University of New Brunswick

CS2704

Project Report

Social Bytes And Wellness

Data Analysis project to test relation between social media use and mental well-being

Sumitr Banik

3712956

December 10, 2023

Introduction

The profound impact of social media on individuals' lives has become undeniable in the ever-changing landscape of the digital age. As I navigate this complex relationship between technology and human experience, there is an urgent need to examine the potential consequences for mental health. This study delves into the intricate relationship between social media engagement patterns and a range of mental health indicators.

"My Hypothesis states that increased social media usage is associated with specific mental health outcomes, influencing factors such as stress levels, anxiety, and overall well-being"

The widespread nature of digital connectivity, the constant flow of information, and the complex dynamics of online interactions compel me to investigate the links between social media usage and mental health. My motivation for embarking on this exploration stems from a desire to unravel the complex threads that connect the digital and psychological realms, thereby contributing valuable insights to the broader discussion of technology's impact on human health.

Data Overview

The dataset under examination is a valuable collection of responses from 791 adults in Bangladesh, gathered by the University of Asia Pacific, Dhaka. It is sourced from a study conducted between February 4, 2021, and March 18, 2021, utilizing a self-made questionnaire administered via the Google survey tool. The primary objective of this dataset is to establish a paradigm for understanding the relationship between the use of social networking sites (SNS) and various dimensions of psychological distress, encompassing depression, anxiety, loneliness, and sleep disturbances.

Key Information:

- **Survey Structure**: The dataset is structured into five sections, covering (i) sociodemographic information, (ii) patterns of SNS usage, and (iii) the assessment of mental health problems.
- **Participants**: The study encompasses responses from 669 Facebook users and 122 non-Facebook users, aged between 15 to 40 years, acknowledging the prominence of Facebook as the most popular social media platform in Bangladesh.

- **Data Analysis**: The collected responses were analyzed using Microsoft Excel (version 2016), with results presented as frequencies and percentages based on the entire survey.
- **Mental Health Assessment**: The survey questionnaire incorporates sections aligned with internationally validated scales: the UCLA Loneliness Scale-8 (UCLA-8), Patient Health Questionnaire-9 (PHQ-9), 7-item Generalized Anxiety Disorder (GAD-7) Scale, and Pittsburgh Sleep Quality Index (PSQI). These scales provide a standardized framework for evaluating mental health aspects.

This is what the Google form looks like:

Please tick one box for each statement

Section 1: Demographic questions

1. *Express your consent to participate in the research and processing of anonymous data for scientific purposes. Also, please confirm that you are a current used of social networking sites and your age is within 40 years.
   - ☐ Agree
2. *Age in years
   - ☐ Below 18
   - ☐ 18-25
   - ☐ 26-40
3. *Sex
   - ☐ Male
   - ☐ Female
4. *Weight and height ratio/BMI score
   - ☐ Below 18.5 (underweight)
   - ☐ 18.5-25 (normal)
   - ☐ Above 25 (overweight)
5. *Marital status
   - ☐ Unmarried
   - ☐ Married
   - ☐ Others
6. *Education level

The database contains the following columns:

