Student Depression Analysis

Final Report

Why This Project Matters ?

◇ *Rising student depression rates are a global concern, especially in higher education.*

◇ *Data-driven insights can help identify academic, lifestyle, and financial risk factors.*

◇ *This project explores patterns in student well-being and suggests actionable recommendations.*

Agenda

This project is important because student mental health is a critical issue. wanted to use data to move beyond assumptions and find concrete, actionable insights.

By analyzing responses from over 27,000 students, this project does three important things

Identifies Real Drivers

- This analysis demonstrates how lifestyle and experience are more important than grades!

Pinpoints Vulnerable Groups

- Financial Stress combined with Academic Stress is an Amplifier of Depression Rates (85% – 95%)

From Insights to Action

- Academic Support
- Financial & Social Support
- Lifestyle Interventions

Contents

Questions & Hypotheses

Q1.What is the relationship between CGPA and the likelihood of depression?

•*Hypothesis : Students with lower academic performance (CGPA) are more likely to report depression.*

Q2.How do 'Academic Pressure' and 'Study Satisfaction' relate to depression?

•*Hypothesis: Higher academic pressure and lower study satisfaction are associated with higher rates of depression.*

Q3.Which lifestyle factors (Sleep Duration, Dietary Habits) show the strongest correlation with depression?

•*Hypothesis : Poorer lifestyle habits, such as insufficient sleep and unhealthy diets, are linked to a higher prevalence of depression.*

Q4.What is the impact of a family history of mental illness on student depression?

•*Hypothesis : Students with a family history of mental illness will have a higher rate of depression.*

Q5.Does 'Financial Stress' amplify the negative effect of 'Academic Pressure' on depression?

•*Hypothesis : Yes, the negative impact of academic pressure is more severe for students also experiencing high financial stress.*

Q6.Which academic degrees are associated with the highest rates of depression?

•*Hypothesis : Certain high-stress fields of study will show higher depression rates than others.*

Q7. Among students with suicidal thoughts, what's the depression prevalence?

Hypothesis : Students reporting depression are more likely to report suicidal thoughts.

Approach and Analysis

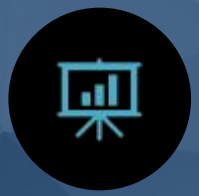
1. Load dataset from CSV (Kaggle source)

2. Clean and preprocess: handle missing values, encode categories

3. Identify key features: Sleep, Dietary Habits, Financial and Academic Stress, CGPA

4. Perform EDA using Python Libraries

5. Apply statistical analysis to verify key results(chi-square tests)



6. Use visualization for trend discovery (bar charts, boxplots, heatmaps)

7. Prepare a final report with all the key findings and recommendations

Technical Challenges

Self-reported data → risk of bias

Potential class imbalance (Depressed vs Not)

Encoding categorical variables (Sleep Duration, Dietary Habits, Family History)

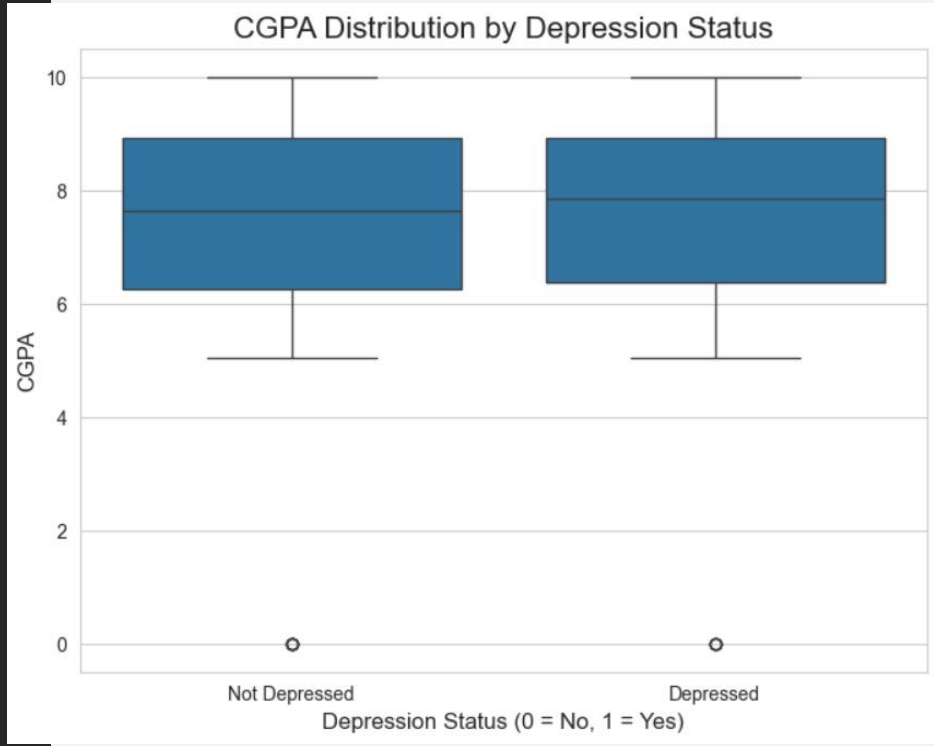
Data formatting issues → e.g., extra quotes around Sleep Duration values ('5-6 hours') created parsing problems.

