EGYPTIAN RUSSIAN UNIVERSITY

Statistical Insights & Predictive Models on Depression in Students

Student Depression Dataset

Team Members

Nada Saad Shawer (234003)

Sama El-Sayed (234078)

Malak Sameeh (234073)

Rana Kishty (234215)

Shimaa Mohamed (234253)

Introduction to Business Analytics (BUA201)

Dr. Rowaida Ali

T.A. Salem Adel

May 25, 2025

TABLE OF CONTENTS

INTRODUCTION

This dataset contains comprehensive information about students' mental health and related factors. It is designed to analyze the trends and predictors of depression among students. The data includes demographic details, academic and work-related pressures, lifestyle habits, and specific mental health indicators. Researchers can use this dataset to identify risk factors for depression and develop strategies for early intervention, while ensuring ethical handling of sensitive information.

Dataset Description:

- **Total Rows:** 27,901
- **Total Columns:** 18

Field Descriptions:

CATEGORICAL VARIABLES:

1. **Id:** A unique identifier assigned to each student record in the dataset.
2. **Gender:** The gender of the student (e.g., Male, Female, Other). This helps in analyzing gender-specific trends in mental health.
3. **Profession:** The field of work or study of the student, which may offer insights into occupational or academic stress factors.
4. **City:** The city or region where the student resides, providing geographical context for the analysis.
5. **Dietary Habits:** An assessment of the student's eating patterns and nutritional habits, potentially impacting overall health and mood.
6. **Degree:** The academic degree or program that the student is pursuing.
7. **Have you ever had suicidal thoughts?:** A binary indicator (Yes/No) that reflects whether the student has ever experienced suicidal ideation.

8. **Family History of Mental Illness:** Indicates whether there is a family history of mental illness (Yes/No), which can be a significant factor in mental health predispositions.
9. **Depression:** The target variable that indicates whether the student is experiencing depression (Yes/No). This is the primary focus of the analysis.

NUMERICAL VARIABLES:

1. **Age:** The age of the student in years.
2. **Academic Pressure:** A measure indicating the level of pressure the student faces in academic settings. This could include stress from exams, assignments, and overall academic expectations.
3. **Work Pressure:** A measure of the pressure related to work or job responsibilities, relevant for students who are employed alongside their studies.
4. **CGPA:** The cumulative grade point average of the student, reflecting overall academic performance.
5. **Study Satisfaction:** An indicator of how satisfied the student is with their studies, which can correlate with mental well-being.
6. **Job Satisfaction:** A measure of the student's satisfaction with their job or work environment, if applicable.
7. **Sleep Duration:** The average number of hours the student sleeps per day, which is an important factor in mental health.
8. **Work/Study Hours:** The average number of hours per day the student dedicates to work or study, which can influence stress levels.
9. **Financial Stress:** A measure of the stress experienced due to financial concerns, which may affect mental health.

This dataset compiles a wide range of information aimed at understanding, analyzing, and predicting depression levels among students. It is designed for research in psychology, data science, and education, providing insights into factors that contribute to student mental health challenges and aiding in the design of early intervention strategies.

PROBLEMS & SOLUTIONS (DATA CLEANING)

Problem #1: Irrelevant columns (Work Pressure, Job Satisfaction)

- **Solution:** Removing both columns

Problem #2: Presence of quotation marks in some records (Profession, Degree, Sleep Duration)

- **Solution:** Removing the quotation marks

Problem #3: Having unstandardized values for binary variables (Have you ever had suicidal thoughts, Family History of Mental Illness)

- **Solution:** Standardizing all the values into 0s and 1s

Problem #4: Records with non-numeric values (Sleep Duration)

- **Solution:** Removing the words (less than, more than and hours) and replacing each range of values with its average (7-8 → 7.5)

Problem #5: Having random values in some fields (Financial Stress, City)

- **Solution:** Replacing the random value with null cells

Problem #6: Having null/empty values for some cells

- **Solution:** Removing the records that contain null fields (City) and replacing the null cells with the average value for the column (Sleep Duration)

SELECTING A RANDOM SAMPLE

Given the size of the dataset, more than 27,900 records, we created a column named "random," assigned a random value to each row, sorted the data based on that column, and then created a new sheet specifically for our random sample. In order to conduct our study, we chose to use a sample of around 1,000 people, or 3.5% of the population. 1000 is the size of our sample (n), with a 3.5 margin of error.

ANALYSIS QUESTIONS

DESCRIPTIVE ANALYSIS

Demographic Insights:

1. What is the distribution of depression among genders?
2. Does age correlate with depression levels?
3. Are there geographic patterns in student depression (by city)?

Academic & Study Factors:

4. Is there a relationship between academic pressure and depression?
5. How does CGPA correlate with depression or suicidal thoughts?
6. Does study satisfaction reduce depression symptoms?
7. Are students in certain degree programs more prone to depression?

Mental Health Indicators:

8. What proportion of students have had suicidal thoughts?
9. Is there a link between family history of mental illness and depression?
10. Does sleep duration impact depression levels?

Lifestyle & Stress:

11. Does financial stress significantly correlate with depression?
12. Are longer work/study hours associated with higher depression levels?

Multivariate Analysis:

13. Which variables have statistically significant relationships with depression?

INFERENTIAL ANALYSIS

Categorical Variables:

14. Is depression related to gender?
15. Is suicidal ideation linked to depression?
16. Is profession associated with depression?
17. Do dietary habits have a direct effect on depression?
18. Does family history of mental illness affect depression?

Numerical Variables:

19. Do depressed students sleep less than non-depressed students?
20. Do depressed students experience more financial stress?
21. Is academic performance different between depressed/non-depressed students?
22. Do depressed students study more or less?
23. Is academic pressure higher in depressed students?
24. Do students with suicidal thoughts and depression sleep less?
25. Are suicidal thoughts linked to higher financial stress?
26. Do suicidal students perform worse academically?
27. Does academic pressure impact how long students sleep?
28. Does academic performance differ based on sleep duration range?
29. Are study hours different between stress levels?

PREDICTIVE ANALYSIS

Single Linear Regression:

30. Does sleep duration predict depression levels?
31. Does financial stress negatively affect academic performance?

Multiple Linear Regression:

32. How well do sleep, stress, and satisfaction predict depression?
33. Do gender and being a high school senior predict academic performance?

CONCLUSIONS FROM ANALYSIS

Note:

Some of the charts are "Combo Charts" and are done using an online Excel program because the version we had was an older excel version that didn't support Combo Charts.

Link to the online excel workbook where the charts were created:

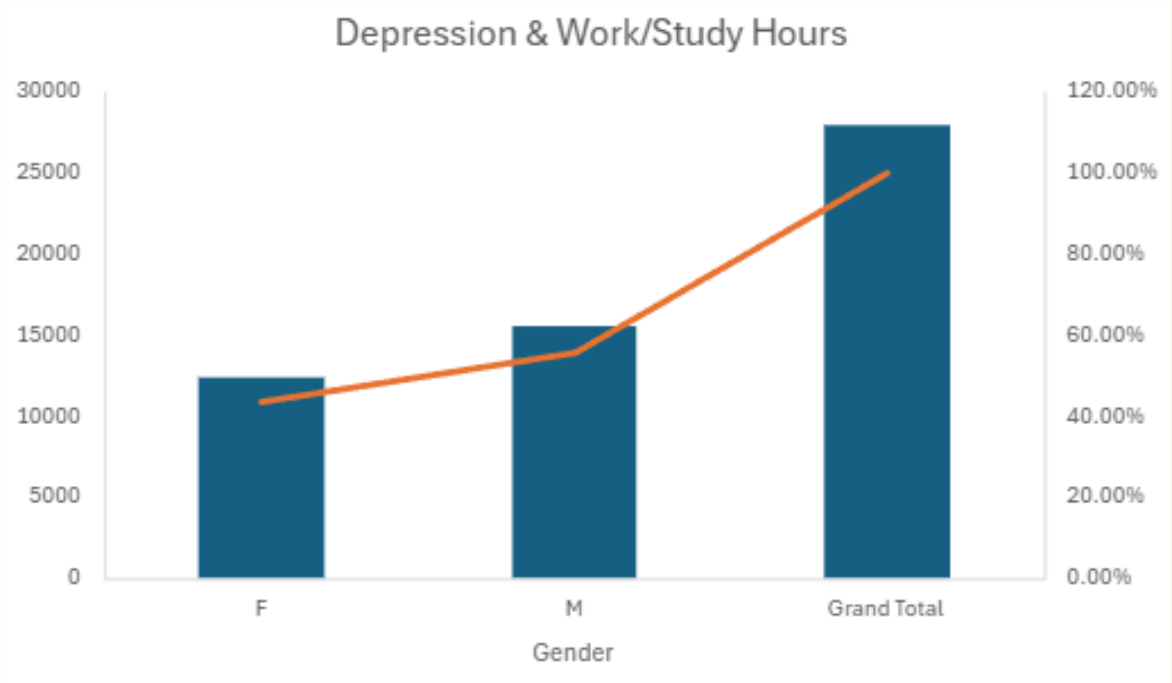
https://eruedueg-my.sharepoint.com/:x:/g/personal/234003_eru_edu_eg/EfWhmvRSSf9OszFBJU8PVTgBMFb2Ln6WdhFntfAZHp--ew?e=DAv13n

---

Demographic Insights

1. What is the distribution of depression among genders?

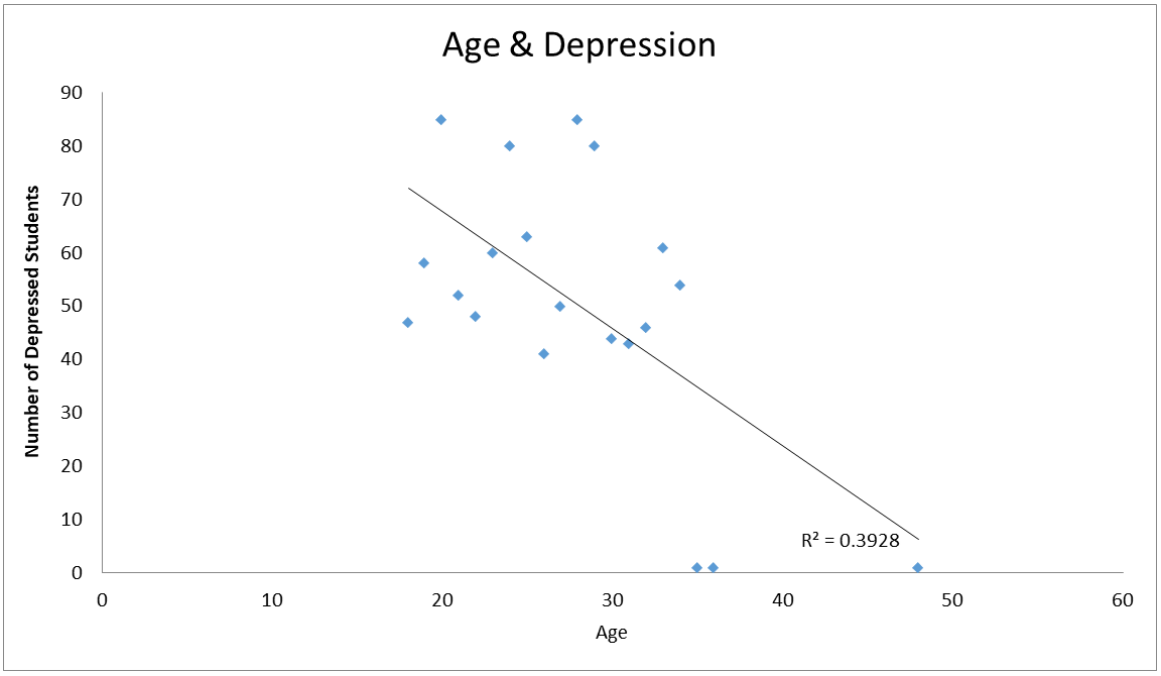
Gender	Count of Depression	Sum of Work/Study Hours
F	12354	43.94%
M	15547	56.06%
Grand Total	**27901**	**100.00%**



