ST. XAVIER'S COLLEGE (AUTONOMOUS), KOLKATA

DEPARTMENT OF STATISTICS

NAME: SREYASI DEY

ROLL NUMBER: 0418

SEMESTER: 6

SESSION: 2022-2025

SUPERVISOR: DR. SURABHI DASGUPTA

PROJECT TITLE - Social media engagement and its psychological influence on the mental health of university students: A Statistical Exploration using Categorical Data Analysis

DECLARATION: I affirm that I have identified all my sources and that no part of my dissertation paper uses unacknowledged materials.

Sreyasi Dey

Signature

TABLE OF CONTENTS

Introduction to the project

Humans are inherently social beings and building social connections is an integral part of human nature. The development of social media platforms over the past decade has significantly altered the way in which individuals establish and maintain relationships with each other. Social media has emerged as a dominant mode of communication, interaction, and self-expression, particularly among young adults or university students. Although these platforms offer numerous beneficial networking opportunities, they also raise concerns regarding their psychological impact on students. The effect of social media engagement on the mental health of students is an important area of research that requires statistical exploration. In this project, we aim to analyse and study about this relationship through a data-driven approach. The insights from this study can help to raise awareness about the psychological impact of social media and guide efforts to create a healthier digital environment for students.

Literature review

Previous studies on social media use and mental health highlight both its benefits and drawbacks. While platforms like Facebook, Instagram and LinkedIn facilitate social bonding and academic support [2], excessive usage of these platforms has been associated with loneliness, sadness, and anxiety [3]. Two major causes of detrimental psychological effects include social comparison and FOMO (fear of missing out) [4]. Indicators of mental health, however, are frequently found to be uncorrelated with a number of characteristics, including gender, average usage time, and number of platforms.

Objective of the project

Using categorical data analysis (CDA), this study aims to investigate the connection between social media usage and mental health attributes, particularly for university students. Our goal is to identify the key factors which might influence mental well-being, such as usage time, negative emotions, social comparisons, etc. and how they relate to students' mental health. The study focuses on providing data-driven insights, which shall help to identify healthier digital practices and encourage a balanced digital environment for students.

Data Description

A survey was conducted using an online questionnaire that was sent as a Google Form to students of different colleges and universities in India. 172 students responded to the questionnaire.

Link to questionnaire:
https://drive.google.com/drive/folders/10UlNgULUAjK9aymBfIoQPIrIaDh1N_fE?usp=sharing

The questionnaire was divided into three sections **Personal details**, **Social media usage** and **Emotional and psychological aspects** and data was collected for the following variables.

Variable	**Scale of measurement**
Age of the respondent	continuous variable measured in ratio scale
Sex of the respondents	nominal variable with 2 categories “Male” and “Female”

Number of social media platforms actively used	ordinal variable with 3 categories “1 or 2”,“3 or 4” and “5 or more”
Average time spent daily on social media	ordinal variable with 4 categories “Less than 1 hour”, “1 - 3 hours”, “4 - 6 hours” and “More than 6 hours”
Preferred time of the day for using social media	ordinal variable with 4 categories “Morning”, “Afternoon”, “Evening” and “Late night”
Academic or professional workload	ordinal variable measured on Likert scale(1-5), 1:Very Light to 5: Very heavy
Frequency of using social media without a specific purpose	ordinal variable measured on Likert scale(1-5), 1:Not at all to 5: Always
Tendency of getting distracted when busy	ordinal variable measured on Likert scale(1-5), 1:Not at all to 5: Always
Feeling restless or suffering from FOMO	ordinal variable measured on Likert scale(1-5), 1:Not at all to 5: Always
Comparison with other people through the use of social media	ordinal variable measured on Likert scale(1-5), 1:Not at all to 5: Always
Feeling pressure to conform to social norms	nominal variable with 3 categories “Yes”, “No”, “Maybe”
Faced cyberbullying	nominal variable with 2 categories “Yes” and “No”
Experiencing negative emotions like stress, anxiety after using social media	ordinal variable measured on Likert scale(1-5), 1:Not at all to 5: Always
Disrupted sleep patterns	ordinal variable measured on Likert scale(1-5), 1:Not at all to 5: Always
Opinion on setting boundaries on social media	nominal variable with 3 categories “Yes”, “No”, “Maybe”
Considered taking a break or permanently leaving social media	nominal variable with 2 categories “Yes” and “No”
Sought therapy or counselling	nominal variable with 2 categories “Yes” and “No”

A glimpse of the dataset is attached below.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Timestamp	Email Address	Name	Age	Sex	Level of Education currently pursuing	Univerity/ College	Discipline	Number of platforms	Platform names	Averarge time	Primary purpose	Prefered time of day
2	10/13/2024 1	agarwalar	Rasika	19	F	Ug	St Xavier's	Statistics	5 or more	Instagram,	4 - 6 hours	Sharing yo	Late night
3	10/13/2024 1	mananarc	Manan	18	Female	Undergraduate	St. Xavier's Colle	Statistics	5 or more	Instagram,	1 - 3 hours	Sharing yo	Evening
4	10/13/2024 1	rounakgh	Rounal	19	Male	Undergraduate	St Xavier's Colleg	Statistics	3 or 4	YouTube	1 - 3 hours	Connectin	Afternoon
5	10/13/2024 1	priyangba	Priyan	18	Female	Undergraduate	St.Xavierâ€™s Cc	Statistics	1 or 2	Instagram,	1 - 3 hours	Sharing yo	Afternoon
6	10/13/2024 1	anushach	Anusha	18	Female	Undergraduate	St Xavier's Colleg	Statistics	3 or 4	Instagram,	4 - 6 hours	Connectin	Afternoon
7	10/13/2024 1	angelaqm	Angela	18	F	undergraduate	st. xavier's colleg	statistics	1 or 2	Instagram,	1 - 3 hours	Connectin	Evening
8	10/13/2024 1	chattopac	Saman	18	Female	Undergraduate	St Xavier's Colleg	Mathematics	3 or 4	Instagram,	4 - 6 hours	Sharing yo	Evening
9	10/13/2024 1	jeetjohng	Jeet Jo	18	Male	Undergraduate	St. Xavier's Colle	Statistics	1 or 2	Facebook,	Less than 1	Connectin	Evening
10	10/13/2024 1	diyakund	Diya Ku	18	Female	Undergraduate	St. Xavier's Colle	Commerce	1 or 2	Instagram,	4 - 6 hours	Entertainn	Evening
11	10/13/2024 1	zainabpri	Zainab	22	Female	Undergraduate	St. Xavier's(autor	Statistics	1 or 2	Instagram,	1 - 3 hours	Connectin	Afternoon

1	Workload	Usage without sepecific purpose	Distracted by social media	FOMO	Comparison	Pressure to conform to social norms	Setting boundaries	Cyberbullying	Experienced negative emotions	Disrupted sleep patterns	Taking a break or permanently leaving social media	Sought help (therapy, counselling, etc.)	Opinion on if social media adversely affects the mental health of students
2	4	3	3	2	2	No	No	No	3	4	No	No	
3	4	4	4	3	3	Maybe	Maybe	No	3	3	No	No	yes
4	4	3	3	1	2	No	No	No	3	1	No	No	At times it does depending on the
5	3	4	4	4	3	Maybe	Maybe	No	3	1	Yes	No	Yes
6	3	4	5	3	5	Maybe	Maybe	No	4	3	No	No	
7	3	3	3	2	2	No	Yes	No	2	1	No	No	no
8	3	5	4	5	4	Yes	Yes	No	3	1	Yes	No	Yes it does, well we live in a worl
9	3	3	3	2	4	No	Yes	No	1	1	Yes	No	Yes, dopamine detox
10	4	4	4	2	3	Maybe	Yes	No	3	2	Yes	No	
11	4	3	3	1	2	No	Maybe	No	2	3	No	No	No idea

Methodology

The fundamental objective of this project is to analyse the relationship between social media engagement and its influence on the mental health of university students. The methodology consists of the following -

i) **Pearsonian chi square test of association**: We have used chi square test of association to check if there exists any dependence between experiencing negative emotions after using social media and other explanatory variables.

ii) **Logistic regression**: We have regressed our response variable (experiencing negative emotions after using social media) after converting it to binary variable, on the explanatory variables which have significant association with the response.

iii) **Interpretation of regression parameters**: The estimated regression coefficients were analyzed to determine how different levels of categorical predictors influence the likelihood of experiencing negative emotions. This enabled us to interpret the direction and magnitude of these effects.

iv) **Classification statistics**: We have applied statistical tools like Confusion matrix, TPR and FPR and ROC curve to assess if our model acts as a good classifier or not.

v) **Kendall's τ_b measure**: We also wanted to check if different factors related to social media have any influence on the sleep patterns of an individual. To understand this association, we have used Kendall's τ_b measure for ordinal variables.

Graphical representation of collected data

The primary data collected on different variables through the survey are represented below graphically.

Survey characteristics

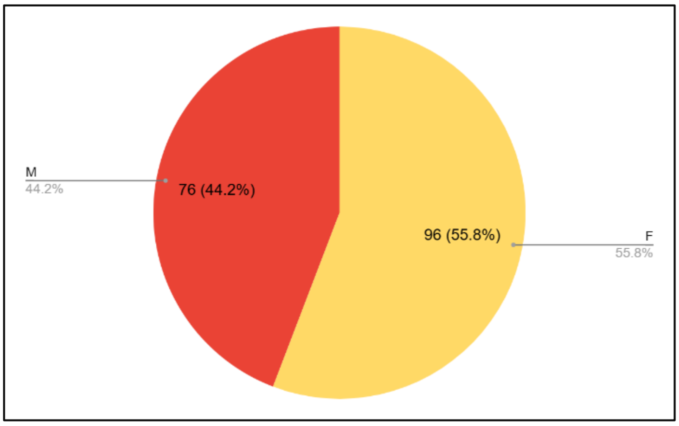


Figure 1: Pie chart showing the % of male and female respondents

Data was collected for 76 male and 96 female students.