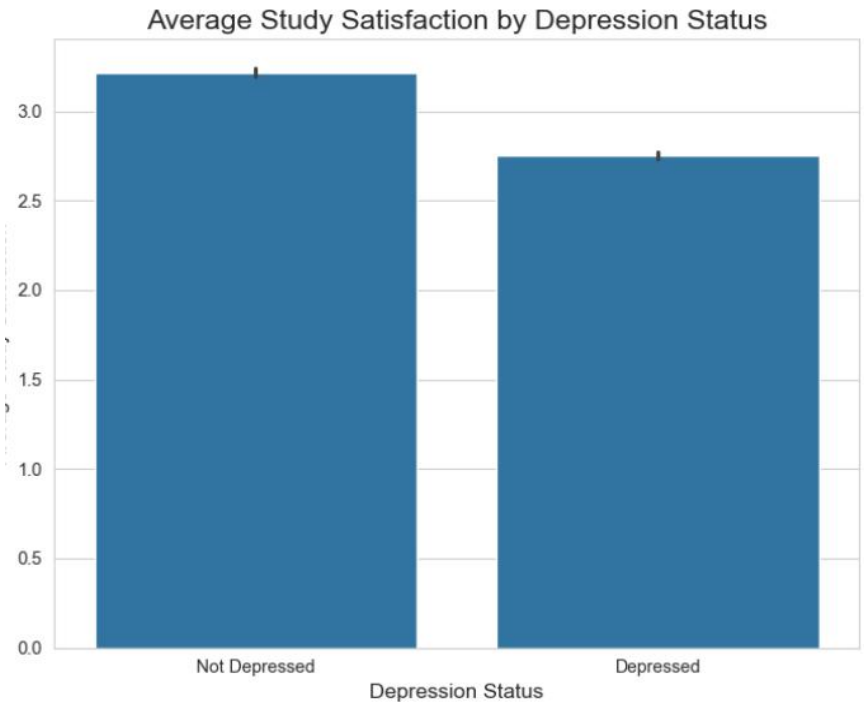
Handling high-cardinality features (Cities, Degrees)

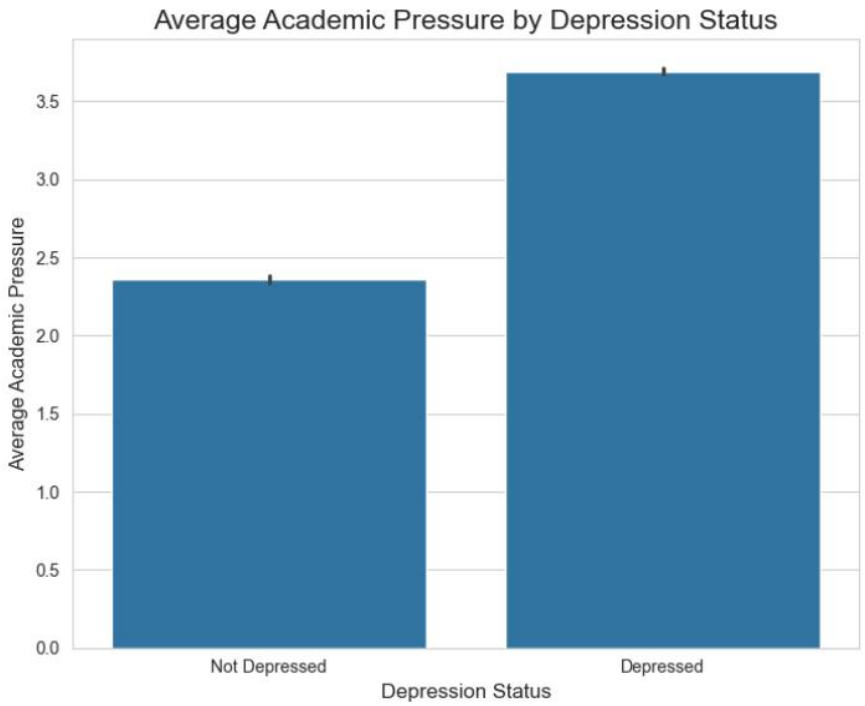
Power BI integration → building DAX measures (e.g., depression rate %) and risk factor KPIs was time-consuming but crucial for storytelling.