A breakdown of depression counts by gender and their corresponding percentages in relation to the total is shown in the pivot table. With a count of 1.2354, the data indicates that women (F) make up 44.28% of all depression cases, while men (M) make up 55.72% of cases, with a

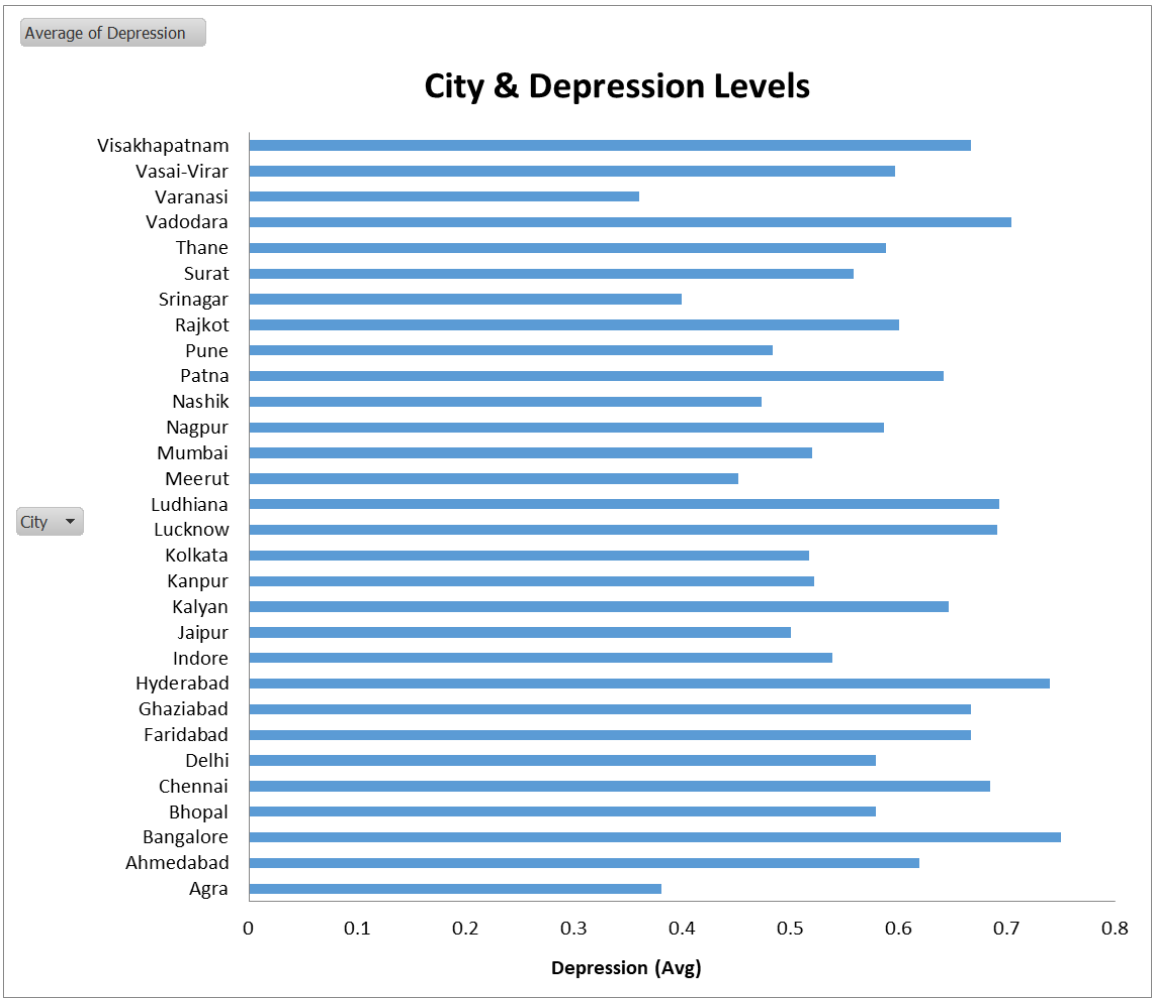
count of 1.5547. The overall sample size or aggregated measure for the dataset is 27,901 in total. This implies that, as indicated by the absolute numbers and the percentages, ***men in this dataset have a larger prevalence or count of depression than women****. However, the context of the data, whether they reflect averages, weighted values, or another metric, determines how the counts (1.2354 and 1.5547) should be interpreted. The figures demonstrate the gender gap in depression rates, with* ***men accounting for a higher share of all cases****.*

2. Does age correlate with depression levels?



Age and depression appear to be somewhat ***correlated negatively****, according to the trendline fitted to the data, which suggests that* ***depression levels may gradually decline with increasing age****. Age only accounts for around 40% of the variation in depressive state, according to the R^2 value of roughly 0.39, which indicates a* ***moderate association****. This implies that* ***age is somewhat a reliable indicator of depression****, even though there is a trend for* ***younger students to report higher levels of stress****.*

3. Are there geographic patterns in student depression (by city)?

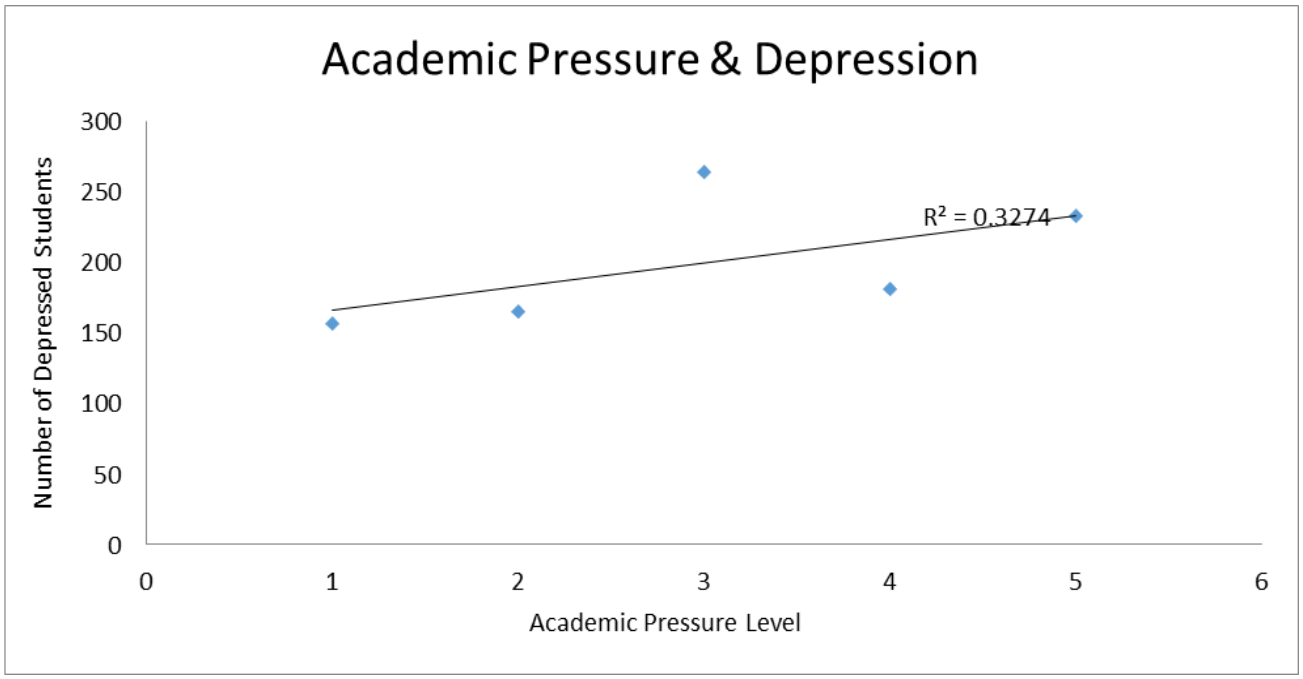


While most cities fall between 0.4 and 0.6, indicating ***moderate levels of depression****, the chart highlights* ***clear geographic variability****. This suggests that local factors, such as educational pressures, social environment, or access to mental health resources, may influence depression levels differently across cities, pointing to the need for city-specific mental health interventions. The graph shows distinct regional variation, even though the majority of cities have modest levels of depression (0.4–0.6). The necessity for city-specific mental health interventions is indicated by the possibility that local characteristics, such as social environments, educational constraints, or access to mental health facilities, may have varying effects on depression levels across different locations. Banlgore has the highest depression rates, while Varanasi has the lowest.*

Academic & Study Factors

4. Is there a relationship between academic pressure and depression?

Academic Pressure	Count of Depression
1	157
2	165
3	264
4	181
5	233
Grand Total	**1000**



*The R^2 value of 0.3274 indicates a **moderate positive association** between depression and academic pressure, and the chart features a trendline. This indicates that academic pressure alone accounts for about 32.7% of the variation in depression levels. The trendline's slope suggests that the **risk of depression tends to climb in relation with increased academic stress**. Overall, the chart illustrates the complexity of mental health and the necessity of taking into account a number of contributing factors, even though it offers evidence of a moderate link between academic pressure and depression.*

5. How does CGPA correlate with depression or suicidal thoughts?

- **Suicidal Thoughts & CGPA Correlation (r_{pb}):** -0.00083

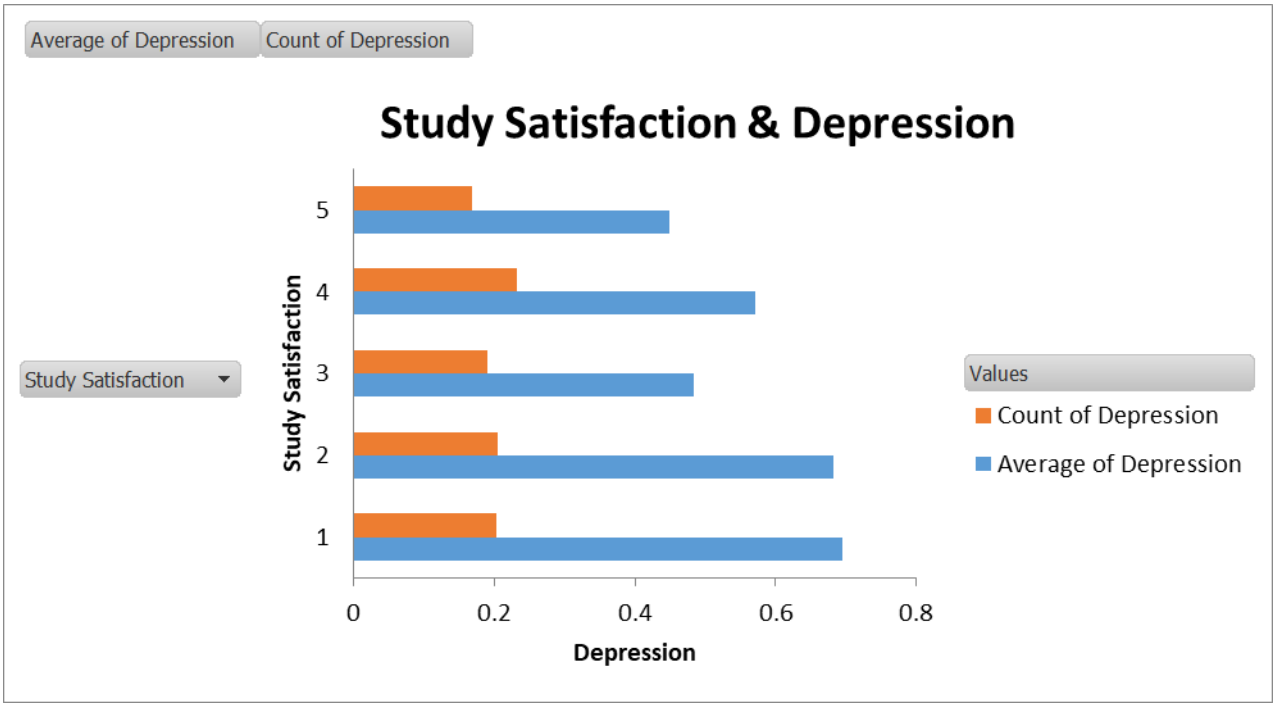
*According to the graph and statistical results, there is **no significant correlation** between the investigated dataset's CGPA and suicidal thoughts. Since the correlation value of -0.00083 is so near to zero, there is very little evidence of a linear relationship between these variables. While some individual cases show higher or lower values, the overall distribution **does not suggest that academic performance, as measured by CGPA, has any meaningful impact on reported suicidal thoughts** in this sample. This finding implies that other psychological, social, or environmental factors, rather than academic achievement, may play a more critical role in influencing suicidal thoughts.*