#	Column	Non-Null Count	Dtype
---	------	-------------	-----
0	Timestamp	791 non-null	datetime64[ns]
1	1. Do you have a social media account? (e.g., Facebook, Twitter, etc.)	791 non-null	object
2	2. Which social media account do you use usually?	791 non-null	object
3	3. Which device do you usually use to connect social media?	791 non-null	object
4	4. Which type of internet connection do you use?	791 non-null	object
5	5. How long have you been using a social media account?	791 non-null	object
6	6. How frequently do you post (upload status or add photos/videos) on social media?	791 non-null	object
7	7. How much time do you spend daily in social media?	791 non-null	object
8	8. When do you usually use social media?	791 non-null	object
9	9. How many friends do you have on social media?	791 non-null	object
10	10. How many friends do you know personally in social media?	791 non-null	object
11	11. How many groups you are tagged in social media?	791 non-null	object
12	12. What is your main purpose for using social media (e.g. Facebook)?	791 non-null	object
13	13. What contents do you mainly look for in your social media news feed?	791 non-null	object
14	14.Do you believe social media is a good thing?	791 non-null	object
15	15. When you see something in social media, do you instantly believe it?	791 non-null	object
16	16. Have you ever experienced peer pressure due to social media?	791 non-null	object
17	17. Does your emotion get influenced by other's posts (success, failure, loss)?	791 non-null	object
18	18. Have you ever compared yourself with other's success or luxurious life?	791 non-null	object
19	19. Do you think, your mental wellbeing would be better if you do not use social media?	791 non-null	object
20	20. If answer is yes, are you trying to control that thing and trying to reduce the use of social media?	791 non-null	object
21	21. Please write your age in years (number).	791 non-null	float64
22	22. Gender	791 non-null	object
23	23. Marital Status	791 non-null	object
24	24. Religion	791 non-null	object
25	25. Education	789 non-null	object
26	26. Profession	791 non-null	object
27	27. Monthly income	791 non-null	object
28	28. Area of residence	791 non-null	object
29	29. Living with-	791 non-null	object
30	30. Body weight (Kg)	791 non-null	float64
31	31. Height (m)	791 non-null	float64
32	32. Smoking habit	791 non-null	object
33	1. In the past 30 days, do you feel lack of companionship?	791 non-null	object
34	2. In the past 30 days, there is no one I can turn to	791 non-null	object
35	3. In the past 30 days, I feel left out.	791 non-null	object
36	4. In the last 30 days, I feel isolated from others.	791 non-null	object
37	5. In the last 30 days, I am unhappy being so withdrawn.	791 non-null	object
38	6. In the last 30 days, people are around me but not with me.	791 non-null	object
39	7. In the last 30 days, I am an outgoing person.	791 non-null	object
40	8. In the last 30 days, I can find companionship when I want it.	791 non-null	object
41	1 In the last 30 days, little interest or pleasure in doing things.	791 non-null	object
42	2. In the last 30 days, feeling down, depressed or hopeless.	791 non-null	object
43	3. In the last 30 days, trouble falling or staying asleep, sleeping too much	791 non-null	object
44	4. In the last 30 days, Feeling tired or having little energy.	791 non-null	object
45	5. In the last 30 days, poor appetite or over-eating.	791 non-null	object
46	6. In the last 30 days, feeling bad about yourself-or that you are a failure or have let yourself or your family down.	791 non-null	object
47	7. In the last 30 days, trouble concentrating on things, such as reading the newspaper or watching television.	791 non-null	object
48	8. In the last 30 days, moving or speaking so slowly or the opposite-moving around a lot more than usual.	791 non-null	object
49	9. In the last 30 days, thoughts that you would be better off dead, or of hurting yourself.	791 non-null	object
50	1. In the last 30 days, I am feeling nervous, anxious, or on edge	791 non-null	object
51	2. In the last 30 days, I an not being able to stop or control worrying	791 non-null	object
52	3. In the last 30 days, I am worrying too much about different things.	791 non-null	object
53	4. In the last 30 days, I felt trouble in relaxing.	791 non-null	object
54	5. In the last 30 days, I am being so restless that it's hard to sit still	791 non-null	object
55	6. In the last 30 days, I becoming easily annoyed or irritable.	791 non-null	object
56	7. In the last 30 days, I am feeling afraid as if something awful might happen.	791 non-null	object
57	1. When I am usually gone to bed?	791 non-null	object
58	2. How long (in minutes) has it taken you to fall asleep each night?	791 non-null	object
59	3. When have you usually gotten up in the morning?	791 non-null	object
60	4. How many hours of actual sleep do you get at night?	791 non-null	object
61	5. How many hours were you in bed?	791 non-null	object
62	6. In last 30 days, How many times, I cannot get to sleep within 30 minutes?	791 non-null	object
63	7. In last 30 days, How many times, I wake up in the middle of the night or early morning?	791 non-null	object
64	8. In last 30 days, How many times, I had to get up to use the bathroom?	791 non-null	object
65	9. In last 30 days, How many times, I cannot breathe comfortably?	791 non-null	object
66	10. In last 30 days, How many times, I cough or snore loudly?	791 non-null	object
67	11. In last 30 days, How many times, I feel too cold?	791 non-null	object
68	12. In last 30 days, How many times, I feel too hot?	791 non-null	object
69	13. In last 30 days, How many times, I saw bad dreams?	791 non-null	object
70	14. In last 30 days, How many times, I have pain during sleep?	791 non-null	object
71	15. In last 30 days, How many times, I having trouble sleeping for any other reason?	791 non-null	object
72	16. In last 30 days, In last month, have you take medicines for sleep?	791 non-null	object
73	17. In last month, how many times you cannot sleep due to any program or other important case?	791 non-null	object
74	18. In last month, how many times you face problems to maintain program or other important case?	791 non-null	object
75	19. During the past month, how would you rate your sleep quality overall?	791 non-null	object

