # Assign final ranking
+   mutate(Final_Rank = row_number())
> # Select relevant columns
> ranked_students <- ranked_data %>%
+   select(id, CGPA, CGPA_Rank, Depression, Final_Rank)
> # Print top 10 ranked students
> print(head(ranked_students, 10))
# A tibble: 10 × 5
      id  CGPA CGPA_Rank Depression Final_Rank
   <dbl> <dbl>     <int>      <dbl>      <int>
 1 13170    10         1          1          1
 2 15800    10         1          1          2
 3 22499    10         1          1          3
 4 24975    10         1          1          4
 5 25353    10         1          1          5
 6 25482    10         1          1          6
 7 26892    10         1          1          7
 8 32697    10         1          1          8
 9 34831    10         1          1          9
10 38077    10         1          1         10
>
```

Interpretation:

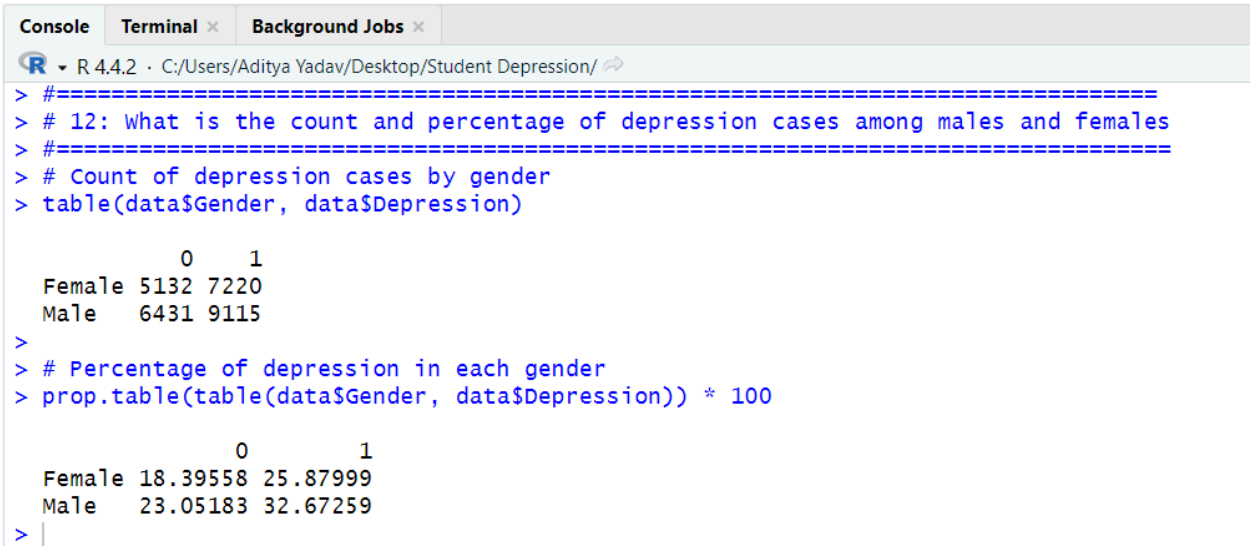
- Top-ranked students (CGPA = 10) are depressed, proving that academic excellence ≠ mental well-being.
- CGPA alone does not determine depression; other factors like pressure and stress contribute.
- High achievers might face extreme expectations, leading to mental health struggles.
- Universities should focus on student well-being, not just grades.

12. What is the count and percentage of depression cases among males and females?

```
# Count of depression cases by gender
table(data$Gender, data$Depression)

# Percentage of depression in each gender
prop.table(table(data$Gender, data$Depression)) * 100
```

Output



Interpretation:

- **Total Cases:** More males (9115) suffer from depression compared to females (7220).
- **Percentage:**
  - **Females:** 25.88% have depression, while 18.39% do not.
  - **Males:** 32.67% have depression, while 23.05% do not.
- **Insight:** Males have a higher percentage of depression cases than females in this dataset.

13. How many students from each dietary habit category (Healthy, Moderate, Unhealthy) have depression, and what is their percentage?