- **Depression & CGPA Correlation (r_{pb}):** 0.025397

*There is a **very weak correlation** between depression and CGPA, according to the scatterplot and statistical results. There is hardly any linear link between these two variables, as indicated by the correlation coefficient of 0.025, which is pretty low. The indicates that **academic success, as determined by CGPA, does not significantly affect depression levels** in this population, despite a few outliers that may exhibit minor differences. This result suggests that depression is probably caused by psychological, social, or environmental factors in addition to academic success.*

6. Does study satisfaction reduce depression symptoms?

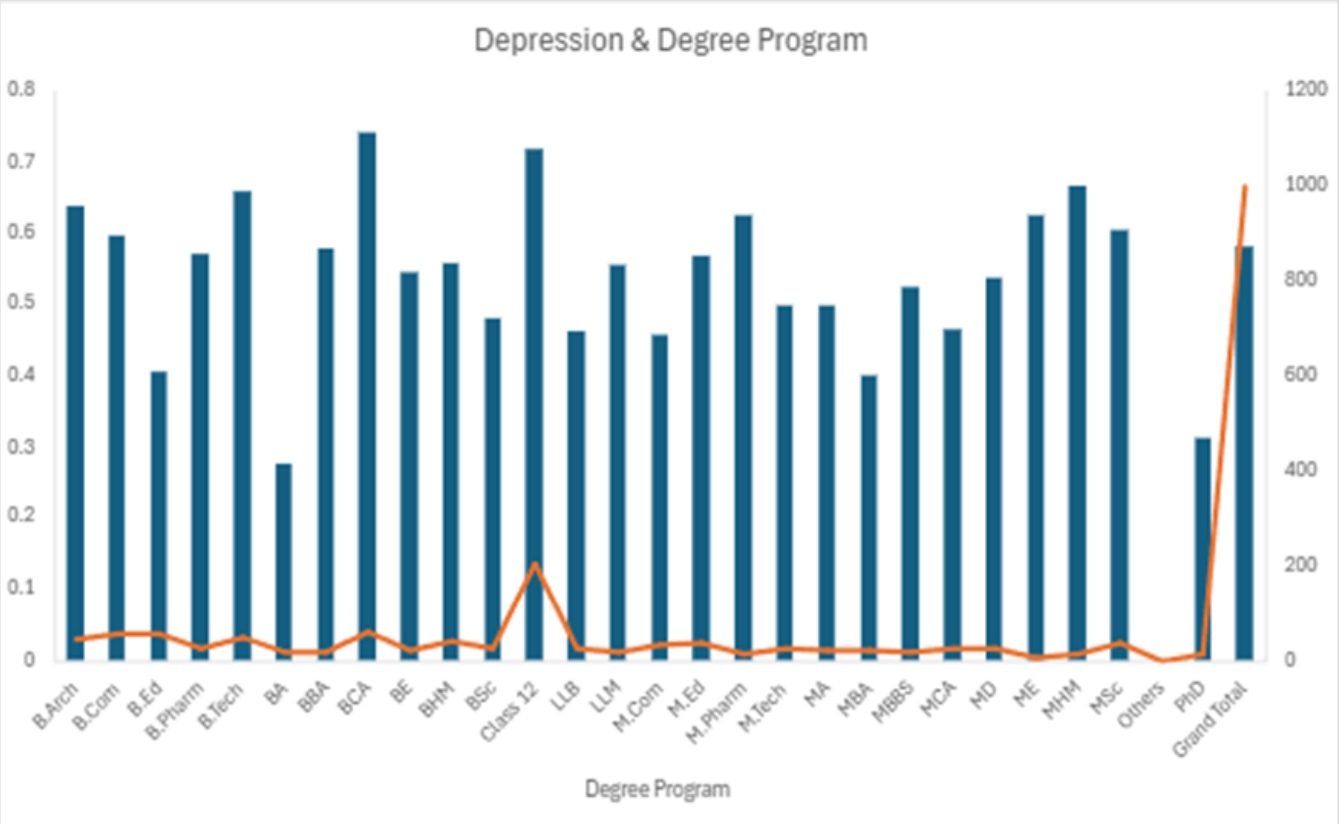
Study Satisfaction	Average of Depression	Count of Depression
1	0.694581281	20.30%
2	0.682926829	20.50%
3	0.484210526	19.00%
4	0.570815451	23.30%
5	0.449704142	16.90%
Grand Total	**0.582**	**100.00%**



The lowest average depression score (0.484) is reported by students with mid-range satisfaction (Level 3), according to the bar graph and counts of depression values, indicating that moderate satisfaction may be associated with better mental health outcomes. People who are most satisfied (Level 5) have a depression score that is marginally higher (0.450) than Level 3 but still lower than Levels 1, 2, and 4. There ***may be a connection between lower mental health and depression*** *because students with low satisfaction (Levels 1 and 2) have the highest average depression scores (0.695 and 0.683). The percentage distribution shows that most respondents fall into Level 4 (23.3%), followed closely by Levels 2 (20.5%) and 1 (20.3%). This suggests that while a significant portion of students report moderate to high satisfaction, a notable subset (over 40%) experiences lower satisfaction (Levels 1–2), which aligns with their higher depression rates.*

7. Are students in certain degree programs more prone to depression?

Degree Program	Average of Depression	Count of Depression
B.Arch	0.638297872	47
B.Com	0.596491228	57
B.Ed	0.406779661	59
B.Pharm	0.571428571	28
B.Tech	0.66	50
BA	0.277777778	18
BBA	0.578947368	19
BCA	0.741935484	62
BE	0.545454545	22
BHM	0.558139535	43
BSc	0.481481481	27
Class 12	0.718446602	206
LLB	0.461538462	26
LLM	0.555555556	18
M.Com	0.457142857	35
M.Ed	0.567567568	37
M.Pharm	0.625	16
M.Tech	0.5	28
MA	0.5	24
MBA	0.4	25
MBBS	0.523809524	21
MCA	0.464285714	28
MD	0.538461538	26
ME	0.625	8
MHM	0.666666667	15
MSc	0.605263158	38
Others	0	1
PhD	0.3125	16
Grand Total	**0.582**	**1000**

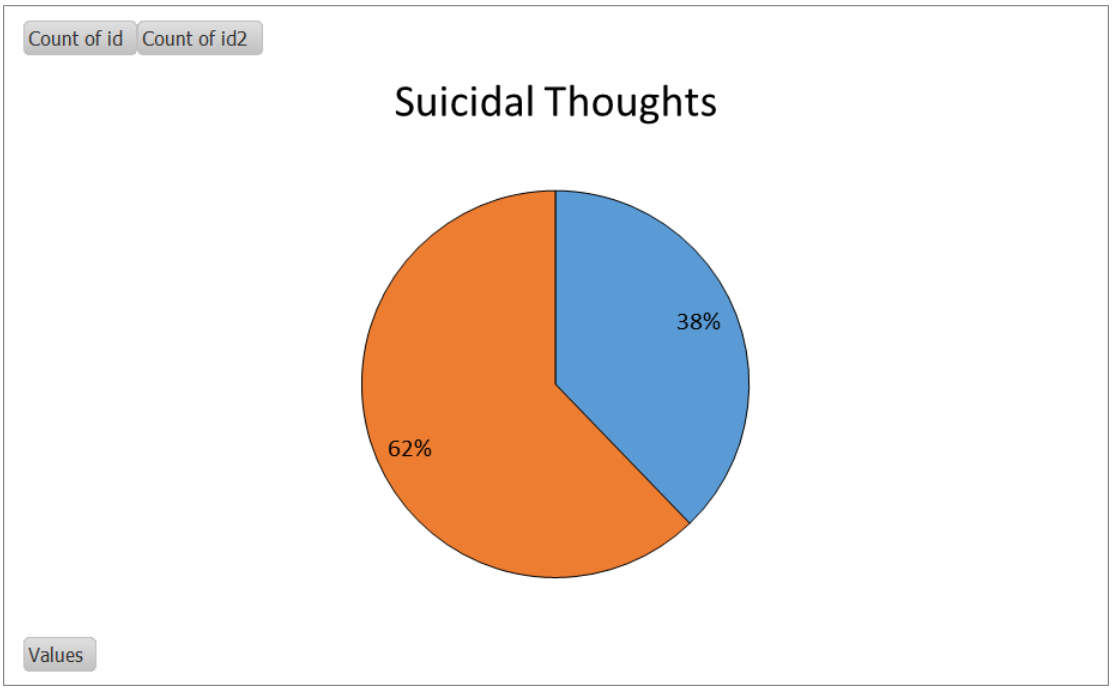


According to the data, BCA students have the greatest rates of depression (6.2%) and study satisfaction (0.74), indicating that ***academic satisfaction may have an effect on mental health****. Students in class 12 have a particularly high risk of depression (20.6%), which is probably caused by stress before college. Perhaps as a result of less pressure, BA students report the lowest levels of satisfaction (0.28) and depression (1.8%). PhD students have low levels of depression (1.6%) but low levels of satisfaction (0.31), which may indicate underreporting or resiliency.*

Mental Health Indicators

8. What proportion of students have had suicidal thoughts?

Have you ever had suicidal thoughts?	Count of id	Count of id2
0	378	37.80%
1	622	62.20%
Grand Total	**1000**	**100.00%**



*Suicidal thoughts were reported by **the majority of respondents** (62%), suggesting a serious mental health issue in this sample. Less than 38% of respondents said they didn't have such thoughts. For the majority of those impacted, this emphasizes the necessity of mental health care and intervention.*

9. Is there a link between family history of mental illness and depression?

Count of id	Observed Suicidal Thoughts		
Gender	0	1	Grand Total
F	162	274	436
M	216	348	564
Grand Total	**378**	**622**	**1000**

Expected Suicidal Thoughts			
Gender	0	1	Grand Total
F	164.808	271.192	436
M	213.192	350.808	564
Grand Total	378	622	1000

