

Egypt University of Informatics

Computer and Information Systems

Data Analysis Course

"Examining the Impact of Social Media on Mental Health"

**Submitted by:**

Basmala Yasser (22-101098)

Abeer Sherif (22-101136)

Renad Sameh (22-101225)

Nagham Mohamed (22-101286)

**Supervised by:**

Dr. Mohamed Taher ElRefaei PhD

Eng. Nadine ElSaeed

Eng. Mai Ismail

# Introduction

Social media has become a constant presence in our daily lives, significantly impacting our mental health. This report dives into the effects of social media based on user data, examining everything from how long we spend online to how it influences our emotions and behaviours. We explore the common feelings of distraction and restlessness when we're away from our screens, the tendency to compare ourselves to others, and the pursuit of likes and comments for validation. The findings highlight how these habits can affect our mood, concentration, and sleep. By understanding the influence of social media, we can make informed decisions to foster healthier interactions and improve our mental well-being.

# Research Question

* Is there a significant association between social media usage and mental health issues among the participants?

# Hypothesis

### First Hypothesis

* Null Hypothesis (H0): Increased time spent on social media does not lead to an increase in mental health issues among different groups.
* Alternative Hypothesis (H1): Increased time spent on social media leads to an increase in mental health issues among different groups.

### Second Hypothesis

* Null Hypothesis (H0): There is no positive correlation between any of the pairs of mental health issues (Anxiety, Depression, and Self-Esteem) among participants.
* Alternative Hypothesis (H1): There is a positive correlation between at least one pair of mental health issues (Anxiety, Depression, and Self-Esteem) among participants.

# Population of Interest:

All individuals identified as University Students, School Students, Salaried Workers, or Retired.

# Sampling Method:

We chose random sampling because it ensures that every respondent has an equal chance of being selected, reducing selection bias and increasing the representativeness and generalizability of the results. Our sample consists of respondents from different demographics and social media usage patterns.

Bias Identification:

We recognized potential biases in using an existing dataset for our study on social media and mental health. Our confidence in the dataset's accuracy may influence our analysis, and using data from an external website could introduce further bias. To address this, we ensured a diverse sample, conducted through data cleaning, and examined the original survey for measurement bias.

Survey Questions/Collected Data/Dataset:

The dataset used in this study contains responses from a survey focused on the relationship between social media usage and mental health among participants. The dataset contains 21 columns: Age, Gender, Relationship Status, Occupation, Affiliations, Social Media User, Platforms Used, Time Spent, ADHD, Anxiety, Self Esteem, Depression and 481 entire (rows)

**Age:** Age of the participants.

**Gender:** Female or Male

**Relationship Status:** Single or Married or In a relationship or divorced.

**Occupation:** School Student or University Student or Salaried Worker or Retired

**Affiliations:** School or university or Government or Company

**Social Media** **User:** Indicates if the participant uses social media.

**Social Media** **Platforms:** Social media platforms the participant uses.

**Time Spent:** Amount of time spent on social media.

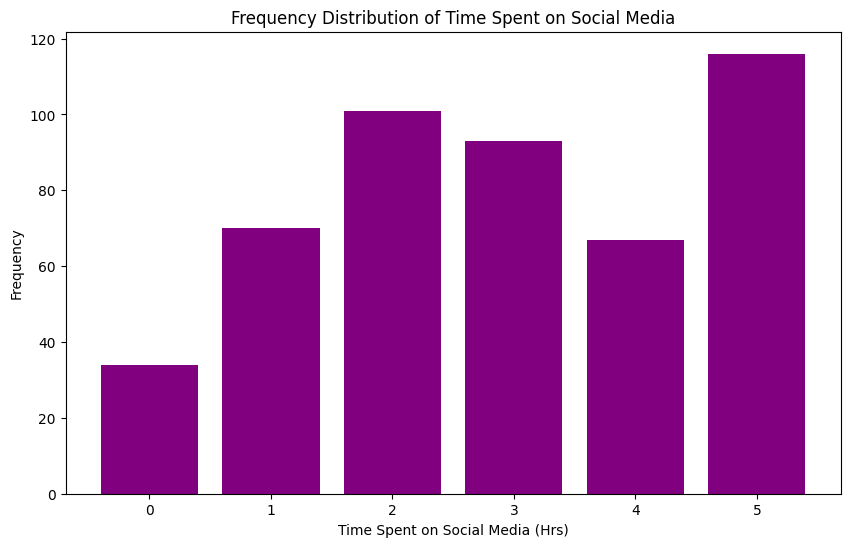
**ADHD:** Indicates if the participant has ADHD.