Q1: CGPA and Depression

- This plot gives us a clear answer to our first question.
- Insight: The data suggests that students at the higher end of the CGPA scale are also reporting depression.
- **CGPA Doesn't play a significant role in causing depression!**



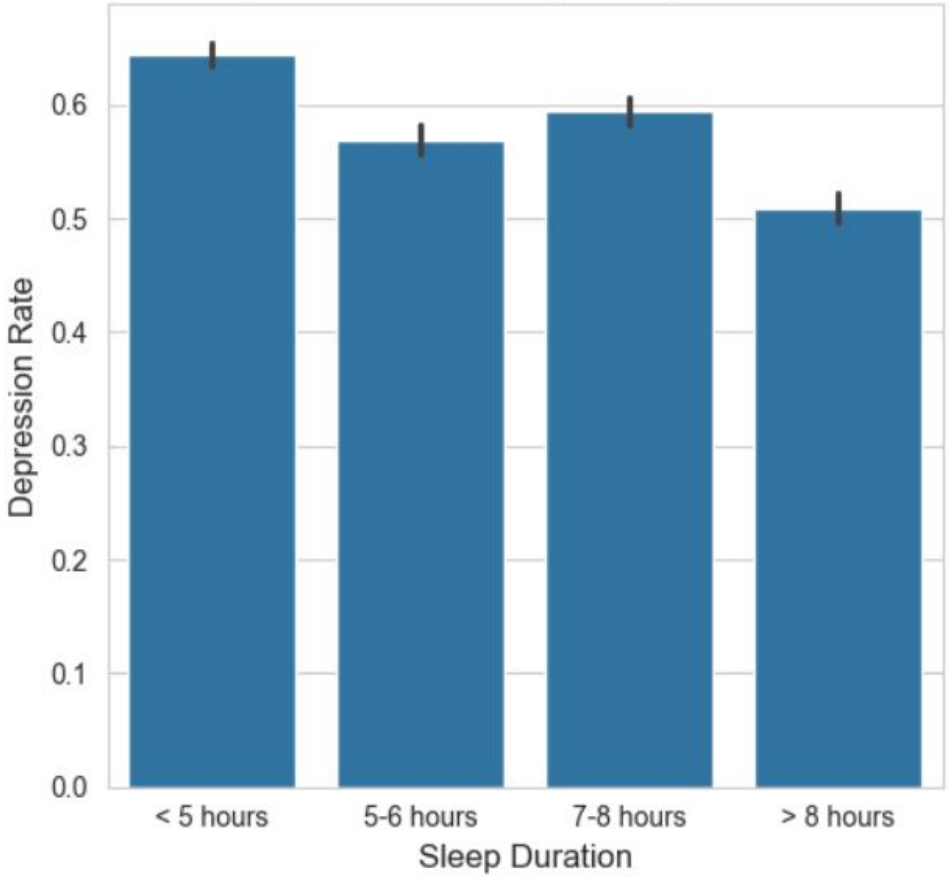


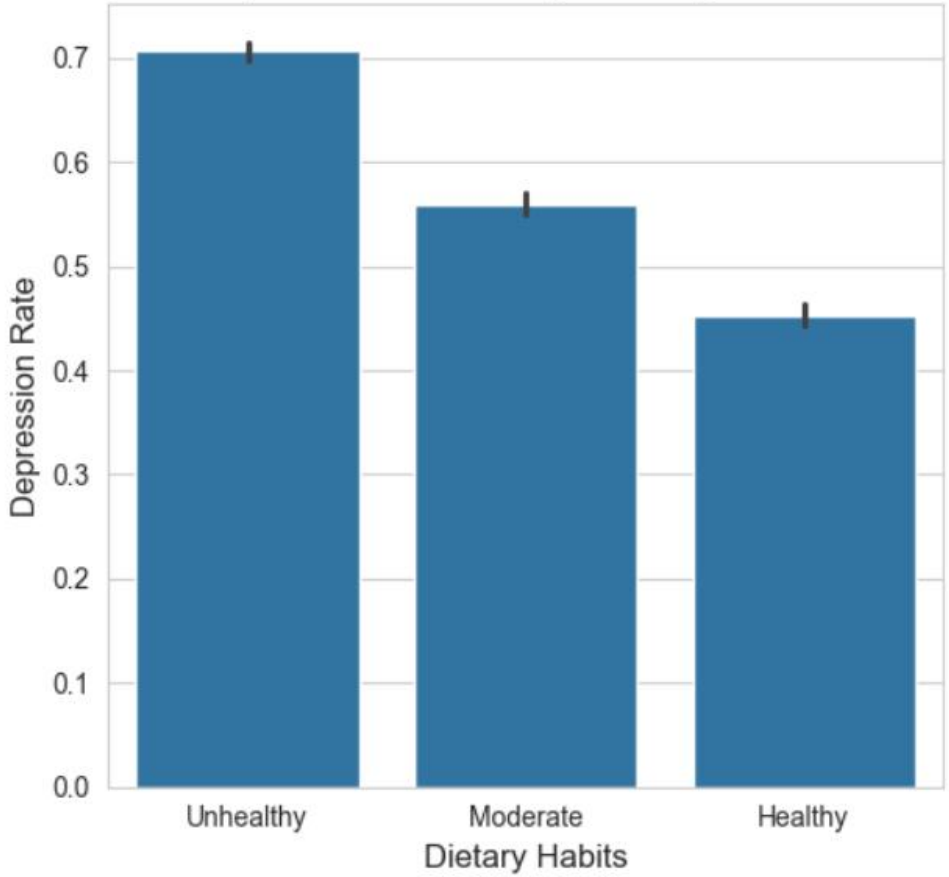


Q2: Academic Pressure vs Study Satisfaction

- Depressed students show a notably lower satisfaction with their studies, averaging 2.75 out of 5, while their non-depressed peers are higher at 3.21

