Social Media and Mental Health

Social research (2023/2024)

Content

- Introduction
- Methodology, data and type of analysis
- Findings (Data Visualiation)
- Discussion
- Conclusion
- Bibliography

Introduction.
Does time spent on Social Media affect Mental Health?

Social media has become a prominent fixture in the lives of many individuals facing the challenges of mental illness. Social media refers broadly to web and mobile platforms that allow individuals to connect with others within a virtual network (such as Facebook, Twitter, Instagram, Snapchat, or LinkedIn), where they can share, co-create, or exchange various forms of digital content, including information, messages, photos, or videos (Ahmed et al. 2019). Studies have reported that individuals living with a range of mental disorders, including depression, psychotic disorders, or other severe mental illnesses, use social media platforms at comparable rates as the general population, with use ranging from about 70% among middle-age and older individuals to upwards of 97% among younger individuals (Aschbrenner et al. 2018b; Birnbaum et al. 2017b; Brunette et al. 2019; Naslund et al. 2016).The current correlational study examined 480 young adults for their time spent using social media, importance of social media in their lives and tendency to engage in vaguebooking (posting unclear but alarming sounding posts to get attention). Outcomes considered included general mental health symptoms, loneliness, social anxiety and decreased empathy.

Journal of Technology in Behavioral Science

Methodology, data and type of analysis

- This dataset was originally collected for a data science and machine learning project that aimed at investigating the potential correlation between the amount of time an individual spends on social media and the impact it has on their mental health.
- The project involves conducting a survey to collect data, organizing the data, and using machine learning techniques to create a predictive model that can determine whether a person should seek professional help based on their answers to the survey questions.
- Dataset consist of It consists of 7 variables, and 12 Likert scale based questions giving us points that measure either frequency or intensity of various aspects of Mental Health. A low score of 0 generally indicates low frequency or intensity, and a high score of 5 typically indicates high frequency or intensity. The main objective of this study is to investigate whether there is a correlation between Social Media usage and Mental health, and to explore and try to predict whether the individual is suffering from mental health symptoms and should be recommended a mental health checkup, based on multivariate predictive modelling.

Dataset link

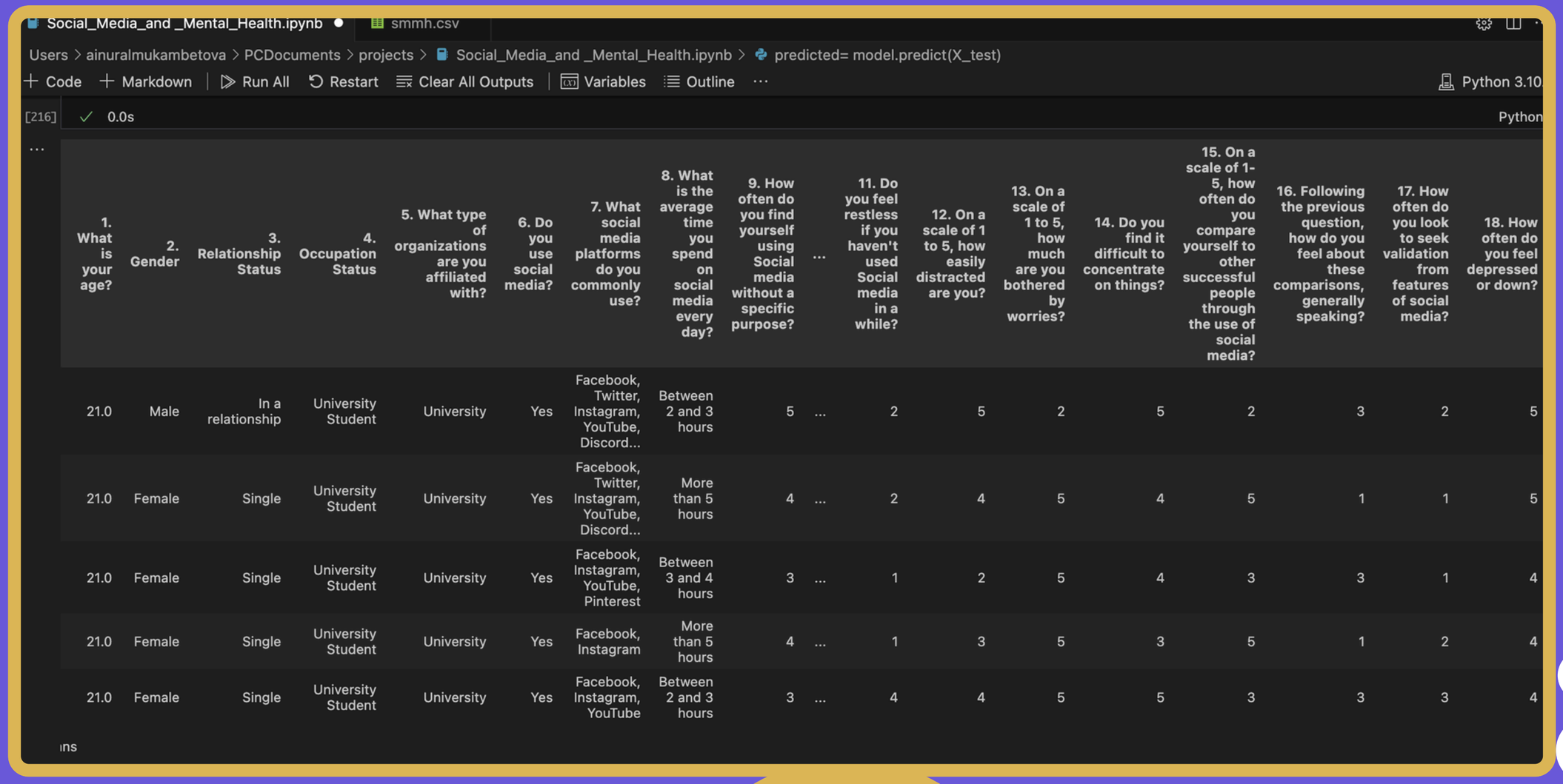
Variables

- Age
- Gender
- Relationship Status
- Occupation Status
- Affiliated Organizations
- Social Medias Used
- Time spent - social media use, in hours