**Anxiety:** Anxiety level of the participant.

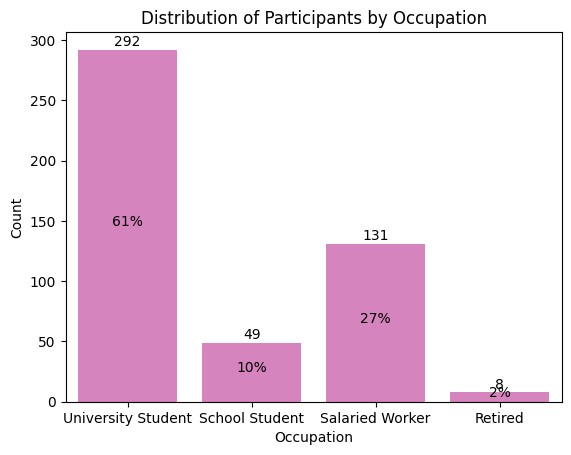
**Self-Esteem:** Self-esteem level of the participant.

**Depression:** Depression level of the participant.

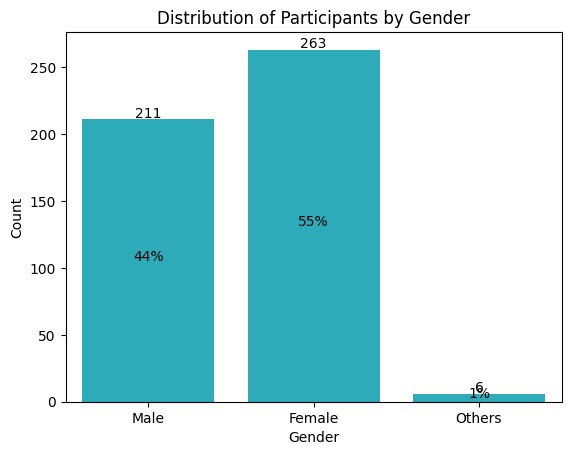
# Analysis:



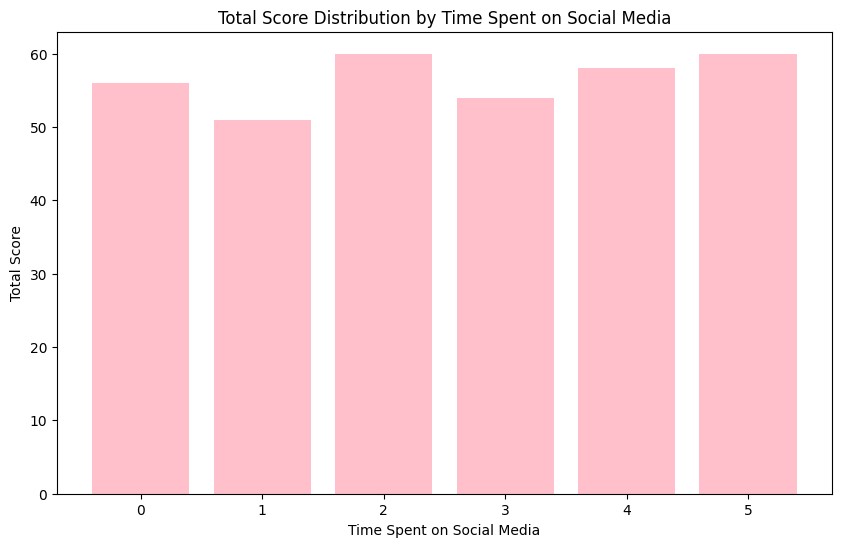
From the above plot, we can infer that in the sample, there are less than 40 people who have an average social media use of less than an hour. The other groups each have 60 to 120 people with average social media use of 1 to 5 hours or more.