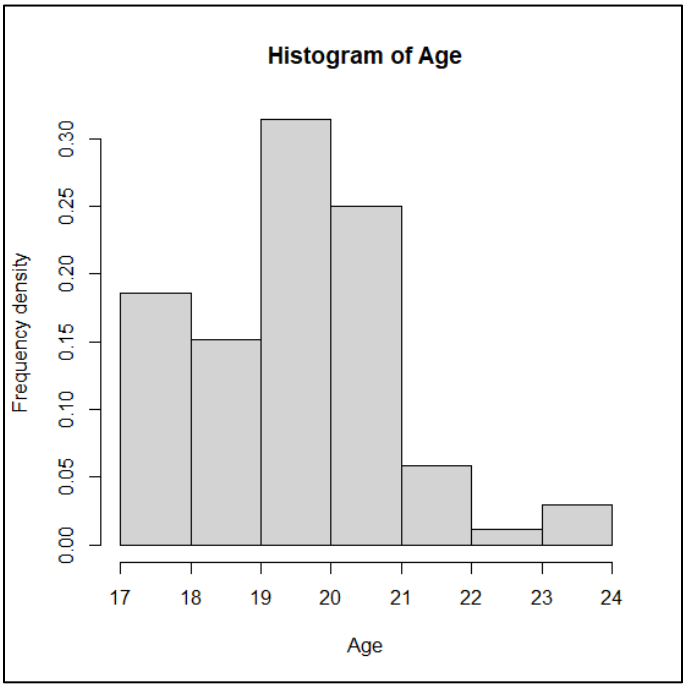


Figure 2: Histogram showing the distribution of age of respondents

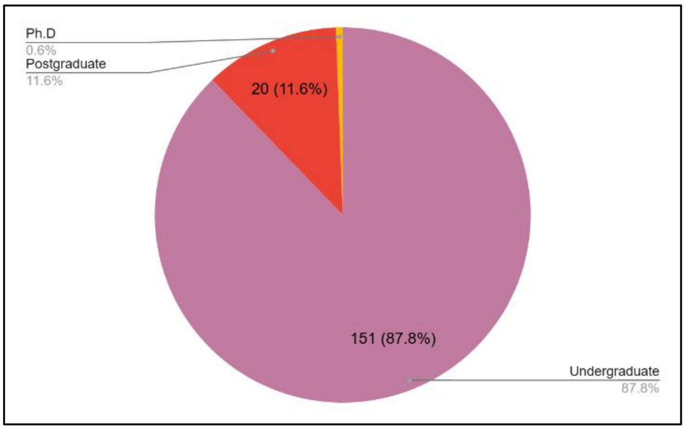


Figure 3: Pie chart showing the % of respondents categorized by their current level of education

Students from different colleges and universities across India from several disciplines filled up the questionnaire.

<u>**Social media usage patterns**</u>

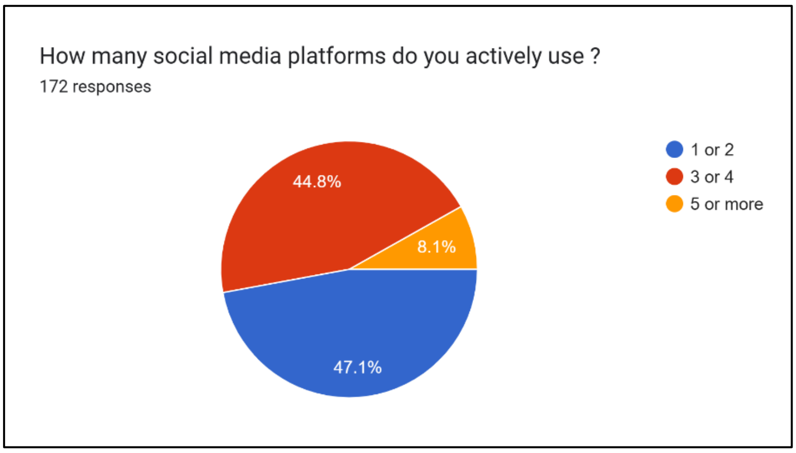


Figure 4: Pie diagram showing the % of respondents categorized by the number of social media platforms used by them

We observe from this chart that the percentage of people using 1 or 2 platforms and 3 or 4 platforms are quite close to each other.

From the survey, we also found that the top three most used social media platforms are Instagram/Thread, YouTube and LinkedIn.

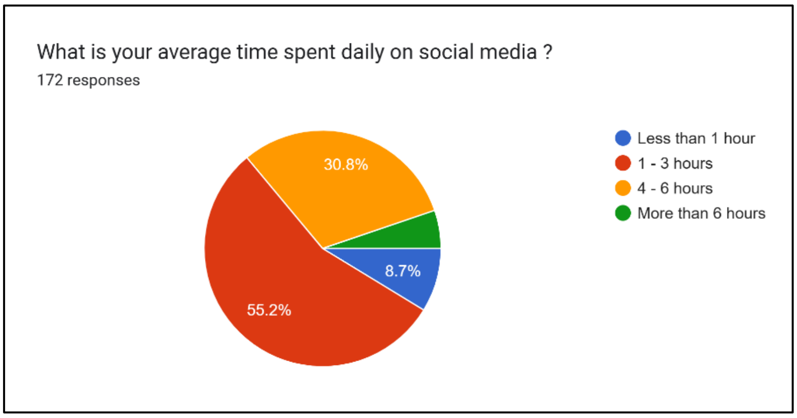


Figure 5: Pie diagram showing the % of respondents categorized by their average time spent daily on social media

In the light of this data, it seems that that approximately 55.2% people use social media for 1-3 hours and 30.8% people used social media for 4-6 hours daily. This signifies that social media occupy a significant portion of students' daily screen time.

According to the data collected, the primary purpose behind using social media is Entertainment, followed by new & information, connecting with friends and family etc.

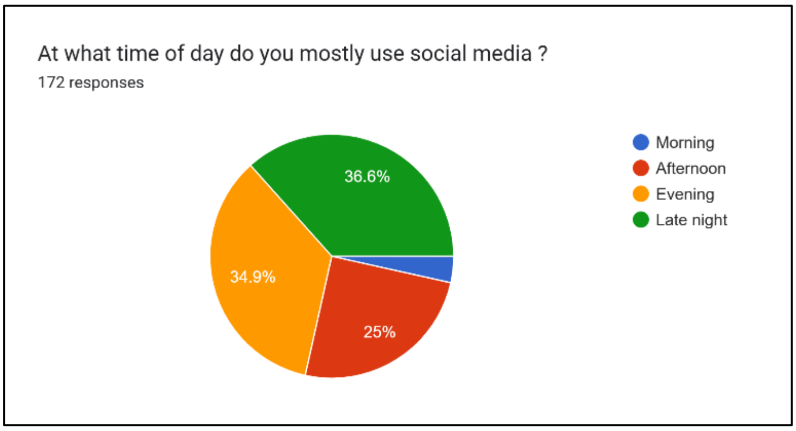


Figure 6: Pie diagram showing the % of respondents according to their preferred time of using social media

We can observe that approximately 34.9% students use social media at evening and 36.6% students use social media at late night. Late night usage of social media might have an association with sleep deprivation or insomnia.

<u>Behavioral patterns</u>

A Likert scale(1-5) was used to collect the opinions of people on different questions related to their social media usage and mental health.

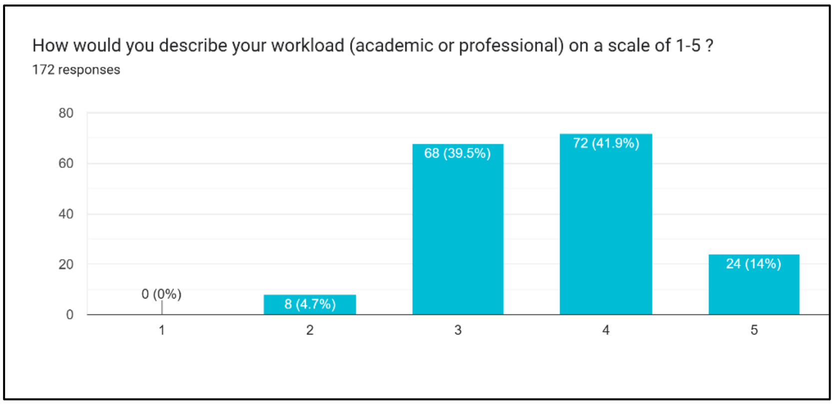


Figure 7: Bar Diagram showing the frequency of respondents according to their workload

[1: Very Light, 2: Light, 3: Moderate, 4: Heavy, 5: Very heavy]

It seems that university students tend to have a moderate to extreme academic/professional workload. We want to analyze how workload may affect social media usage and mental health of students.

For the following graphs, the numbers on x-axis of this scale denote:

1: Not At All, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

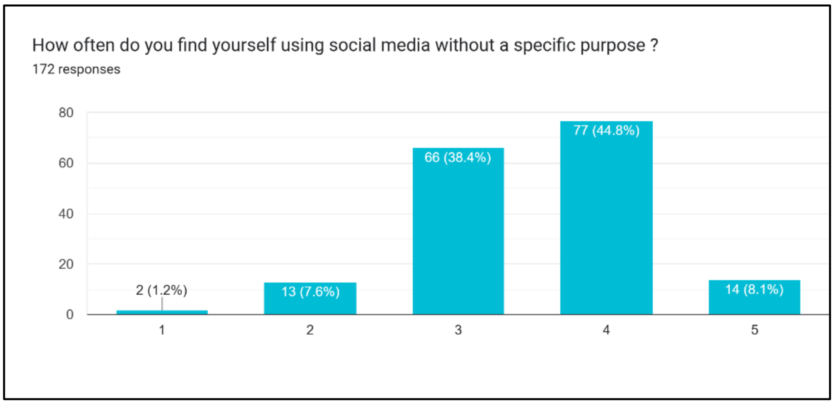


Figure 8: Bar diagram showing frequency of respondents according to their usage of social media without a specific purpose

From the diagram, it seems that almost 44.8% students frequently use social media without a specific purpose.

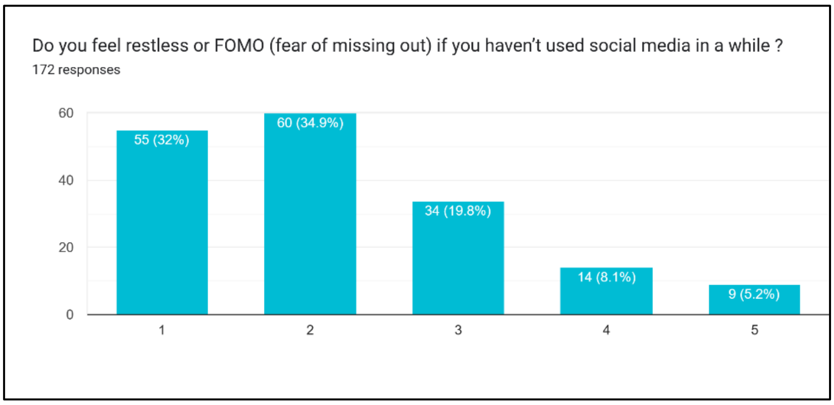


Figure 9: Bar diagram showing frequency of respondents suffering from FOMO (fear of missing out) if they haven't used social media for a while

Although, we observe that majority of students do no suffer from FOMO, approximately 33% ((34+14+9)/172) students sometimes or frequently face FOMO.