- Students who report depression experience significantly higher levels of academic pressure



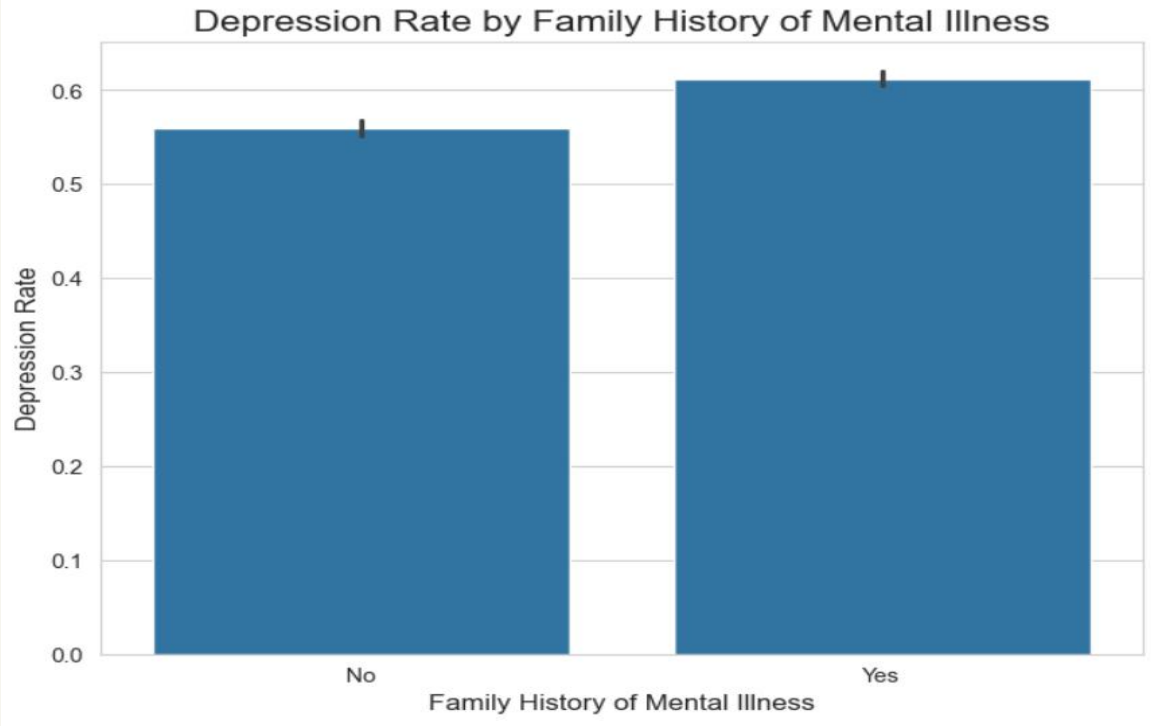


Q3: Lifestyle factors

- Students with unhealthy eating habits show the highest rate of depression. **70.7%** of students in this group report being depressed (**7,289 out of 10,309**)

- The depression rate improves for those getting **less than 5 hours of sleep**, dropping to around **64%**