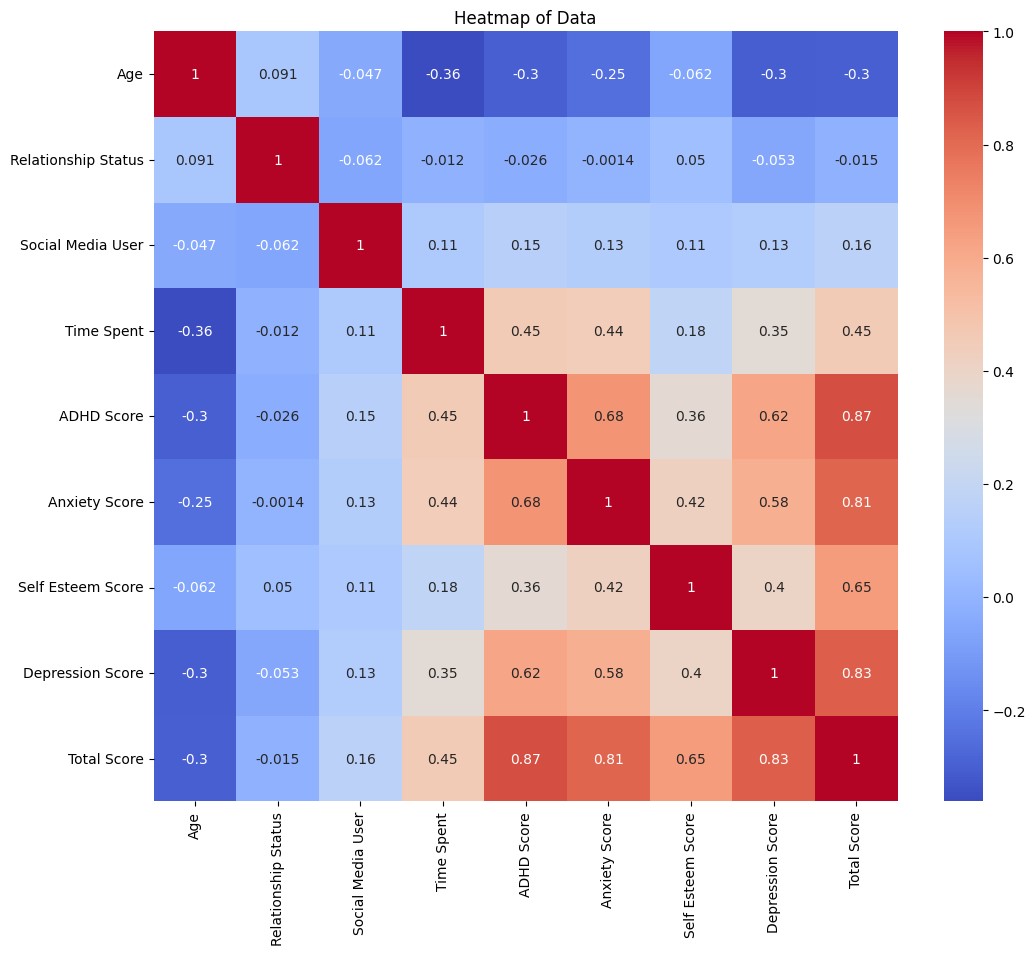
The sample is over-representated by University students, making up an overwhelming 62% of the sample.



Approximately 263 participants out of 473 are female, making up the majority in the sample. 'Others' make up approximately 1% of the sample size, which makes it impossible to make statistical inferences based on the "Other" category specifically.



The bar chart shows that individuals who spend 2, 4, and 5 units of time on social media have higher total scores compared to those spending 1 and 3 units



The heat map shows correlations between various factors and mental health scores, with higher correlations shown in darker blue and lighter red indicating negative correlations. It highlights strong interrelations among ADHD score, anxiety score, depression score, and total mental health score, indicating that high scores in one are associated with high scores in the others. Time spent on social media is moderately positively correlated with these mental health issues, suggesting increased usage causes higher levels of ADHD, anxiety, and depression. Age negatively correlates with these issues, meaning younger individuals tend to have worse mental health. Relationship status has minimal impact on mental health. Overall, the heat map effectively visualizes these relationships, emphasizing significant connections between social media usage and mental health, as well as the varying impacts of age and relationship status.