```
# 13: Count and percentage of depressed students by dietary habit
dietary_depression <- data %>%
  group_by(`Dietary Habits`, Depression) %>%
  summarise(count = n(), .groups = "drop") %>%
  mutate(Percentage = (count / sum(count)) * 100)

# Print result
print(dietary_depression)
```

Output

```
> print(dietary_depression)
# A tibble: 8 × 4
  `Dietary Habits` Depression count Percentage
  <chr>                 <dbl> <int>      <dbl>
1 Healthy                   0  4177     15.0
2 Healthy                   1  3472     12.4
3 Moderate                  0  4363     15.6
4 Moderate                  1  5558     19.9
5 Others                    0     4      0.0143
6 Others                    1     8      0.0287
7 Unhealthy                 0  3019     10.8
8 Unhealthy                 1  7297     26.2
>
```

Interpretation:

- **Unhealthy diet**: Highest depression cases (**26.2%**).
- **Moderate diet**: Significant depression cases (**19.9%**).
- **Healthy diet**: Lowest depression cases (**12.4%**).
- **Key insight**: Poor dietary habits may be linked to higher depression risks.
- **Suggestion**: Encouraging a healthy diet could help improve mental well-being.

14. How many students who are depressed have also had suicidal thoughts?

```
suicidal_depressed <- data %>%
  filter(Depression == 1, `Have you ever had suicidal thoughts ?` == "Yes") %>%
  summarise(count = n())

print(suicidal_depressed)
```

Output

```
Console  Terminal  Background Jobs
R 4.4.2 · C:/Users/Aditya Yadav/Desktop/Student Depression/
> suicidal_depressed <- data %>%
+   filter(Depression == 1, `Have you ever had suicidal thoughts ?` == "Yes") %>%
+   summarise(count = n())
> print(suicidal_depressed)
# A tibble: 1 × 1
  count
  <int>
1 13957
>
```

Interpretation:

- **Total 13,957 students** who are **depressed** have also reported having **suicidal thoughts**.
- This suggests a **strong link between depression and suicidal thoughts** among students.
- Identifying and **supporting these students** through mental health initiatives is crucial.

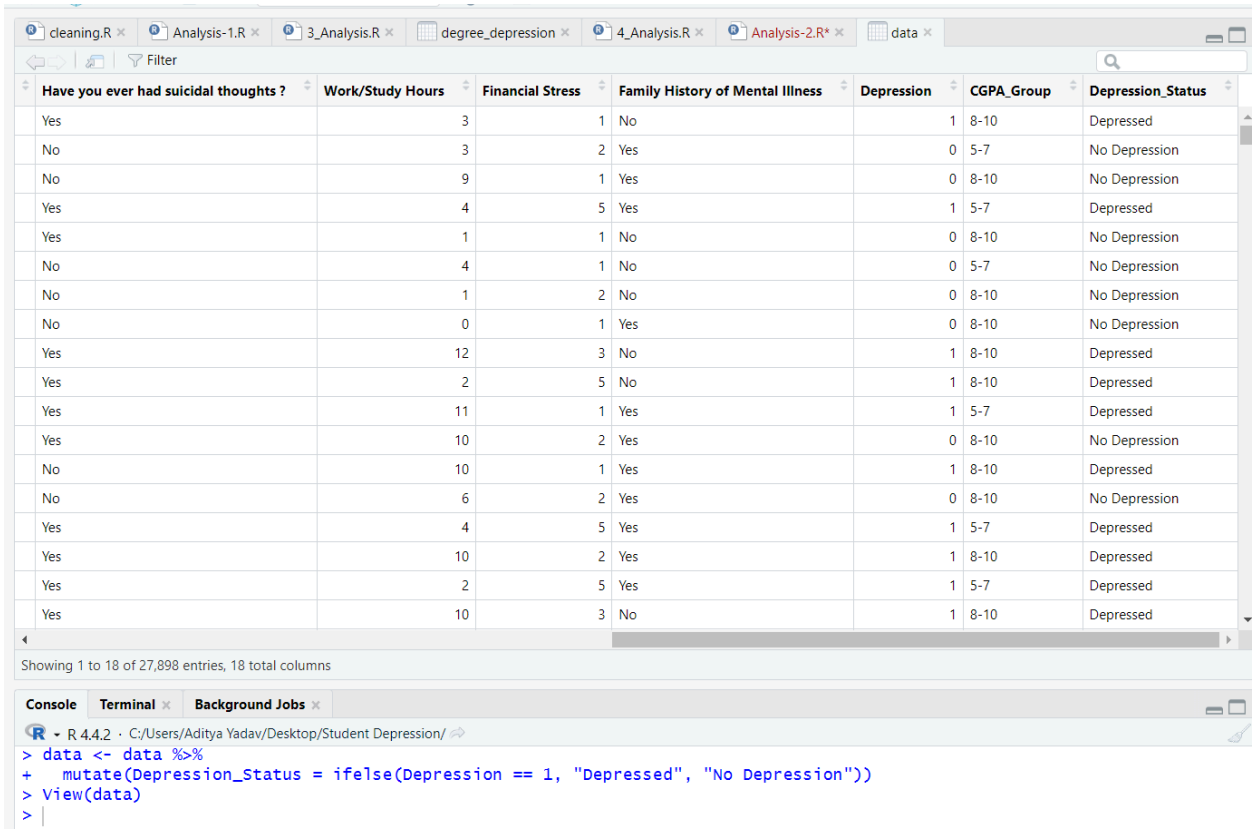
Level 5: Creating New Insights

15. Add a new column "Depression_Status" based on the depression column.

- "No Depression" (0)
- "Depressed" (1)