Q4: Family history of Mental Illness

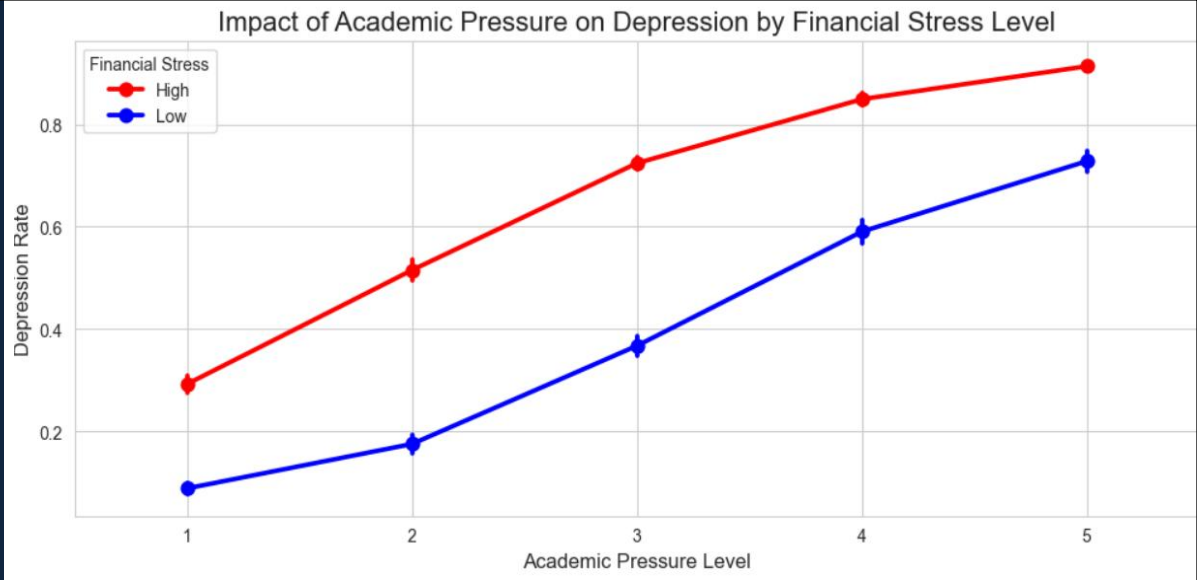


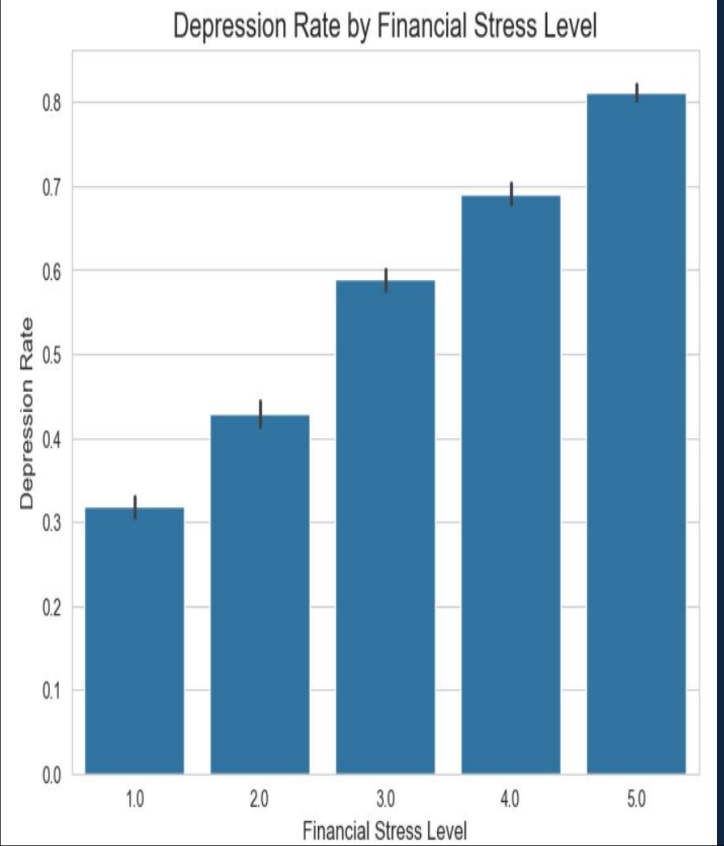
Students with a family history of mental illness have a depression rate of **61.2%**

Students without a family history have a depression rate of **56.0%**.

