Student Depression Analysis Report

Data Science Workflow

1. Dataset Source and Description

The dataset used in this project was obtained from Kaggle:
Student Depression Dataset

It contains survey responses from students across India, covering multiple factors that may relate to depression. Key variables include:

- Academic pressure
- Financial stress
- Sleep duration
- CGPA
- Degree type
- Gender, age, and dietary habits
- Family history of mental illness
- Depression symptoms and labels (Yes/No)

---

2. Data Exploration and Cleaning (CSV to Excel)

The dataset was initially explored in CSV format using spreadsheet tools. Key steps included:

- Checking for **missing values**, **duplicate entries**, and **data type inconsistencies**
- Replacing or correcting invalid entries
- Standardizing categorical values (e.g., financial stress levels, stress levels, depression label)

- Verifying column headers and formatting

After cleaning, the dataset was saved in **Excel (.xlsx)** format to preserve structure and ease the transition into Python-based analysis.

---

3. Data Analysis and Visualization (Python - Jupyter Notebook)

The cleaned Excel file was imported into a Jupyter Notebook for further analysis using Python. The steps and key outputs included:

3.1 Libraries Used

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

3.2 Basic Exploration

- Displayed shape, column names, and value counts
- Checked for null and duplicate values (none present after initial cleaning)
- Generated descriptive statistics for numerical features like age, CGPA, sleep hours, and study time

[3]:
```
import pandas as pd

# Load the uploaded dataset
file_path = r"E:\student_depression_dataset\clean_student_depression_dataset.xlsx"
df = pd.read_excel(file_path)

# Display the first few rows and column names to understand the structure
df.head()
```

[3]:

	id	Gender	Age	City	Profession	Academic Pressure	Academic Pressure label	Work Pressure	Work Pressure2	CGPA	...	Dietary Habits	Degree	Have you ever had suicidal thoughts ?	Work/Study Hours	Work/Study Hours2
0	2	Male	33	Visakhapatnam	Student	5	Strongly Agree	1	Strongly Disagree	8.97	...	Healthy	B.Pharm	Yes	3	Neutral
1	8	Female	24	Bangalore	Student	2	Disagree	1	Strongly Disagree	5.90	...	Moderate	BSc	No	3	Neutral

```
#Basic Info
print("Shape:", df.shape)

print("\nInfo:\n")
print(df.info())
```

```
Shape: (27901, 25)

Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27901 entries, 0 to 27900
Data columns (total 25 columns):
 #   Column                                 Non-Null Count  Dtype
---  ------                                 --------------  -----
 0   id                                     27901 non-null  int64
 1   Gender                                 27901 non-null  object
 2   Age                                    27901 non-null  int64
 3   City                                   27901 non-null  object
 4   Profession                             27901 non-null  object
 5   Academic Pressure                      27901 non-null  int64
 6   Academic Pressure label                27901 non-null  object
 7   Work Pressure                          27901 non-null  int64
 8   Work Pressure2                         27901 non-null  object
 9   CGPA                                   27901 non-null  float64
```

```
# General descriptive stats
df.describe(include='number')
```

	id	Age	Academic Pressure	Work Pressure	CGPA	Study Satisfaction	Job Satisfaction	Work/Study Hours	Financial Stress	Depression
count	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000	27901.000000
mean	70442.149421	25.822300	3.141214	1.000323	7.656104	2.944196	1.000681	7.156984	3.139744	0.585499
std	40641.175216	4.905687	1.381465	0.034390	1.470707	1.360504	0.044394	3.707642	1.437318	0.492645
min	2.000000	18.000000	0.000000	1.000000	0.000000	1.000000	1.000000	0.000000	1.000000	0.000000
25%	35039.000000	21.000000	2.000000	1.000000	6.290000	2.000000	1.000000	4.000000	2.000000	0.000000
50%	70684.000000	25.000000	3.000000	1.000000	7.770000	3.000000	1.000000	8.000000	3.000000	1.000000
75%	105818.000000	30.000000	4.000000	1.000000	8.920000	4.000000	1.000000	10.000000	4.000000	1.000000
max	140699.000000	59.000000	5.000000	5.000000	10.000000	5.000000	5.000000	12.000000	5.000000	1.000000

```
# Check for duplicates
duplicates = df.duplicated()
print("\nDuplicate rows (True indicates a duplicate):")
display(duplicates)
```

```
Duplicate rows (True indicates a duplicate):
0        False
1        False
2        False
3        False
4        False
         ...
27896    False
27897    False
27898    False
27899    False
27900    False
Length: 27901, dtype: bool
```

```
df["Depression label"].value_counts()
```

```
Depression label
YES    16336
NO     11565
Name: count, dtype: int64
```

3.3 Correlation Heatmap

To identify how numerical features relate to each other and depression symptoms:

Python Code:

import seaborn as sns

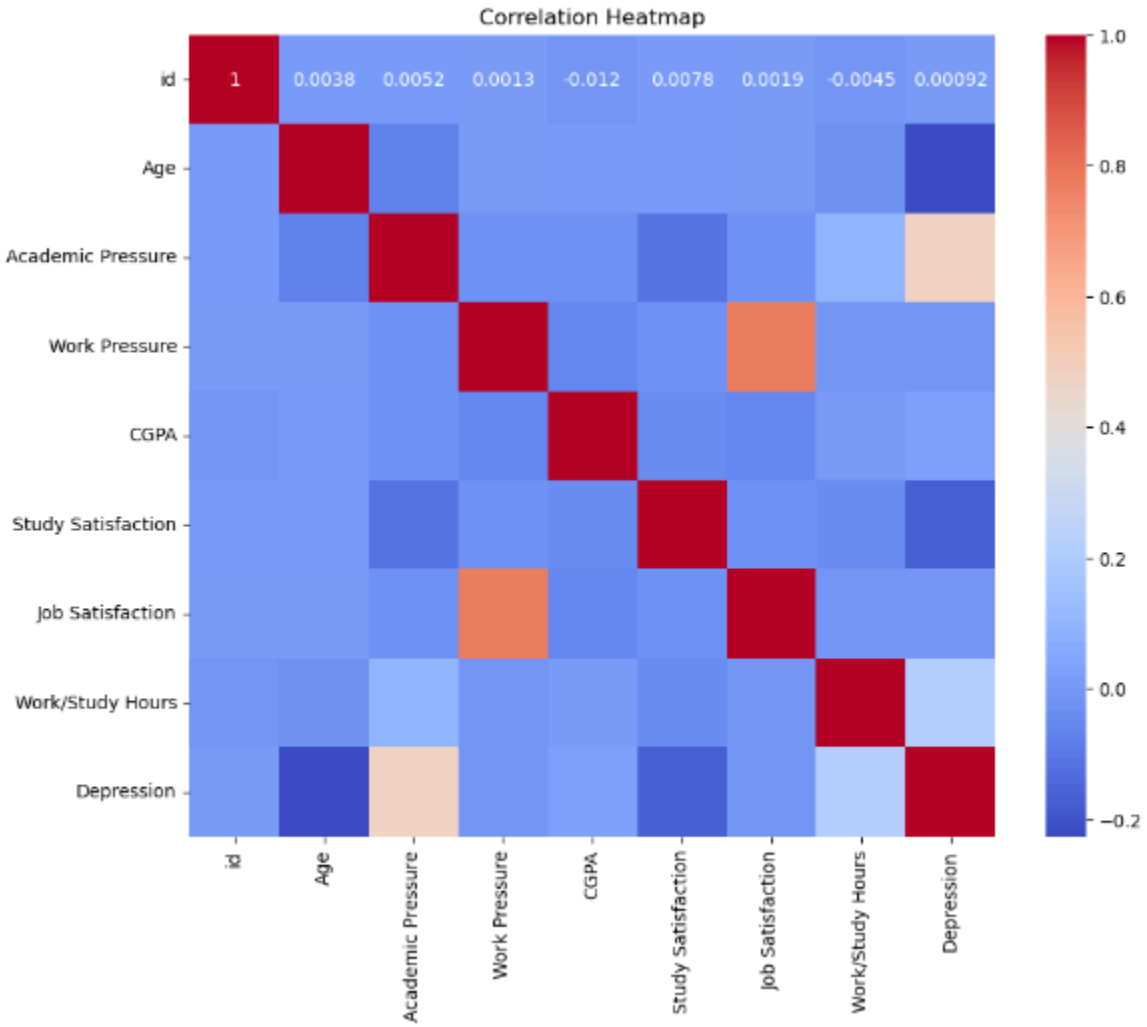
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 8))