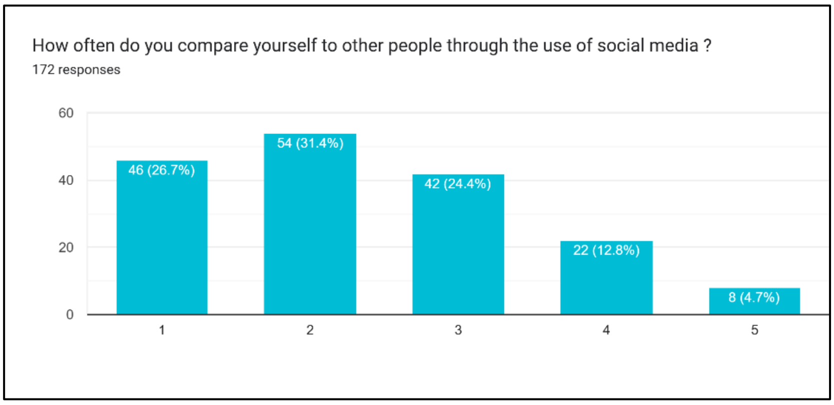


Figure 10: Bar diagram showing frequency of respondents comparing themselves with other people through the use of social media

Although, we observe that most of the students do not compare themselves with other people through the use of social media, approximately 42% ((42+22+8)/172) students sometimes or frequently compare themselves with others. This might affect their self-esteem and confidence.

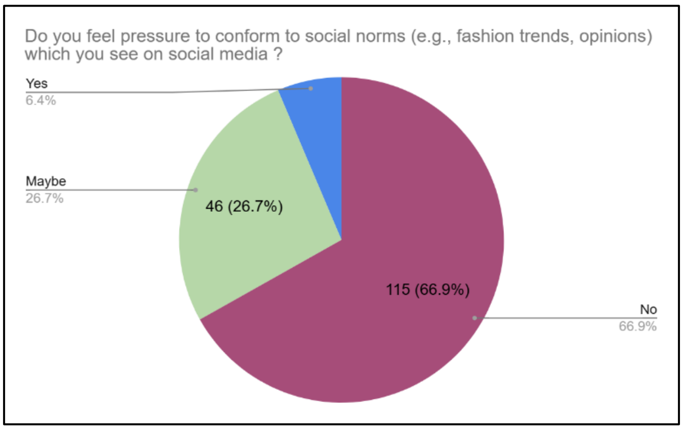


Figure 11: Pie diagram showing the % of respondents who feel pressure to conform to social media

Although, we observe that majority of students do not feel pressure to conform to social media (like fashion trends, opinion etc.), approximately 33% ((46+11)/172) students feel some pressure to conform to ongoing trends and are influenced by social media.

Apart from these factors, students were also asked if they have ever faced **cyberbullying** on social media. According to the data collected, 5% students have faced cyberbullying while using social media. This might also have an adverse effect on their mental health.

Effects on physical and mental health

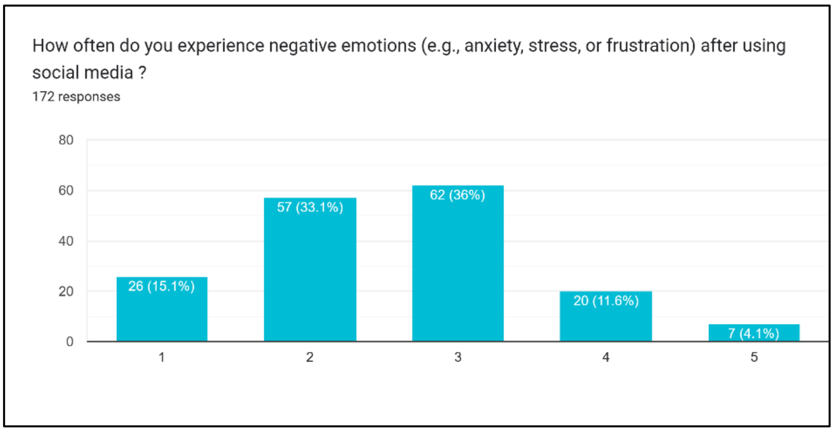


Figure 12: Bar diagram showing the frequency of respondents experiencing negative emotions after using social media

We observe that approximately 52% ((62+20+7)/172) students face negative emotions sometimes or frequently after using social media. This will be our **response variable** and we will try to find out its association with other factors.

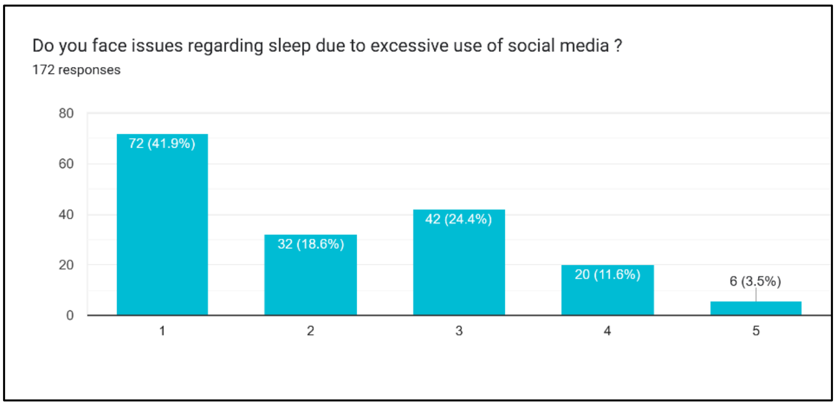


Figure 13: Bar diagram showing the frequency of respondents facing sleep issues due to excessive use of social media.

Although, we observe that most of the students do not face sleep issues, approximately 40% ((42+20+6)/172) students have faced sleep issues sometimes or frequently due to excessive social media usage.

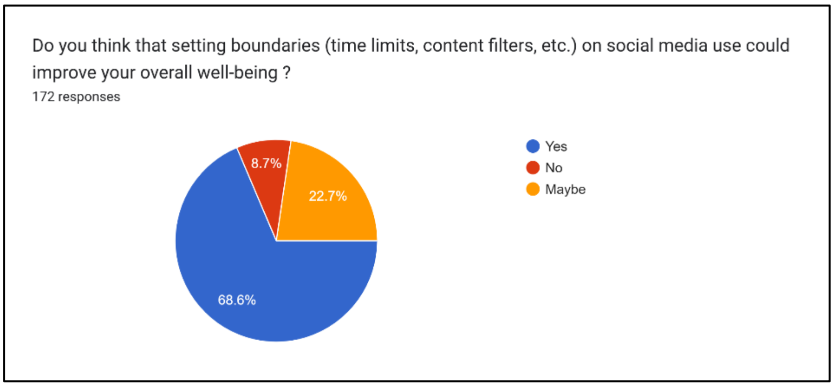


Figure 14: Pie diagram showing the % of respondents according to their opinions on setting constraints on social media usage

We observe that 68.6% respondents think that setting boundaries (time limits, content filters etc.) could improve their overall well-being. This suggests that a majority of users recognize the need for moderation in social media usage.

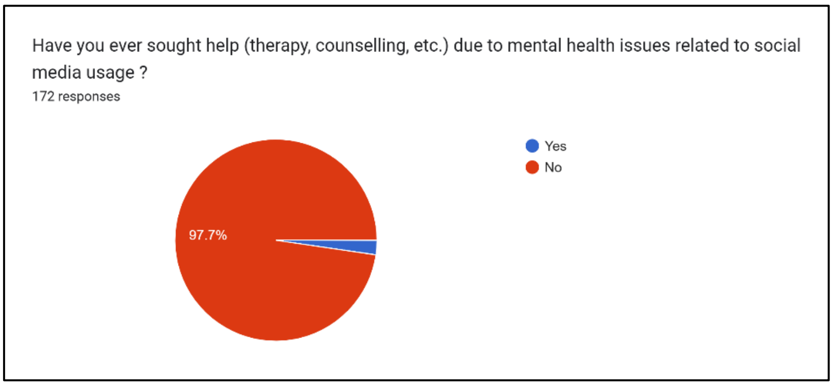


Figure 15: Pie diagram showing the % of respondents who sought help due to mental health issues related to social media usage

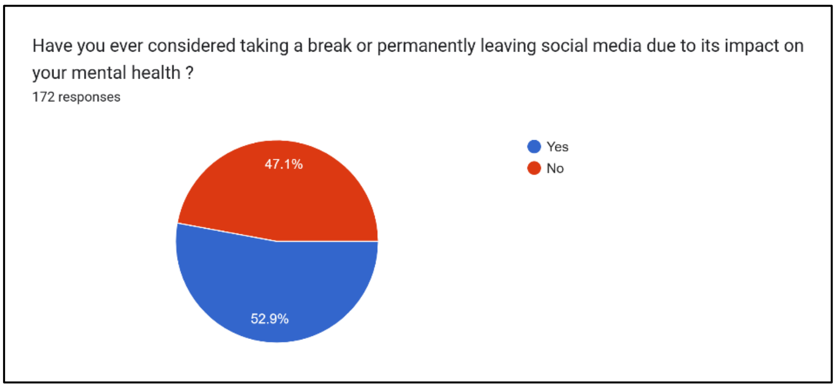


Figure 16: Pie diagram showing the % of respondents according to their opinions on taking a break or permanently leaving social media

We observe that 52.9% respondents considered taking a break or permanently leaving social media due to its effect on their mental health. This suggests that social media's impact on mental health is a significant concern for many people. The idea of taking breaks or quitting reflects a growing awareness of digital well-being.

Analysis of data

The primary objective of our study is to analyse the relationship between social media engagement and the mental health of an individual. Here, we consider "**Experiencing negative emotions like stress, anxiety after using social media** " to be our response variable.

This variable has 5 categories on the basis of extent of negative emotions faced by an individual after using social media – 1: Not At All, 2: Rarely, 3: Sometimes, 4: Often, 5: Always

Since facing negative emotions like anxiety or depression may be a sensitive issue, there is a possibility of bias. This occurs when people adjust their answers to appear more socially acceptable. In the case of negative emotions, respondents may avoid extreme answers (4 or 5) and choose 3 to seem unaffected. This can lead to underreporting of distress, even if they experience anxiety or comparison due to social media.

We convert this variable into a binary variable Y by amalgamating responses "1" and "2" and responses "3", "4" and "5" into another group. Y takes the value 0 for the first group and 1 for the second group. The first group represents individuals who experiences minimal or almost no negative emotions after using social media. The second group represents individuals who experiences moderate or high negative emotions after using social media.