	Non-Depressed	Depressed	Total	Depressed %	Non-Depressed %
No Family History	6334	8050	14384	55.964961	44.035039
Family History	5228	8258	13486	61.233872	38.766128

Q5. Financial Stress





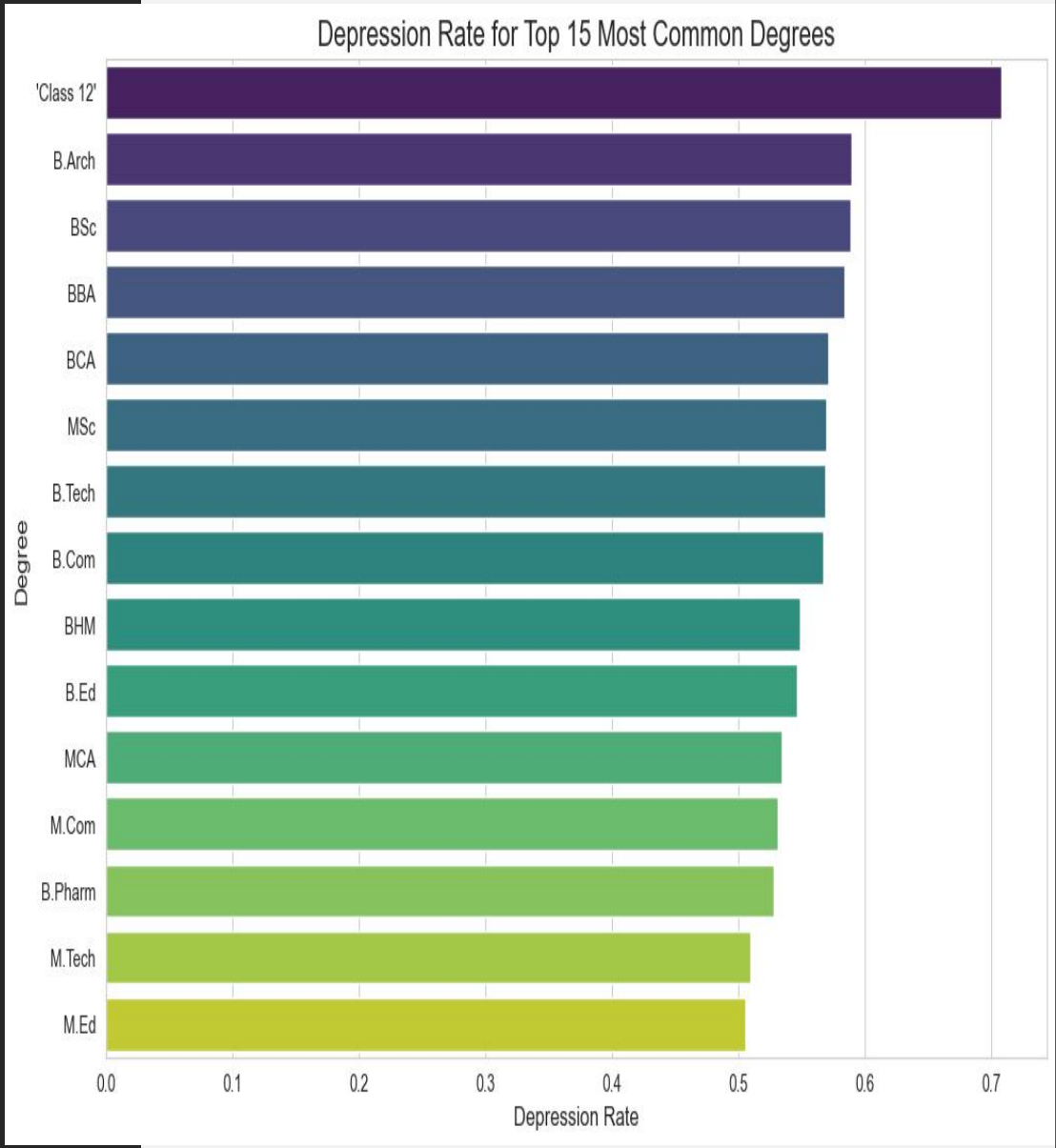
- **Financial Stress Significantly Amplifies the Impact of Academic Pressure**
- **At every single level of academic pressure, the red line (High Financial Stress) is significantly above the blue line (Low Financial Stress).**
- Students with high financial stress have a much higher baseline rate of depression, regardless of how much academic pressure they feel.