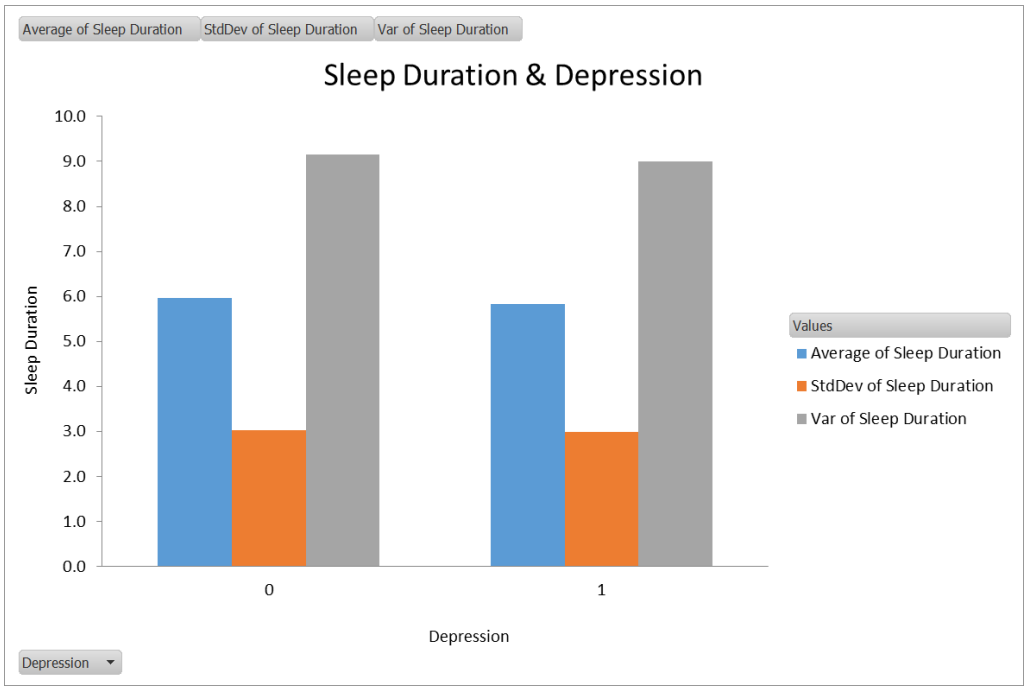
- **Statistical Values:**

p-value	0.711907987
Chi-square	0.136378663
Chi-critical	3.841458821

At the standard 0.05 significance level, the chi-square statistic (0.136) is well ***below*** *the critical value (3.841). This indicates that the* ***data shows a statistically significant correlation between gender and suicidal thoughts****. Additionally, the p-value (0.712) is* ***more than*** *our alpha (0.05). We* ***fail to reject (accept) the null hypothesis*** *that these variables are independent since the test statistic falls above the critical value. There is* ***no statistically significant evidence of a relationship between gender and suicidal thoughts*** *in this data. Thus* ***there is no statistical evidence that suicidal thoughts rates differ between males and females****.*

10. Does sleep duration impact depression levels?

Depression	Average of Sleep Duration	StdDev of Sleep Duration	Var of Sleep Duration
0	6.0	3.0	9.1
1	5.8	3.0	9.0
Grand Total	**5.9**	**3.0**	**9.1**

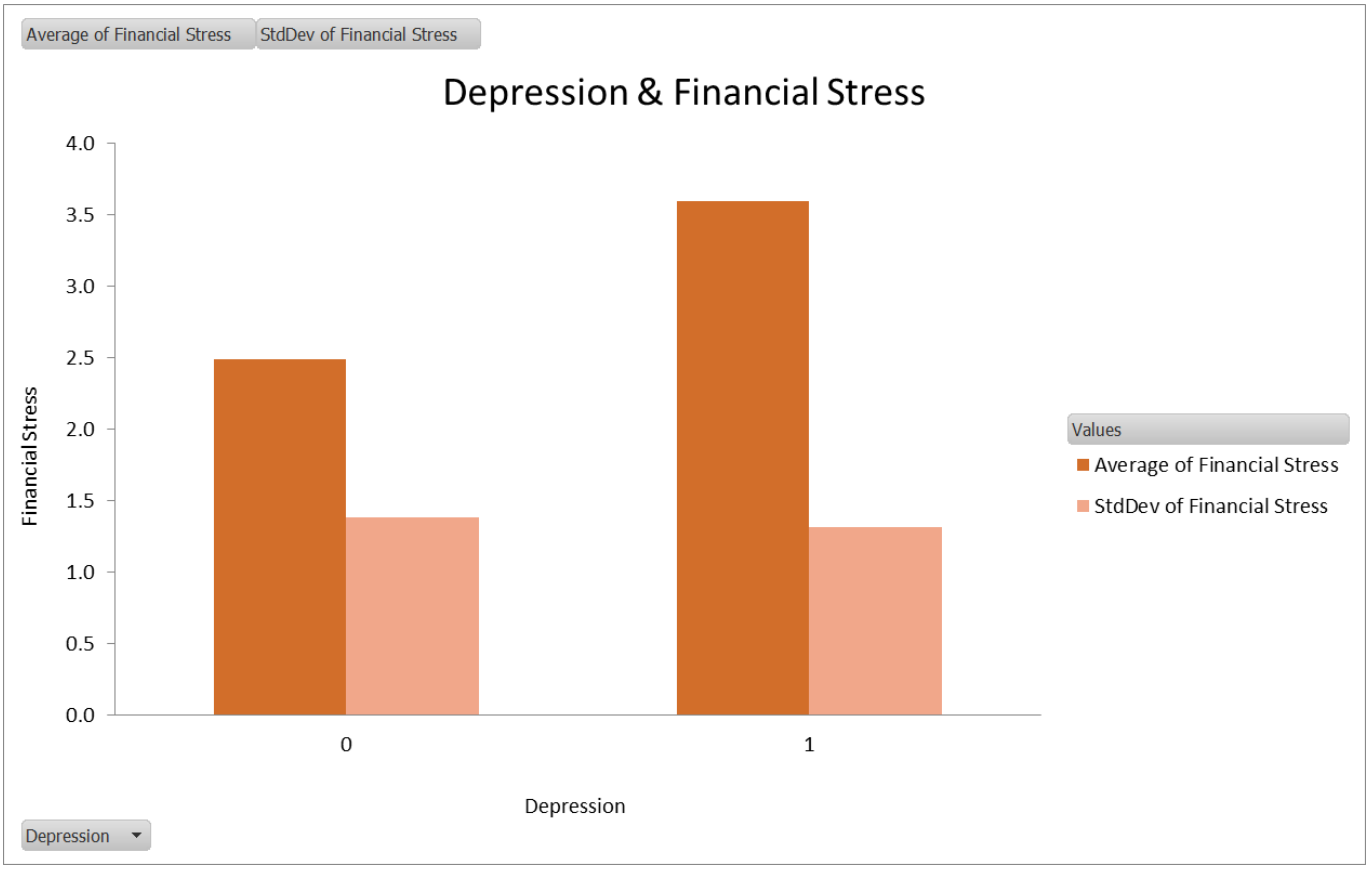


Students with depression (1) sleep an average of 5.8 hours, somewhat less than those without depression (0), who sleep an average of 6.0 hours. The two groups' identical standard deviations (3.0 hours) suggest that their sleep patterns vary similarly. ***Depression may not have a substantial effect on sleep duration****, as seen by the very insignificant 0.2 hour sleep difference between depressed and undepressed students. The* ***consistent spread*** *is confirmed by the near equivalent variance values (~9.0). Sleep duration varies greatly regardless of depressive level, as indicated by the identical standard variation (3 hours) in both groups.*

Lifestyle & Stress

11. Does financial stress significantly correlate with depression?

Depression	Count of Financial Stress	Average of Financial Stress	StdDev of Financial Stress
0	417	2.5	1.4
1	582	3.6	1.3
Grand Total	**999**	**3.1**	**1.5**

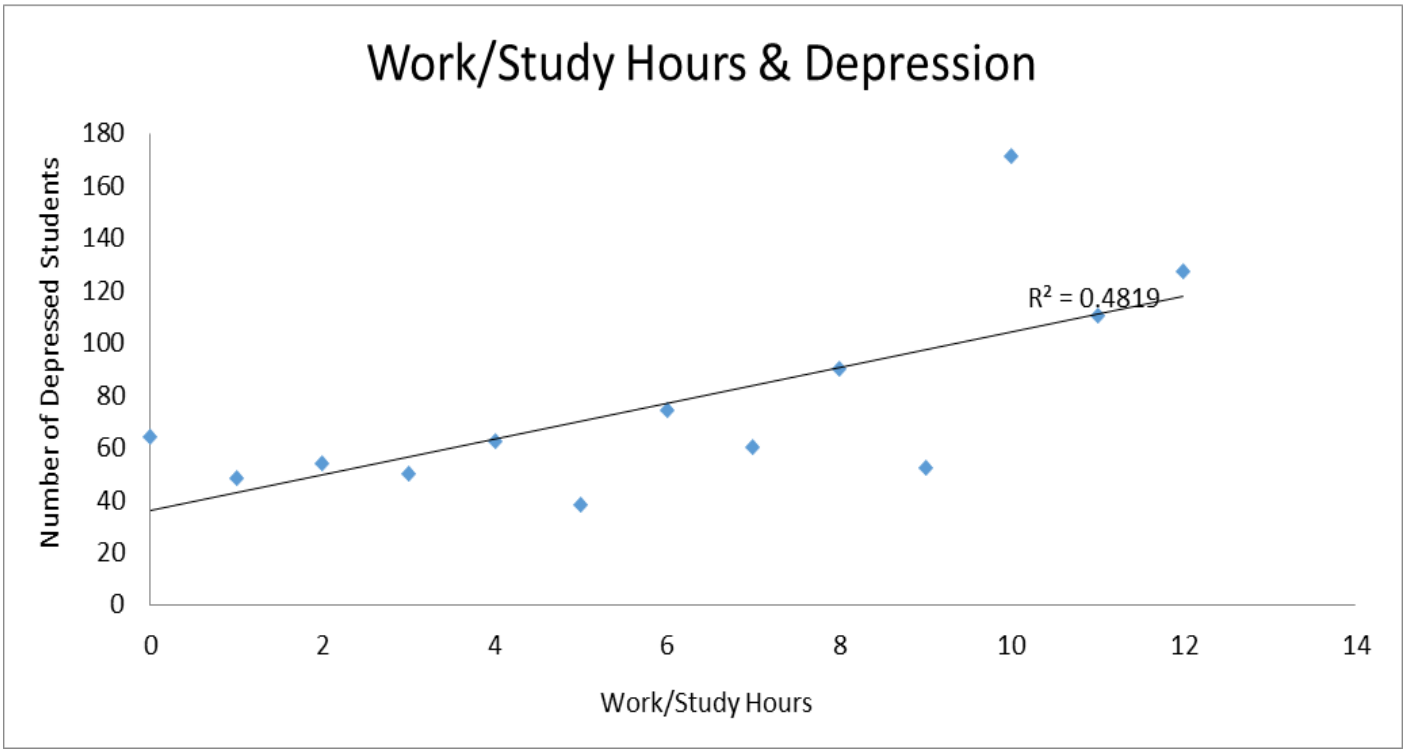


- **Point-biserial Correlation Value (r_{pb}):** 0.374762

*A **moderately positive association** between financial stress and depression is indicated by the point-biserial correlation of 0.37, which shows that the chance of depression increases with financial stress. According to the PivotTable, students who are depressed (coded as 1) report an average of 3.6 times more financial stress than students who are not depressed (0, average = 2.5), with comparable variability (StdDev ~1.3–1.4). Given that r_{pb} = 0.37 only accounts for around 14% of the variance, this shows that **financial stress may increase the likelihood of depression**, but other factors probably also play a part. In behavioral research, r_{pb} > 0.3 is generally regarded as practically significant.*