Measurements of Frequency or Intensity of mental health symptoms is take through Likert Scale questions

1. Purposeless use of Social Media [ADHD] - Question 9
2. Distracted by Social Media [ADHD] - Question 10
3. Restlessness if Social Media not used [Anxiety] - Question 11
4. Ease of Distraction by Social Media [ADHD] - Question 12
5. Bothered by worries [Anxiety] - Question 13
6. Difficulty in concentrating [ADHD] - Question 14
7. Comparison of self to peers [Self Esteem] - Question 15
8. Feelings about above comparison [Self Esteem] - Question 16
9. Validation sought from Social Media [Self Esteem] - Question 17
10. Feelings of Depression [Depression] - Question 18
11. Fluctuation of interest [Depression] - Question 19
12. Sleep Issues [Depression] - Question 20

First 5 raws of dataframe


Social_Media_and _Mental_Health.ipynb ● smmh.csv

Users > ainuralmukambetova > PCDocuments > projects > Social_Media_and _Mental_Health.ipynb > predicted= model.predict(X_test)

+ Code + Markdown | Run All Restart Clear All Outputs | Variables Outline ··· Python 3.10

[216] ✓ 0.0s Python

	1. What is your age?	2. Gender	3. Relationship Status	4. Occupation Status	5. What type of organizations are you affiliated with?	6. Do you use social media?	7. What social media platforms do you commonly use?	8. What is the average time you spend on social media every day?	9. How often do you find yourself using Social media without a specific purpose?	...	11. Do you feel restless if you haven't used Social media in a while?	12. On a scale of 1 to 5, how easily distracted are you?	13. On a scale of 1 to 5, how much are you bothered by worries?	14. Do you find it difficult to concentrate on things?	15. On a scale of 1-5, how often do you compare yourself to other successful people through the use of social media?	16. Following the previous question, how do you feel about these comparisons, generally speaking?	17. How often do you look to seek validation from features of social media?	18. How often do you feel depressed or down?
	21.0	Male	In a relationship	University Student	University	Yes	Facebook, Twitter, Instagram, YouTube, Discord...	Between 2 and 3 hours	5	...	2	5	2	5	2	3	2	5
	21.0	Female	Single	University Student	University	Yes	Facebook, Twitter, Instagram, YouTube, Discord...	More than 5 hours	4	...	2	4	5	4	5	1	1	5
	21.0	Female	Single	University Student	University	Yes	Facebook, Instagram, YouTube, Pinterest	Between 3 and 4 hours	3	...	1	2	5	4	3	3	1	4
	21.0	Female	Single	University Student	University	Yes	Facebook, Instagram	More than 5 hours	4	...	1	3	5	3	5	1	2	4
	21.0	Female	Single	University Student	University	Yes	Facebook, Instagram, YouTube	Between 2 and 3 hours	3	...	4	4	5	5	3	3	3	4

ıns

Steps before to get the findings

- Data Pre-processing and Cleaning
  - Renaming the Columns
- Re-arranging the Columns
- Missing Value Detection and Treatment
- Data Transformation
- Summation of Scores of different aspects of mental well being
- Adding an "Outcome" column