sns.heatmap(correlation, annot=True, cmap='coolwarm')

plt.title("Correlation Heatmap")

plt.show()



- **Insights from Correlation heatmap:**

1. Depression & Academic Pressure

- **Correlation: Positive**
- Students who experience higher academic pressure are more likely to report symptoms of depression.
- This may be attributed to **high expectations** from themselves, parents, or faculty, as well as fear of failure and **constant comparisons** with peers. **Competitive environments,** such as merit-based scholarships and limited seating, can also contribute to this issue.

2. Depression & Study Satisfaction

- **Correlation: Negative**
- Students who are more satisfied with their studies generally experience lower levels of depression.
- Higher satisfaction may lead students to **feel more in control** of their education and enjoy the learning process. They often benefit from **better support from faculty** and engage in more interesting coursework. Additionally, a **sense of achievement** helps protect against feelings of low mood.

3. Depression & Work/Study Hours

- **Correlation: Positive**
- Longer study/work hours are linked with higher depression
- Long work hours leave little time for **self-care, rest, or socializing**. Students often struggle to balance their **studies with part-time jobs** due to financial stress. **This can lead to burnout** from an unbalanced life.

4. Depression & CGPA

- **Correlation: Weak Positive**
- CGPA doesn't show a strong relation with depression in this dataset. CGPA had a weak but present inverse relationship with depression level.
- Mental health issues can impact **all students,** including those who perform well academically. Some students may maintain their grades while hiding certain problems, such as family issues or the constant pressure to achieve high marks.

---

5. Academic Pressure & Study Satisfaction

- **Correlation: Strong Negative**
- Students experiencing **greater academic pressure** often report **lower satisfaction with their studies**. Strategies to reduce this pressure, such as a balanced curriculum and flexible deadlines, can enhance satisfaction.

---

6. Academic Pressure & Financial Stress

- **Correlation: Moderate Positive**
- Students who face financial stress often feel more academic pressure. They may not have access to important resources like tutoring, textbooks, and technology. This lack of resources makes their situation even more stressful.

---

7. Study Satisfaction & Job Satisfaction

- **Correlation: Moderate Positive**
- Students who are satisfied with their study life are also more likely to feel content with their job or work life. A sense of stability or success in one area often influences other parts of life.

---

8. Sleep Duration & Study Satisfaction

- **Correlation: Weak to Moderate Positive**
- Students who sleep more tend to be more satisfied with their studies. Well-rested students may have better focus, motivation, and engagement in their academic life.

---

9. Work/Study Hours & Academic Pressure

- **Correlation: Moderate to High Positive**
- The longer students engage in study or work, the more academic pressure they report. This may indicate overwork or poor time management.

---

10. Financial Stress & Job Satisfaction

- **Correlation: Moderate Negative**
- Students experiencing **financial stress** are **likely to have lower job satisfaction**, possibly due to low-paying or overly demanding jobs.

---

11. CGPA & Study Satisfaction

- **Correlation: Weak Positive**
- Students who have **higher CGPAs** typically report **greater satisfaction** with their studies.

3.4 Histograms

Histograms are used to show the distribution of data by grouping values into ranges (bins) and displaying how frequently they occur.

Python Code:

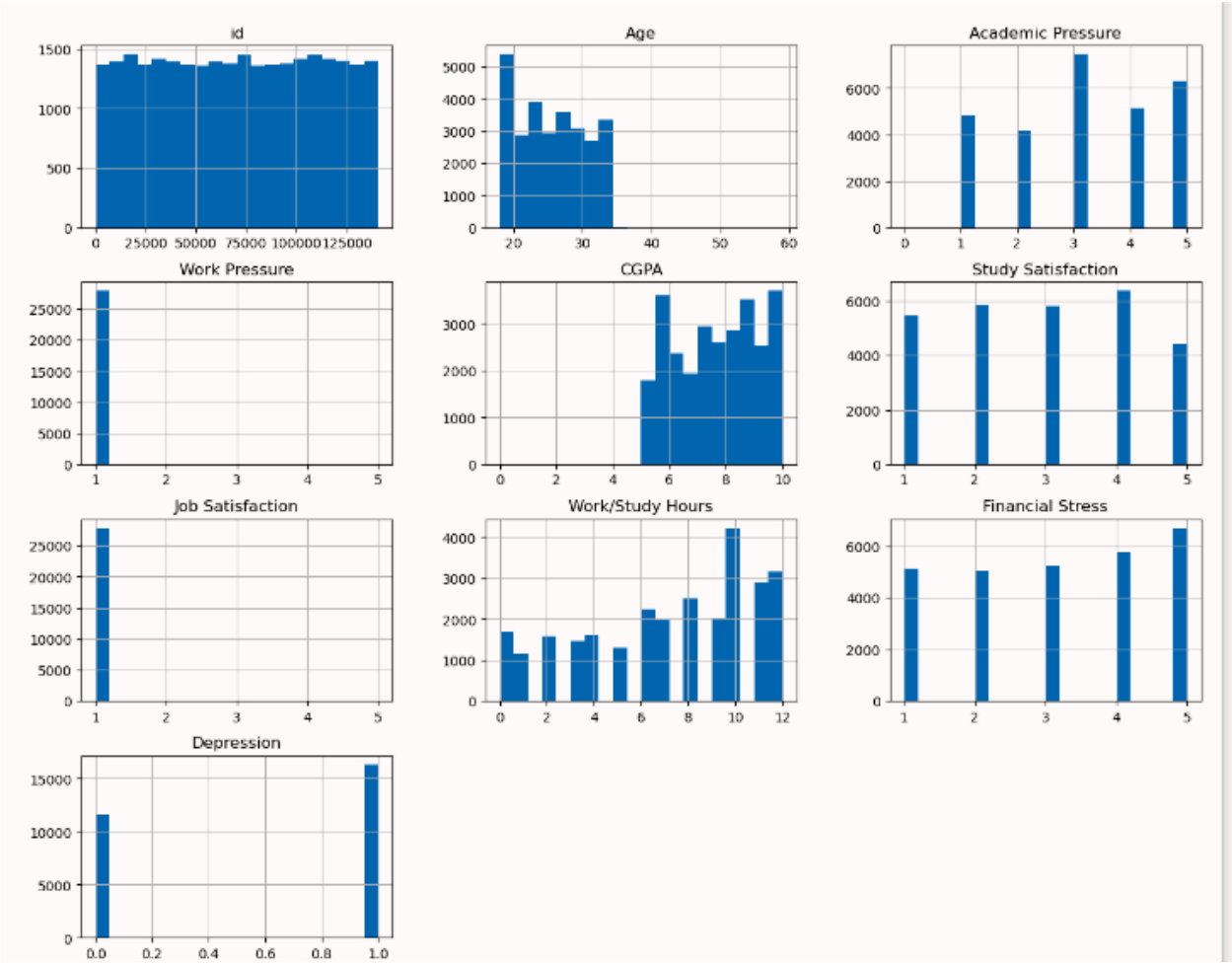
```
# Histograms (Numeric Columns)

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

df[numeric_cols].hist(figsize=(15, 12), bins=20)

plt.suptitle("Histograms of Numeric Columns")

plt.show()
```