Descriptive Analysis

The initial phase of our analysis involved extensive data cleaning and preprocessing to ensure the dataset's integrity and relevance. This encompassed several critical steps, including the simplification of column names, the conversion of all responses to numeric values, and the handling of null values.

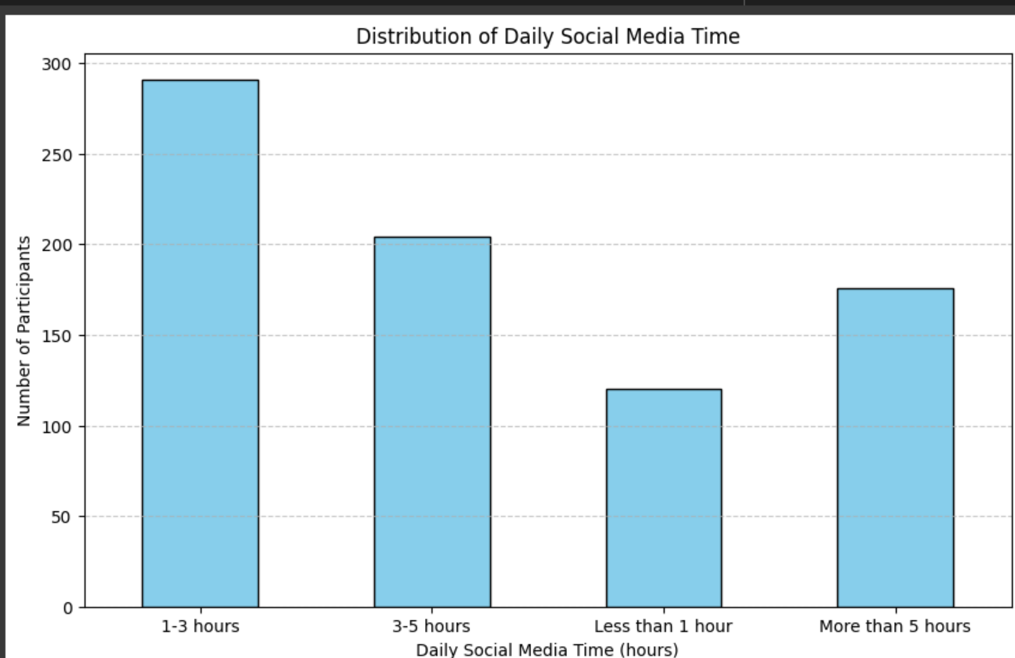
Data Cleaning and Processing:

- **Column Names**: To enhance clarity and simplicity, we streamlined the column names, promoting a more intuitive understanding of the dataset's structure.
- **Numeric Conversion**: Respondent answers were converted into numeric values, a pivotal step for facilitating statistical analyses and modelling.
- **Handling Null Values**: Rigorous efforts were invested in identifying and addressing null values, fortifying the dataset's completeness.

