

Depression Among College Students: A Data-Driven Analysis

Domenic Cabibbo and Janav Bansal

Introduction

Depression has become a significant concern among college students, affecting their academic success and overall well-being. Our study aims to understand what factors contribute most significantly to depression in college students, with particular attention to academic pressure, lifestyle factors, and personal circumstances. Using data from 502 students, we examine various aspects of student life to identify patterns and potential intervention points.

Research Questions

Our study focuses on three key questions:

1. What specific factors affect depression among college students?
2. Does gender play a significant role in depression rates?
3. To what extent does the college environment contribute to depression?

Data Description

Our dataset comprises information from 502 college students and includes both numerical and categorical variables:

Numerical Variables: - Age - Academic Pressure (scale 1-5) - Study Satisfaction (scale 1-5) - Study Hours - Financial Stress (scale 1-5)

Categorical Variables: - Gender - Sleep Duration - Dietary Habits - Suicidal Thoughts - Family History - Depression Status

Graph is split roughly in half with an equal amount of men and women. Roughly half of each gender reported Yes when asked if they are depressed. See [Figure 1](#)

Pie Chart of Count by Gender and Depression

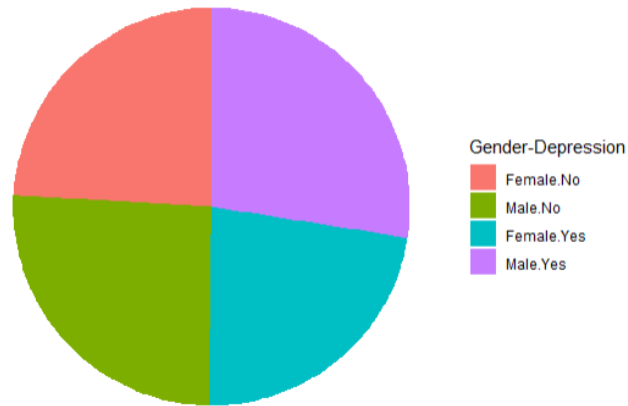


Figure 1: Depression/Gender Split

	Correlation.to.Depression <dbl>
Gender	0.03168323
Age	-0.21562453
Academic.Pressure	0.47625918
Study.Satisfaction	-0.29348374
Sleep.Duration	-0.04996768
Dietary.Habits	-0.18668522
Have.you.ever.had.suicidal.thoughts..	0.46629616
Study.Hours	0.20044134
Financial.Stress	0.29684613
Family.History.of.Mental.Illness	0.05608687

Figure 2: Correlation of all Factors

Analysis of Key Factors

Gender and Depression:



Figure 3: Gender Count

Students were asked to report yes or no to depression. Male students, 51.7% reported having depression; female students, the rate was 48.5%. There is a difference of 3.2 percentage points between the genders. Correlation analysis resulted in no correlation. See Figure 3.

Academic Pressure and Depression:

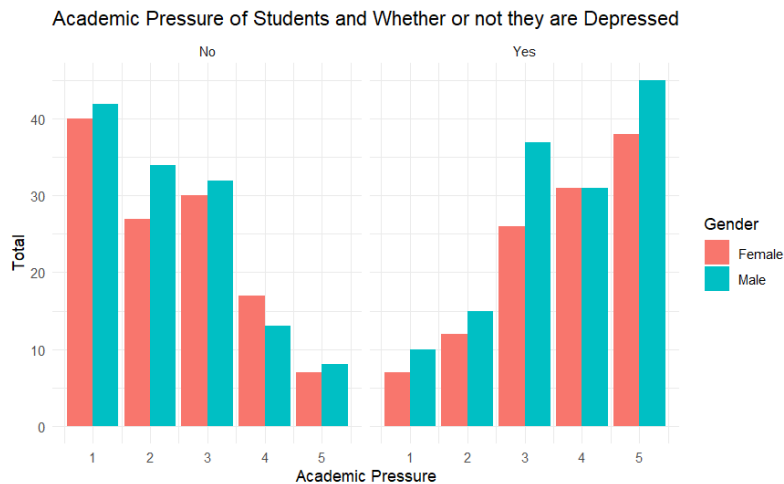


Figure 4: Academic Pressure (1-5)