Insights from the Histograms

1. Age
Most students are between 18–25 years old. Very few are above 30.
This shows the dataset mainly represents young students.

2. Academic Pressure
Values range from 1 to 5. Many students report medium to high levels of academic pressure.
Academic stress is common among students.

3. Work Pressure
Most students are at the lowest level (1). Very few have higher work pressure.
This suggests that students generally do not have much job-related stress.

4. CGPA
The majority of students have a CGPA between 6 and 10, with fewer students at very low CGPA levels.
Academically, many students are performing reasonably well.

5. Study Satisfaction
Responses are spread across the scale, but many students fall around 3 and 4.
This shows moderate levels of satisfaction with studies.

6. Job Satisfaction
Almost all values are at 1 (very low).
This may be because most students do not have jobs while studying.

7. Work/Study Hours
Most students study between 4 and 10 hours daily, with some going up to 12 hours.
This indicates a significant time commitment to studies.

8. Financial Stress
Values are spread across the scale, with many students reporting higher levels (3–5).
This shows that financial stress is a major issue for students.

9. Depression
The data is mostly split between 0 and 1 (0 = no depression, 1 = depression). A large portion of students fall into the depression category.
Depression is a significant concern in this dataset.

3.5 Skewness and Kurtosis:

Skewness measures the asymmetry of data distribution, showing whether values are more spread out on the left or right side of the mean.
Kurtosis measures the "tailedness" of a distribution, indicating whether data have heavy or light tails compared to a normal distribution.

```
# Skewness & Kurtosis
for col in numeric_cols:
    print(f"{col}: Skewness={df[col].skew():.2f}, Kurtosis={df[col].kurt():.2f}")
```

```
id: Skewness=-0.01, Kurtosis=-1.21
Age: Skewness=0.13, Kurtosis=-0.85
Academic Pressure: Skewness=-0.14, Kurtosis=-1.16
Work Pressure: Skewness=113.66, Kurtosis=13140.85
CGPA: Skewness=-0.11, Kurtosis=-1.02
Study Satisfaction: Skewness=0.01, Kurtosis=-1.23
Job Satisfaction: Skewness=74.11, Kurtosis=5927.00
Work/Study Hours: Skewness=-0.45, Kurtosis=-1.00
Financial Stress: Skewness=-0.13, Kurtosis=-1.32
Depression: Skewness=-0.35, Kurtosis=-1.88
```

Insights:

1. Age (Skewness = 0.13, Kurtosis = -0.85)

- Slightly right-skewed but very close to symmetric.
- Flatter than a normal distribution, with fewer extreme values.

2. Academic Pressure (Skewness = -0.14, Kurtosis = -1.16)

- Slight left skew, almost balanced.
- Flatter distribution, meaning responses are spread across the scale.

3. Work Pressure (Skewness = 113.66, Kurtosis = 13140.85)

- Extremely right-skewed with very high kurtosis.
- Data is highly concentrated at the lowest values, with a few extreme cases pulling the distribution far to the right.

4. CGPA (Skewness = -0.11, Kurtosis = -1.02)

- Almost symmetric with a slight left skew.
- Flatter than normal, values are spread without heavy tails.

5. Study Satisfaction (Skewness = 0.01, Kurtosis = -1.23)

- Perfectly symmetric.
- Flatter than normal, spread evenly across categories.

6. Job Satisfaction (Skewness = 74.11, Kurtosis = 5927.00)

- Extremely right-skewed with very high kurtosis.
- Almost all students report very low job satisfaction, with very few extreme responses creating long tails.

7. Work/Study Hours (Skewness = -0.45, Kurtosis = -1.00)

- Moderate left skew → more students are on the higher side of study hours.
- Flatter than normal, responses spread across the range.

8. Financial Stress (Skewness = -0.13, Kurtosis = -1.32)

- Nearly symmetric with slight left skew.
- Flatter distribution, students' responses cover the whole range.

9. Depression (Skewness = -0.35, Kurtosis = -1.88)

- Some left skew → more students reporting depression (value 1).
- Very flat distribution because data is concentrated in two categories (0 and 1) rather than spread.

3.6 Quantiles:

Quantiles

Quantiles are used to divide continuous variables into equal-sized groups, enabling the comparison of patterns across different ranges of values. This method is useful for examining how outcomes, such as depression, vary across levels of academic performance or study effort.

Python Code:

```python
import pandas as pd

# Load cleaned dataset
file_path = r"E:\student_depression_dataset\clean_student_depression_dataset.xlsx"
df = pd.read_excel(file_path)

# CGPA quantiles
df["CGPA_Quantile"] = pd.qcut(df["CGPA"], q=4, labels=["Lowest 25%", "25-50%", "50-75%", "Top 25%"])

# Study/Work Hours quantiles
df["Hours_Quantile"] = pd.qcut(df["Work/Study Hours"], q=4, labels=["Lowest 25%", "25-50%", "50-75%", "Top 25%"])

# Depression counts by CGPA quantile
cgpa_counts = df.groupby("CGPA_Quantile")["Depression label"].value_counts(normalize=True).unstack().fillna(0) * 100

# Depression counts by Study Hours quantile
hours_counts = df.groupby("Hours_Quantile")["Depression label"].value_counts(normalize=True).unstack().fillna(0) * 100

print("Depression % by CGPA Quantile:\n", cgpa_counts)
print("\nDepression % by Study Hours Quantile:\n", hours_counts)
```

```
Depression % by CGPA Quantile:
 Depression label         NO        YES
CGPA_Quantile
Lowest 25%        43.188612  56.811388
25-50%            43.273192  56.726808
50-75%            37.307581  62.692419
Top 25%           41.950798  58.049202

Depression % by Study Hours Quantile:
 Depression label         NO        YES
Hours_Quantile
Lowest 25%        56.459330  43.540670
25-50%            41.164783  58.835217
50-75%            33.019169  66.980831
Top 25%           31.909631  68.090369
```

Insights:

The quantile analysis reveals important patterns. For CGPA, students in the lowest 25% exhibit a depression rate of 56.8%. The percentage increases slightly in the middle ranges, reaching 62.7% in the 50–75% quantile, before declining to 58.0% in the top 25%. This indicates that depression is not restricted to low academic performers but is also present among middle and higher achievers. For study hours, a clear upward trend is observed: only 43.5% of students in the lowest 25% of study hours report depression, compared to 68.1% in the top 25%. This suggests that extended study hours are strongly associated with higher levels of depression.