Statistical Exploration:

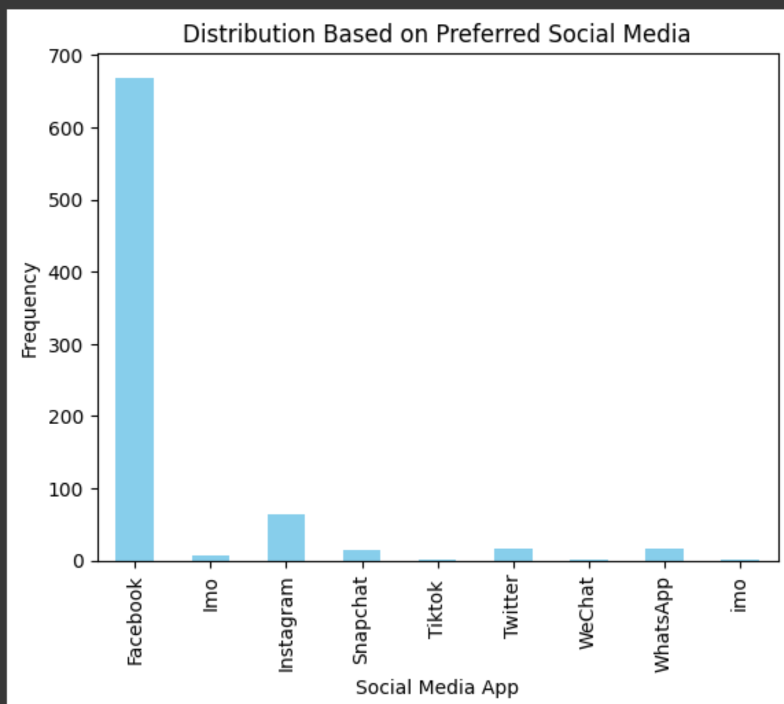
- **Time Spent on Social Media**: A bar plot was employed to visualize the distribution of participants across different timeframes of social media usage. This provided insights into the predominant patterns of engagement.

```
[ ] # For Starters, let's understand the distribution of data for 791 participants and their time spent on Social media
    data['Daily Social Media Time'].value_counts().sort_index().plot.bar(rot=0, color='skyblue', edgecolor='black', figsize=(10, 6))
    plt.title('Distribution of Daily Social Media Time')
    plt.xlabel('Daily Social Media Time (hours)')
    plt.ylabel('Number of Participants')
    plt.grid(axis='y', linestyle='--', alpha=0.7)
    plt.show()
```



- **Preferred Social Media Platforms**: A subsequent bar plot highlighted the distribution of individuals based on their preferred social media platforms. It can be seen Facebook is the most popular SNS. One of the reasons that it is preferred much more as it is always pre-installed in most of the new mobile devices bought in Bangladesh

```
[ ] # Let's understand the distribution of data for participants based on the social media app they use.
    data.groupby('Preferred Social Media').size().plot(kind='bar', color='skyblue')
    plt.title('Distribution Based on Preferred Social Media')
    plt.xlabel('Social Media App')
    plt.ylabel('Frequency')
    plt.show()
```