12. Are longer work/study hours associated with higher depression levels?

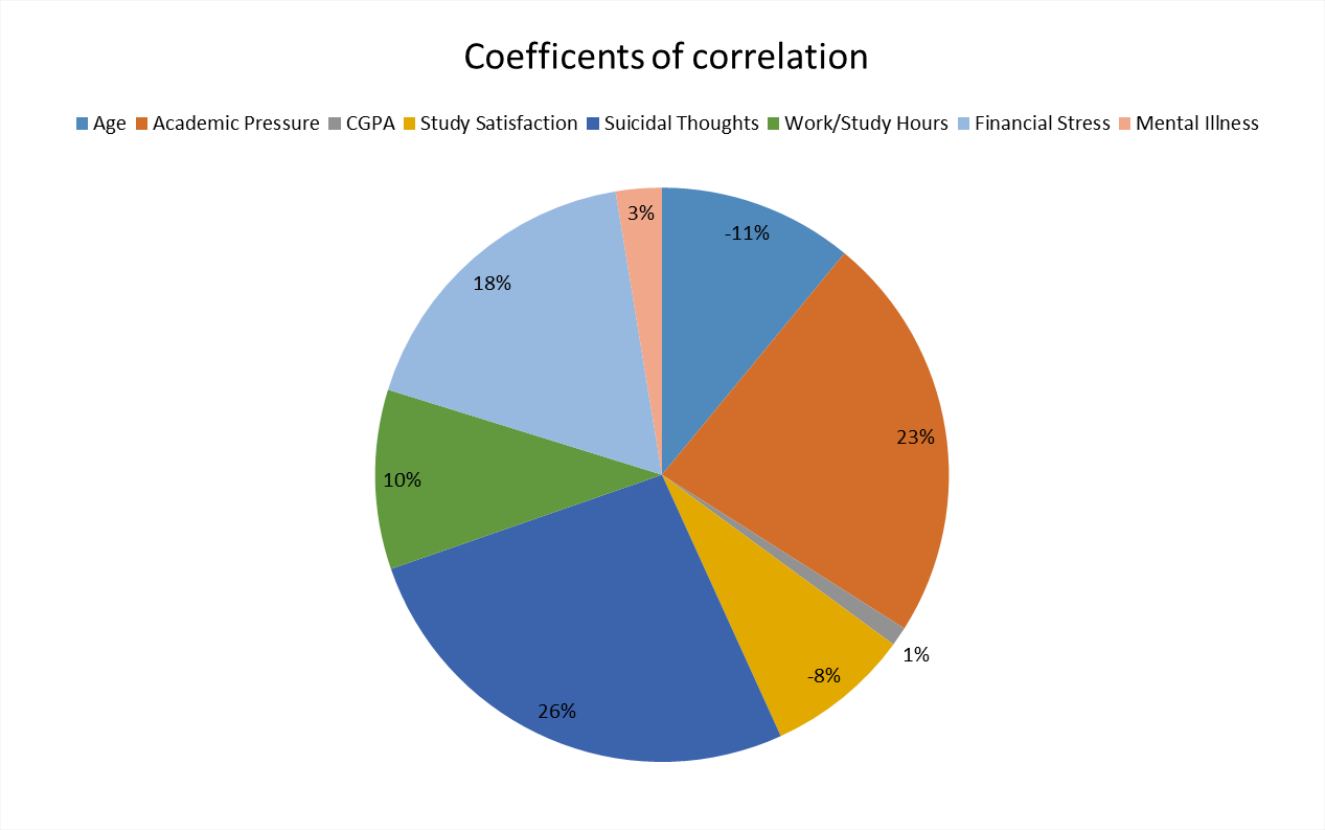
Work/Study Hours	Count of Depression
0	64
1	48
2	54
3	50
4	62
5	38
6	74
7	60
8	90
9	52
10	171
11	110
12	127



*There seems to be a **semi-linear association** between depression and work/study hours. Students with the longest hours (12–14) report the greatest depression counts (74 and 90), indicating a possible threshold effect, whereas students with moderate hours (6–10 hours) show varying depression counts (50–62). The shortest hours (0–2) do, however, also exhibit higher counts (64, 48), which may be a sign of stress brought on by under-engagement or outside influences. Although outliers (such as 171 & 127) suggest subgroups with substantial consequences, the average depression score of 1.71 (compared to a baseline of 1.10) further suggests a **weak overall connection**. Although the tendency isn't quite linear, **extreme hours, either very low or high, may be associated with an increased risk of depression**.*

13. Which variables have statistically significant relationships with depression?

Factors That Affects Depression	Coefficent of correlation
Age	-0.226422041
Academic Pressure	0.474834944
CGPA	0.02221047
Study Satisfaction	-0.167971384
Suicidal Thoughts	0.546276684
Work/Study Hours	0.208562831
Financial Stress	0.363591429
Mental Illness	0.053430172



DESCRIPTIVE STATISTICS

CATEGORICAL VARIABLES

Descriptive Measures (Frequency Distributions):

Gender	Count of Depression	Count of Depression2
F	436	43.6%
M	564	56.4%
Grand Total	**1000**	**100.0%**

Suicidal Thoughts	Count of Depression	Count of Depression2
0	378	37.8%
1	622	62.2%
Grand Total	**1000**	**100.0%**

Family History	Count of Depression	Count of Depression2
0	521	52.1%
1	479	47.9%
Grand Total	**1000**	**100.0%**

City	Count of Depression	Count of Depression2
Agra	42	4.2%
Ahmedabad	42	4.2%
Bangalore	32	3.2%
Bhopal	38	3.8%
Chennai	19	1.9%
Delhi	19	1.9%
Faridabad	18	1.8%
Ghaziabad	30	3.0%
Hyderabad	46	4.6%
Indore	26	2.6%
Jaipur	34	3.4%
Kalyan	48	4.8%
Kanpur	23	2.3%
Kolkata	29	2.9%
Lucknow	42	4.2%
Ludhiana	39	3.9%
Meerut	31	3.1%
Mumbai	25	2.5%
Nagpur	29	2.9%
Nashik	19	1.9%
Patna	39	3.9%
Pune	31	3.1%
Rajkot	30	3.0%
Srinagar	50	5.0%
Surat	43	4.3%
Thane	34	3.4%
Vadodara	27	2.7%
Varanasi	25	2.5%
Vasai-Virar	57	5.7%
Visakhapatnam	33	3.3%
Grand Total	**1000**	**100.0%**

Profession	Count of Depression	Count of Depression2
Content Writer	1	0.1%
Pharmacist	1	0.1%
Student	997	99.7%
Teacher	1	0.1%
Grand Total	**1000**	**100.0%**

Dietary Habits	Count of Depression	Count of Depression2
Healthy	288	28.8%
Moderate	351	35.1%
Unhealthy	361	36.1%
Grand Total	**1000**	**100.0%**

Degree	Count of Depression	Count of Depression2
B.Arch	47	4.7%
B.Com	57	5.7%
B.Ed	59	5.9%
B.Pharm	28	2.8%
B.Tech	50	5.0%
BA	18	1.8%
BBA	19	1.9%
BCA	62	6.2%
BE	22	2.2%
BHM	43	4.3%
BSc	27	2.7%
Class 12	206	20.6%
LLB	26	2.6%
LLM	18	1.8%
M.Com	35	3.5%
M.Ed	37	3.7%
M.Pharm	16	1.6%
M.Tech	28	2.8%
MA	24	2.4%
MBA	25	2.5%
MBBS	21	2.1%
MCA	28	2.8%
MD	26	2.6%
ME	8	0.8%
MHM	15	1.5%
MSc	38	3.8%
Others	1	0.1%
PhD	16	1.6%
Grand Total	**1000**	**100.0%**

NUMERICAL VARIABLES

1. Descriptive Measures:

Academic Pressure	
Mean	3.168
Standard Error	0.043355
Median	3
Mode	3
Standard Deviation	1.371005
Sample Variance	1.879656
Kurtosis	-1.16004
Skewness	-0.12733
Range	4
Minimum	1
Maximum	5
Sum	3168
Count	1000

Sleep Duration	
Mean	5.886887
Standard Error	0.095189
Median	5.5
Mode	2
Standard Deviation	3.008625
Sample Variance	9.051822
Kurtosis	-1.36155
Skewness	-0.09336
Range	8
Minimum	2
Maximum	10
Sum	5881
Count	999

Financial Stress	
Mean	3.134134
Standard Error	0.045939
Median	3
Mode	5
Standard Deviation	1.451979
Sample Variance	2.108243
Kurtosis	-1.34065
Skewness	-0.11843
Range	4
Minimum	1
Maximum	5
Sum	3131
Count	999

CGPA	
Mean	7.75346
Standard Error	0.046138
Median	7.88
Mode	8.04
Standard Deviation	1.459011
Sample Variance	2.128714
Kurtosis	-1.20578
Skewness	-0.15013
Range	4.92
Minimum	5.08
Maximum	10
Sum	7753.46
Count	1000

Age	
Mean	25.88
Standard Error	0.153509
Median	26
Mode	20
Standard Deviation	4.854396
Sample Variance	23.56517
Kurtosis	-0.79159
Skewness	0.142293
Range	30
Minimum	18
Maximum	48
Sum	25880
Count	1000

Work/Study Hours	
Mean	7.24
Standard Error	0.120528
Median	8
Mode	10
Standard Deviation	3.811421
Sample Variance	14.52693
Kurtosis	-1.04732
Skewness	-0.48452
Range	12
Minimum	0
Maximum	12
Sum	7240
Count	1000

Study Satisfaction	
Mean	2.96
Standard Error	0.04389
Median	3
Mode	4
Standard Deviation	1.387922
Sample Variance	1.926326
Kurtosis	-1.28047
Skewness	-0.00488
Range	4
Minimum	1
Maximum	5
Sum	2960
Count	1000

*According to the descriptive analysis, students have **high mental health issues**. There appears to be a mental health issue in this sample, as over 58% of respondents reported feeling depressed and almost 62% acknowledged having suicide thoughts. Both academic and financial demands have moderate averages (means of 3.17 and 3.13, respectively), suggesting that they are contributing factors. Students report a pretty high average CGPA (~ 7.8/10), adequate study hours (~7hrs) and sleep (~6hrs) despite the expectations, all of which may have a detrimental effect on their academic performance and general well-being. The majority of students are in their middle twenties (mean: ~26), though age varies. Even while study satisfaction is neutral, the confluence of high levels of stress, low academic achievement, and mental health issues presents a troubling image that highlights the urgent need for interventions and support networks in student environments.*

2. Correlation Matrix:

	Academic Pressure	*CGPA*	*Study Satisfaction*	*Sleep Duration*	*Work/Study Hours*	*Financial Stress*	*Age*
Academic Pressure	1						
CGPA	-0.053874844	1					
Study Satisfaction	-0.112156443	-0.085936725	1				
Sleep Duration	0.045812536	0.013166136	-0.013267439	1			
Work/Study Hours	0.094693955	-0.004412421	-0.018345082	-0.037192958	1		
Financial Stress	0.181360108	-0.001535087	-0.043965695	-0.014761237	0.084087364	1	
Age	-0.105084775	-0.01821871	0.032853743	0.01642212	-0.044183159	-0.062742825	1