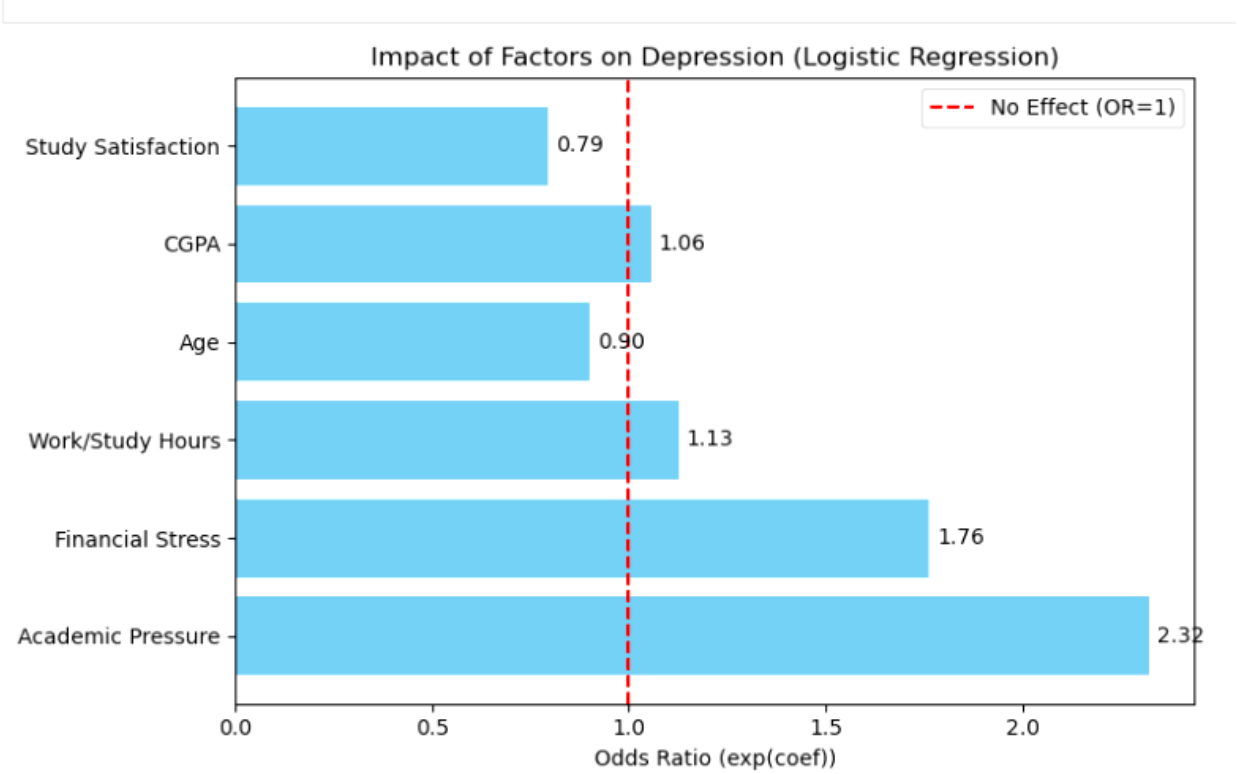
3.7 Logistic Regression

Logistic regression is used for binary outcomes, such as predicting whether a student is depressed (1) or not (0). It helps identify factors like academic pressure, financial stress, study hours, CGPA, study satisfaction, and age that are significantly linked to the likelihood of depression. The model provides coefficients, p-values, and odds ratios, showing the strength and direction of these relationships.

```
import statsmodels.api as sm

X = df[['Academic Pressure','Financial Stress','Work/Study Hours','Age','CGPA','Study Satisfact
X = sm.add_constant(X)
y = df['Depression']   # numeric 0/1
model = sm.Logit(y, X).fit()
print(model.summary())
```

```
Optimization terminated successfully.
         Current function value: 0.457489
         Iterations 6
                           Logit Regression Results
==============================================================================
Dep. Variable:             Depression   No. Observations:                27901
Model:                          Logit   Df Residuals:                    27894
Method:                           MLE   Df Model:                            6
Date:                Wed, 17 Sep 2025   Pseudo R-squ.:                  0.3257
Time:                        19:28:13   Log-Likelihood:                -12764.
converged:                       True   LL-Null:                       -18930.
Covariance Type:            nonrobust   LLR p-value:                     0.000
==========================================================================================
                             coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------
const                  -1.7514      0.137    -12.816      0.000      -2.019      -1.484
Academic Pressure       0.8424      0.013     66.158      0.000       0.817       0.867
Financial Stress        0.5670      0.011     49.698      0.000       0.545       0.589
Work/Study Hours        0.1188      0.004     28.023      0.000       0.111       0.127
Age                    -0.1058      0.003    -32.435      0.000      -0.112      -0.099
CGPA                    0.0546      0.011      5.169      0.000       0.034       0.075
Study Satisfaction     -0.2307      0.012    -19.898      0.000      -0.253      -0.208
==========================================================================================
```



Insights:

Academic Pressure (OR = 2.32, p < 0.001): Students with higher academic pressure are more than twice as likely to experience depression compared to those with lower academic pressure. This is the strongest predictor.

Financial Stress (OR = 1.76, p < 0.001): Students experiencing financial stress are 1.76 times more likely to be depressed, showing a strong positive effect.

Work/Study Hours (OR = 1.13, p < 0.001): Longer work or study hours slightly increase the chances of depression.

CGPA (OR = 1.06, p < 0.001): Higher CGPA is weakly but positively associated with depression, suggesting that academic achievement alone does not protect students from mental health challenges.