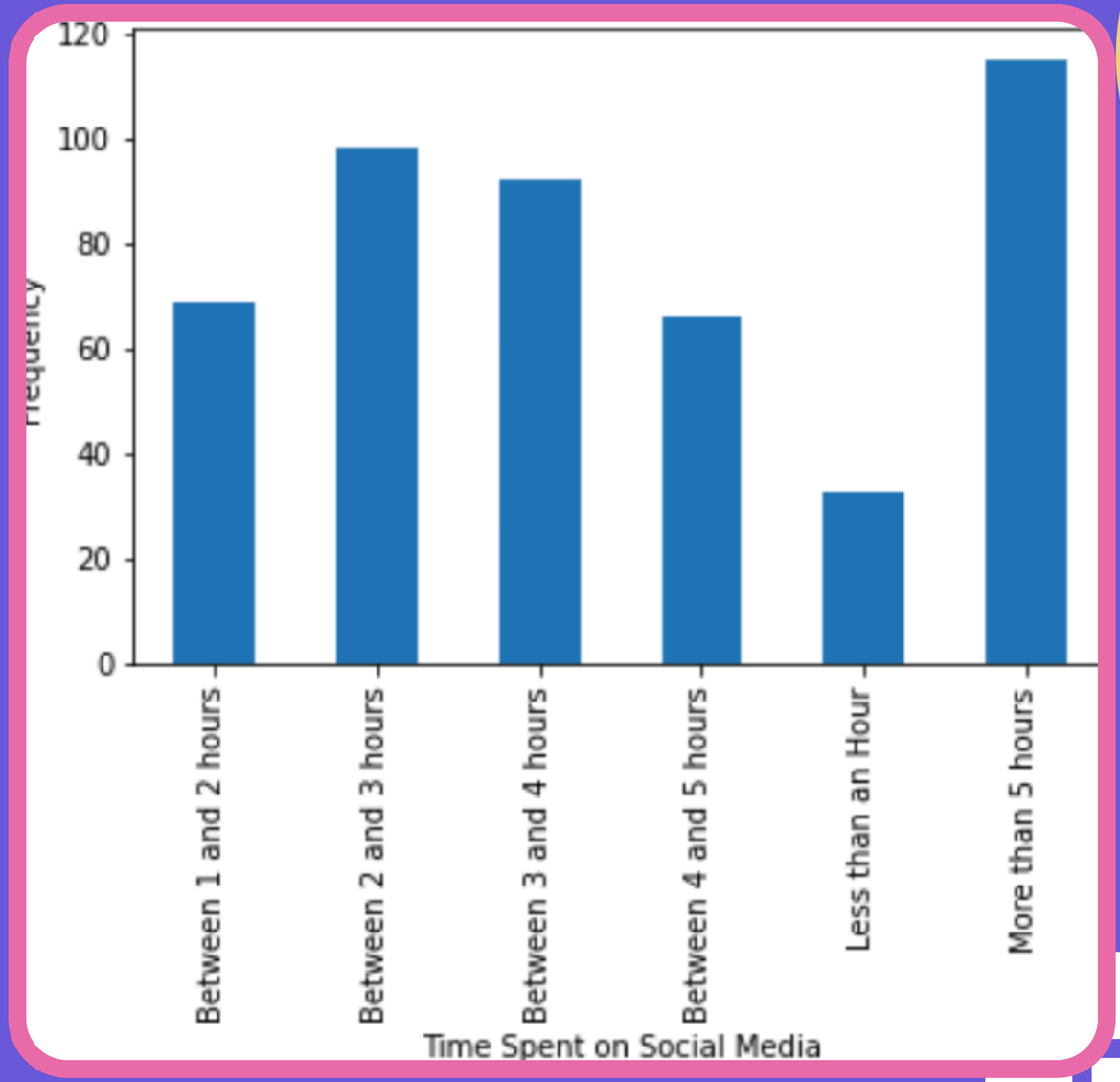
Adding an "Outcome" column

- In this section, we consider adding an "Outcome" variable based on the "Total Score" of the participant.
- Previously, it was established that the "Total score" indicates magnitude of individual experiencing negative symptoms of mental health. An accumulated total score of 59 is the highest an individual can obtain from the questionnaire, which would indicate that the individual is definitely experiencing negative symptoms in some aspect of mental health, based on binary classification.
- The Outcome variable is defined as whether we think that the individual is experiencing mental health disease symptoms in some aspect, and whether we are recommending the participant to get a mental health checkup or not.
- A participant with a score of 3 out of 5 on every question on the questionnaire suggests that they are are experiencing slight to moderate symptoms in every aspect of mental health, but they may not be severe or frequent enough, giving a score of 35 (12 questions, scores of 3 on each question except for self esteem question #2 with score of 2 = 35 score total score).
- Therefore, we assign a rational value of 40 to be the point where we can reliably say that the individual is very likely to be suffering severely and extremely frequently from some symptoms, and thus we highly recommend a mental health checkup.
- An Outcome of 0 means that individual is not confirmed to be experiencing severe mental health symptoms. Therefore we do not think the individual needs to go to get a mental health check up.
- An Outcome of 1 means that the individual definitely experiencing some severe negative symptoms of mental health. They are recommended to go to get a mental health check up.
- Note that the Total score variable will be dropped later when we use logistic regression to train and predict data.

Data Visualisation

Time spent

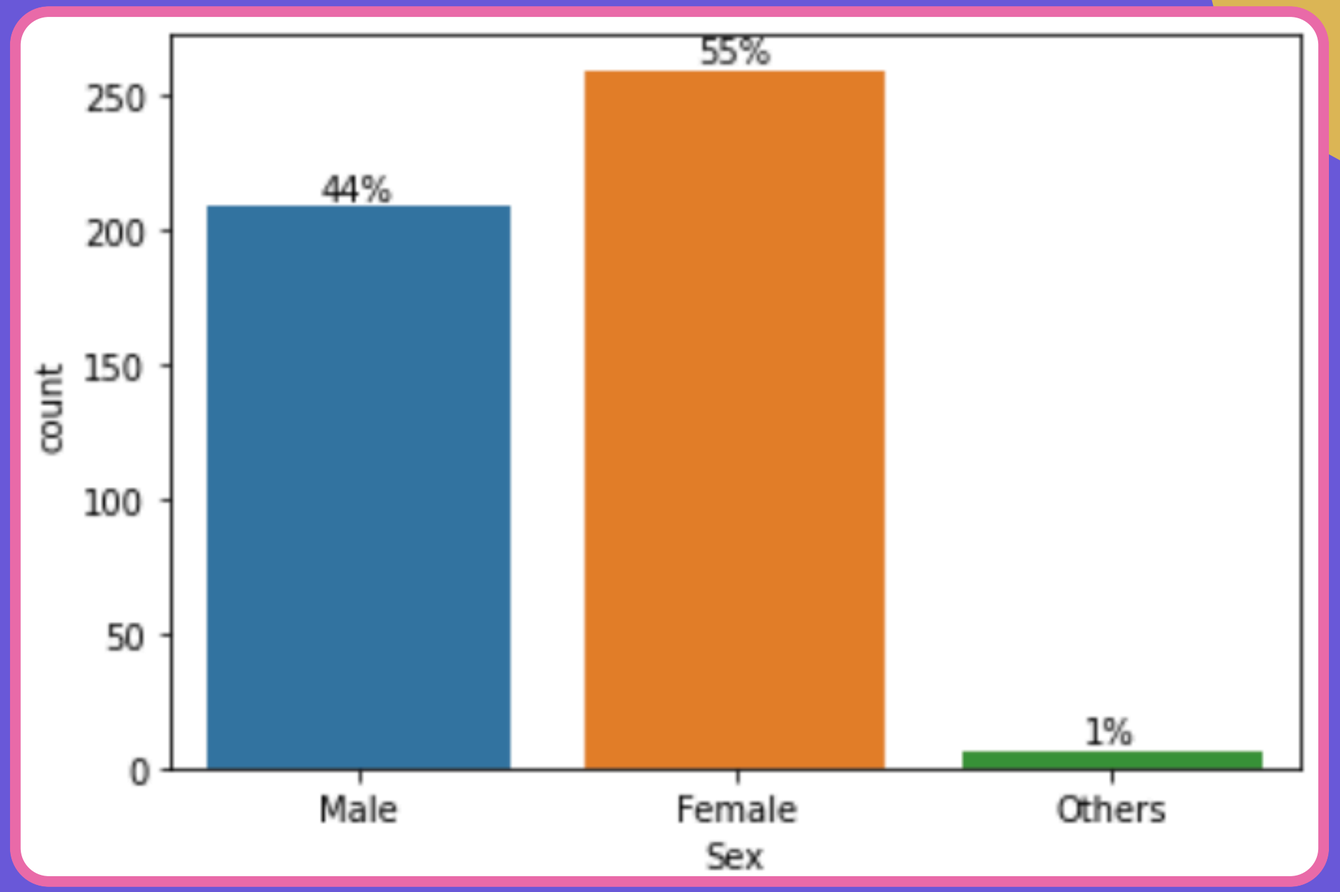
The distribution of data for 480 participants and their time spent on Social media