Outcome Model

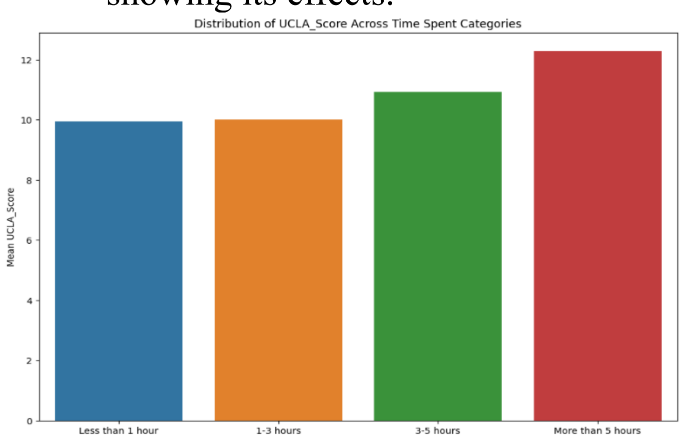
- *Outcome Determination*: An outcome model was introduced, wherein individuals with a total score exceeding 80% of the maximum score (100/128) were labelled as having an outcome of 1. This decision was influenced by the desire to identify participants experiencing severe mental health symptoms.
- *Outcome Classification*: Individuals with an outcome of 0 were deemed not to exhibit severe mental health symptoms, thus not requiring an immediate mental health checkup. Conversely, those with an outcome of 1 were recommended to seek a mental health checkup, indicating a high likelihood of experiencing significant mental health challenges.

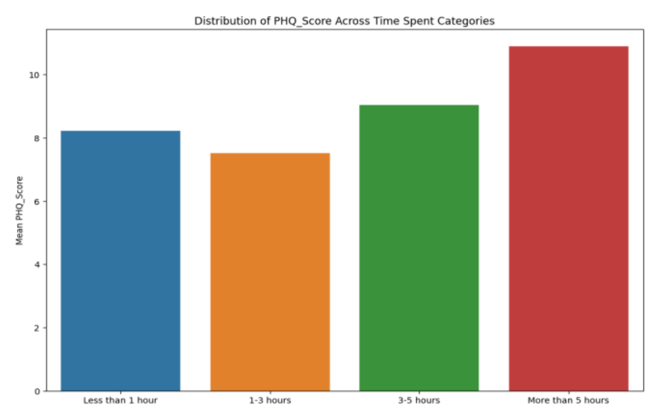
- *Outcome Analysis*: Among the 791 participants, a meticulous examination of the outcome variable revealed a noteworthy distribution. Remarkably, only 4 individuals, a mere fraction of the total respondents, exhibited an outcome value of 1.
- *Imbalance Awareness*: The limited occurrences of an outcome of 1 underscore the relative rarity of severe mental health symptoms within our dataset. This imbalance necessitates a nuanced approach in our predictive modelling and subsequent interpretations.

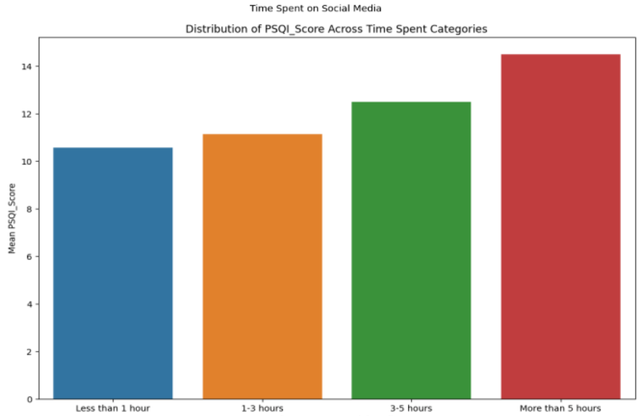
	Timestamp	Social Media Account	Preferred Social Media	Device Used for Social Media	Internet Connection Type	Years on Social Media	Post Frequency	Daily Social Media Time	Preferred Social Media Time	Number of Friends on Social Media	...	Medication for Sleep (Last 30 Days)	Sleep Disturbances Due to Events (Last 30 Days)	Problems Maintaining Routine (Last 30 Days)	Overall Sleep Quality (Last 30 Days)	UCLA_Score	PHQ_Score	GAD_Score	PSQI_Score	Total_Score	Outcome
391	2021-02-21 13:37:15.107	Yes	Facebook	Mobile Phone	Broadband (Wi-Fi)	5-10 years	3-5 per day	3.0	Frequently at anytime	2000-4000	...	3	3	1	1	20	24	21	38	103	1
441	2021-02-22 16:38:40.331	Yes	Facebook	Mobile Phone	Broadband (Wi-Fi)	2-5 years	1-2 per day	2.0	Frequently at anytime	500-2000	...	0	3	3	3	22	25	21	38	106	1
605	2021-02-22 17:07:26.423	Yes	Instagram	Mobile Phone	Broadband (Wi-Fi)	More than 10 years	Less than 1 per day	3.0	Frequently at anytime	500-2000	...	3	3	3	3	24	27	21	42	114	1
690	2021-03-06 20:05:02.777	Yes	Facebook	Mobile Phone	Broadband (Wi-Fi)	2-5 years	1-2 per day	2.0	Evening	Less than 500	...	3	2	2	3	18	24	21	37	100	1