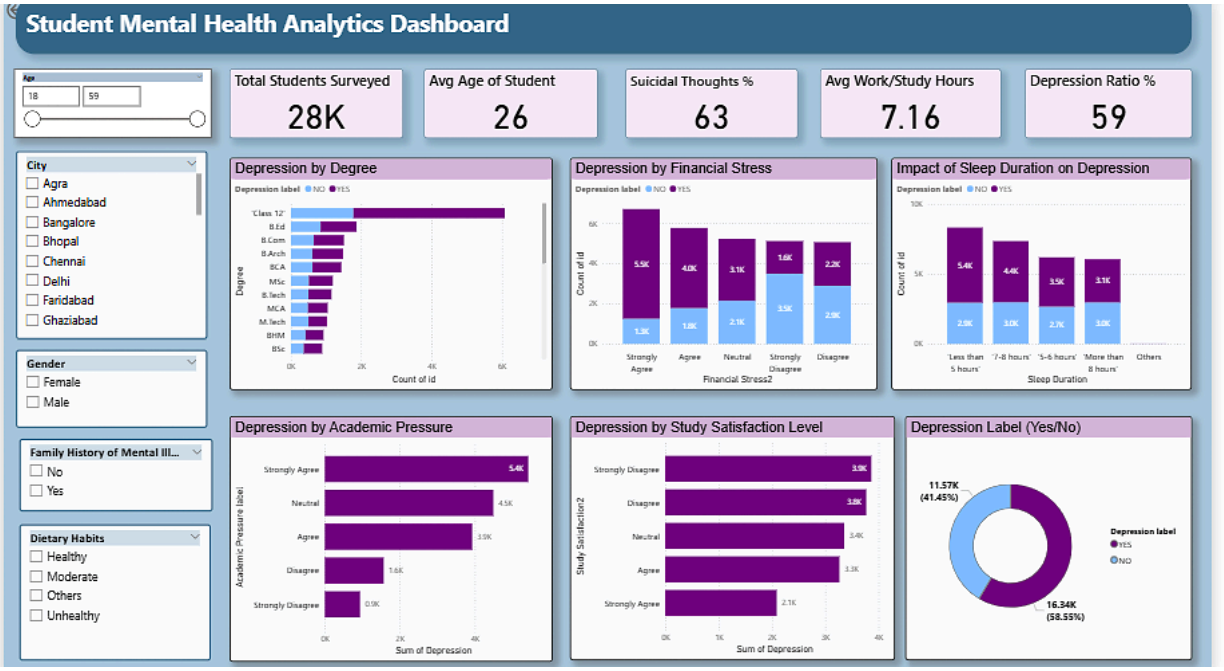
Age (OR = 0.90, p < 0.001): Older students are slightly less likely to experience depression compared to younger ones.

Study Satisfaction (OR = 0.79, p < 0.001): Higher study satisfaction reduces the likelihood of depression, indicating a protective effect.

4. Advanced Visualization and Insights (Power BI)

1. Dashboard Overview

The dashboard presents a comprehensive analysis of depression among 28,000 surveyed students, using various academic, personal, and demographic variables. The visuals explore correlations between depression and factors such as academic pressure, financial stress, sleep duration, satisfaction with studies, and degree programs. Interactive slicers (filters) allow for breakdowns by **city, gender, profession, and sleep duration**.



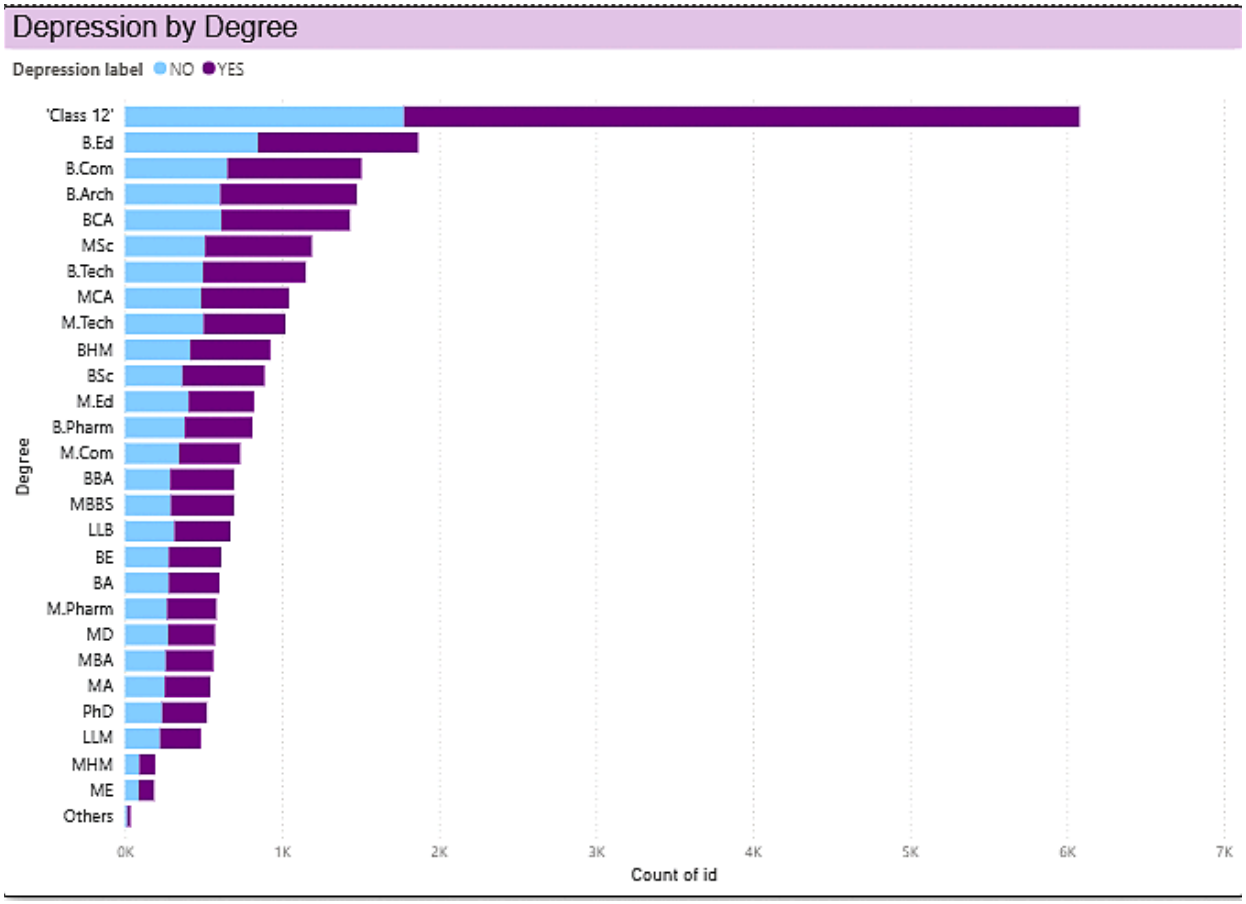
5. Visual-wise Insight Breakdown

- **KPI Cards (Top Metrics)**

- **Total Students Surveyed:** 28,000
- **Average Age:** 26 years
- **Depression Ratio:** 59%
- **Average Work/Study Hours:** 7.16 hours/day
- **Suicidal Thoughts:** 63%

Insight: Most **students (59%) report experiencing depression**, and **63% have had suicidal thoughts**, highlighting serious mental health issues. The average workload exceeds 7 hours, which could add to mental stress.