# **First Hypothesis Analysis:**

F-value: 3.677861625970449

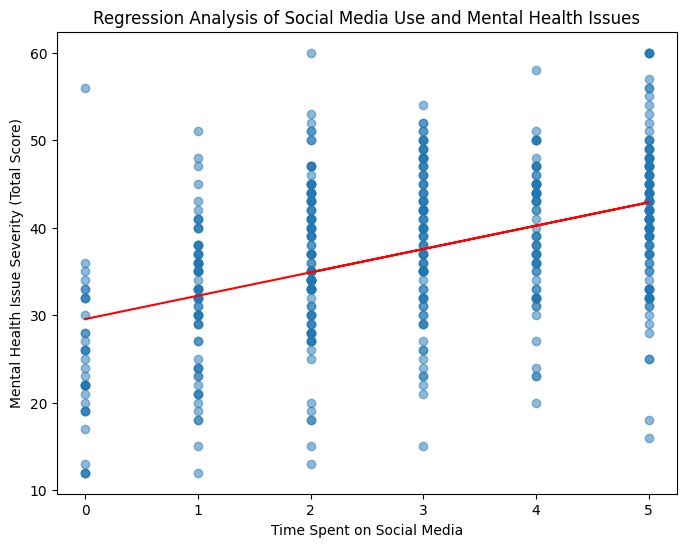
P-value: 6.39865151740502e-13

There is a statistically significant difference in Time Spent across Mental health issues

Coefficient for Time Spent on Social Media: 2.667994020020533

P-value: 9.144738799299639e-26

R-squared: 0.205981586322048



**The analysis shows a significant positive correlation between time spent on social media and mental health issues. Each additional unit of social media use is associated with a 2.67-point increase in mental health severity, with a highly significant P-value (9.14e-26) and an R-squared of 0.206. The scatter plot confirms this relationship, indicating that more time on social media leads to greater mental health problems.**

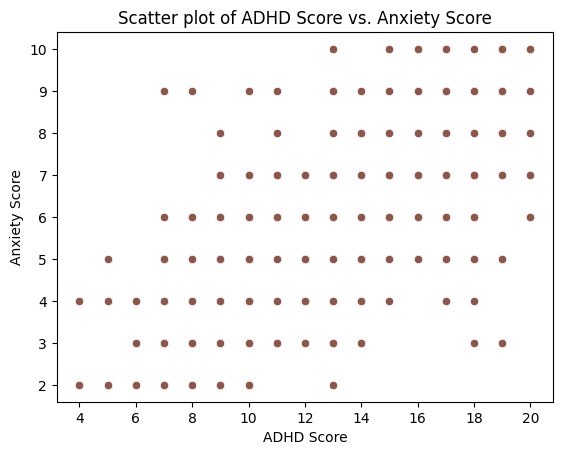
# **Second Hypothesis Analysis:**

**1-There is a statistically significant association between ADHD Score and Anxiety Score.**

Correlation coefficient: 0.6762068183656027

P-value: 1.991265780653397e-65

Chi-Square Test: chi2=591.2830626873806, p-value=1.1724077531307608e-60



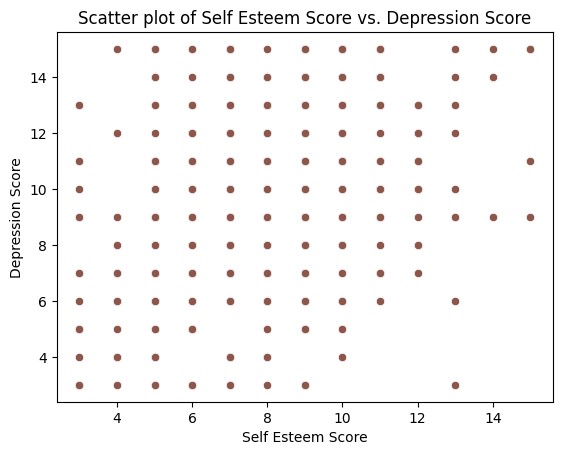
The scatter plot shows a clear link between ADHD scores and anxiety scores. A strong positive correlation of 0.676 means that as ADHD scores increase, anxiety scores also tend to increase. This relationship is statistically significant, as indicated by the very low p-value (1.99e-65). The Chi-Square test also supports this finding, showing a significant association between ADHD and anxiety scores with a chi2 value of 591.283 and a p-value of 1.17e-60. This means that people with higher ADHD are likely to have higher anxiety.