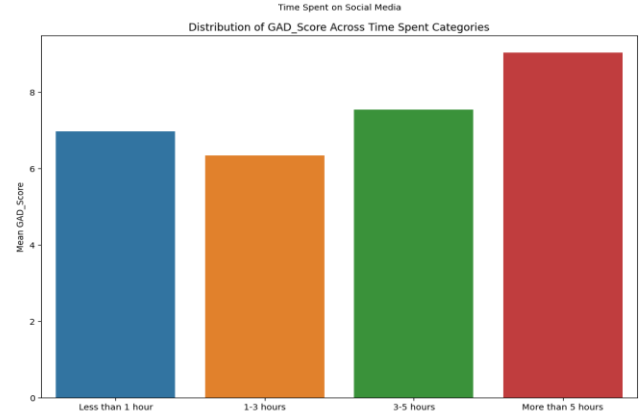
Predictive Analytics

In-depth Exploration: To discern potential patterns and relationships, we initiated our predictive analytics journey with an insightful examination of the association between each mental health score and the daily time spent on social media. It is visible that people with usage of more than 5 hours daily scored higher in each test, showing its effects.





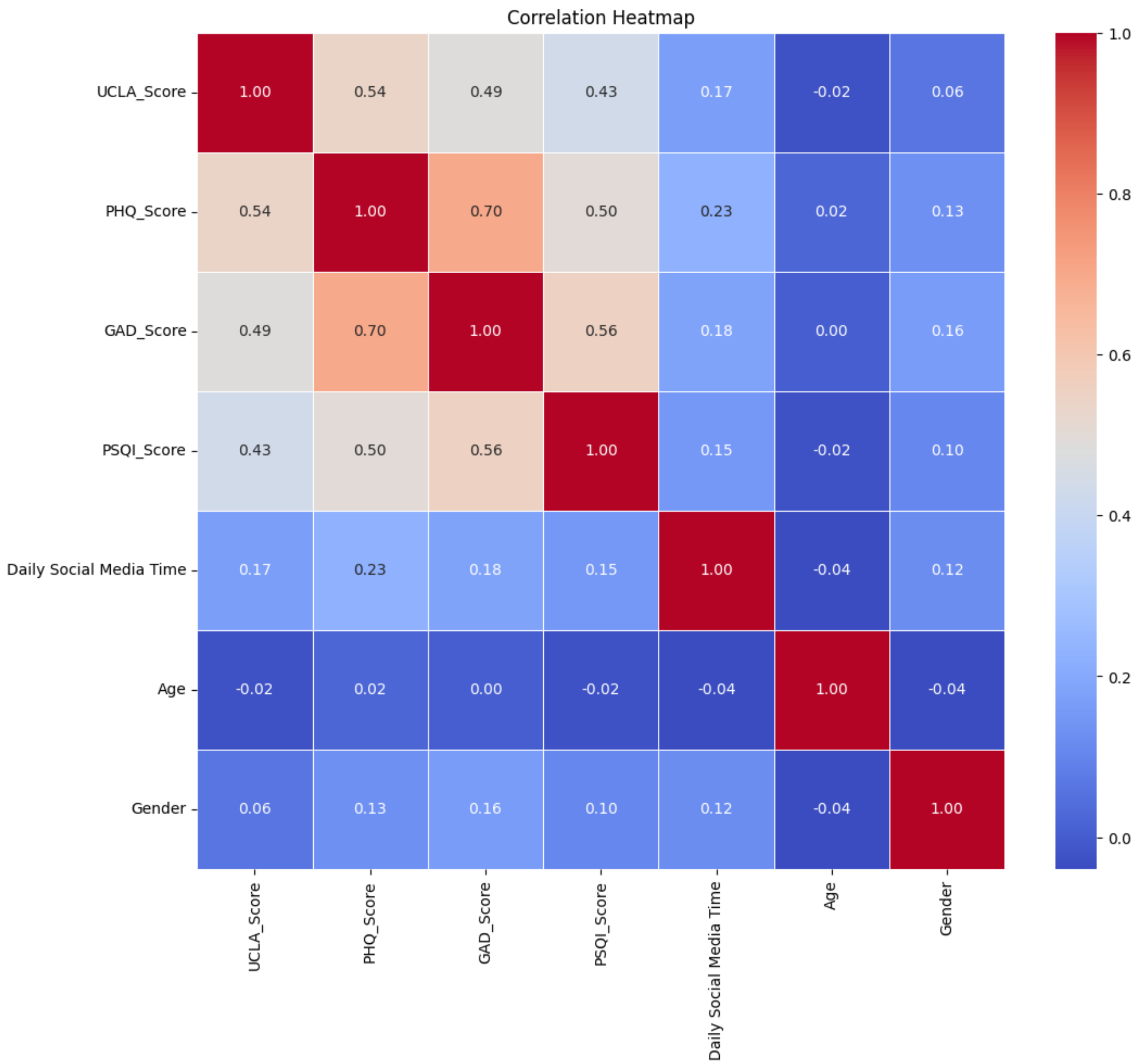




Correlation Map Interpretation

Before delving into the correlation map, categorical columns like 'Time Spent' and 'Gender' underwent necessary transformations to enable numerical representations.

The correlation map, a heatmap representation of correlation coefficients, unveiled subtle relationships between mental health scores and predictor variables. Positive values around 0.17-0.2 suggest a modest positive correlation, indicating that as daily social media time increases, there might be a slight increase in certain mental health scores.



In the context of our analysis, it's imperative to recognize the unique backdrop of the COVID-19 pandemic during which the data was collected. The pandemic, characterized by widespread uncertainties, lockdowns, and social disruptions, has been associated with profound impacts on mental health. Participants' responses may reflect the challenges posed by the pandemic, introducing a potential confounding factor in our exploration of the relationship between social media usage and mental health scores. The observed correlations should be interpreted with an awareness of the broader socio-environmental context, considering the unprecedented circumstances that participants navigated during the data collection period.

As we delve into the results, it's advisable to approach the findings with a nuanced understanding of the external influences brought about by the pandemic. Future investigations may benefit from comparative analyses with pre-pandemic data or the inclusion of variables capturing pandemic-specific stressors, offering a more comprehensive examination of the intricate interplay between social media engagement and mental health in the context of a global health crisis.

Predictive Model

In this phase of our analysis, we employed a predictive model to discern the potential relationship between daily social media engagement and mental health outcomes. Utilizing logistic regression, a widely used method for binary classification, we endeavoured to predict the likelihood of individuals experiencing mental health issues based on their reported daily social media time.

The dataset was divided into training and testing sets, with 80% of the data allocated for training the model and the remaining 20% for assessing its predictive performance. The logistic regression model, a powerful tool in binary classification tasks, was employed to understand the association between daily social media time and the binary outcome variable, which indicates whether an individual is likely to be experiencing mental health symptoms.