Depression by Degree

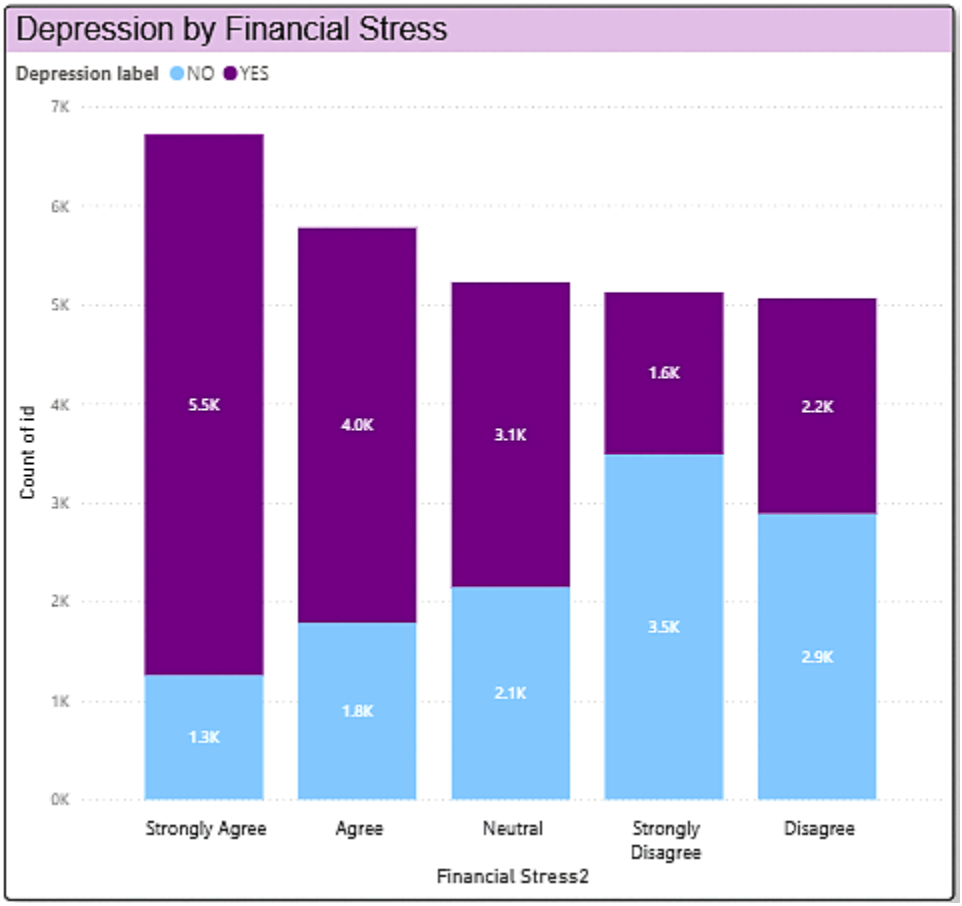


Visual Type: Clustered Column Chart
Observation:

- **B.Tech**, **B.Sc.**, and **M.Tec.h** students report **higher levels of depression**
- Students in **BHM(Bachelor of Hotel Managment), BCA(Bachelor of Computer Applications) , and B.Com(Bachelor of Commerce)** show comparatively lower depression

Insight: Technical degrees correlate with higher depression rates, likely due to heavy workloads or competitive environments.

Depression by Financial Stress

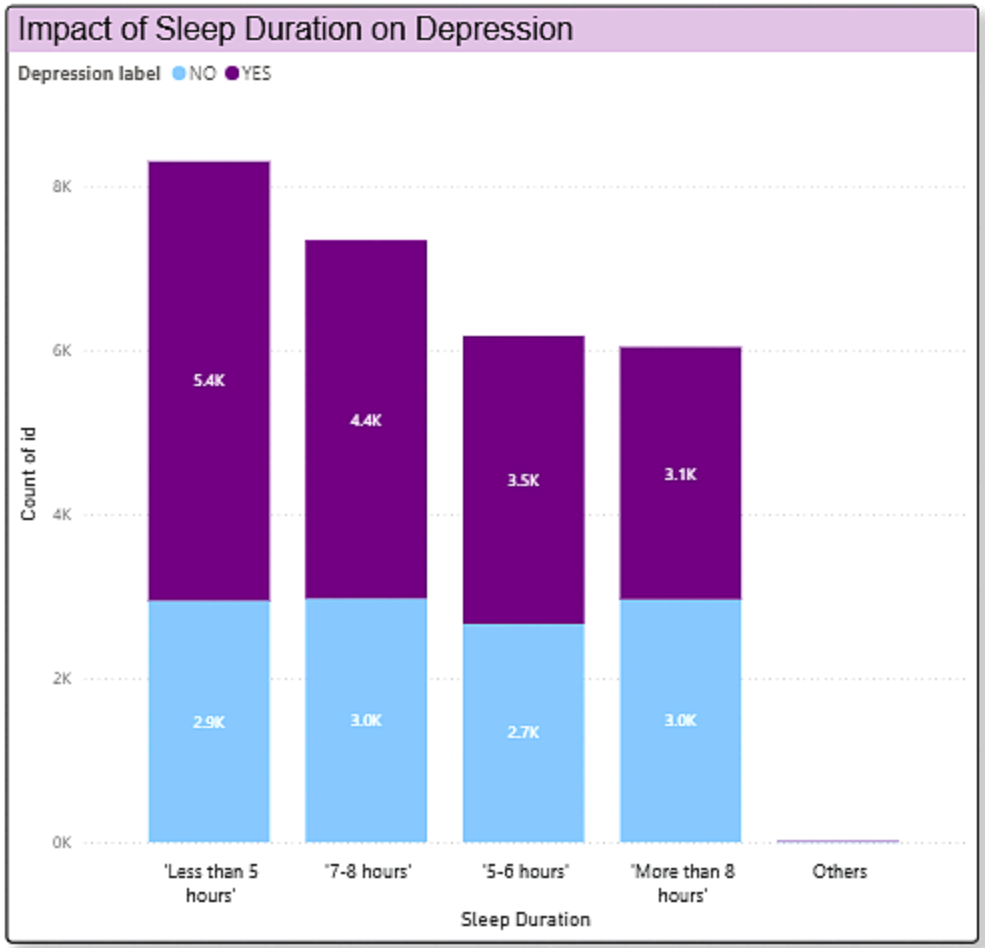


Visual Type: Clustered Column Chart
Observation:

- **Strongly Agree** (Financial Stress): 5.5K depressed
- Depression decreases as stress decreases

Insight: Financial stress is a strong predictor of depression among students.