**2- There is a statistically significant association between Self Esteem Score and Depression Score.**

Correlation coefficient: 0.40224336191936266

P-value: 4.2960255338312135e-20

Chi-Square Test: chi2=337.1642474707411, p-value=1.56311286026107e-17



The scatter plot shows the relationship between self-esteem scores and depression scores. There is a statistically significant association between the two with a correlation coefficient of 0.402, indicating a moderate positive relationship. This means that as self-esteem scores increase, depression scores also tend to increase. The extremely low p-value (4.30e-20) confirms that this relationship is statistically significant. Additionally, the Chi-Square test (chi2 = 337.164, p-value = 1.56e-17) supports this finding, further indicating a significant association between self-esteem and depression scores.

#### 

# Hypothesis Testing Steps

#### 1st Hypothesis

* **Step 1:** The null and alternative hypotheses were defined as mentioned earlier.
* **Step 2:** The count of data points in each group was verified, and groups with no data were filtered out.
* **Step 3:** ANOVA was performed to test for differences in the mean time spent on social media across the Total Score groups.
* **Step 4:** The results were interpreted:
  + If the p-value is less than 0.05, it was concluded that there is a statistically significant difference in time spent across Total Score groups.
  + If the p-value is greater than or equal to 0.05, it was concluded that there is no statistically significant difference in time spent across Total Score groups.

**2nd Hypothesis**

* **Step 1:** The null and alternative hypotheses were defined as mentioned earlier.
* Step 2: The Pearson correlation coefficients between the pairs of mental health scores were calculated:
  + ADHD Score and Anxiety Score
  + Self Esteem Score and Depression Score
* **Step 3:** Scatter plots were created to visualize the correlations between pairs of mental health scores.
* Step 4: Chi-Square tests were performed to check for associations between pairs of mental health scores:
  + ADHD Score and Anxiety Score
  + Self Esteem Score and Depression Score
* **Step 5:** The Chi-Square test results were interpreted:
  + If the p-value is less than 0.05, it was concluded that there is a statistically significant association between the mental health scores in question.
  + If the p-value is greater than or equal to 0.05, it was concluded that there is not enough evidence to support a statistically significant association between the mental health scores in question.
* **Step 6:** The Pearson correlation coefficients were calculated for further verification of the relationships, and the results were interpreted:
  + The correlation coefficient and p-value for each pair of mental health scores were reported.
  + It was determined if there is a significant positive correlation based on the p-values.

# Conclusion

# This report explores the significant effects of extended social media use on mental health, revealing a strong link between heavy social media engagement and increased mental health disorders. We hypothesized that increased social media time correlates with higher rates of mental health issues, including anxiety, ADHD symptoms, depression, and lower self-esteem. Our analysis confirmed this hypothesis, showing that individuals who spend a lot of time on social media report higher levels of distraction and anxiety, along with lower self-esteem. Visual evidence from our charts and graphs supports these findings, showing the harmful impact of extensive social media use. These results highlight the importance of using social media in moderation to reduce negative effects on mental health.

# Any potential issues

1. Inconsistent Data Types: Ensure that all entries in columns meant to contain numerical data (like age or scales of 1-5) are indeed numeric and not mixed with text. If there are non-numeric entries, they will need to be cleaned or converted.
2. Missing Values: It's common to have missing values in survey data. You may need to decide whether to fill these missing values, drop them, or ignore them during analysis depending on the extent and nature of the missingness.
3. Data Privacy: The dataset contains potentially sensitive information, such as details about personal mental health states. Ensure that data handling complies with relevant data protection laws and ethical guidelines.
4. Column Renaming: Simplify column names, originally in the form of questions, to one-word names for easier reference and analysis.