We define Y as follows –

$$Y = \begin{cases} 1 & \text{if the indiviudal faces negative emotions after using social media} \\ 0 & \text{if the indiviudal does not negative emotions after using social media} \end{cases}$$

Here we have data for n=172 students.

We consider the following variables as our explanatory variables for occurrence of negative emotions after using social media.

1. Age of the respondent
2. Sex of the respondent
3. Number of social media platforms actively used
4. Average time spent daily on social media
5. Preferred time of the day for using social media
6. Academic or professional workload
7. Frequency of using social media without a specific purpose
8. Tendency of getting distracted when busy
9. Feeling restless or suffering from FOMO
10. Comparison with other people through the use of social media
11. Feeling pressure to conform to social norms
12. Faced cyberbullying on social media

We find out the association between Y and each of the explanatory variables using Pearsonian Chi square test for independence/association of attributes.

Pearsonian Chi square test for independence/association of attributes

Suppose a population is classified (simultaneously) into k classes $A_1, A_2, \dots A_k$ and l classes $B_1, B_2, \dots B_l$ respectively with respect to a pair of attributes A and B.

Let p_{ij} denote the proportion of members in the population who belong simultaneously to the ith class of A and jth class of B. i=1(1)k, j=1(1)l

p_{ij}'s are unknown.

Table 1: Population contingency table

B \ A	B_1	B_2	….	B_j	….	B_l	Row totals
A_1	p_{11}	p_{12}	….	p_{1j}	….	p_{1l}	p_{10}
A_2	p_{21}	p_{22}	….	p_{2j}	….	p_{2l}	p_{20}
. .	. .	. .	….	. .	….	. .	. .
Ai	p_{i1}	p_{i2}	….	p_{ij}	….	p_{il}	p_{i0}
. .	. .	. .	….	. .	….	. .	. .
A_k	p_{k1}	p_{k2}	….	p_{kj}	….	p_{kl}	p_{k0}
Column totals	p_{01}	p_{02}	….	p_{0j}	….	p_{0l}	1

To test: H_0: $p_{ij} = p_{i0}*p_{0j}$ for all (i,j) vs H_1: Not H_0

Suppose we draw a random sample of size n from the population.

Let f_{ij} denote the number of members in the sample who belong simultaneously to the ith class of A and jth class of B. i=1(1)k, j=1(1)l

Table 2: Sample contingency table

B \ A	B_1	B_2	….	B_j	….	B_l	Row totals
A_1	f_{11}	f_{12}	….	f_{1j}	….	f_{1l}	f_{10}
A_2	f_{21}	f_{22}	….	f_{2j}	….	f_{2l}	f_{20}
. .	. .	. .	….	. .	….	. .	. .
Ai	f_{i1}	f_{i2}	….	f_{ij}	….	f_{il}	f_{i0}
. .	. .	. .	….	. .	….	. .	. .
A_k	f_{k1}	f_{k2}	….	f_{kj}	….	f_{kl}	f_{k0}
Column totals	f_{01}	f_{02}	….	f_{0j}	….	f_{0l}	n

We define,

$$\chi^2 \quad = \quad \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{(f_{ij}-np_{ij})^2}{np_{ij}}$$

Now under H_0, $p_{ij} = p_{i0}.p_{0j}$

$$\chi^2 \quad = \quad \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{\left(f_{ij}-np_{i0}p_{0j}\right)^2}{np_{i0}p_{0j}}$$

Since, p_{i0} and p_{0j} are unknown, we estimate them as $\hat{p}_{i0} = f_{i0}/n$ and $\hat{p}_{0j} = f_{0j}/n$

Under H_0, our test statistic is given by,

$$\chi^2 \quad = \quad \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{\left(f_{ij}-\frac{f_{i0}f_{0j}}{n}\right)^2}{\frac{f_{i0}f_{0j}}{n}} \quad = \quad n\sum_{i=1}^{k}\sum_{j=1}^{l}\left(\frac{{f_{ij}}^2}{f_{i0}f_{0j}}\right)-n$$

$\chi^2 \sim \chi^2_{(k-1)(l-1)}$ under H_0

Reject H_0 against H_1 at α level of significance if $\chi^2_{obs} > \chi^2_{\alpha,(k-1)(l-1)}$

where χ^2_{obs} is the observed value of the test statistic and $\chi^2_{\alpha,(k-1)(l-1)}$ is the upper α% point of a χ^2 distribution with degrees of freedom (k-1)(l-1).

If H_0 is rejected, we conclude that the attributes A and B are associated with each other.

All approximations considered with reference to the Pearsonian Chi square statistic are valid under the restriction that expected frequency of any class is greater than or equal to 5.

Expected frequency = $\frac{\text{Row Total x Column Total}}{\text{Grand total}}$

For the following tests for association, the expected frequencies are written in bracket in the cells of the contingency table.

Here, we test at 10% level of significance.

1. Association of negative emotions with sex of the respondent

Table 3: Contingency table of negative emotions and sex of the respondent

Negative emotions \ Sex	M	F	Total
1	38 (39.33)	51 (49.67)	89
0	38 (36.67)	45 (46.33)	83
Total	76	96	172

Observed value of chi square test statistic = 0.1659

p-value = 0.6838 > 0.1

2. Association of negative emotions with age of the respondent

We have data for individuals of age 17, 18, 19, 20, 21, 22, 23, 24. We divide them into two classes 17-20 and 21-24.

Table 4: Contingency table of negative emotions and age group of the respondent

Negative emotions \ Age	17-20	21-24	Total
1	60 (57.95)	29(31.05)	89
0	52 (54.05)	31(28.95)	83
Total	112	60	172

Observed value of chi square test statistic = 0.42932

p-value = 0.5123 > 0.1

3. Association of negative emotions with number of social media platforms actively used

Table 5: Contingency table of negative emotions and number of social media platforms actively used

Negative emotions \ Number of platforms	1 or 2	3 or 4	5 or more	Total
1	40 (41.91)	40 (39.84)	9 (7.24)	89
0	41 (39.09)	37 (37.16)	5 (6.76)	83
Total	81	77	14	172

Observed value of chi square test statistic = 1.0641

p-value = 0.5874 > 0.1

4. Association of negative emotions with average time spent on social media daily

Table 6.1: Contingency table of negative emotions and average time spent on social media daily

Negative emotions \ Avg time	Less than 1 hour	1 - 3 hours	4 - 6 hours	More than 6 hours	Total
1	5(7.762)	47(49.157)	34(27.424)	3(4.657)	89
0	10(7.238)	48(45.843)	19(25.576)	6(4.343)	83
Total	15	95	53	9	172

Since, the expected frequencies of two cells is < 5, we amalgamate the classes "4-6 hours" and "More than 6 hours" and we obtain the following table.

Table 6.2: Contingency table of negative emotions and average time spent on social media daily

Avg time / Negative emotions	Less than 1 hour	1 - 3 hours	4 hours or more	Total
1	5(7.762)	47(49.157)	37(32.081)	89
0	10(7.238)	48(45.843)	25(29.919)	83
Total	15	95	62	172

Observed value of chi square test statistic = 3.7951

p-value = 0.1499 > 0.1

5. Association of negative emotions with preferred time of day for using social media

Table 7.1: Contingency table of negative emotions and preferred time of day for using social media

Time of day / Negative emotions	Morning	Afternoon	Evening	Late night	Total
1	2 (3.1)	18 (22.25)	30 (31.05)	39 (32.6)	89
0	4 (2.9)	25 (20.75)	30 (28.95)	24 (30.4)	83
Total	6	43	60	63	172

Since, the expected frequencies of two cells are < 5, we amalgamate the classes "Morning" and "Afternoon" and we obtain the following table.

Table 7.2: Contingency table of negative emotions and preferred time of day for using social media

Time of day / Negative emotions	Daytime	Evening	Late night	Total
1	20 (25.35)	30 (31.05)	39 (32.6)	89
0	29 (23.65)	30 (28.95)	24 (30.4)	83
Total	49	60	63	172

Observed value of chi square test statistic = 5.0213

p-value = 0.08122<0.1

6. Association of negative emotions with academic or professional workload

Table 8.1: Contingency table of negative emotions and academic or professional workload

Workload / Negative emotions	1	2	3	4	5	Total

1	0	3(4.14)	34(35.19)	40(37.26)	12(12.42)	89
0	0	5(3.86)	34(32.81)	32(34.74)	12(11.58)	83
Total	0	8	68	72	24	172

Since, the expected frequencies of two cells are < 5 and the frequency of two classes are 0, we amalgamate the classes “1”, “2” and “3” or workload and we obtain the following table.

Table 8.2: Contingency table of negative emotions and academic or professional workload

Workload / Negative emotions	Less than or equal to 3	4	5	Total
1	37(39.33)	40(37.26)	12(12.42)	89
0	39(36.67)	32(34.74)	12(11.58)	83
Total	76	72	24	172

Observed value of chi square test statistic = 0.73311

p-value = 0.6931 > 0.1

7. Association of negative emotions with usage without a specific purpose

Table 9.1: Contingency table of negative emotions and usage without a specific purpose

Usage without specific purpose / Negative emotions	1	2	3	4	5	Total
1	1(1.035)	7(6.727)	27(34.151)	46(39.843)	8(7.244)	89
0	1(0.965)	6(6.273)	39(31.849)	31(37.157)	6(6.756)	83
Total	2	13	66	77	14	172

Since, the expected frequencies of two cells are < 5, we amalgamate the classes “1” and “2” of usage without a specific purpose and we obtain the following table.

Table 9.2: Contingency table of negative emotions and usage without a specific purpose

Usage without specific purpose / Negative emotions	Less than or equal to 2	3	4	5	Total
1	8(7.762)	27(34.151)	46(39.843)	8(7.244)	89
0	7(7.238)	39(31.849)	31(37.157)	6(6.756)	83
Total	15	66	77	14	172

Observed value of chi square test statistic = 5.2534 p-value = 0.1542 > 0.1

8. Association of negative emotions with getting distracted when busy

Table 10.1: Contingency table of negative emotions and getting distracted when busy

Tendency of getting distracted when busy / Negative emotions	1	2	3	4	5	Total
1	4(4.14)	14(17.59)	28(31.56)	37(28.98)	6(6.73)	89
0	4(3.86)	20(16.41)	33(29.44)	19(27.02)	7(6.27)	83
Total	8	34	61	56	13	172

Since, the expected frequencies of two cells are < 5, we amalgamate the classes "1" and "2" of getting distracted while busy and we obtain the following table.