Gender

The distribution of data for 480 participants based on their Gender.

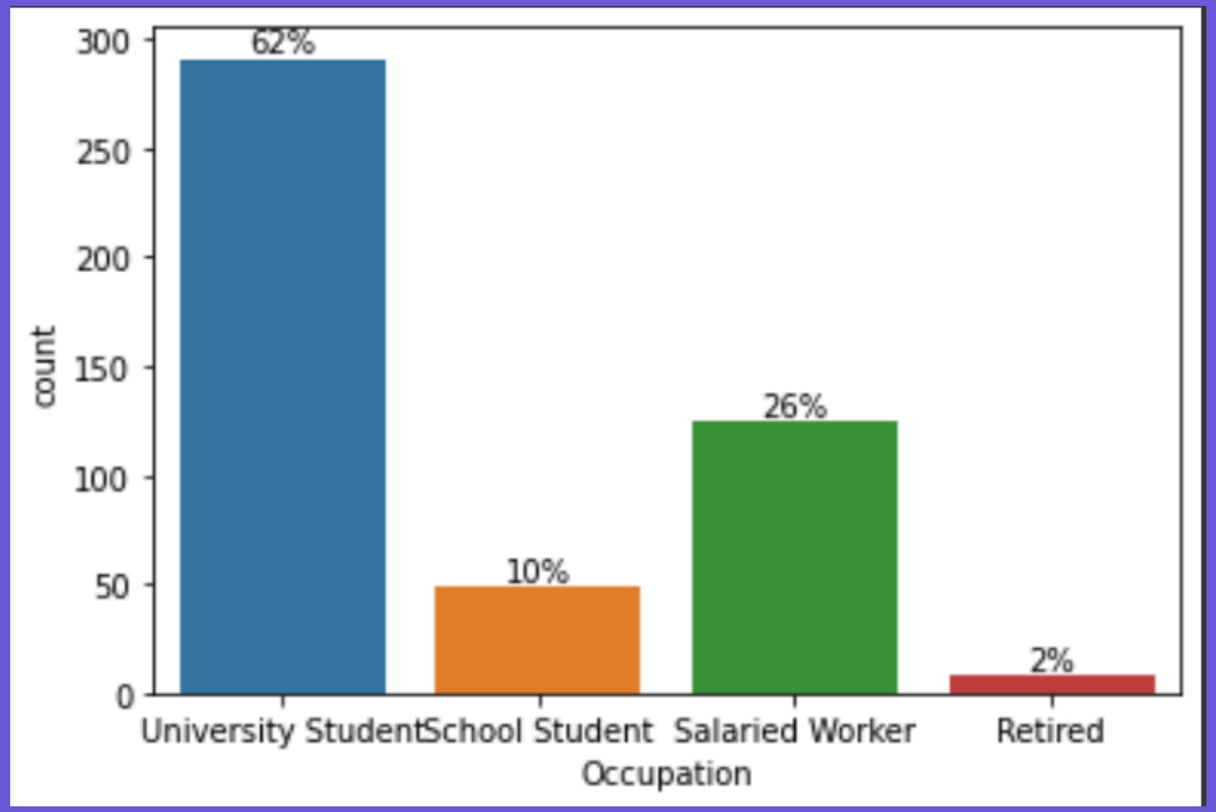
Approximately 260 participants out of 473 are female, making up the majority in the sample. 'Others' make up approximately 1% of the sample size, which makes it impossible to make statistical inferences based on the "Other" category specifically.



Occupation

The distribution of data for 480 participants based on their Occupation.