Students were asked to report their perceived level of academic pressure on a likert scale 1 to 5, with one being the lowest level and five being the highest level. Analysis reveals that 32% of

students reporting depression rated academic pressure a 5 out of 5 (highest) and 6% students reporting depression rated academic pressure level 1 out of 5 (lowest). The correlation analysis resulted in moderate correlation ($r=0.47$). See Figure 4

Sleep Patterns:

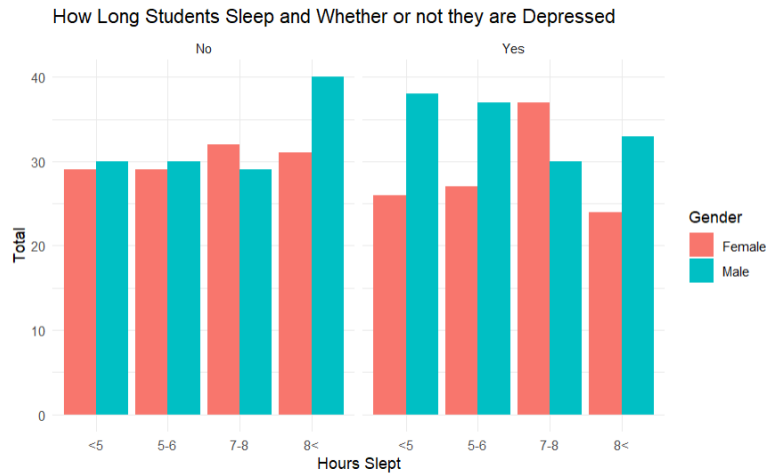


Figure 5: Sleep Patterns

Students were asked to respond to the hours they sleep per night. Analysis reveals that 26% of students reported depression and reported 7-8 hours of sleep; subsequently 25% of students reported depression and reported less than 5 hours of sleep. Correlation analysis resulted in no correlation with $r= -0.04$. See Figure 5.

Dietary Habits:

Students were asked to rate their eating habits as healthy, moderately, and unhealthy. Analysis states 25% of students reported depression and reported eating healthy while 42% of students who reported depression and reported eating unhealthy. 33% of students reporting a positive result for depression reported moderately healthy eating. Correlation analysis resulted in no correlation. See Figure 6.

Suicidal Thoughts:

Students were asked to answer yes or no to suicidal thoughts. 75% of students reporting depression also reported suicidal thoughts while 25% answered no to suicidal thoughts. This correlation resulted in a moderate correlation ($r=0.46$). See Figure 7.

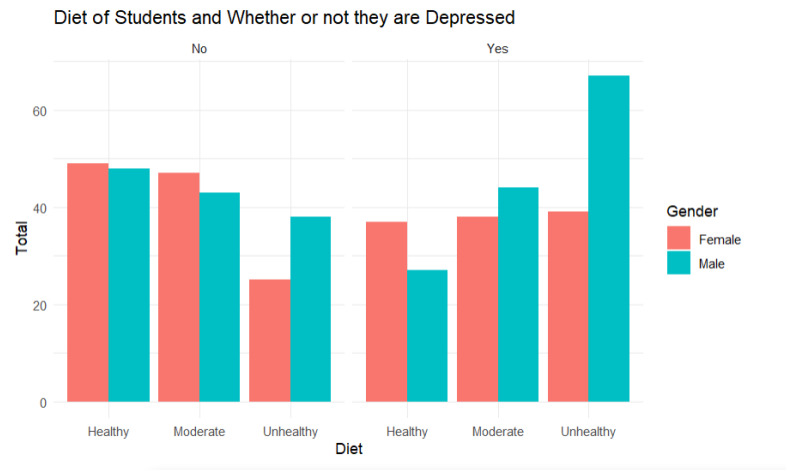


Figure 6: Diet



Figure 7: Suicidal Thoughts

Financial Stress:

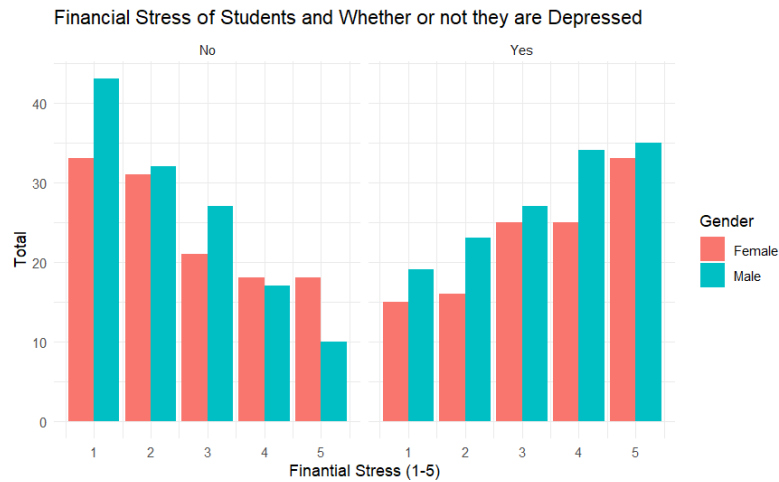


Figure 8: Financial Stress (1-5)

Students were asked to report their perceived level of financial stress on a likert scale 1 to 5, with one being the lowest level and five being the highest level. Analysis reveals that 26% of students reporting depression rated financial stress a 5 out of 5 (highest) and 13% students reporting depression rated financial stress level 1 out of 5 (lowest). The correlation analysis resulted in a weak correlation ($r=0.29$). See Figure 8.

Conclusion

The first question of the study was to examine factors that may contribute to depression among college students. The data set provided multiple variables to analyze. The second question of the study was to examine gender differences and reporting of depression. The results of this study did not show a difference in gender and reporting of depression as it was close to equally reported between genders. The final question was to examine college environmental factors and depression. Of these reported variables the following resulted in a weak to moderate correlation were financial stress, academic pressure respectively. Results indicate that there is a moderate correlation of reporting depression and having suicidal thoughts.

Recommendations

With gender not being a differentiator, it is important for education, support, interventions and other strategies to be applicable to all genders when seeking to prevent or support college

students who may be feeling depressed.

Limitations and Future Research

Our study was limited to 502 students and relied on self-reported data. There are many other variables that could impact feeling depressed. Future research could benefit from gaining more information such as but not limited to family history, social relationships, illicit drugs, alcohol consumption, etc. In addition, longitudinal studies track assessing throughout the academic terms through graduation..