Table 10.2: Contingency table of negative emotions and getting distracted when busy

Tendency of getting distracted when busy / Negative emotions	Less than or equal to 2	3	4	5	Total
1	18(21.73)	28(31.56)	37(28.98)	6(6.73)	89
0	24(20.27)	33(29.44)	19(27.02)	7(6.27)	83
Total	42	61	56	13	172

Observed value of chi square test statistic = 6.9287

p-value = 0.0742 < 0.1

9. Association of negative emotions with FOMO

Table 11.1: Contingency table of negative emotions and FOMO

FOMO / Negative emotions	1	2	3	4	5	Total
1	18(28.46)	31(31.05)	26(17.59)	11(7.24)	3(4.66)	89
0	37(26.54)	29(28.95)	8(16.41)	3(6.76)	6(4.34)	83
Total	55	60	34	14	9	172

Since, the expected frequencies of two cells are < 5, we amalgamate the classes "4" and "5" of FOMO and we obtain the following table.

Table 11.2: Contingency table of negative emotions and FOMO

FOMO / Negative emotions	1	2	3	Greater than or equal to 4	Total
1	18(28.46)	31(31.05)	26(17.59)	14(11.9)	89
0	37(26.54)	29(28.95)	8(16.41)	9(11.1)	83
Total	55	60	34	23	172

Observed value of chi square test statistic = 17.058,

p-value = 0.0006876 < 0.1

10. Association of negative emotions with comparison on social media

Table 12.1: Contingency table of negative emotions and comparison on social media

Comparison / Negative emotions	1	2	3	4	5	Total
1	15(23.8)	28(27.94)	24(21.73)	16(11.38)	6(4.14)	89
0	31(22.2)	26(26.06)	18(20.27)	6(10.62)	2(3.86)	83
Total	46	54	42	22	8	172

Since, the expected frequencies of two cells are < 5, we amalgamate the classes “4” and “5” of comparison on social media and we obtain the following table.

Table 12.2: Contingency table of negative emotions and comparison on social media

Comparison / Negative emotions	1	2	3	Greater than or equal to 4	Total
1	15(23.8)	28(27.94)	24(21.73)	22(15.52)	89
0	31(22.2)	26(26.06)	18(20.27)	8(14.48)	83
Total	46	54	42	30	172

Observed value of chi square test statistic = 12.836,

p-value = 0.005005 < 0.1

11. Association of negative emotions with feeling pressure to conform to social norms

Table 13: Contingency table of negative emotions and pressure to conform to social norms

Social Norms / Negative emotions	No	Maybe	Yes	Total

1	33(23.8)	46(59.51)	10(5.69)	89
0	13(22.2)	69(55.49)	1(5.31)	83
Total	46	115	11	172

Observed value of chi square test statistic = 20.475,

p-value = 0.0000358< 0.1

12. Association of negative emotions with cyberbullying

Table 14: Contingency table of negative emotions and cyberbullying

Negative emotions \ Faced cyberbullying	Yes	No	Total
1	5 (4.657)	84 (84.343)	89
0	4 (4.343)	79 (78.657)	83
Total	9	163	172

Since, the expected frequency of 2 classes are < 5, and this is a 2x2 contingency table, we cannot amalgamate cells here. Using Yate's correction for continuity, we get

Observed value of chi square test statistic = $\frac{172.\left\{|5\times79-84\times4|-\frac{172}{2}\right\}^2}{(89\times83\times9\times163)} = 0.01157$ which is quite low in magnitude.

We also check for this association using Odds ratio.

Odds of facing negative emotions among those who faced cyberbullying

= $\widehat{O_1}$

= $\frac{\hat{P}(Negative\ Emotions=1|Cyberbullying=Yes)}{\hat{P}(Negative\ Emotions=0|Cyberbullying=Yes)} = \frac{\frac{5}{9}}{\frac{4}{9}} = \frac{5}{4}$

Odds of facing negative emotions among those who did not face cyberbullying

= $\widehat{O_2}$

= $\frac{\hat{P}(Negative\ Emotions=1|Cyberbullying=No)}{\hat{P}(Negative\ Emotions=0|Cyberbullying=No)} = \frac{\frac{84}{163}}{\frac{79}{163}} = \frac{84}{79}$

The odds ratio is given by = $\widehat{\mathrm{OR}} = \frac{\widehat{O_1}}{\widehat{O_2}} =$ 1.175595 ≈ 1.18

This implies that odds of negative emotions among students who have faced cyberbullying is approximately 1.18 times that odds of negative emotions among students who have not faced cyberbullying. However, since this ratio is very close to 1, in the light of the given data, it seems that there is no significant association between negative emotions and cyberbullying.

However, in our dataset only 9 out of 172 individuals reported cyberbullying. The effect of cyberbullying on mental health might be more pronounced if larger and more diverse datasets are considered.

Thus, summarizing the results, in the light of the given data, it seems that

- **Age, sex, number of platforms used, academic workload, usage without a specific purpose** and **cyberbullying** do not show a significant association with negative emotions after using social media.
- **Average time spent daily on social media** might have a weak association with negative emotions after using social media (since p value = 0.1499 which is slightly greater than 0.1).
- **Preferred time of day for using social media** and **getting distracted when busy** has a significant but weak association with negative emotions after using social media.
- **FOMO (Fear of Missing Out), social comparison,** and **pressure to conform to social norms** have the strongest association with negative emotions after using social media.

<u>Regression</u>

Following the results of chi square tests of association, we shall evaluate the relationship between Y and the associated variables by fitting a multiple logistic regression model.

We consider the following predictors.

a) Average time spent daily on social media (x1)
b) Preferred time of the day for using social media (x2)
c) Tendency of getting distracted when busy (x3)
d) Feeling restless or suffering from FOMO (x4)
e) Comparison with other people through the use of social media (x5)
f) Feeling pressure to conform to social norms (x6)

Now x1, x2, … x6 are categorical predictors with more than 2 levels. We cannot incorporate them using a single regression coefficient. This is because the impact on the logit function when moving from one category to another is not necessarily uniform across all levels. Hence, we define the following dummy predictors.

Average time spent daily on social media (x1)

Reference category : Less than 1 hour

$$x_{11} = \begin{cases} 1 & \text{if the indiviudal uses social media for "1 } - \text{ 3 hours"} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{12} = \begin{cases} 1 & \text{if the indiviudal uses social media for "4 } - \text{ 6 hours"} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{13} = \begin{cases} 1 & \text{if the indiviudal uses social media for "more than 6 hours"} \\ 0 & \text{otherwise} \end{cases}$$

Preferred time of the day for using social media (x2)

Reference category : Morning

$$x_{21} = \begin{cases} 1 & \text{if the indiviudal uses social media at afternoon} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{22} = \begin{cases} 1 & \text{if the indiviudal uses social media at evening} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{23} = \begin{cases} 1 & \text{if the indiviudal uses social media at late night} \\ 0 & \text{otherwise} \end{cases}$$

Tendency of getting distracted when busy (x3)

Reference category : 1 (Not at all)

$$x_{31} = \begin{cases} 1 & \text{if the indiviudal rarely gets distracted by social media when busy} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{32} = \begin{cases} 1 & \text{if the indiviudal sometimes gets distracted by social media when busy} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{33} = \begin{cases} 1 & \text{if the indiviudal often gets distracted by social media when busy} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{34} = \begin{cases} 1 & \text{if the indiviudal always gets distracted by social media when busy} \\ 0 & \text{otherwise} \end{cases}$$

Feeling restless or suffering from FOMO (x4)

Reference category : 1 (Not at all)

$$x_{41} = \begin{cases} 1 & \text{if the indiviudal rarely suffers from FOMO} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{42} = \begin{cases} 1 & \text{if the indiviudal sometimes suffers from FOMO} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{43} = \begin{cases} 1 & \text{if the indiviudal often suffers from FOMO} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{44} = \begin{cases} 1 & \text{if the indiviudal always suffers from FOMO} \\ 0 & \text{otherwise} \end{cases}$$

Comparison with other people through the use of social media (x5)

Reference category : 1 (Not at all)

$$x_{51} = \begin{cases} 1 & \text{if the indiviudal rarely compares their life to others} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{52} = \begin{cases} 1 & \text{if the indiviudal sometimes compares their life to others} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{53} = \begin{cases} 1 & \text{if the indiviudal often compares their life to others} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{54} = \begin{cases} 1 & \text{if the indiviudal always compares their life to others} \\ 0 & \text{otherwise} \end{cases}$$

Feeling pressure to conform to social norms (x6)

Reference category : No

$$x_{61} = \begin{cases} 1 & \text{if the indiviudal chose the option "Maybe"} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{62} = \begin{cases} 1 & \text{if the indiviudal chose the option "Yes"} \\ 0 & \text{otherwise} \end{cases}$$

Checking for multicollinearity

Before proceeding with fitting the regression model, we need to check if multicollinearity exists among the predictors.

Multicollinearity occurs when two or more explanatory variables in a regression model are highly correlated. This can cause instability in the coefficient estimates, making it difficult to determine the individual effect of each predictor on the dependent variable. Severe multicollinearity can inflate standard errors, leading to unreliable statistical significance tests.

Variance inflation factor

The Variance Inflation Factor (VIF) is a statistical measure used to detect multicollinearity in regression models. It quantifies how much the variance of a regression coefficient is inflated due to the presence of correlation among independent variables.

Suppose $\widehat{\beta_m}$ is the estimate of the regression coefficient for a predictor x_m. Then VIF_m is defined as:

$$VIF_m = \frac{1}{1-{R_m}^2}, \; m=1(1)6$$

R_m is the unadjusted coefficient of determination for regressing the mth independent predictor on the remaining ones.

If ${R_m}^2$ is high, then VIF_m will be high indicating high multicollinearity.

If ${R_m}^2$ is low, then VIF_m will be low indicating low multicollinearity.

Since, here we have predictors with more than 2 levels, we will use Generalized Variance Inflation Factor (GVIF).

$$GVIF_m = (VIF_m)^{\frac{1}{2*df_m}} = \left(\frac{1}{1-{R_m}^2}\right)^{\frac{1}{2*df_m}}, \; m=1(1)6$$

where df_m (degrees of freedom) of the mth predictor = Number of levels – 1

= number of dummy variables created for mth categorical predictor in the regression model.

Generally, a GVIF value greater than 5 indicates high multicollinearity among the predictor variables.

Using R, the generalized variance inflation factors for our predictors are

Table 15: Table showing the degrees of freedom and GVIF values for the categorical predictors

Predictor	df	GVIF
Average time spent daily on social media (x1)	3	1.0936
Preferred time of the day for using social media (x2)	3	1.0734
Tendency of getting distracted when busy (x3)	4	1.1112
Feeling restless or suffering from FOMO (x4)	4	1.0959
Comparison with other people through the use of social media (x5)	4	1.1143
Feeling pressure to conform to social norms (x6)	2	1.1623

Since the GVIF values are close to 1, we can say that no significant multicollinearity is present in the predictors of the model.