```
data <- data %>%
mutate(Depression_Status = ifelse(Depression == 1, "Depressed", "No Depression"))
head(data)
```

Output



cleaning.R | Analysis-1.R | 3_Analysis.R | degree_depression | 4_Analysis.R | Analysis-2.R* | data

Filter

Have you ever had suicidal thoughts ?	Work/Study Hours	Financial Stress	Family History of Mental Illness	Depression	CGPA_Group	Depression_Status
Yes	3	1	No	1	8-10	Depressed
No	3	2	Yes	0	5-7	No Depression
No	9	1	Yes	0	8-10	No Depression
Yes	4	5	Yes	1	5-7	Depressed
Yes	1	1	No	0	8-10	No Depression
No	4	1	No	0	5-7	No Depression
No	1	2	No	0	8-10	No Depression
No	0	1	Yes	0	8-10	No Depression
Yes	12	3	No	1	8-10	Depressed
Yes	2	5	No	1	8-10	Depressed
Yes	11	1	Yes	1	5-7	Depressed
Yes	10	2	Yes	0	8-10	No Depression
No	10	1	Yes	1	8-10	Depressed
No	6	2	Yes	0	8-10	No Depression
Yes	4	5	Yes	1	5-7	Depressed
Yes	10	2	Yes	1	8-10	Depressed
Yes	2	5	Yes	1	5-7	Depressed
Yes	10	3	No	1	8-10	Depressed

Showing 1 to 18 of 27,898 entries, 18 total columns

Console | Terminal | Background Jobs

R 4.4.2 · C:/Users/Aditya Yadav/Desktop/Student Depression/

```
> data <- data %>%
+   mutate(Depression_Status = ifelse(Depression == 1, "Depressed", "No Depression"))
> View(data)
>
```

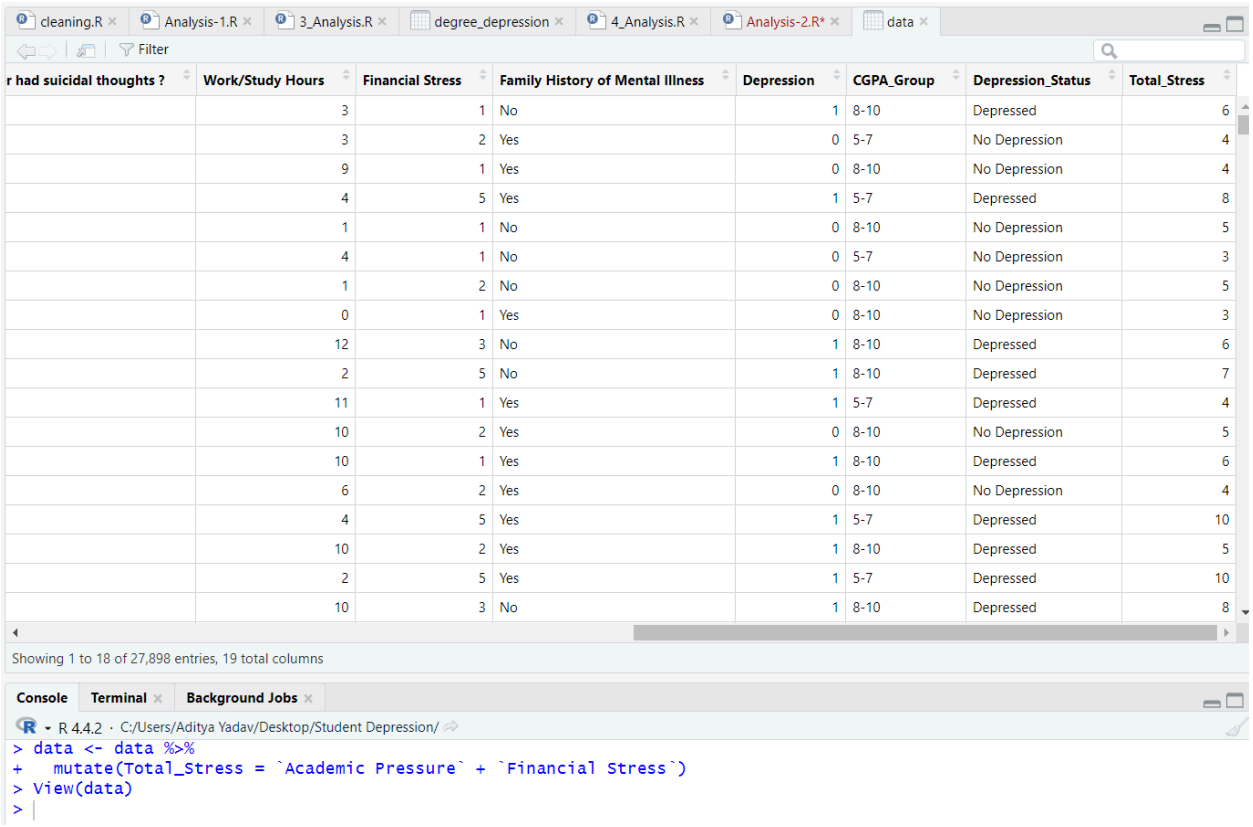
Interpretation:

- A new column named Depression_Status has been added to the dataset.
- It converts the numeric values in the Depression column (0 or 1) into more readable labels:
  - If Depression == 1, it shows "Depressed"
  - If Depression == 0, it shows "No Depression"
- This makes the data easier to understand and analyze, especially for reporting or visualization purposes.

16. What is the total stress level of each student by combining academic and financial stress?