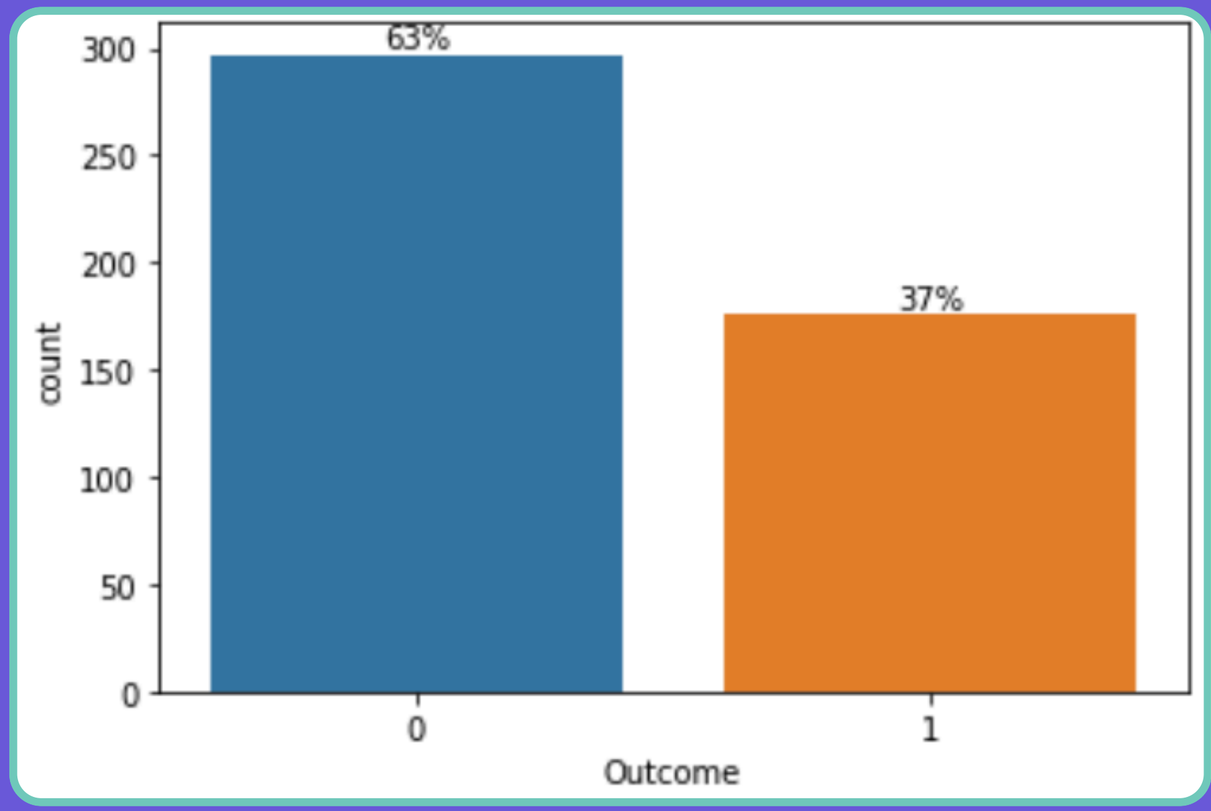
The sample is over-representated by University students, making up an overwhelming 62% of the sample.



Outcome

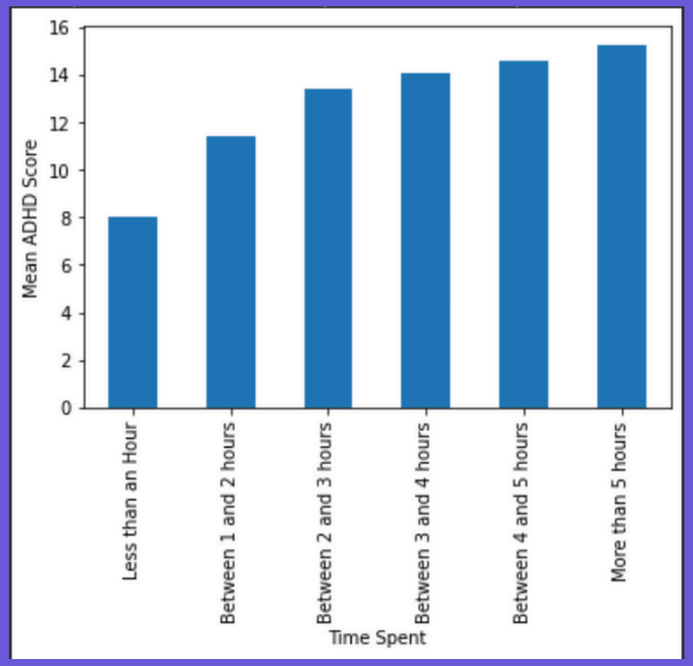
The the distribution of data for 480 participants based on "Outcome" - whether or not the individual is experiencing severe mental health issues and therefore whether we recommend the individual to go to get a mental health check up.

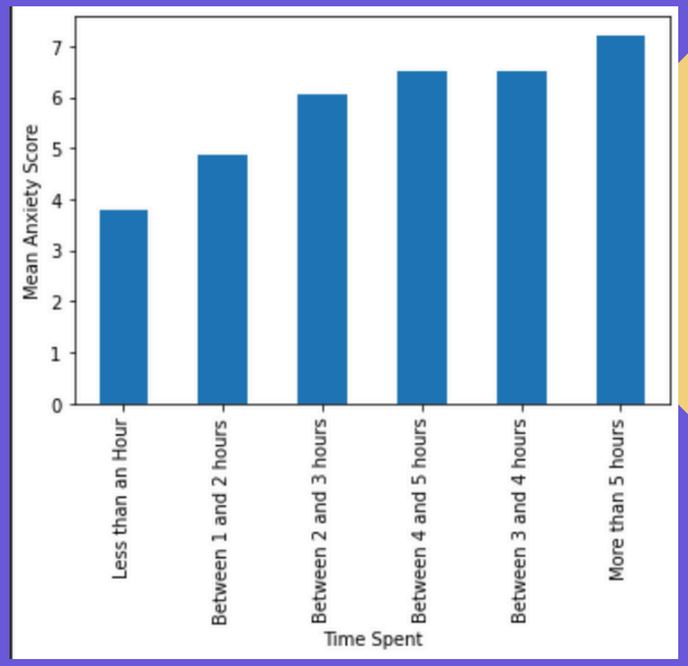
Approximately 37% of the sample of 473 participants meet criteria of scoring 40 points and above, are experiencing severe mental health symptoms and are recommended to go get their mental health evaluated by a professional.