Now, we proceed to fit the multiple logistic regression equation of Y on x1,x2,x3,x4,x5,x6.

Logistic regression

Logistic regression is a statistical generalized lineal model used for binary classification. It is used to model responses which are binary in nature.

Our response variable is

$$Y = \begin{cases} 1 & \text{if the indiviudal faces negative emotions after using social media} \\ 0 & \text{if the indiviudal does not negative emotions after using social media} \end{cases}$$

which is binary in nature i.e., it can take two values 0 and 1.

Let Y ~ Bernoulli (π) where 0< π<1.

$\therefore$ P(Y=1)= π and P(Y=0)=1 − π

The multiple logistic regression model is given by,

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_{11}x_{11} + \beta_{12}x_{12} + \beta_{13}x_{13} + \beta_{21}x_{21} + \beta_{22}x_{22} + \beta_{23}x_{23} + \beta_{31}x_{31} + \beta_{32}x_{32} + \beta_{33}x_{33} + \beta_{34}x_{34} + \beta_{41}x_{41} + \beta_{42}x_{42} + \beta_{43}x_{43} + \beta_{44}x_{44} + \beta_{51}x_{51} + \beta_{52}x_{52} + \beta_{53}x_{53} + \beta_{54}x_{54} + \beta_{61}x_{61} + \beta_{62}x_{62}$$

$\beta_0, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{31}, \beta_{32}, \beta_{33}, \beta_{34}, \beta_{41}, \beta_{42}, \beta_{43}, \beta_{44}, \beta_{51}, \beta_{52}, \beta_{53}, \beta_{54}, \beta_{61}$, and β_{62} are unknown real valued parameters to be estimated by Fisher scoring maximum likelihood method.

We have data for n=172 individuals.

For rth individual, we have

$P(Y_r = 1|x_{11}, x_{12}, \ldots\ldots, x_{62}) = \pi_r$ r=1(1)n

$$\ln\left(\frac{\pi_r}{1-\pi_r}\right) = \beta_{0r} + \beta_{11}x_{11r} + \beta_{12}x_{12r} + \beta_{13}x_{13r} + \beta_{21}x_{21r} + \beta_{22}x_{22r} + \beta_{23}x_{23r} + \beta_{31}x_{31} + \cdots \ldots + \beta_{61}x_{61r} + \beta_{62}x_{62r}$$

$$= \eta_r$$

$=> \ln(\frac{\pi_r}{1-\pi_r}) = \eta_r$

$=> \pi_r = \frac{e^{\eta_r}}{1+e^{\eta_r}}$,   r=1(172)

Let L be the log likelihood function. The score equations are given by,

$\frac{\partial L}{\partial \beta_0} = 0 => \sum_{r=1}^{n}(y_r - \pi_r) = 0$

$\frac{\partial L}{\partial \beta_{11}} = 0 => \sum_{r=1}^{n}(y_r - \pi_r)x_{11r} = 0$

$\frac{\partial L}{\partial \beta_{12}} = 0 => \sum_{r=1}^{n}(y_r - \pi_r)x_{12r} = 0$

.

.

.

$\frac{\partial L}{\partial \beta_{62}} = 0 => \sum_{r=1}^{n}(y_r - \pi_r)x_{62r} = 0$

We get 21 score equations. We solve these equations simultaneously and get the estimates of the parameters.

From R, the fitted logistic regression eq. is given by,

$$\widehat{\eta_r} = -2.88216 + 1.02568x_{11} + 1.31107x_{12} - 0.76388x_{13} + 1.65819x_{21} + 1.69302x_{22} + 2.44061x_{23} - 1.31835x_{31} - 1.44503x_{32} - 1.18476x_{33} - 1.85572x_{34} + 0.60781x_{41} + 1.77002x_{42} + 1.64705x_{43} - 0.72577x_{44} + 0.70807x_{51} + 0.05637x_{52} + 0.37951x_{53} + 1.06074x_{54} + 1.11915x_{61} + 3.17757x_{62}$$

$\widehat{\pi_r} = \frac{e^{\widehat{\eta_r}}}{1+e^{\widehat{\eta_r}}}$ r=1(1)n

Interpretation

I. **Intercept:** $\widehat{\beta_0}$= -2.88216
The log odds of experiencing negative emotions when all the predictors are at their reference level is -2.88216. This implies that the probability of experiencing negative emotions when all the predictors are at their reference level is
$\frac{e^{-2.88216}}{1+e^{-2.88216}}$ = 0.05304254 which is quite low.

II. **Average time spent daily on social media**

i) **1-3 hours:** $\widehat{\beta_{11}}$= 1.02568

As we move from the reference category "Less than 1 hour" to "1-3 hours" of daily social media usage time keeping other predictors constant, the log odds of experiencing negative emotions increases by 1.02568.

ii) **4-6 hours:** $\widehat{\beta_{12}}$= 1.31107

As we move from the reference category "Less than 1 hour" to "4-6 hours" of daily social media usage time keeping other predictors constant, the log odds of experiencing negative emotions increases by 1.31107.

iii) **More than 6 hours:** $\widehat{\beta_{13}}$= -0.76388

As we move from the reference category "Less than 1 hour" to "More than 6 hours" of daily social media usage time keeping other predictors constant, the log odds of experiencing negative emotions decreases by 0.76388.

The results indicate that moderate social media use (1-3 hours or 4-6 hours) increases the likelihood of experiencing negative emotions as compared to using it for less than 1 hour. Possible reasons behind these might be overexposure to unhealthy content, social comparisons etc. However, for excessive usage (more than 6 hours), we observe that the likelihood of experiencing negative emotions is less as compared to the reference category. A possible explanation might be that some users may become habituated and develop a higher tolerance to social media content, which reduces its emotional impact on them.

III. Preferred time of the day for using social media

i) **Afternoon:** $\widehat{\beta_{21}}$= 1.65819

As we move from the reference category "Morning" to "Afternoon" for preferred time of day for social media use, keeping other predictors constant, the log odds of experiencing negative emotions increases by 1.65819.

ii) **Evening:** $\widehat{\beta_{22}}$= 1.69302

As we move from the reference category "Morning" to "Evening" for preferred time of day for social media use, keeping other predictors constant, the log odds of experiencing negative emotions increases by 1.69302.

iii) **Late night:** $\widehat{\beta_{23}}$= 2.44061

As we move from the reference category "Morning" to "Late night" for preferred time of day for social media use, keeping other predictors constant, the log odds of experiencing negative emotions increases by 2.44061.

The results suggest that the likelihood of experiencing negative emotions increases as social media usage shifts from morning to later parts of the day. Compared to using social media in the morning, individuals who engage in social media during the afternoon, evening, or late at night exhibit higher log odds of experiencing negative emotions, with the effect being most pronounced for late-night users. Possible reasons behind this might be mental exhaustion from work, fatigue, disrupted sleep patterns etc.

IV. Tendency of getting distracted while busy

i) **Rarely(2):** $\widehat{\beta_{31}}$= -1.31835

As we move from the reference category "Not at all" to "Rarely" for tendency of getting distracted while busy, keeping other predictors constant, the log odds of experiencing negative emotions decreases by 1.31835.

ii) **Sometimes(3):** $\widehat{\beta_{32}}$= -1.44503
As we move from the reference category "Not at all" to "Sometimes" for tendency of getting distracted while busy, keeping other predictors constant, the log odds of experiencing negative emotions decreases by 1.44503.

iii) **Often(4):** $\widehat{\beta_{33}}$= -1.18476
As we move from the reference category "Not at all" to "Often" for tendency of getting distracted while busy, keeping other predictors constant, the log odds of experiencing negative emotions decreases by 1.18476.

iv) **Always(5):** $\widehat{\beta_{34}}$= -1.85572
As we move from the reference category "Not at all" to "Often" for tendency of getting distracted while busy, keeping other predictors constant, the log odds of experiencing negative emotions decreases by 1.85572.

The results indicate that as the tendency to get distracted while busy increases (from "Not at all" to higher order categories, the log odds of experiencing negative emotions decrease. This suggests that individuals who frequently get distracted while busy are less likely to experience negative emotions compared to those who do not get distracted at all. A possible explanation might be that social media provides a temporary mental break from work, and engaging with entertaining content may help relieve work-related stress.

V. Feeling restless or suffering from FOMO

i) **Rarely(2):** $\widehat{\beta_{41}}$= 0.60781
As we move from the reference category "Not at all" to "Rarely" for suffering from FOMO, keeping other predictors constant, the log odds of experiencing negative emotions increases by 0.60781.

ii) **Sometimes(3):** $\widehat{\beta_{42}}$= 1.77002
As we move from the reference category "Not at all" to "Sometimes" for suffering from FOMO, keeping other predictors constant, the log odds of experiencing negative emotions increases by 1.77002.

iii) **Often(4):** $\widehat{\beta_{43}}$= 1.64705
As we move from the reference category "Not at all" to "Often" for suffering from FOMO, keeping other predictors constant, the log odds of experiencing negative emotions increases by 1.64705.

iv) **Always(5):** $\widehat{\beta_{44}}$= -0.72577
As we move from the reference category "Not at all" to "Often" for suffering from FOMO, keeping other predictors constant, the log odds of experiencing negative emotions decreases by 0.72577.

The results indicate that as experiencing FOMO increases from "Not at all" to higher order categories, the log odds of experiencing negative emotions increase.

This suggests that individuals who rarely, sometimes, or often experience FOMO are more likely to experience negative emotions compared to those who do not experience FOMO at all. However, those who always experience FOMO show a decrease in negative emotions as compared to those who do not experience FOMO at all. A possible reason behind this might be that continuous exposure to FOMO-related thoughts may lead to emotional habituation resulting in lower negative emotions.

VI. Comparison with others through the use of social media

i) **Rarely(2):** $\widehat{\beta_{51}}$= 0.70807
As we move from the reference category "Not at all" to "Rarely" for comparison with others, keeping other predictors constant, the log odds of experiencing negative emotions increases by 0.70807.

ii) **Sometimes(3):** $\widehat{\beta_{52}}$= 0.05637
As we move from the reference category "Not at all" to "Sometimes" for comparison with others, keeping other predictors constant, the log odds of experiencing negative emotions increases by 0.05637.

iii) **Often(4):** $\widehat{\beta_{53}}$= 0.37951
As we move from the reference category "Not at all" to "Often" for comparison with others, keeping other predictors constant, the log odds of experiencing negative emotions increases by 0.37951.

iv) **Always(5):** $\widehat{\beta_{54}}$= 1.06074
As we move from the reference category "Not at all" to "Often" for comparison with others, keeping other predictors constant, the log odds of experiencing negative emotions increases by 1.06074.