Q6. Academic Degrees

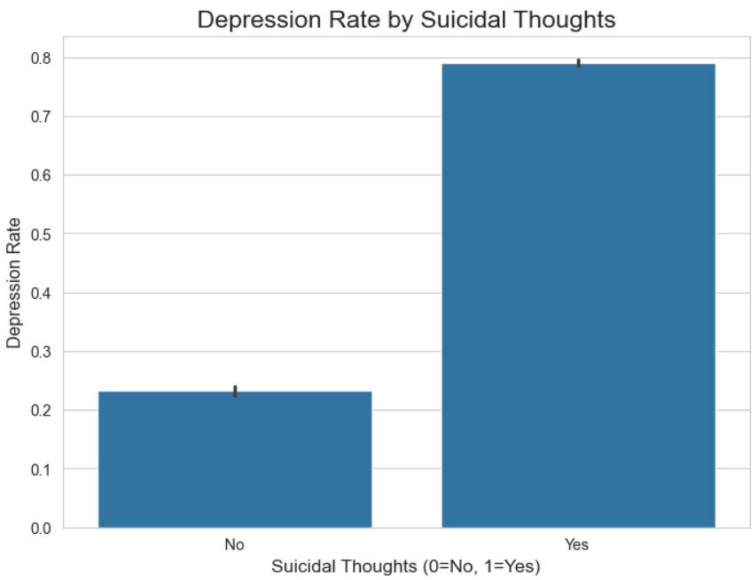
Class 12' has the highest depression rate among the top 15 most common degrees.

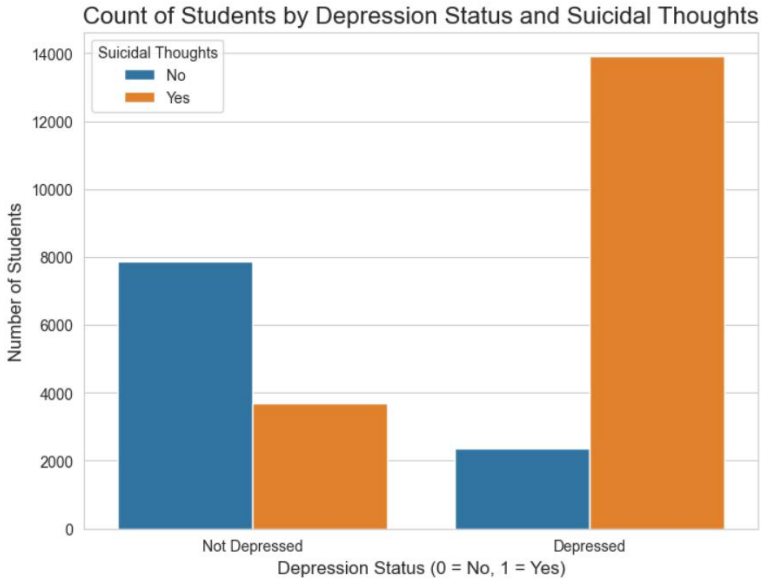
Several Bachelor's degrees like B.Arch, BSc, BBA, and BCA show relatively high depression rates, clustering together.

The **depression rates for the top 15 degrees generally range from around 50% to over 70%**



Q7. Suicidal Thoughts





The orange bar ('Yes' suicidal thoughts) is much taller for the 'Depressed' group than the 'Not Depressed' group.

Similarly, the blue bar ('No' suicidal thoughts) is much taller for the 'Not Depressed' group.

Out of the 17,631 students with suicidal thoughts, 13,934 are depressed.

- **79.03%** of depressed students have suicidal thoughts
- Only **23.19%** of non-depressed students have suicidal thoughts

Deeper Analysis

We performed additional statistical and SQL-based analyses to validate patterns and uncover hidden relationships:

1. Statistical Validation (Python)

Chi-square tests showed strong associations between **sleep duration**, **financial stress**, and **depression**.

A sample for the **chi - square tests for (sleep duration vs depression) and Work Hours Vs Depression is shown in the two snapshots.**

Null Hypothesis (H_0): 👉 Sleep Duration and Depression are independent.

Alternative Hypothesis (H_1): 👉 Sleep Duration and Depression are not independent..

```
contingency_table = pd.crosstab(df_final_eda['Sleep Duration'], df_final_eda['Depression'])
chi2, p, dof, expected = chi2_contingency(contingency_table)
```

RESULT

Chi2 Statistic: ~ 276.83

That's a huge Chi-square value. It means the differences between your observed counts and what you'd expect under independence are very large.
Translation: your observed distribution of depression varies quite a bit depending on sleep duration.

p-value: 1.03e-59

A p-value this small → reject the null hypothesis.

Interpretation: There is a significant association between Sleep Duration and Depression.

how long someone sleeps is not random with respect to whether they're depressed.

Null Hypothesis (H_0):
Study Hours Category and Depression are **independent** (no relationship).