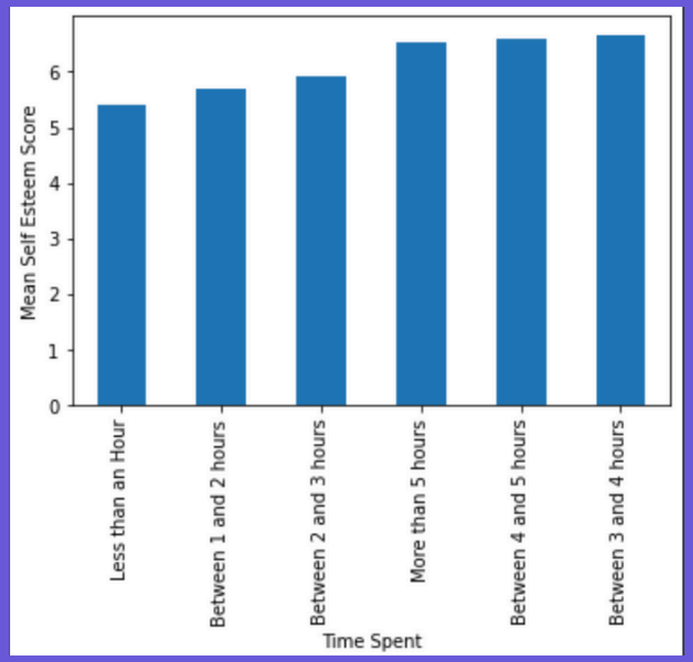


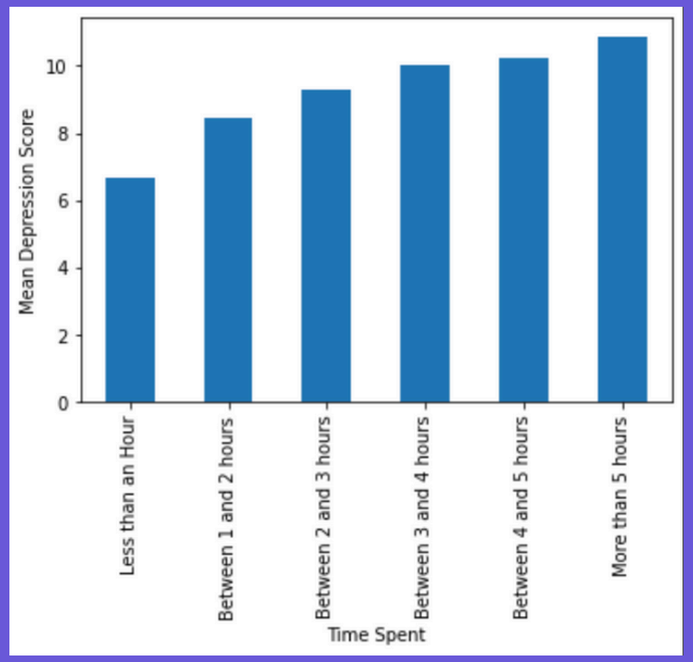
The mean:

- ADHD score of each Time group of participants
- Anxiety score of each Time group of participants
- Self Esteem score of each Time group of participants
- Depression score of each Time group of participants









Data Correlation

```
data.corr()
```

[265] ✓ 0.0s Python

	Age	Sex	Relationship Status	Occupation	Social Media User?	Platforms Used	Time Spent	ADHD Score	Anxiety Score	Self Esteem Score	Depression Score	Outcome
Age	1.000000	-0.134974	0.620397	0.691140	-0.046849	-0.287499	-0.361333	-0.301063	-0.253629	-0.072147	-0.304066	-0.189515
Sex	-0.134974	1.000000	-0.026187	-0.038985	0.087492	-0.050672	0.215704	0.102384	0.150707	0.127576	0.102340	0.116448
Relationship Status	0.620397	-0.026187	1.000000	0.532216	-0.063507	-0.247744	-0.295474	-0.235377	-0.204677	-0.091253	-0.285028	-0.140457
Occupation	0.691140	-0.038985	0.532216	1.000000	-0.057995	-0.239455	-0.277251	-0.203962	-0.191960	-0.055126	-0.220591	-0.111445
Social Media User?	-0.046849	0.087492	-0.063507	-0.057995	1.000000	0.077840	0.111780	0.145141	0.132418	0.084385	0.127166	0.060885
Platforms Used	-0.287499	-0.050672	-0.247744	-0.239455	0.077840	1.000000	0.249160	0.189266	0.196072	0.075348	0.188903	0.160030
Time Spent	-0.361333	0.215704	-0.295474	-0.277251	0.111780	0.249160	1.000000	0.453670	0.443020	0.130091	0.346333	0.345216
ADHD Score	-0.301063	0.102384	-0.235377	-0.203962	0.145141	0.189266	0.453670	1.000000	0.676207	0.280042	0.621464	0.645753
Anxiety Score	-0.253629	0.150707	-0.204677	-0.191960	0.132418	0.196072	0.443020	0.676207	1.000000	0.340021	0.580797	0.615793
Self Esteem Score	-0.072147	0.127576	-0.091253	-0.055126	0.084385	0.075348	0.130091	0.280042	0.340021	1.000000	0.332410	0.520263
Depression Score	-0.304066	0.102340	-0.285028	-0.220591	0.127166	0.188903	0.346333	0.621464	0.580797	0.332410	1.000000	0.659422
Outcome	-0.189515	0.116448	-0.140457	-0.111445	0.060885	0.160030	0.345216	0.645753	0.615793	0.520263	0.659422	1.000000

From the above correlation table, it can be inferred that the time spent on various social media platforms has a moderate positive correlation with ADHD, Anxiety and Total Scores, with r values of 0.45, 0.438 and 0.44, respectively.