*Numerous significant connections between the variables influencing students' well-being are shown by the correlation matrix. The **highest correlations are seen in depression**, which is positively connected with study hours (r = 0.22), financial stress (r = 0.37), suicidal thoughts (r = 0.51), and academic pressure (r = 0.51). This implies that **greater degrees of stress, academic pressure, and even study time are associated with higher depression levels**. Remarkably, depression has a negative correlation with age (r = -0.24) and study satisfaction (r = -0.16), suggesting that **older or more satisfied students are less likely to be depressed**. The idea that these stressors are related is further supported by the correlations between academic pressure and financial stress (r = 0.18), as well as suicidal thoughts (r = 0.27). Interestingly, **CGPA has a weak correlation with all variables**, indicating that, at least in this group, **stress levels and mental health do not directly affect GPA performance**.*

INFERENTIAL STATISTICS

CATEGORICAL VARIABLES

1. Is depression related to gender?

-	Observed Depression		
Gender	0	1	Grand Total
F	177	259	436
M	241	323	564
Grand Total	**418**	**582**	**1000**

	Expected Depression		
Gender	**0**	**1**	**Grand Total**
F	182.248	253.752	436
M	235.752	328.248	564
Grand Total	418	582	**1000**

- **Statistical Values:**

p-value	0.497443292
Chi-square	0.460386726
Chi-critical	3.841458821

At the standard 0.05 significance level, the chi-square statistic (0.46) is well ***below*** *the critical value (3.841). This indicates that the* ***data does not show a statistically significant correlation*** *between gender and depression. Additionally the p-value (0.497) is* ***greater than*** *our alpha (0.05). We* ***fail to reject (accept) the null hypothesis*** *that these variables are independent since the test statistic falls below the critical value. There is* ***no statistically significant evidence of an association between gender and depression*** *in this data. Thus, at a 5% confidence level,* ***there is no statistical evidence that depression rates differ between males and females****. According to these findings, researchers should look into factors other than gender that may predict depression.*

2. Is suicidal ideation linked to depression?

-	Observed Depression		
Suicidal Thoughts	0	1	Grand Total
0	281	97	378
1	137	485	622
Grand Total	**418**	**582**	**1000**

	Expected Depression		
Suicidal Thoughts	**0**	**1**	**Grand Total**
0	158.004	219.996	**378**
1	259.996	362.004	**622**
Grand Total	**418**	**582**	**1000**

- **Statistical Values:**

p-value	1.80715E-59
Chi-square	264.4846822
Chi-critical	3.841458821

*At the standard 0.05 significance level, the chi-square statistic (264.48) is **far above** the critical value (3.841). This indicates that the **data shows a statistically significant correlation** between gender and depression. Additionally, the very low p-value (1.807E-59) is far **less than** our alpha (0.05). We **reject the null hypothesis** that these variables are independent since the test statistic falls above the critical value. There **is statistically significant evidence of an association between suicidal thoughts and depression** in this data. Thus, at a 5% confidence level, **there is statistical evidence that depression rates differ with suicidal thoughts**.*

3. Is profession associated with depression?

-	Observed Depression		
Profession	0	1	Grand Total
Content Writer		1	1
Pharmacist		1	1
Student	418	579	997
Teacher		1	1
Grand Total	**418**	**582**	**1000**

-	Expected Depression		
Profession	**0**	**1**	**Grand Total**
Content Writer	0.418	0.582	1
Pharmacist	0.418	0.582	**1**
Student	416.746	580.254	997
Teacher	0.418	0.582	1
Grand Total	418	582	**1000**

- **Statistical Values:**

p-value	0.823708644
Chi-square	2.161122543
Chi-critical	7.814727903

At the standard 0.05 significance level, the chi-square statistic (2.16) is well ***below*** *the critical value (7.81). This indicates that the* ***data does not show a statistically significant correlation*** *between profession and depression. Additionally the p-value (0.824) is* ***greater than*** *our alpha (0.05). We* ***fail to reject (accept) the null hypothesis*** *that these variables are independent since the test statistic falls below the critical value. There is* ***no statistically significant evidence of an association between profession and depression*** *in this data. Thus, at a 5% confidence level,* ***there is no statistical evidence that depression rates differ with profession****. According to these findings, researchers should look into factors other than profession that may predict depression.*

4. Do dietary habits have a direct effect on depression?

-	Observed Depression		
Dietary Habit	**0**	**1**	**Grand Total**
Healthy	155	133	288
Moderate	161	190	351
Unhealthy	102	259	361
Grand Total	**418**	**582**	**1000**

-	Expected Depression		
Dietary Habit	**0**	**1**	**Grand Total**
Healthy	120.384	167.616	288
Moderate	146.718	204.282	351
Unhealthy	150.898	210.102	361
Grand Total	418	582	**1000**

- **Statistical Values:**

p-value	7.17079E-11
Chi-square	46.71684151
Chi-critical	5.991464547

*At the standard 0.05 significance level, the chi-square statistic (46.72) is **far above** the critical value (5.99). This indicates that the **data shows a statistically significant correlation** between dietary habits and depression. Additionally, the very low p-value (7.17E-11) is far **less than** our alpha (0.05). We **reject the null hypothesis** that these variables are independent since the test statistic falls above the critical value. There **is statistically significant evidence of an association between dietary habits and depression** in this data. Thus, at a 5% confidence level, **there is statistical evidence that depression rates differ with dietary habits** of students.*

5. Does family history of mental illness affect depression?

-	Observed Depression		
Family History	**0**	**1**	**Grand Total**
0	234	287	521
1	184	295	479
Grand Total	**418**	**582**	**1000**

	Expected Depression		
Family History	**0**	**1**	**Grand Total**
0	217.778	303.222	**521**
1	200.222	278.778	**479**
Grand Total	**418**	**582**	**1000**

- **Statistical Values:**

p-value	0.037347978
Chi-square	4.33447289
Chi-critical	3.841458821

*At the standard 0.05 significance level, the chi-square statistic (4.33) is **above** the critical value (3.84). This indicates that the **data shows a statistically significant correlation** between family history of mental illness and depression. Additionally, the p-value (0.037) is*

less than *our alpha (0.05). We* ***reject the null hypothesis*** *that these variables are independent since the test statistic falls above the critical value. There* ***is statistically significant evidence of an association between family history of mental illness and depression*** *in this data. Thus, at a 5% confidence level,* ***there is statistical evidence that depression rates differ with family history of mental illness****.*

NOTE: *City and degree measure's relation with depression could be measured in a similar way*

NUMERICAL VARIABLES

NOTE:

For this numerical variable inferential analysis we used Welch's t-test (unequal variances) since it is the more reliable and advised method for comparing group means—particularly in situations when population variances are unknown or perhaps unequal—we went with it. Welch's adaption, in contrast to Student's t-test, does not presume equal variances, which lowers the possibility of Type I errors (false positive) and yields trustworthy results even in cases when variances vary. This approach is preferred by current statistical recommendations since it does not require pre-testing variance equality, which could inflate error rates. We made sure the analysis was accurate and cautious by using Welch's t-test by default, irrespective of variance homogeneity.

1. Do depressed students sleep less than non-depressed students?

t-Test: Two-Sample Assuming Unequal Variances		
	Sleep Duration (Not Depressed)	*Sleep Duration (Depressed)*
Mean	5.96875	5.835051546
Variance	9.127936747	8.995120393
Observations	416	582
Hypothesized Mean Difference	0	
df	890	
t Stat	0.691359606	
P(T<=t) one-tail	0.244759938	
t Critical one-tail	1.646567521	
P(T<=t) two-tail	0.489519875	
t Critical two-tail	1.962633025	

*There was **no statistically significant difference** in the amount of sleep between the groups that were depressed (M=5.84, SD=√8.99≈3.00) and those that were not (M=5.97, SD=√9.13≈3.02) (t(0.05, 890)=0.69, p-value=0.49). We **do not have enough evidence to reject the null hypothesis** and thus **accept it** since the p-value (0.49) is significantly higher than the conventional alpha threshold of 0.05. At a 5% confidence level, there is **no significant difference in the mean sleep duration among different depression subgroups** The virtually **identical means** (5.84 vs. 5.97hrs) and **comparable variances** (9.13 vs. 8.99) indicate that the sample's **depression is not significantly impacted by sleep duration**.*

2. Do depressed students experience more financial stress?

t-Test: Two-Sample Assuming Unequal Variances		
	Financial Stress (Not Depressed)	*Financial Stress (Depressed)*
Mean	2.495192308	3.594501718
Variance	1.91804912	1.738902591
Observations	416	582
Hypothesized Mean Difference	0	
df	867	
t Stat	-12.61118823	
P(T<=t) one-tail	6.68494E-34	
t Critical one-tail	1.646613037	
P(T<=t) two-tail	1.33699E-33	
t Critical two-tail	1.962703927	

*Financial stress levels in the most and less depressed groups **differ statistically significantly**, according to the analysis (t(0.05, 867) = -12.61, p-value < 0.001). In comparison to the depressed group (M = 3.59, SD = √1.74 ≈ 1.32), the **undepressed people reported lower mean financial stress** (M = 2.49, SD = √1.92 ≈ 1.39). This directional link is confirmed by the **negative t-stat**. Since the p-value is significantly less than alpha, at a 5% confidence level, we **reject the null hypothesis** and come to the conclusion that **there are significant differences in the mean level of financial stress among different depression subgroups**. Further research into potential mediating factors is warranted in light of this unexpected conclusion, which shows a correlation between lesser financial stress and more severe depression.*

3. Is academic performance different between depressed/non-depressed students?

t-Test: Two-Sample Assuming Unequal Variances		
	CGPA (Not Depressed)	*CGPA (Depressed)*
Mean	7.707601918	7.78266323
Variance	2.155510101	2.10670873
Observations	417	582
Hypothesized Mean Difference	0	
df	890	
t Stat	-0.80066306	
P(T<=t) one-tail	0.211770205	
t Critical one-tail	1.646567521	
P(T<=t) two-tail	0.42354041	
t Critical two-tail	1.962633025	

The undepressed (M=7.71, SD=$\sqrt{2.16}$≈1.47) and depressed (M=7.73, SD=$\sqrt{2.11}$≈1.45) groups ***did not differ statistically significantly in their mean CGPA****, according to the t-test (t(0.05, 890)=-0.80, p-value=0.42). The* ***overlapping distributions*** *(equal variances) and nearly* ***comparable means*** *(7.71 vs. 7.78) indicate that* ***depression levels are not significantly correlated with CGPA differences*** *in this population. We* ***accept the null hypothesis of equal means*** *at a 5% confidence level since p-value is greater than our alpha. This pattern demonstrated clear non-significance for both the one-tailed (p-value=0.21) and two-tailed (p-value=0.42) tests.*

4. Do depressed students study more or less?

t-Test: Two-Sample Assuming Unequal Variances		
	Work/Study Hours (Depressed)	*Work/Study Hours (Not Depressed)*
Mean	6.230215827	7.967353952
Variance	16.03341173	12.22784806
Observations	417	582
Hypothesized Mean Difference	0	
df	820	
t Stat	-7.123998345	
P(T<=t) one-tail	1.14819E-12	
t Critical one-tail	1.646713996	
P(T<=t) two-tail	2.29639E-12	
t Critical two-tail	1.962861201	

The results show that the work/study hours of the depressed (M=6.23, SD=√16.03≈4.00) and non-depressed (M=7.97, SD=√12.23≈3.50) groups ***differ statistically significantly*** *(t(0.05, 820)=-7.12, p<.001). The* ***negative t-stat*** *shows that* ***depressed students reported less work/study hours than their non-depressed colleagues****. We* ***reject the null hypothesis of equal means*** *at a 5% confidence level since p-value is less than our alpha. This substantial difference (p<0.001 for both one- and two-tailed tests) implies that* ***having depression may be linked to less time spent being productive****. Nonetheless, the output's odd variance ratios (16.03 vs. 12.23) call for confirmation of the measurement scales and data cleansing.*