The results indicate that as comparison from "Not at all" to higher order categories, the log odds of experiencing negative emotions increase. This suggest that engaging in social comparison through social media is associated with an increase in negative emotions. Notably, those who always compare themselves to others show the highest increase in negative emotions.

VII. Feeling pressure to conform to social norms

i) **Maybe:** $\widehat{\beta_{61}}$= 1.11915
As we move from the reference category "No" to "Maybe" for feeling pressure to conform to social norms, keeping other predictors constant, the log odds of experiencing negative emotions increases by 1.11915.

ii) **Yes:** $\widehat{\beta_{62}}$= 3.17757
As we move from the reference category "No" to "Yes" for feeling pressure to conform to social norms, keeping other predictors constant, the log odds of experiencing negative emotions increases by 3.17757.

The results indicate that as feeling pressured to conform to social norms is strongly associated with increased negative emotions. The log odds of experiencing negative emotions rise significantly as individuals feel more compelled to conform to ongoing trends and social norms.

Test for significance of the parameters

Let β be a regression parameter. Now, we want to test for the significance of the regression parameters. Our hypothesis is given by,

$$H_0 : \beta = 0 \; ag \; H_1 : \beta \neq 0$$

For sufficiently large sample size, we apply Central Limit Theorem

$$\frac{\sqrt{n}(\hat{\beta} - \beta)}{\sqrt{se(\hat{\beta})}} \sim \mathrm{AN}(0,1)$$

where $\hat{\beta}$ represents the estimate of β and $se(\hat{\beta})$ represents the estimated standard error of $\hat{\beta}$.

Under H_0, $\beta = 0$, so the test statistic becomes,

$$Z = \frac{\sqrt{n} \;.\; \widehat{\beta}}{\sqrt{se(\widehat{\beta})}} \sim \mathrm{AN}(0,1)$$

We reject H_0 at 10% level of significance iff $|Z_{obs}| > \tau_{\frac{0.1}{2}}$

where Z_{obs} is the observed value of the test statistic and $\tau_{\frac{0.1}{2}}$ is the upper 10% point of a N(0,1) distribution.

$$\tau_{\frac{0.1}{2}} = 1.644854$$

We perform this test for the parameters $\beta_0, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{31}, \beta_{32}, \beta_{33}, \beta_{34}, \beta_{41}$, $\beta_{42}, \beta_{43}, \beta_{44}, \beta_{51}, \beta_{52}, \beta_{53}, \beta_{54}, \beta_{61}$, and β_{62}.

We obtain the following table from R.

Table 16: Table showing the results of the tests of significance for regression parameters

Category of predictor	Parameter (regression coefficient)	Estimates	Estimated standard error	Absolute value of test statistic	p-value	Significant at 10% level
Intercept	β_0	-2.88216	1.58357	1.820	0.06875	Yes
Average time (1 - 3 hours)	β_{11}	1.02568	0.76913	1.334	0.18235	No
Average time (4 - 6 hours)	β_{12}	1.31107	0.81530	1.608	0.10781	No
Average time (More than 6 hours)	β_{13}	-0.76388	1.17628	0.649	0.51608	No

Time of day (Afternoon)	β_{21}	1.65819	1.25091	1.326	0.18498	No
Time of day (Evening)	β_{22}	1.69302	1.22549	1.382	0.16712	No
Time of day (Late night)	β_{23}	2.44061	1.25820	1.940	0.05241	Yes
Distracted when busy (2)	β_{31}	-1.31835	0.91193	1.446	0.14827	No
Distracted when busy (3)	β_{32}	-1.44503	0.90183	1.602	0.10908	No
Distracted when busy (4)	β_{33}	-1.18476	0.93778	1.263	0.20646	No
Distracted when busy (5)	β_{34}	-1.85572	1.26320	1.469	0.14181	No
FOMO (2)	β_{41}	0.60781	0.45965	1.322	0.18606	No
FOMO (3)	β_{42}	1.77002	0.64843	2.730	0.00634	Yes
FOMO (4)	β_{43}	1.64705	0.83647	1.969	0.04895	Yes
FOMO (5)	β_{44}	-0.72577	0.96831	0.750	0.45354	No
Comparison with others (2)	β_{51}	0.70807	0.49552	1.429	0.15302	No
Comparison with others (3)	β_{52}	0.05637	0.57504	0.098	0.92191	No
Comparison with others (4)	β_{53}	0.37951	0.73095	0.519	0.60361	No
Comparison with others (5)	β_{54}	1.06074	1.23672	0.858	0.39106	No
Conform to social norms (Maybe)	β_{61}	1.11915	0.47748	2.344	0.01908	Yes
Conform to social norms (Yes)	β_{62}	3.17757	1.36061	2.335	0.01952	Yes

Observations

We observe that at 10% level of significance, usage of social media at late night, facing FOMO sometimes or very often and conformation to social norms show strong statistical significance for predicting occurrence of negative emotions after using social media.

- Using social media late at night is associated with a higher likelihood of negative emotions. This could be due to various reasons, such as watching content that triggers overthinking, increased social comparisons, or disrupted sleep patterns caused by prolonged screen time.
- Individuals who sometimes or often experience FOMO are more likely to report negative emotions. This suggests that constantly seeing updates about others' activities, achievements, or social events can lead to feelings of exclusion, dissatisfaction, or anxiety.
- The pressure to fit into social norms on social media can lead to negative emotions by making people feel the need to maintain a perfect image, follow trends, and seek validation through likes and comments. This can cause stress and self-doubt, and over time, it may lower self-esteem and increase feelings of anxiety.
- Average time (4 - 6 hours) and distracted when busy (3) are not statistically significant, but their p-values are very close to 0.1 It seems that they might have an effect on the occurrence of negative emotions after using social media. Their effects can be studied further with larger datasets.

<u>Goodness of fit</u>

Here we observe that

Null deviance = 238.23

Residual deviance = 183.06

The reduction in deviance (from 238.23 to 183.06) indicates that the model explains a reasonable portion of the variation in the dependent variable, indicating that our fit is moderately good.

<u>Classification statistics</u>

Logistic regression is a widely used statistical method for binary classification tasks. It estimates the probability of an instance belonging to a particular category based on a set of independent variables. In our model, the response variable Y represents whether an individual experiences negative emotion after using social media: yes (1) and no (0). Our objective is to assess how accurately our logistic regression model predicts these outcomes.

Having fit the model on the data, we obtain the estimate of π_r as $\widehat{\pi_r} = \frac{e^{\widehat{\eta_r}}}{1+e^{\widehat{\eta_r}}}$ r=1(1)n

Now, our aim is to estimate $\widehat{Y}$.

To obtain $\widehat{Y}$, we first define a threshold say c, and compare each estimated probability to c. If the estimated probability $\hat{\pi}$ exceeds c, then we set the predicted variable to be 1 and 0 otherwise. Now, our question is, how to choose the threshold.

<u>Confusion matrix</u>

A confusion matrix is a fundamental tool for evaluating the performance of classification models. It provides a detailed breakdown of a model's predictions by comparing the actual vs. predicted outcomes. The matrix consists of four key components: True Positives (TP), True

Negatives (TN), False Positives (FP), and False Negatives (FN), which help to assess the accuracy of the logistic regression model.

Y \ $\hat{Y}$	1	0
1	TP	FN
0	FP	TN

For each threshold, we construct a 2x2 confusion matrix of matches and mismatches.

<u>TPR & FPR</u>

The probability that the predicted outcome is a success given the observed outcome is a success is known as True Positive Rate(TPR) or sensitivity.

TPR = $\hat{P}(\hat{Y} = 1|Y = 1) = \frac{TP}{TP+FN}$

The probability that the predicted outcome is a success given the observed outcome is a failure is known as False Positive Rate(FPR) or (1 – Specificity).

TPR = $\hat{P}(\hat{Y} = 1|Y = 0) = \frac{FP}{FP+TN}$

We consider each $\widehat{\pi_r}$ r=1(1)n to be the threshold value and calculate TPR and FPR for each case.

A good model is one which yields maximum TPR and minimum FPR. Based on this, we choose that value as the threshold for which TPR(1-FPR) is maximum.

Using R, we get the optimal threshold value as 0.4505884.

The confusion matrix using threshold = 0.4505884 is given by,

Y \ $\hat{Y}$	1	0
1	70	19
0	23	60

TPR = $\frac{70}{70 + 19} = 0.7865169$

This implies that the model correctly identifies 78.65% of actual positive cases.

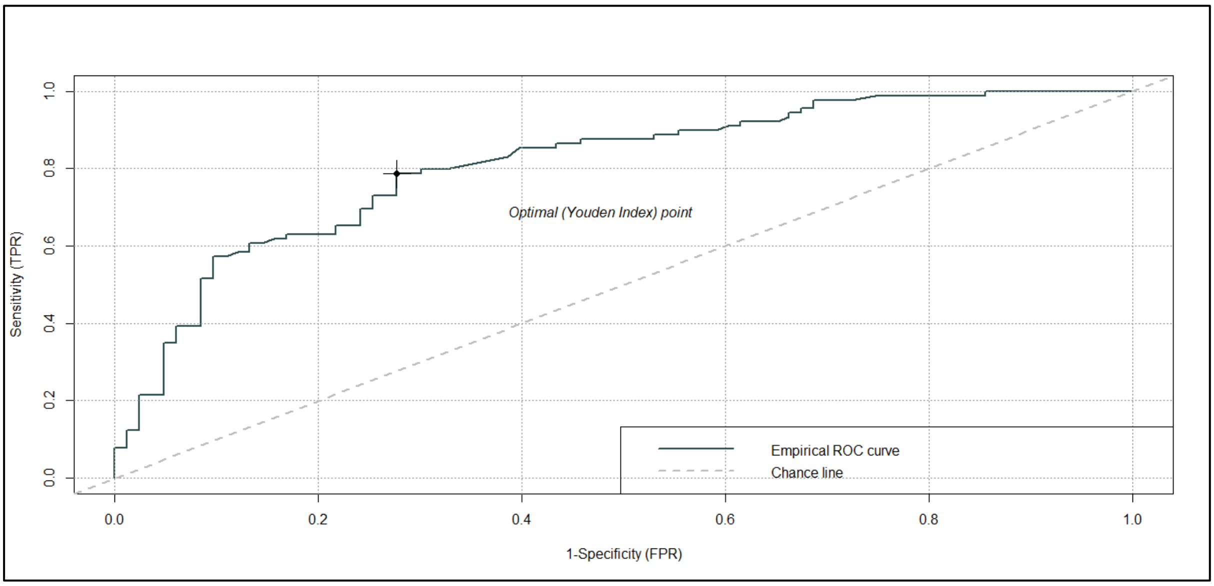
FPR = $\frac{23}{23+60} = 0.2771084$

This implies that the model classified 27.71% of actual negative cases as positive.