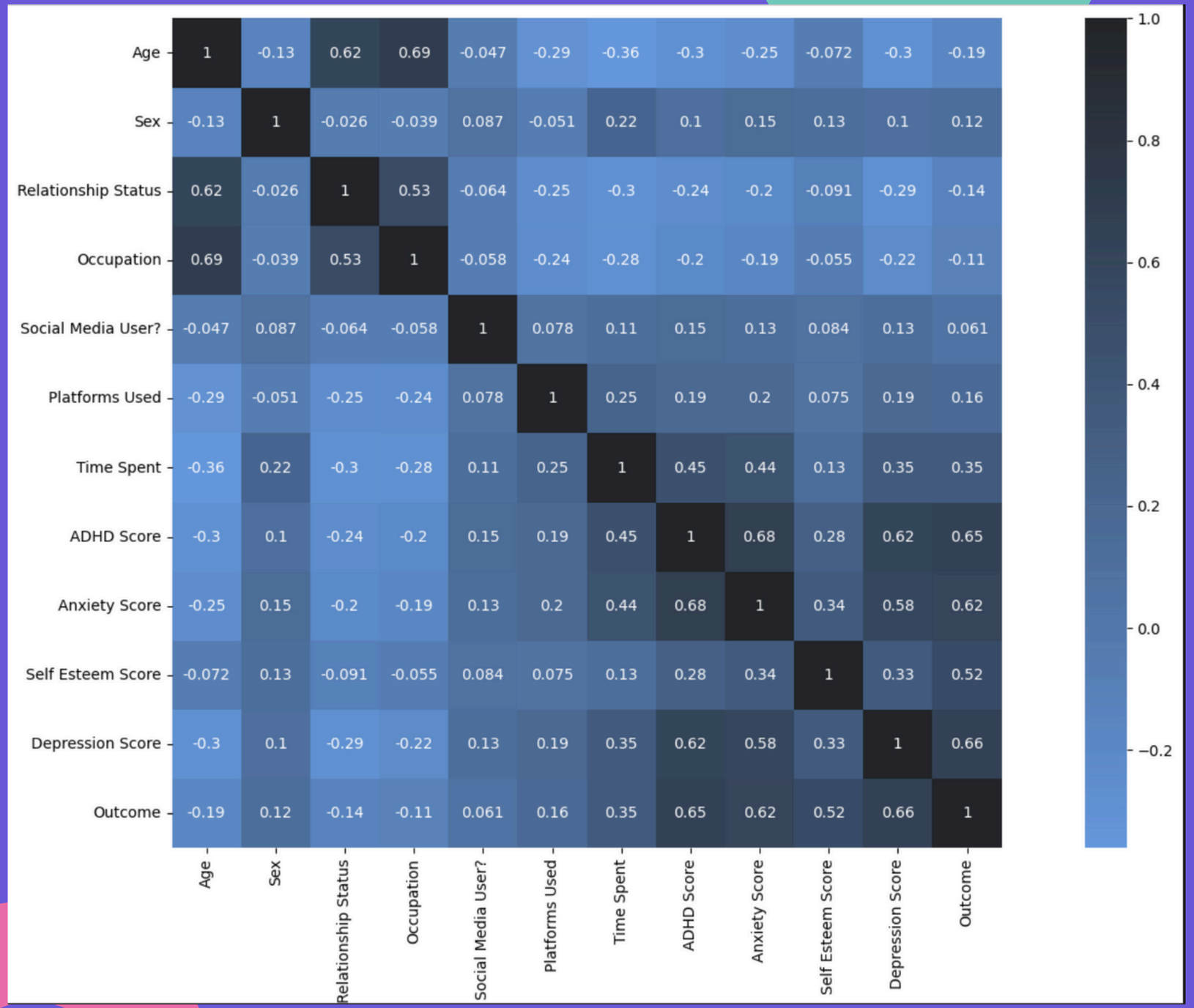
Correlation between Time Spent on social media and Self esteem scores and Depression scores are on the positive weaker side, with r values of 0.138 and 0.35 respectively.

There is a negative weak correlation between Age and all the other variables. The interpretation may be that the higher the participant's age is, the lower their social media usage and mental health scores will be. Note that this is a weak correlation, with r values between -0.35 and 0 for all variables.

From the this Correlation table and heatmap, it can be inffered that multi-collinearity between different supposedly independent variables exist in ranges from low to medium ($0.11 < r < 0.68$). The independent variables are Time spent, Sex, Age, scores of ADHD, anxiety, self esteem and depression.