Impact of Sleep Duration on Depression

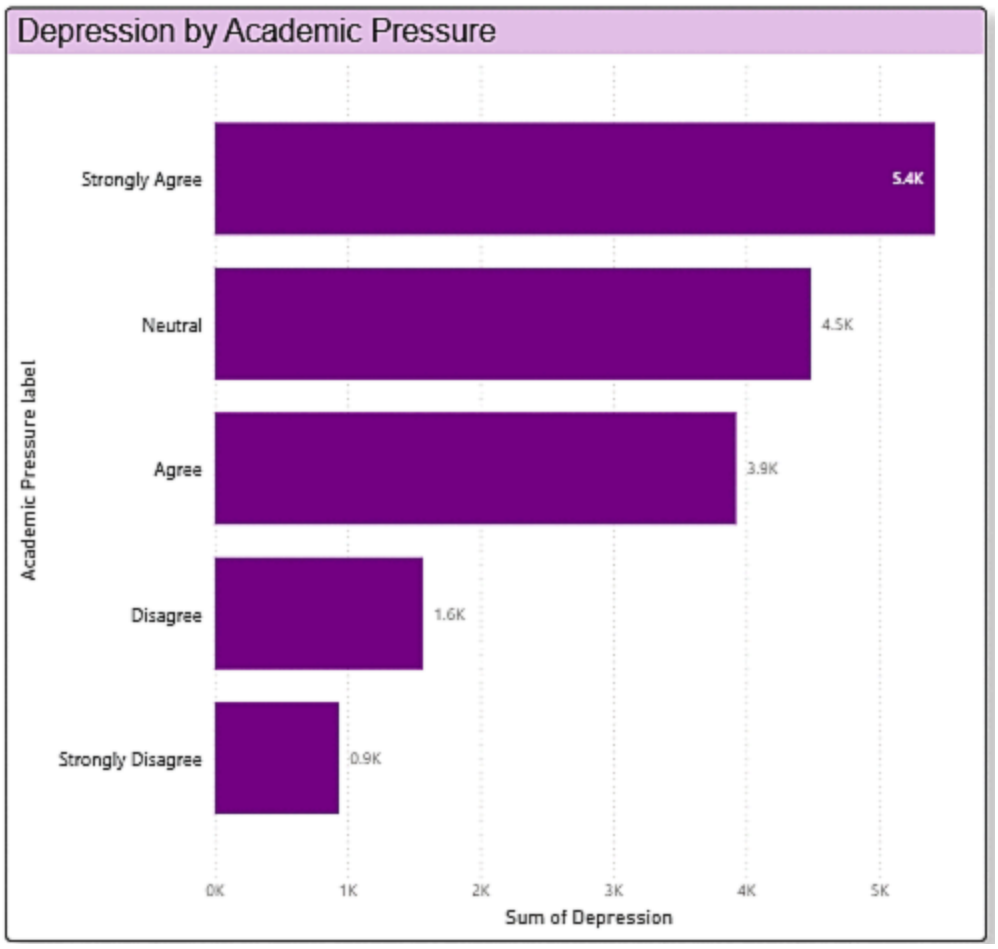


Visual Type: Clustered Column Chart
Observation:

- Students with **<5 hrs sleep** show high depression levels (5.4K)
- Students with **>7 hrs sleep** have relatively **lower depression counts**
- **Highest "No Depression" counts** among those sleeping 7–8 hours

Insight: There is a clear connection between sleep duration and depression. Students who get less than **5 hours** of sleep exhibit the highest levels of depression, while those who sleep more tend to have better mental health.

Depression by Academic Pressure

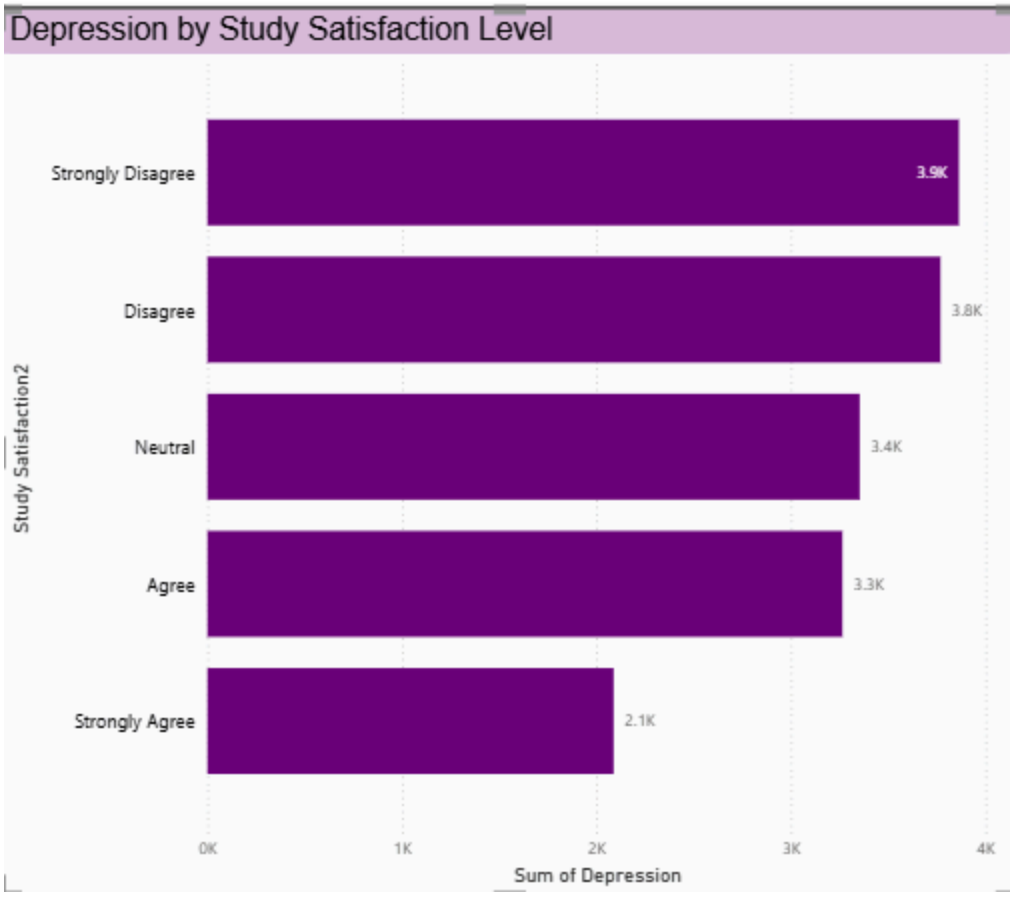


Visual Type: Column Chart
Observation:

- **Strongly Agree:** 5.4K students are depressed
- **Agree:** 3.9K
- **Neutral:** 4.5K
- **Disagree/Strongly Disagree:** Less than 1.6K combined

Insight: There is a **positive correlation between academic pressure and depression.** Students who strongly agree that academic pressure is high tend to experience higher levels of depression.