Receiver Operating Characteristic curve

The Receiver Operating Characteristic curve is obtained by plotting the TPR on the y axis and FPR on the x axis. The line TPR=FPR indicates the chance line which implies that for a threshold with equal TPR and FPR, prediction is left to chance. An ideal situation is one for which TPR=1 and FPR=0. However, in practice, it is not possible to get such a predicted model. Generally, we get a curve which lies above the chance line. The **area under the ROC curve** serves as a measure to evaluate the overall classification performance of the model.

Figure 17: ROC Curve for our logistic regression model



From R, we obtain that area under curve = 0.8089

This indicates that the model has a good ability to distinguish between positive and negative cases, significantly better than random guessing (AUC = 0.5). There is an 80.89% probability that the model will correctly distinguish between a randomly chosen positive and negative case. Since AUC values between 0.8 and 0.9 are considered strong, hence we can say that the model demonstrates high predictive accuracy.

Impact of social media usage on sleep patterns on students

In addition to studying the relationship between social media usage and mental health of students, we are also interested in how this engagement influences their sleep patterns. Possible variables from our study related to social media engagement which might disrupt an individual's sleep patterns are –

i) Average time spent daily on social media
ii) Preferred time of the day for using social media
iii) Academic or professional workload
iv) Usage without a specific purpose
v) Tendency of getting distracted when busy
vi) Feeling restless or suffering from FOMO
vii) Comparison with other people through the use of social media
viii) Facing negative emotions after using social media

We shall use Kendall's τ_b measure to analyze these associations.

Kendall's τ_b measure

Kendall's Tau-b is a measure of rank correlation that assesses the strength and direction of association between two ordinal variables. It accounts for both concordant and discordant pairs while adjusting for ties in both variables, making it particularly effective in datasets where ties are common. It ranges from -1 to +1. +1 indicates perfect positive association, -1 indicates perfect negative association, 0 indicates no association.

Let (U_1,V_1), (U_2,V_2), ….. (U_n,V_n) be n pairs of observations for two ordinal variables U and V. Consider two pairs of observations (U_i,V_i) and (U_j,V_j). i=1(1)n, j=1(1)n

These pairs are said to be **concordant** if

i) $u_i > u_j$ and $v_i > v_j$
ii) $u_i < u_j$ and $v_i < v_j$

These pairs are said to be **discordant** if

i) $u_i > u_j$ and $v_i < v_j$
ii) $u_i < u_j$ and $v_i > v_j$

These pairs are said to be **ties on U only** if

i) $u_i = u_j$ and $v_i < v_j$
ii) $u_i = u_j$ and $v_i > v_j$

These pairs are said to be **tied on V only** if

i) $u_i > u_j$ and $v_i = v_j$
ii) $u_i < u_j$ and $v_i = v_j$

These pairs are said to be **tied on both U and V** if $u_i = u_j$ and $v_i = v_j$.

We define the following:

2C = Total number of concordant pairs

2D = Total number of discordant pairs

$2T_U$ = Total number of pairs tied on U only

$2T_V$ = Total number of pairs tied on V only

$2T_{UV}$ = Total number of pairs tied on both U and V

Then Kendall τ_b measure is given by,

$$\tau_b = \frac{C - D}{\sqrt{C + D + T_U}\sqrt{C + D + T_V}}$$

1. **Association between sleep problems and average time spent daily on social media**

Table 17: Contingency table of sleep problems and average time spent daily on social media

Average time / Sleep problems	Less than 1 hour	1-3 hours	4-6 hours	More than 6 hours	Total
1	8	40	23	1	72
2	1	20	8	3	32
3	3	26	10	3	42
4	2	9	8	1	20
5	1	0	4	1	6
Total	15	95	53	9	172

Kendall τ_b = 0.091469

2. **Association between sleep problems and preferred time of the day for using social media**

Table 18: Contingency table of sleep problems and preferred time of the day for using social media

Time of day / Negative emotions	Morning	Afternoon	Evening	Late night	Total
1	3	22	27	20	72
2	0	8	14	10	32
3	2	8	13	19	42
4	0	5	4	11	20
5	1	0	2	3	6
Total	6	43	60	63	172

Kendall τ_b = 0.157219

3. **Association between sleep problems and academic or professional workload**

Table 19: Contingency table of sleep problems and academic or professional workload

Workload / Sleep problems	1	2	3	4	5	Total
1	0	2	33	30	7	72
2	0	4	11	10	7	32
3	0	2	14	22	4	42
4	0	0	9	8	3	20
5	0	0	1	2	3	6
Total	0	8	68	72	24	172

Kendall τ_b = 0.092796

4. **Association between sleep problems and frequency of using social media without a specific purpose**

Table 20: Contingency table of sleep problems and frequency of using social media without a specific purpose

Usage without a specific purpose / Sleep problems	1	2	3	4	5	Total
1	2	7	32	25	6	72
2	0	4	15	11	2	32
3	0	0	11	28	3	42
4	0	1	8	10	1	20
5	0	1	0	3	2	6
Total	2	13	66	77	14	172

Kendall τ_b = 0.181405

5. Association between sleep problems and tendency of getting distracted when busy

Table 21: Contingency table of sleep problems and tendency of getting distracted when busy

Distracted when busy / Sleep problems	1	2	3	4	5	Total
1	7	16	24	19	6	72
2	0	11	12	9	0	32
3	1	5	20	14	2	42
4	0	1	5	11	3	20
5	0	1	0	3	2	6
Total	8	34	61	56	13	172

Kendall τ_b = 0.212920

6. Association between sleep problems and feeling restless or suffering from FOMO

Table 22: Contingency table of sleep problems and feeling restless or suffering from FOMO

FOMO / Sleep problems	1	2	3	4	5	Total
1	31	27	6	2	6	72
2	12	11	3	5	1	32
3	10	13	14	4	1	42
4	1	7	8	3	1	20
5	1	2	3	0	0	6
Total	55	60	34	14	9	172

Kendall τ_b = 0.237187

7. Association between sleep problems and comparison with other people through the use of social media

Table 23: Contingency table of sleep problems and comparison with other people through the use of social media

Comparison / Sleep problems	1	2	3	4	5	Total
1	23	24	17	7	1	72
2	8	8	10	4	2	32
3	11	14	12	3	2	42
4	4	7	3	6	0	20
5	0	1	0	2	3	6
Total	46	54	42	22	8	172

Kendall τ_b = 0.150128

8. Association between sleep problems and facing negative emotions after using social media

Table 24: Contingency table of sleep problems and facing negative emotions after using social media

Negative emotions / Sleep problems	1	2	3	4	5	Total
1	18	30	20	3	1	72
2	4	11	14	3	0	32
3	4	11	19	6	2	42
4	0	5	8	5	2	20
5	0	0	1	3	2	6
Total	26	57	62	20	7	172

Kendall τ_b = 0.363315

From the above measures of association, we can conclude that in the light of the given data, it seems that

- **Average time spent daily on social media** and **academic or professional workload** do not seem to significantly affect sleep patterns.
- **Usage of social media at evening or late night** , **usage without a specific purpose** and s**ocial comparison** show weak positive association with development of sleep problems.
- **Getting distracted while busy, facing negative emotions after using social media,** and **FOMO** have moderately strong positive association with sleep problems faced by individuals.

Conclusion

In the modern landscape of evolving digital connectivity, social media emerges as a powerful force that shapes communication, emotional well-being, and lifestyle patterns. Following a data-driven approach, we analyzed the influence of different social media factors on the mental health of university students in India. Through different statistical techniques, we found that several factors, like age, sex, workload and purposeless usage do not show any considerable association with causing negative emotions among students. The insights also reveal that getting distracted by social media when busy acts as a stress relief for many students. However, factors like FOMO, social comparison and pressure to conform to social norms play a significant role in intensifying negative emotions like stress and anxiety. The posts on social media are often particularly focused on visually appealing or perfect moments, which might arouse feelings of jealousy or inadequacy.

Factors like experiencing negative emotions, FOMO and social comparison also influence the sleep patterns of any individual. Constant scrolling through social media, especially at late night, can lead to overthinking and delayed sleep time. The fear of missing out may also encourage late-night scrolling, reducing sleep duration and quality.

As a remedy to tackle these problems, students should restrict their social media usage by setting time constraints and content moderation filters. They should avoid consuming negative or sensitive content and should shift their focus to healthy and informative posts. They should also try to decrease the usage during free time and focus on socializing offline, spending time with friends and family, and engaging in extracurricular activities.

Limitations of the study

Although this study identifies some major factors related to social media engagement that affect the mental health of students, there are certain limitations. Firstly, the primary data collected via survey may be biased. Participants might not accurately disclose their social media usage or honestly answer questions related to their mental health. There may exist several other factors which affect the mental health of students apart from social media. The findings might be more accurate if larger and more diverse datasets are considered. Finally, to keep pace with the rapid development of social media, continuous research is required to ensure up-to-date conclusions.

Acknowledgment

I would like to express my sincere thanks of gratitude to my supervisor Dr. Surabhi Dasgupta for her constant support and guidance in my project work.

I also want to express my sincere gratitude to all the esteemed professors of Department of Statistics, St. Xavier's College (Autonomous), Kolkata, who provided me with the opportunity to explore different fields of statistics and motivated me to complete this research work.

I would like to thank my friends for their valuable advice during the project. In conclusion, I want to thanks my family for their care, guidance and inspiration.

References

1. Khalaf A M, Alubied A A, Khalaf A M, et al. (August 05, 2023) The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review. Cureus 15(8): e42990. doi:10.7759/cureus.42990
2. Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. Journal of Computer-Mediated Communication, 12(4), 1143–1168.
3. Keles, B., McCrae, N., & Grealish, A. (2020). A Systematic Review: The Influence of Social Media on Depression, Anxiety and Psychological Distress in Adolescents. International Journal of Adolescence and Youth, 25(1), 79–93.
4. Appel, H., Gerlach, A. L., & Crusius, J. (2016). The Interplay Between Facebook Use, Social Comparison, Envy, and Depression. Current Opinion in Psychology, 9, 44–49.
5. Karim F, Oyewande AA, Abdalla LF, Chaudhry Ehsanullah R, Khan S. Social Media Use and Its Connection to Mental Health: A Systematic Review. Cureus. 2020 Jun 15;12(6):e8627. doi: 10.7759/cureus.8627. PMID: 32685296; PMCID: PMC7364393.
6. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. J Technol Behav Sci. 2020 Sep;5(3):245-257. doi: 10.1007/s41347-020-00134-x. Epub 2020 Apr 20. PMID: 33415185; PMCID: PMC7785056.