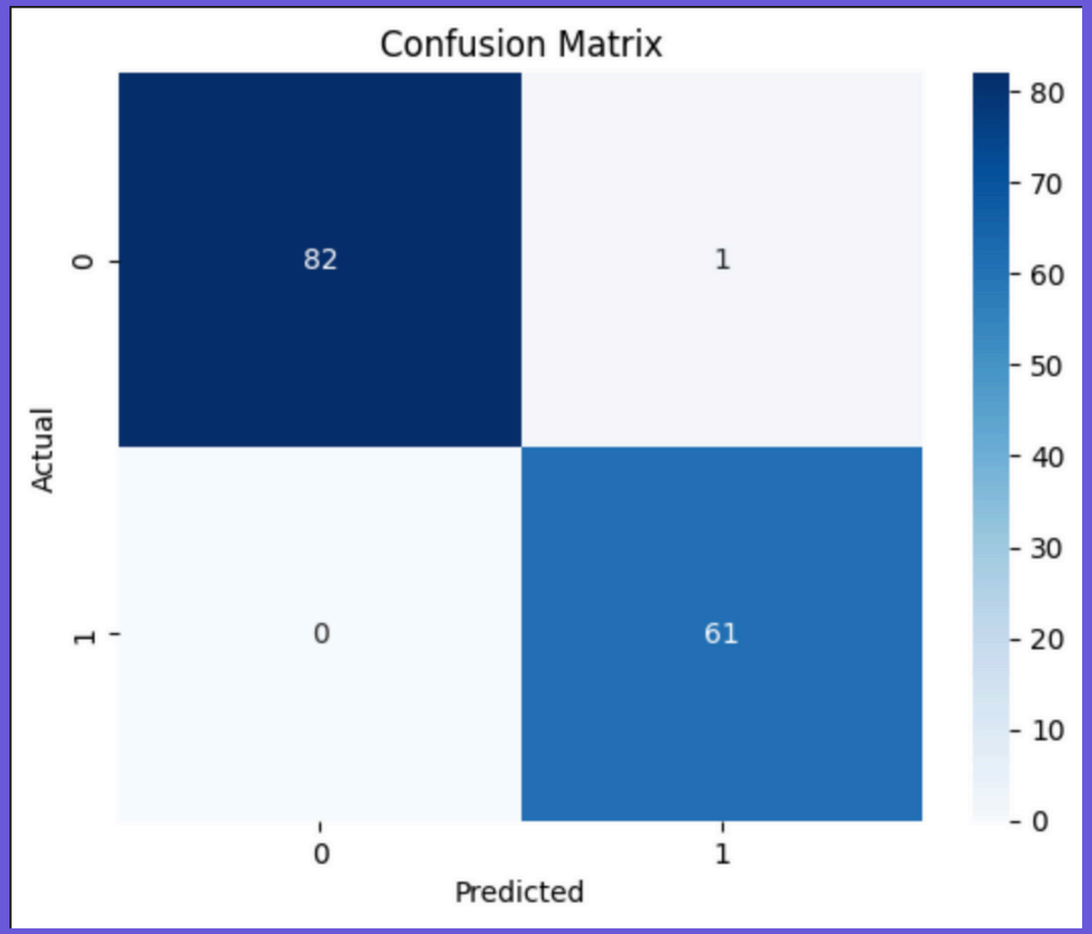
Age seems to have negative low colinearity values against the other variables.

Sex seems to have low colinearity values for all other variables.



Logistic Regression

1. True Negative (TN):
 • The model correctly predicted 82 instances as class 0 (negative class).
 2. False Positive (FP):
 • The model incorrectly predicted 1 instances as class 1 (positive class) when they were actually class 0. These are also known as Type I errors.
 3. False Negative (FN):
 • The model incorrectly predicted 0 instances as class 0 when they were actually class 1. These are also known as Type II errors.
 4. True Positive (TP):
 • The model correctly predicted 61 instances as class 1 (positive class).



Accuracy score

```
print("Accuracy of Logistic Regression: ",metrics.accuracy_score(y_test, predicted))
```

[49] ✓ 0.0s Python

```
Accuracy of Logistic Regression:  0.9930555555555556
```

```
# Model Accuracy
print('Accuracy score using the Logistic regression model: ', accuracy*100,'%')
```

[50] ✓ 0.0s Python

```
Accuracy score using the Logistic regression model:  99.30555555555556 %
```

Discussion

- This study provides important insights into the link between social media use and mental health in young adults:

- 1. ADHD and Anxiety: We found a moderate positive correlation between social media use and symptoms of ADHD and anxiety. More time on social media is associated with higher restlessness and distraction levels.
- 2. Self-Esteem and Depression: There is a weaker correlation between social media use and self-esteem and depression, suggesting these conditions are influenced by a broader range of factors.
- 3. Age Differences: Younger individuals are more affected by social media use, as shown by the negative correlation between age and mental health scores. Older individuals might use social media differently or have better coping mechanisms.
- 4. Intervention Potential: Our predictive model accurately identifies individuals at risk of severe mental health issues. This could guide targeted interventions like mental health awareness programs and tools to manage social media use.
- 5. Multi-Collinearity: The interrelation among various mental health symptoms and demographic factors highlights the complexity of predicting mental health outcomes.
- 6. Limitations and Future Research: This study is cross-sectional and cannot establish causality. Future research should explore longitudinal data and include diverse demographics for better generalizability.
-

Conclusion

- In conclusion, our study highlights a significant correlation between time spent on social media and the presence of mental health symptoms among young adults. Through the analysis of survey data from 467 participants, we observed that excessive social media usage is moderately associated with increased symptoms of ADHD, anxiety, and overall negative mental health outcomes.

- Key findings include:
- - A moderate positive correlation between social media use and symptoms of ADHD, anxiety, and overall mental health scores.
- - Weaker correlations with self-esteem and depression.
- - A negative weak correlation between age and social media use/mental health scores, suggesting younger individuals are more affected.

- Using logistic regression, our predictive model demonstrated high accuracy (99.3%) in identifying individuals at risk of severe mental health issues, recommending professional mental health evaluations for 37% of the sample who scored above the threshold.

- These findings emphasize the need for awareness and potential interventions to mitigate the adverse mental health effects of social media use, particularly among younger populations. Future research could further refine these predictive models and explore causal relationships to better inform public health strategies and individual practices regarding social media consumption.

Bibliography

- Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice
- Role of social media in tackling challenges in mental health
- Does time spent using social media impact mental health?: An eight year longitudinal study
- Dataset