Alternative Hypothesis (H_1):
Study Hours Category and Depression are **not independent** (there is a relationship).

```
contingency_table = pd.crosstab(df_final_eda['Study_Hours_Category'], df_final_eda['Depression'])
chi2, p, dof, expected = chi2_contingency(contingency_table)
```

Test Results:

- Chi^2 Statistic = **1116.29**
- p-value = **1.06×10^{-241}** (essentially 0)
- Degrees of Freedom = **3**

Conclusion:

Since the p-value is far below 0.05, we **reject the null hypothesis**.
There is a **very strong and statistically significant association** between study hours and depression.
The distribution of depression differs greatly across study hour categories.

```sql
SELECT
    Degree,
    COUNT(*) AS number_of_students,
    AVG(CASE WHEN Depression = 1 THEN 1 ELSE 0 END) AS depression_rate,
    AVG("Academic Pressure") AS avg_academic_pressure,
    AVG(CGPA) AS avg_cgpa,
    AVG(CASE
        WHEN "Sleep Duration" = "'Less than 5 hours'" THEN 0
        WHEN "Sleep Duration" = "'5-6 hours'" THEN 1
        WHEN "Sleep Duration" = "'7-8 hours'" THEN 2
        WHEN "Sleep Duration" = "'More than 8 hours'" THEN 3
    END) AS avg_sleep_score
FROM
    students_depression_raw
WHERE
    -- We only analyze students from the top 15 most
    -- common degrees for statistical significance
    Degree IN (SELECT Degree FROM students_depression_raw
               GROUP BY Degree ORDER BY COUNT(*) DESC LIMIT 15)
    AND Profession = 'Student'
GROUP BY
    Degree
ORDER BY
    depression_rate DESC;
```

	Degree	number_of_students	depression_rate	avg_academic_pressure	avg_cgpa	avg_sleep_score
0	'Class 12'	6080	0.707730	3.359375	7.594184	1.423026
1	B.Arch	1477	0.589032	3.062288	7.615579	1.431864
2	BSc	885	0.588701	3.045198	7.578531	1.425339
3	MBBS	695	0.579856	3.153957	7.550345	1.369784
4	BCA	1432	0.571229	3.126397	7.770803	1.321229
5	MSc	1186	0.569140	3.056492	7.513524	1.370464
6	B.Tech	1152	0.568576	3.078993	7.636513	1.400174
7	B.Com	1506	0.566401	3.083001	7.671826	1.297674
8	BHM	922	0.548807	3.131236	7.647213	1.419740
9	B.Ed	1864	0.546137	3.116953	7.852017	1.349436
10	MCA	1042	0.534549	2.952015	7.560768	1.497598
11	M.Com	733	0.530696	2.959072	7.790177	1.373806
12	B.Pharm	809	0.527812	3.185414	7.710148	1.454883
13	M.Tech	1022	0.509785	2.982387	7.592896	1.425490
14	M.Ed	821	0.505481	3.045067	7.707491	1.385366

Deeper Analysis

2. SQL Based Aggregations

Top 15 degrees analyzed: depression rates varied significantly across fields of study.

Many other such aggregations are performed while doing the deeper analysis.

Results - What the Data Revealed ?

Academic Factors	High **academic pressure** strongly linked with higher depression. Surprisingly, **CGPA** had a weak correlation with depression → performance ≠ mental health.
Lifestyle Factors	Students with **<5 hours of sleep** and an **unhealthy diet** had depression rates in the range of **70% - 80%**. Balanced sleep & diet reduced risk significantly.
Financial	**Financial stress** amplifies depression risk. Financial condition of a student act as a critical amplier and when combined with academic stress can increase the depression rates in the range of 85%-95%.
Family history of mental illness	Students with a family history of mental illness have a depression rate of **61.2%** Students without a family history have a depression rate of **56.0%**.
Suicidal thoughts	Suicidal thoughts are common in the depressed category. **79.03%** of depressed students have suicidal thoughts Only **23.19%** of non-depressed students have suicidal thoughts

Recommendations
Moving from Data to Action

Academic Support	Reduce unnecessary academic pressure by encouraging flexible grading and counseling. Provide **study satisfaction surveys** → act early if students feel overwhelmed.
Lifestyle Interventions	Promote **sleep hygiene programs** (awareness campaigns, workshops). Provide access to **healthy & affordable meals** on campus.
Financial & Social Support	Offer **financial aid schemes** or micro-grants for stressed students. Provide **free/affordable mental health counseling** services. Create **peer support groups** to reduce stigma.
Early Detection	Use surveys/dashboards to flag at-risk groups (e.g., *high stress + poor sleep* cluster). Run periodic screenings for depression and suicidal thoughts.
Workload Management	Promote **balanced workloads**, flexible deadlines, and stress management sessions.

THANK YOU