5. Is academic pressure higher in depressed students?

t-Test: Two-Sample Assuming Unequal Variances		
	Academic Pressure (Depressed)	*Academic Pressure (Not Depressed)*
Mean	2.338129496	3.764604811
Variance	1.455105147	1.336920584
Observations	417	582
Hypothesized Mean Difference	0	
df	873	
t Stat	-18.75225328	
P(T<=t) one-tail	1.72169E-66	
t Critical one-tail	1.646600932	
P(T<=t) two-tail	3.44339E-66	
t Critical two-tail	1.96268507	

The findings show that students who are depressed (M=2.34, SD=√1.46≈1.21) and those who are not (M=3.76, SD=√1.34≈1.16) have ***significantly different levels of academic pressure*** *(t(0.05, 873)=-18.75, p-value<0.001). We* ***reject the null hypothesis of equal means*** *since p-value is far less than our alpha. The* ***negative t-stat*** *shows that* ***depressed students reported significantly less academic pressure than their peers who were not depressed****. In this sample and at a 5% confidence level,* ***depression and perceived academic pressure appear to be inversely related****, as indicated by the very low significant difference (p-value<0.001 for both tests).*

6. Do students with suicidal thoughts and depression sleep less?

t-Test: Two-Sample Assuming Unequal Variances		
	Sleep Duration (Suicidal Thoughts, Depressed)	*Sleep Duration (No Suicidal Thoughts, Undepressed)*
Mean	5.907216495	5.820618557
Variance	9.767343213	8.859284741
Observations	97	485
Hypothesized Mean Difference	0	
df	133	
t Stat	0.251076211	
P(T<=t) one-tail	0.40107134	
t Critical one-tail	1.656391244	
P(T<=t) two-tail	0.80214268	
t Critical two-tail	1.977961264	

The suicidal thoughts & depressed (M=5.91, SD=$\sqrt{9.77}\approx1.94$) and no suicidal thoughts &not depressed (M=5.82, SD=$\sqrt{8.86}\approx1.96$) group means ***did not differ statistically significantly*** *in terms of sleep duration, according to the t-test results (t(0.05, 133)=0.25, p-value=0.80). We* ***do not have enough evidence to reject the null hypothesis*** *and thus* ***accept it*** *because both the one-tailed (p-value=0.40) and two-tailed (p-value=0.80) tests have p-values greater than our alpha. At a 5% confidence level,* ***suicidal thoughts and depression factors, do not appear to have a significant effect on sleep duration*** *in this sample, as seen by the almost* ***comparable means*** *(5.91 vs. 5.82) and* ***similar variances*** *(9.77 vs. 8.86).*

7. Are suicidal thoughts linked to higher financial stress?

t-Test: Two-Sample Assuming Unequal Variances		
	Financial Stress (Suicidal Thoughts)	*Financial Stress (No Suicidal Thoughts)*
Mean	2.781914894	3.350482315
Variance	2.154978723	1.957479638
Observations	376	622
Hypothesized Mean Difference	0	
df	761	
t Stat	-6.034129456	
P(T<=t) one-tail	1.24727E-09	
t Critical one-tail	1.646858406	
P(T<=t) two-tail	2.49454E-09	
t Critical two-tail	1.963086172	

Financial stress levels among those who have suicidal thoughts (M=2.78, SD=√2.15≈1.47) and those who do not (M=3.35, SD=√1.96≈1.40) ***differ statistically significantly****, according to the analysis (t(0.05, 761)=-6.03, p-value<0.001). We* ***reject the null hypothesis of equal means*** *since p-value is far less than our alpha. The* ***negative t-stat*** *shows that* ***those who reported suicidal thoughts had less financial stress than their peers****. At a 5% confidence level,* ***suicidal thoughts and financial stress appear to be inversely related*** *in this population, as indicated by the substantial difference (p<0.001 for both one- and two-tailed tests).*

8. Do suicidal students perform worse academically?

t-Test: Two-Sample Assuming Unequal Variances		
	CGPA (Suicidal Thoughts)	*CGPA (No Suicidal Thoughts)*
Mean	7.752891247	7.750385852
Variance	2.152533906	2.11385589
Observations	377	622
Hypothesized Mean Difference	0	
df	788	
t Stat	0.02625196	
P(T<=t) one-tail	0.489531509	
t Critical one-tail	1.646789633	
P(T<=t) two-tail	0.979063018	
t Critical two-tail	1.962979032	

The suicidal thoughts (M=7.75, SD=√2.15≈1.47) and no suicidal thoughts (M=7.75, SD=√2.11≈1.46) groups ***do not differ statistically significantly in their CGPA****, according to the t-test results (t(0.05, 788)=0.003, p-value=0.98). We* ***do not have enough evidence to reject the null hypothesis*** *and thus* ***accept it*** *because both the one-tailed (p-value=0.49) and two-tailed (p-value=0.98) tests have p-values greater than our alpha. The* ***virtually equal variances*** *(2.15 vs. 2.11) and* ***identical means*** *(7.75) show that these grouping variables, as described here, do not different CGPA outcomes in this sample. At a 5% confidence level, there is* ***no practical difference between the groups****.*

9. Does academic pressure impact how long students sleep?

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Level 1	157	895	5.700637	9.066859		
Level 2	164	917	5.591463	8.387289		
Level 3	263	1594	6.060837	9.423766		
Level 4	181	1055	5.828729	9.023281		
Level 5	233	1418	6.085837	9.078807		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	37.53672	4	9.38418	1.037572	0.386633	2.380894
Within Groups	8981.058	993	9.044369			
Total	9018.595	997				

The ANOVA results show ***no statistically significant difference in sleep duration means across the five academic pressure levels****, (F(0.05, 4, 993) = 1.04, p-value = 0.39). Although the group means varied slightly, from 5.57 hrs (Level 1) to 6.09 hrs (Level 5), the* ***high p-value*** *suggests that these changes were probably random. Additionally,* ***equality of variances among the different groups is guaranteed*** *as shown by the nearly identical variances (~9). The p-value is greater than our alpha (0.05) and F-stat < F crit (1.04 <2.38), and thus we* ***do not have enough evidence to reject the null hypothesis*** *(there is no significant difference between group means) and thus* ***accept it****. Therefore, we come to the conclusion that, at a 5% confidence level,* ***academic pressure levels do not affect how long students sleep.***

10. Does academic performance differ based on sleep duration range?

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
0-4 hrs	312	2406.73	7.713878	2.053221
5-6 hrs	210	1641.72	7.817714	2.081003
7-8 hrs	264	2034.86	7.707803	2.088006
8-12 hrs	212	1652.06	7.792736	2.355426

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.226492	3	0.742164	0.348039	0.790584	2.613857
Within Groups	2119.622	994	2.132416			
Total	2121.848	997				

The ANOVA results show ***no statistically significant difference in academic performance means across the different groups of sleep duration****, (F(0.05, 3, 994) = 0.35, p-value = 0.79). The* ***almost identical group means*** *(~8) verify that. Additionally,* ***equality of variances among the different groups is guaranteed*** *as shown by the nearly identical variances (~2). The p-value is greater than our alpha (0.05) and F-stat > F crit (0.35 < 2.61), and thus we* ***do not have enough evidence to reject the null hypothesis*** *(there is no significant difference between group means) and thus* ***accept it****. Therefore, we come to the conclusion that, at a 5% confidence level,* ***academic performance does not differ based on sleep duration.***

11. Are study hours different between stress levels?

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Level 1	189	1288	6.814815	15.75808		
Level 2	175	1223	6.988571	14.9309		
Level 3	194	1413	7.283505	14.88812		
Level 4	191	1351	7.073298	14.28934		
Level 5	249	1948	7.823293	12.90413		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	135.6168	4	33.90421	2.346271	0.052914	2.380894
Within Groups	14349.1	993	14.45025			
Total	14484.72	997				

The ANOVA results show ***no statistically significant difference in study hour means across the different groups of stress levels****, (F(0.05, 4, 993) = 2.35, p-value = 0.053). The* ***very close group means*** *(~7) verify that. Additionally,* ***equality of variances among the different groups is guaranteed*** *as shown by the very close variances (~14-15). The p-value is slightly greater than our alpha (0.05) and F-stat > F crit (2.35 < 2.38), and thus we* ***do not have enough evidence to reject the null hypothesis*** *(there is no significant difference between group means) and thus* ***accept it****. Therefore, we come to the conclusion that, at a 5% confidence level,* ***study hours does not differ based on stress levels.***

PREDICITVE ANALYTICS

SIMPLE LINEAR REGRESSION

1. Does sleep duration predict depression levels?

- **Dependent/Outcome (Y):** Depression
- **Independent/Predictor (X):** Sleep Duration

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.02192804							
R Square	0.000480839							
Adjusted R Square	-0.000522694							
Standard Error	0.493410946							
Observations	998							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	0.116650183	0.11665	0.479146	0.488970739			
Residual	996	242.4805442	0.243454					
Total	997	242.5971944						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.604352197	0.034361228	17.5882	1.64E-60	0.536923488	0.671780905	0.536923488	0.671780905
X Variable 1	-0.003596444	0.005195643	-0.6922	0.488971	-0.013792107	0.00659922	-0.01379211	0.00659922

1. Model Fit (Regression Statistics Table):

- *Multiple R (0.022): The correlation between sleep duration and depression levels is **extremely weak**.*
- *R Square (0.00048): Only **0.048%** of the variation in depression levels is explained by sleep duration.*
- *Adjusted R Square (-0.00052): The **negative value** indicates the model explains less variance than a horizontal line, suggesting **sleep duration is irrelevant for predicting depression** in this dataset. The **model is worse than just guessing** the average depression score.*
- *Standard Error (0.493): The average distance between observed depression scores and the regression line is ~0.49 units.*

2. ANOVA Table (Significance of the Model):

- *F-stat (0.479) with p-value (0.489): The* ***overall regression model is not statistically significant*** *(p-value > 0.05, we accept H_0). This confirms* ***sleep duration does not predict depression levels****.*
- *Sum of Squares (SS):*
    - *Regression (Explained Variation/SSR) | (0.117):* ***Minimal variance explained*** *by the model (~ 0.048%).*
    - *Residual (Unexplained Variation/SSE) | (242.48):* ***Almost all variance is unexplained*** *by the model (~99.95% noise or other predictors).*

3. Coefficients Table:

- *Intercept (0.604, p-value < 0.001): When* ***sleep duration (X) is zero, the predicted depression score (Y) is 0.60****.*
- *X Variable 1 (Sleep Duration):*
    - *Coefficient (-0.0036): Each* ***additional hour of sleep is associated with a minimal decrease of 0.0036 units in depression*** *(statistically non-significant).*
    - *p-value (0.489): Far above α = 0.05, we* ***accept H_0, there is no evidence that sleep duration affects depression****.*
    - *95% Confidence Interval (-0.014 to 0.0066): Includes zero, reinforcing the null effect.*

In this sample, ***depression levels are not predictable by sleep duration****. The* ***regression model has almost negligible explanatory power*** *(R2 = 0.00048) and is* ***statistically non-significant*** *(F(0.05, 1, 996) = 0.48, p-value = 0.49). At a 5% confidence interval,* ***there is no practical association between sleep duration and depression****, as indicated by the minimal coefficient (-0.0036), and this is confirmed by the confidence interval crossing zero.*

2. Does financial stress negatively affect academic performance?

- **Dependent/Outcome (Y):** CGPA
- **Independent/Predictor (X):** Financial Stress