Depression by Study Satisfaction Level

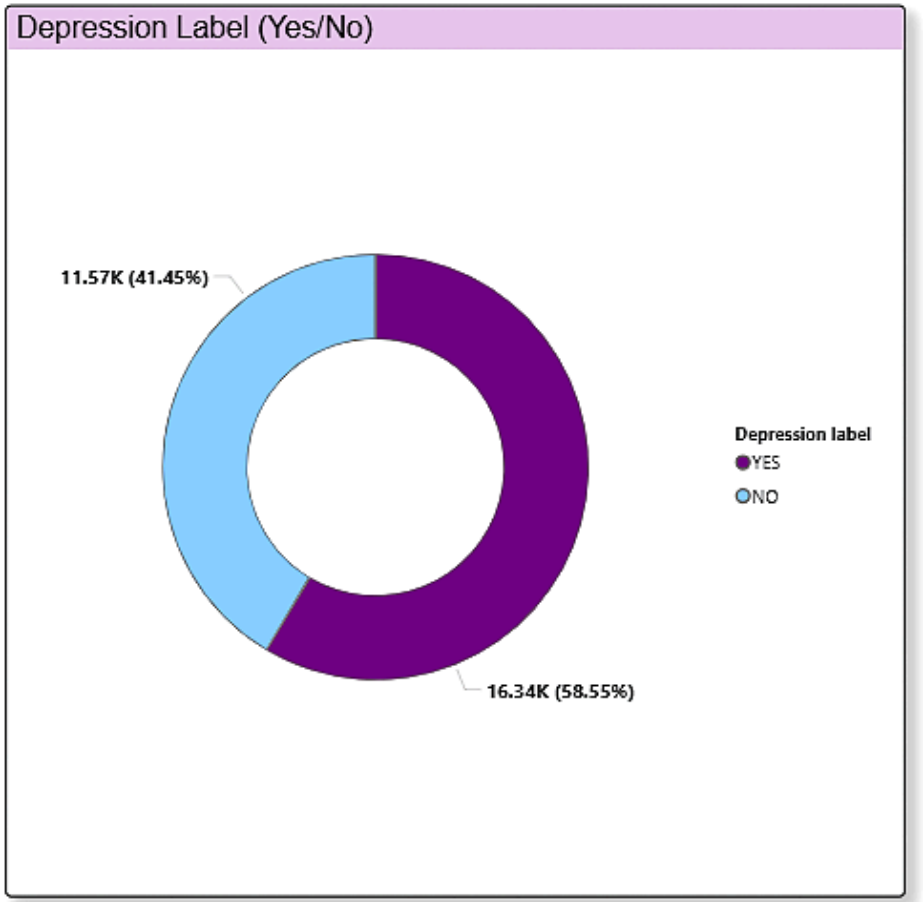


Visual Type: Column Chart
Observation:

- Depression is highest in students who are **dissatisfied** (Strongly Disagree: 3.9K, Disagree: 3.8K)
- It decreases with satisfaction: only 2.1K for "Strongly Agree"

Insight: Increased **study satisfaction is linked to reduced depression levels**, suggesting that student engagement and academic satisfaction have a protective effect.

Depression Label (Yes/No) – Donut Chart



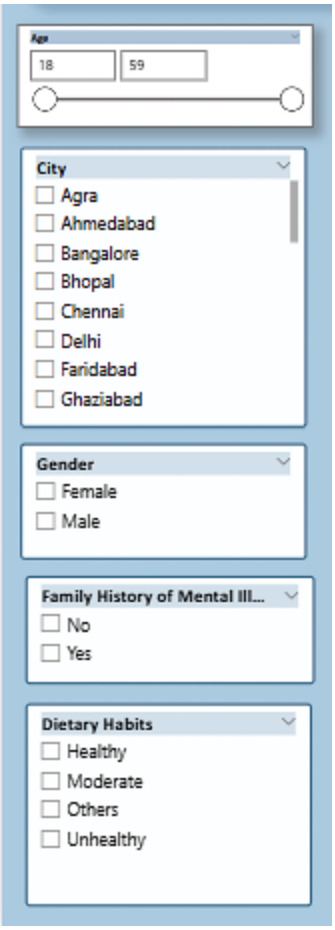
Visual Type: Donut
Observation:

- **Yes:** 58.58%
- **No:** 41.42%

Insight: Over half of the surveyed students are experiencing symptoms of depression.

---

6. Filters Available



These features make the dashboard interactive, enabling **granular analysis**, such as exploring depression among males and females..

Filters allow users to analyze mental health trends by:

- **Age**

  Approximately 14,000 students in the dataset fall within the age group of **18 to 25 years**.

  Among them, **67%** are experiencing symptoms of **depression**, and around **68%** have reported having **suicidal thoughts**. This group also faces high academic and financial stress, and very low study satisfaction, which likely harms their mental health. The pressure of studies and money during this time can be very hard on them.

  About **16,000 students** in the **25 to 40 age** group are represented in the dataset.

  Within this group, around **51%** show signs of depression, and **60%** report suicidal thoughts. This age group also experiences the **highest levels of financial stress,** likely due to responsibilities like supporting families, managing debt, and balancing work with education.

- **City**

Hyderabad reports the highest percentage of students with suicidal thoughts at 68%, accompanied by a depression rate of 67%. **Indore** follows with 62% for suicidal thoughts and 60% for depression, while **Agra** shows 64% and 53%, respectively.

Also, students in other major cities experience depression and suicidal thoughts, typically between 53% and 60%.

The common factor among these students is **sleeping less than 5 hours per night** and facing significant academic and financial stress, contributing to their mental health challenges.

- **Gender**

  Among the surveyed students, about **12,000 are female.** Of these, 63% reported suicidal thoughts and 58% showed signs of depression. Female students face high financial stress, limited sleep, and significant academic pressure, suggesting a vulnerability to mental health challenges due to these combined factors.

  Among the **16,000 male students,** 63% reported suicidal thoughts and 59% showed signs of depression. They also face high financial stress, less sleep, and low study satisfaction, which likely contribute to their poor mental health.

- **Family History of Mental Illness**

  In a study of 14,000 students without a family history of mental illness, 62% reported suicidal thoughts and 56% showed signs of depression. In another group of 14,000 with a family history, 65% had suicidal thoughts and 61% showed signs of depression, suggesting a link between family history and mental health issues.

- **Dietary Habits:**

  Students who consume unhealthy food have a **higher depression rate (71%)** and higher **Suicidal thoughts (70%)**, while students with healthy dietary habits have a lower **depression rate (45%)** and **lower Suicidal thoughts (56%).**

---

7. Recommendations

Mental Health Resources: Launch awareness campaigns, peer support programs, and provide access to on-campus mental health resources.

Academic Reforms: Institutions should consider curriculum moderation, offer academic counseling, and reduce overburdening schedules.

Financial Aid: Scholarships and budgeting education to ease financial anxiety.

Sleep Education: Promote awareness around healthy sleep habits.

Healthy Diet: A healthy diet has a significant impact on students' mental health, so it is essential to encourage students to adopt healthy eating habits.

Degree-specific Interventions: Address challenges specific to technical degree holders.

Study Satisfaction: Improve teaching quality, mentoring, course flexibility, and promote feedback systems to enhance satisfaction.

Sleep Duration: Promote healthy sleep hygiene, discourage late-night academic loads, and encourage mindfulness/sleep workshops.

---

8. Conclusion

The dashboard provides a clear overview of the **mental health crisis impacting students** today. It highlights the need for **structured institutional responses**, including both academic and psychological support. Key interventions can significantly **improve student well-being and academic performance**.