After training the model, we made predictions on the test set to evaluate its accuracy. The evaluation metrics, including accuracy, a confusion matrix, and a classification report, provided a comprehensive understanding of the model's performance. These metrics will guide our interpretation of how well the model predicts mental health outcomes based on daily social media engagement, shedding light on the potential impact of social media use on individuals' mental well-being.

```
features = data[['Daily Social Media Time']]
outcome = data['Outcome']

# Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(features, outcome, test_size=0.2, random_state=42)

# Initialize the logistic regression model
model = LogisticRegression()

# Train the model
model.fit(X_train, y_train)

# Make predictions on the test set
predictions = model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, predictions)
conf_matrix = confusion_matrix(y_test, predictions)
classification_rep = classification_report(y_test, predictions)
```

OUTPUT

```
print(f"Accuracy: {accuracy}")

Accuracy: 0.9850746268656716

print(f"Confusion Matrix:\n{conf_matrix}")

Confusion Matrix:
[[132   0]
 [  2   0]]

print(f"Classification Report:\n{classification_rep}")

Classification Report:
              precision    recall  f1-score   support

           0       0.99      1.00      0.99       132
           1       0.00      0.00      0.00         2

    accuracy                           0.99       134
   macro avg       0.49      0.50      0.50       134
weighted avg       0.97      0.99      0.98       134
```

Model Evaluation

The output of the logistic regression model reveals promising results with an overall accuracy of approximately 98.5%. This indicates that the model correctly classified individuals' mental health outcomes in nearly 99% of cases. The confusion matrix provides a breakdown of the model's predictions. In this case, it correctly identified 132 instances where individuals were not experiencing severe mental health symptoms (Outcome 0), and there were no false positives. However, the model struggled to identify cases where individuals were experiencing severe mental health symptoms (Outcome 1), resulting in 2 false negatives.

The precision, recall and F1-score values provide more nuanced insights into the model's performance. Precision, representing the accuracy of positive predictions, was very high for Outcome 0 (non-severe mental health symptoms), but it was extremely low for Outcome 1 (severe mental health symptoms). Recall, or sensitivity, reflects the model's ability to capture all relevant cases. Here, it excelled in identifying Outcome 0 but struggled with Outcome 1. The F1-score, which balances precision and recall, reflects the model's overall effectiveness, and in our case, it was higher for Outcome 0.

In conclusion, while the model demonstrated impressive accuracy in predicting non-severe mental health symptoms based on daily social media time, its performance in identifying severe mental health symptoms was limited. This highlights the challenges and complexities associated with predicting mental health outcomes solely based on social media engagement, emphasizing the need for further refinement and consideration of additional factors.

The highly imbalanced nature of the dataset, with only 4 individuals out of 791 having a positive outcome (severe mental health symptoms), could significantly impact the performance of the model. In imbalanced datasets, where one class is underrepresented, the model tends to be biased towards the majority class. In this case, since the majority of individuals do not exhibit severe mental health symptoms, the model might become overly inclined to predict the negative outcome (Outcome 0) and struggle with identifying the positive outcome (Outcome 1).

The imbalanced distribution might explain the high accuracy of the model, as it can achieve high accuracy by simply predicting the majority class for most instances. However, this does not necessarily mean that the model is effective in identifying individuals with severe mental health symptoms.

Conclusion

In this project, we investigated the relationship between social media usage and mental health outcomes using a dataset from a study conducted by the University of Asia Pacific in Bangladesh. The analysis included data cleaning, descriptive analytics, and predictive modelling. Descriptive analytics revealed insights into social media usage patterns, while a logistic regression model aimed to predict mental health outcomes based on daily social media time. The model exhibited high accuracy but struggled with the imbalanced dataset, emphasizing the need for caution in interpreting results. Further research and refinement of modelling techniques are recommended for a comprehensive understanding of this complex relationship.

References

Islam, Md. R., Tushar, Md. I., Jannath, S., Moona, A. A., Akter, S., & Islam, S. M. A. (2021). Data set concerning the use of social networking sites and mental health problems among the young generation in Bangladesh. *Data in Brief*, *39*, 107593. https://doi.org/10.1016/j.dib.2021.107593

Banik, S. (2023). Social Media and Mental Health Data Analysis. GitHub. https://github.com/sumitrB/Social-Media_Mental-Health-Data-Analysis