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.00150544							
R Square	2.26635E-06							
Adjusted R Square	-0.001002756							
Standard Error	1.45982304							
Observations	997							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	0.00480564	0.004806	0.002255	0.962134562			
Residual	995	2120.427892	2.131083					
Total	996	2120.432697						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	7.754424251	0.110103523	70.42848	0	7.538362489	7.970486012	7.538362489	7.970486012
X Variable 1	-0.001512945	0.031860135	-0.04749	0.962135	-0.064033714	0.061007825	-0.06403371	0.061007825

1. Model Fit (Regression Statistics Table):

- *Multiple R (0.0015): The correlation between financial stress levels and CGPA is **extremely weak**.*
- *R Square (0.00000227): Only **0.00023%** of the variation in CGPA scores is explained by financial stress levels.*
- *Adjusted R Square (-0.001): The **negative value** indicates the model explains less variance than a horizontal line, suggesting **financial stress levels are irrelevant for predicting CGPA scores** in this dataset. The **model is worse than just guessing** the average CGPA score.*
- *Standard Error (1.46): The average distance between observed CGPA scores and the regression line is ~1.46 units.*

2. ANOVA Table (Significance of the Model):

- *F-stat (0.0023) with p-value (0.96): The* ***overall regression model is not statistically significant*** *(p-value > 0.05, we accept H_0). This confirms* ***financial stress levels do not predict CGPA scores/academic performance****.*
- *Sum of Squares (SS):*
    - *Regression (Explained Variation/SSR) | (0.0048):* ***Almost no variance explained*** *by the model (~ 0.00023%).*
    - *Residual (Unexplained Variation/SSE) | (2120.43): Al****l variance is unexplained*** *by the model (~100% noise or other predictors).*

3. Coefficients Table:

- *Intercept (7.75, p-value = 0): When* ***financial stress level (X) is zero, the predicted CGPA score (Y) is 7.75****.*
- *X Variable 1 (Financial Stress):*
    - *Coefficient (-0.0015): Each* ***additional financial stress level is associated with a very minimal decrease of 0.0015 units in CGPA score*** *(statistically non-significant).*
    - *p-value (0.96): Far above α = 0.05, we* ***accept H_0, there is no evidence that financial stress levels affects CGPA score/academic performance****.*
    - *95% Confidence Interval (-0.064 to 0.061): Includes zero, reinforcing the null effect.*

In this sample, ***CGPA scores are not predictable by financial stress levels****. The* ***regression model has almost negligible explanatory power*** *(R2 = 0.00000227) and is* ***statistically non-significant*** *(F(0.05, 1, 995) = 0.48, p-value = 0.0023). At a 5% confidence interval,* ***there is no practical association between financial stress levels and academic performance (CGPA scores)****, as indicated by the very minimal coefficient (-0.0015), and this is confirmed by the confidence interval crossing zero.*

MULTIPLE LINEAR REGRESSION

3. How well do sleep, stress, and satisfaction predict depression?

- **Dependent/Outcome (Y):** Depression
- **Independent/Predictors (X_1, X_2, X_3):** Sleep Duration, Financial Stress, Study Satisfaction

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.402509346							
R Square	0.162013773							
Adjusted R Square	0.159482093							
Standard Error	0.452149211							
Observations	997							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	39.24893345	13.08298	63.99456	7.85E-38			
Residual	993	203.0078369	0.204439					
Total	996	242.2567703						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.367843166	0.054971071	6.691577	3.68E-11	0.259970363	0.475715969	0.259970363	0.475715969
X Variable 1	-0.003287993	0.00476614	-0.68986	0.49044	-0.012640855	0.00606487	-0.01264086	0.00606487
X Variable 2	0.124678623	0.009878692	12.62096	5.63E-34	0.105293114	0.144064132	0.105293114	0.144064132
X Variable 3	-0.052675087	0.010332106	-5.09819	4.1E-07	-0.072950355	-0.03239982	-0.07295036	-0.03239982

1. Model Fit (Regression Statistics Table):

- *Multiple R (0.403): The correlation between financial stress levels and CGPA is **moderate**.*
- *R Square (0.162): **16.2%** of the variation in depression is explained by sleep duration, financial stress, and study satisfaction.*
- *Adjusted R Square (0.159): Shows modest but **meaningful explanatory power**.*
- *Standard Error (0.45): The average distance between observed depression and the regression line is ~0.45 units.*

2. ANOVA Table (Significance of the Model):

- *F-stat (63.99) with p-value (~0): The* ***overall regression model is highly statistically significant*** *(p-value < 0.05, we* ***reject H_0****). This confirms* ***sleep duration, financial stress, and study satisfaction predict depression****.*
- *Sum of Squares (SS):*
    - *Regression (Explained Variation/SSR) | (39.25):* ***Some variance is explained*** *by the model (~16.2%).*
    - *Residual (Unexplained Variation/SSE) | (~203):* ***Most variance is unexplained*** *by the model (~83.8% noise or other predictors).*

3. Coefficients Table:

- *Intercept (0.37, p-value ~0): When* ***sleep duration, financial stress, and study satisfaction (X) are zero, the predicted depression (Y) is 0.37****.*
- *X Variable 1 (Sleep Duration):*
    - *Coefficient (-0.003): Each* ***additional hour of sleep is associated with a very minimal decrease of 0.003 units in depression*** *(statistically non-significant).*
    - *p-value (0.49): Far above α = 0.05, we* ***accept H_0, there is no evidence that sleep duration alone affects depression****.*
    - *95% Confidence Interval (-0.013 to 0.48): Includes zero, reinforcing the null effect.*
- *X Variable 2 (Financial Stress):*
    - *Coefficient (0.125): Each* ***additional financial stress level is associated with an increase of 0.125 units in depression*** *(statistically significant).*
    - *p-value (~0): Below α = 0.05, we* ***reject H_0, there is evidence that financial stress levels alone affect depression****.*
    - *95% Confidence Interval (0.11 to 0.14): Does not include zero, reinforcing the null effect.*

- *X Variable 3 (Study Satisfaction):*
    - *Coefficient (-0.053): Each* ***additional study satisfaction level is associated with a minimal decrease of 0.053 units in depression*** *(statistically significant).*
    - *p-value (~0): Below α = 0.05, we* ***reject H_0, there is evidence that study satisfaction levels alone affect depression****.*
    - *95% Confidence Interval (-0.073 to -0.032): Does not include zero, reinforcing the null effect.*

In this sample, ***financial stress and study satisfaction significantly predict depression levels,*** *with* ***financial stress showing a stronger detrimental effect****.* ***Sleep duration has no measurable impact on depression****. The* ***regression model has some explanatory power*** *(R2 = 0.162) and is somewhat* ***statistically significant*** *(F(0.05, 3, 993) = 63.99, p-value ~0). At a 5% confidence interval,* ***there is a practical association between financial stress, study satisfaction and depression.*** *While statistically significant,* ***the model explains only about 16%*** *of depression variability, highlighting the need for broader investigations.*

4. Do gender and being a high school senior predict academic performance?

- **Dependent/Outcome (Y):** CGPA
- **Independent/Predictors (X_1, X_2):** Gender (F) , Degree Program (Class 12)

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.102133638							
R Square	0.01043128							
Adjusted R Square	0.008440196							
Standard Error	1.452921109							
Observations	997							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	22.11882701	11.05941	5.238996	0.005453077			
Residual	994	2098.31387	2.11098					
Total	996	2120.432697						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	7.89857077	0.065068956	121.3877	0	7.77088248	8.026259061	7.77088248	8.026259061
X Variable 1	-0.255280354	0.092861226	-2.74905	0.006085	-0.437506901	-0.07305381	-0.4375069	-0.073053808
X Variable 2	-0.181544194	0.113743176	-1.59609	0.110787	-0.404748505	0.041660117	-0.40474851	0.041660117

1. Model Fit (Regression Statistics Table):

- *Multiple R (0.102): The correlation between gender being female and degree program being Class 12 and depression is* ***weak****.*
- *R Square (0.01):* ***1%*** *of the variation in CGPA is explained by gender being female and degree program being Class 12.*
- *Adjusted R Square (0.0084): Shows* ***weak explanatory power****.*
- *Standard Error (1.45): The average distance between observed CGPA and the regression line is ~1.45 units.*

2. ANOVA Table (Significance of the Model):

- *F-stat (5.24) with p-value (~0.005): The* ***overall regression model is somewhat statistically significant*** *(p-value < 0.05, we* ***reject*** H_0*). This confirms* ***gender being female and degree program being Class 12 somewhat predict CGPA/academic performance****.*
- *Sum of Squares (SS):*
  - *Regression (Explained Variation/SSR) | (22.12):* ***Very minimal variance is explained*** *by the model (~1.04%).*
  - *Residual (Unexplained Variation/SSE) | (~2098.31):* ***Almost all variance is unexplained*** *by the model (~98.96% noise or other predictors).*

3. Coefficients Table:

- *Intercept (7.9, p-value = 0): When* ***gender being female and degree program being Class 12 (X) are zero, the predicted CGPA (Y) is 7.9****.*
- *X Variable 1 (Gender = F):*
  - *Coefficient (-0.255): Being a female student* ***is associated with a minimal decrease of 0.255 units in CGPA*** *compared to being male (statistically significant).*
  - *p-value (0.006): Far below α = 0.05, we* ***reject*** H_0***, there is evidence that gender being female alone affects CGPA****.*

- *95% Confidence Interval (-0.44 to -0.07): Does not include zero, reinforcing the null effect.*

- *<u>X Variable 2 (Degree Program = Class 12):</u>*
  - *Coefficient (-0.18): Being a class 12 student **is associated with a minimal decrease of 0.18 units in CGPA** compared to other degree programs (non-statistically significant).*
  - *p-value (0.11): Above α = 0.05, we **accept H_0, there is no evidence that the degree program being class 12 alone affects CGPA**.*
  - *95% Confidence Interval (-0.405 to 0.042): Includes zero, reinforcing the null effect.*

*In this sample, **gender being female significantly predicts academic performance (CGPA)**. **Degree program being Class 12 (senior high schooler) has no measurable impact on academic performance (CGPA)**. The **regression model has minimal explanatory power** (R2 = 0.102) and is **weakly statistically significant** (F(0.05, 2, 994) = 5.24, p-value = 0.005). At a 5% confidence interval, **there is a weak association between gender being female and academic performance (CGPA).** While statistically significant, **the model explains only about 1%** of CGPA variability, highlighting the need for broader investigations.*

CONCLUSION

EXECUTIVE SUMMARY:

Students' mental health issues are caused by a complex interaction of academic, economical, psychological, and demographic factors, according to the Student Depression Dataset analysis. The study found that **financial stress and study satisfaction were the most significant predictors of depression** among students, with both exhibiting statistically significant connections; even though **over 58% of students reported having depression** and **over 60% reported having suicidal thoughts**. It's interesting to note that there was **no significant or reliable association between depression or suicidal thoughts and typical measures of academic achievement**, such as CGPA and study hours.

Age has shown a little negative link with depression in terms of demographics, indicating that **younger students may be at greater risk**. However, **depression levels were not significantly impacted by gender or occupation**, suggesting that stress experiences are more uniformly shared by all groups. In this sample, sleep duration—another widely accepted factor—also showed very little predictive effects.

Significantly, multiple regression results show that although financial stress and study satisfaction with studies are significant predictors of depression, they only account for about 16% of its variability. This shows that **mental health is impacted by a wide range of interrelated factors** that are not included in this dataset.

ACTION PLAN:

Universities and schools should give priority to mental health initiatives that extend beyond academic support in light of these findings. These interventions should include financial aid, programs for emotional well-being, and easily available mental health services. To further understand and address the increased rates of depression among students, future studies should include qualitative data, social support networks, and longitudinal tracking.