```
data <- data %>%
  mutate(Total_Stress = `Academic Pressure` + `Financial Stress`)
View(data)
```

Output



cleaning.R | Analysis-1.R | 3_Analysis.R | degree_depression | 4_Analysis.R | Analysis-2.R* | data

Filter

r had suicidal thoughts ?	Work/Study Hours	Financial Stress	Family History of Mental Illness	Depression	CGPA_Group	Depression_Status	Total_Stress
	3	1	No	1	8-10	Depressed	6
	3	2	Yes	0	5-7	No Depression	4
	9	1	Yes	0	8-10	No Depression	4
	4	5	Yes	1	5-7	Depressed	8
	1	1	No	0	8-10	No Depression	5
	4	1	No	0	5-7	No Depression	3
	1	2	No	0	8-10	No Depression	5
	0	1	Yes	0	8-10	No Depression	3
	12	3	No	1	8-10	Depressed	6
	2	5	No	1	8-10	Depressed	7
	11	1	Yes	1	5-7	Depressed	4
	10	2	Yes	0	8-10	No Depression	5
	10	1	Yes	1	8-10	Depressed	6
	6	2	Yes	0	8-10	No Depression	4
	4	5	Yes	1	5-7	Depressed	10
	10	2	Yes	1	8-10	Depressed	5
	2	5	Yes	1	5-7	Depressed	10
	10	3	No	1	8-10	Depressed	8

Showing 1 to 18 of 27,898 entries, 19 total columns

Console | Terminal | Background Jobs

R 4.4.2 · C:/Users/Aditya Yadav/Desktop/Student Depression/

```
> data <- data %>%
+   mutate(Total_Stress = `Academic Pressure` + `Financial Stress`)
> View(data)
>
```

Interpretation:

- A new column called **Total_Stress** has been created in the dataset.
- It is calculated by adding the values from two existing columns:
    - Academic Pressure
    - Financial Stress
- This helps in understanding the combined effect of both academic and financial pressures on a student.
- You can now analyze how total stress relates to depression, CGPA, sleep, etc., more effectively.

Conclusion

The project successfully analyzed student depression data using R. It included data cleaning, transformation, and filtering steps to make the dataset analysis-ready. Using group-wise summaries and logical filtering, key insights were generated like:

- Percentage of students with depression
- Gender-wise and CGPA-wise depression trends
- Identification of students with suicidal thoughts
- Relationship between financial and academic pressure
- Programs with the highest depression levels

These findings show that data analysis can help in understanding student mental health issues more deeply and suggest focus areas for intervention.

Result

- Dataset cleaned and irrelevant columns removed.
- Depression % among students was calculated.
- Depression was compared with CGPA, age, and gender.
- Suicidal thoughts were analyzed among depressed students.
- “Depression_Status” and “Total_Stress” columns were created for deeper insights.
- A final ranked list of students based on CGPA and depression was generated.