References

1. Depression Student Dataset (2024). Kaggle. Retrieved from <https://www.kaggle.com/datasets/ikynahidw/student-dataset>

```
library(knitr)
library(tinytex)
library(tidyverse)
library(ggplot2)

depression <- read.csv("Depression Student Dataset.csv")
#Creates Pie Chart
result <- depression %>% group_by(Gender, Depression) %>% summarise(Count = n()) %>%
  mutate(Percentage = round((Count / sum(Count)) * 100, 2))

ggplot(result, aes(x = "", y = Count, fill = interaction(Gender, Depression))) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar(theta = "y") +
  labs(
    title = "Pie Chart of Count by Gender and Depression",
    fill = "Gender-Depression"
  ) +
  theme_void()

result %>% select(-Count)
#rewrites all catagoricals into numericals
data <- read.csv("Depression Student Dataset.csv")

data$Depression <- ifelse(data$Depression == "Yes", 1, 0)
data$Have.you.ever.had.suicidal.thoughts.. <- ifelse(data$Have.you.ever.had.suicidal.thoughts.. == "Yes", 1, 0)
```

```

                                == "Yes", 1, 0)
data$Family.History.of.Mental.Illness <- ifelse(data$Family.History.of.Mental.Illness
                                                == "Yes",1, 0)
data$Gender <- ifelse(data$Gender == "Male", 1, 0)

data <- data %>%
  mutate(Sleep.Duration = recode(Sleep.Duration, "5-6 hours" = 1,
                                "7-8 hours" = 2,
                                "More than 8 hours" = 3,
                                "Less than 5 hours" = 0)) %>%
  mutate(Dietary.Habits = recode(Dietary.Habits, "Healthy" = 2 ,
                                "Unhealthy" = 0, "Moderate" = 1))

#Calculates correlation
numeric_data <- data[sapply(data, is.numeric)]

correlation_matrix <- cor(numeric_data)

correlation_with_depression <- data.frame(
  Correlation.to.Depression = correlation_matrix["Depression"]
)

#return table
head(correlation_with_depression, 10)

depression_is_depressed <- depression %>% group_by(Gender,
                                                  Depression) %>% summarise(count = n())

ggplot(depression_is_depressed, aes(x = Depression, y = count, fill = Gender)) +
  geom_bar(stat = "Identity", position = "dodge") +
  labs(title = "Student Gender by Whether or not they are Depressed",
       x = "Are you Depressed?",
       y = "Count") +
  theme_minimal()

depression_press <- depression %>% group_by(Gender, Academic.Pressure,
                                                  Depression) %>% summarise(count=n())

ggplot(depression_press, aes(x = Academic.Pressure, y = count, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Academic Pressure of Students and Whether or not they are Depressed",

```



```

    x = "Academic Pressure",
    y = "Total",
    fill = "Gender"
  ) +
  theme_minimal()+
  facet_wrap(~ Depression)
depression_sleep <- depression %>% group_by(Gender, Sleep.Duration, Depression) %>%
  summarise(count=n())

#shortens categorical data so it's better read on the graph
depression_sleep <- depression_sleep %>%
  mutate(Sleep.Duration = recode(Sleep.Duration,
                                "5-6 hours" = "5-6",
                                "7-8 hours" = "7-8",
                                "More than 8 hours" = "8<",
                                "Less than 5 hours" = "<5"))

ggplot(depression_sleep, aes(x = Sleep.Duration, y = count, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "How Long Students Sleep and Whether or not they are Depressed",
    x = "Hours Slept",
    y = "Total",
    fill = "Gender"
  ) +
  theme_minimal()+
  facet_wrap(~ Depression)

depression_diet <- depression %>% group_by(Gender, Dietary.Habits, Depression) %>%
  summarise(count=n())

ggplot(depression_diet, aes(x = Dietary.Habits, y = count, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Diet of Students and Whether or not they are Depressed",
    x = "Diet",
    y = "Total",
    fill = "Gender"
  ) +
  theme_minimal()+
  facet_wrap(~ Depression)

```

```

depression_suicidal <- depression %>% group_by(Gender,
      Have.you.ever.had.suicidal.thoughts...,
      Depression) %>% summarise(count=n())

ggplot(depression_suicidal, aes(x = Have.you.ever.had.suicidal.thoughts...,
      y = count, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Does the Student have Suicidal Thoughts? and Whether or not they are Depressed.",
    x = "Suicidal Thoughts?",
    y = "Total",
    fill = "Gender"
  ) +
  theme_minimal()+
  facet_wrap(~ Depression)
depression_finance <- depression %>% group_by(Gender, Financial.Stress, Depression) %>%
  summarise(count=n())

ggplot(depression_finance, aes(x = Financial.Stress, y = count, fill = Gender)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(
    title = "Financial Stress of Students and Whether or not they are Depressed",
    x = "Finantial Stress (1-5)",
    y = "Total",
    fill = "Gender"
  ) +
  theme_minimal()+
  facet_wrap(~ Depression)

```