**Exploring Emotional Response to Social Media**

Name: Tina Ma

Course: DESN 210

Date: 4/25/25

## Introduction

With the vast development of social media platforms, more and more people are engaging with digital platforms for communications, entertainment, and expression. While digital platforms bring many conveniences and great impacts to our lives, some people suffer emotionally from consuming too much information online. As people spend more time online, it is important to understand what factors influence emotional responses when using social media. This study explores this research question: **Which factors are strongly associated with emotional responses to social media?**

This question is important because it relates to the public health and digital well-being of using social media. It is also important for social media to understand how some factors have a good or bad influence on people so that they can improve their platforms and create better services. If certain behaviors or any characteristics bring negative emotional responses like depression or sadness, it is significant to reduce those outcomes through different ways, like education.

Studies have found that using social media can have a negative impact on the psychological health of users. Most of them are suffering from anxiety and depression (Saiphoo and Vahedi 2020). This study uses a real-world dataset to explore how behavior and simple traits might align with different emotional states, aiming to contribute some insights into the interaction between digital platforms and emotional health.

## Data

In this project, I used the dataset smmh.csv. I found this data on Kaggle, and it includes users' demographic information and interaction data with different social media platforms. This dataset is survey-based and contains 103 entries recorded from unique users. For each user, it captures valuable information on social media engagement and self-reported dominant emotional state when using those platforms (Bulut n.d.).

**Variables:** The dataset includes the following variables for each user (Bulut n.d.):

- **User_ID:** Unique ID for the user.
- **Age:** Age of the user (numeric).
- **Gender:** Gender of the user (Female, Male, Non-binary).

- **Platform:** Social media platform used (Instagram, Twitter, Facebook, LinkedIn, Snapchat, Whatsapp, Telegram).
- **Daily_Usage_Time (minutes):** Daily time spent on the platform in minutes (numeric).
- **Posts_Per_Day:** Number of posts made per day (numeric).
- **Likes_Received_Per_Day:** Number of likes received per day (numeric).
- **Comments_Received_Per_Day:** Number of comments received per day (numeric).
- **Messages_Sent_Per_Day:** Number of messages sent per day (numeric).
- **Dominant_Emotion:** User's dominant emotional state during the day (Happiness, Sadness, Anger, Anxiety, Boredom, Neutral).

In this study, I defined the dependent variable as the user's dominant emotion, and the independent variables as demographic factors such as gender, and behavioral data such as daily time used and interactions.

## Methodology

**Renaming columns:** To better engage with the dataset, I renamed some columns at first. For example, I changed "Daily_Usage_Time (minutes)" to "Daily_time" and "Likes_Received_Per_Day" to "likes," etc. This allows me to analyze the dataset more easily and simply when writing the code.

**Data Cleaning:** Next, I applied the function to check the missing value. The results showed me that there are no missing values in the dataset. Therefore, I checked if there were any duplicates in the data to improve the accuracy of the numerical calculations. After removing the duplicates, the dataset remains 99 rows of the data, so there are four rows of duplicated data. The cleaned dataset ensures unique user responses, so it can give better results.

**Feature Engineering:** After that, I added two new columns based on all the interactions: one is the total interactions, which is the sum of all interactions such as likes, comments, and messages, and the other is the average interactions, which divide total interactions by the number of posts each user posts per day. In this way, I can normalize each user's engagement and better compare their interactions.

| | User_ID | Age | Gender | Platform | Daily_time | Posts_Per_Day | likes | comments | messages | Dominant_Emotion | total_interactions | average_interactions |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 99.000000 | 99.000000 | 99 | 99 | 99.000000 | 99.00000 | 99.000000 | 99.000000 | 99.00000 | 99 | 99.000000 | 99.000000 |
| **unique** | NaN | NaN | 3 | 7 | NaN | NaN | NaN | NaN | NaN | 6 | NaN | NaN |
| **top** | NaN | NaN | Male | Facebook | NaN | NaN | NaN | NaN | NaN | Neutral | NaN | NaN |
| **freq** | NaN | NaN | 46 | 22 | NaN | NaN | NaN | NaN | NaN | 26 | NaN | NaN |
| **mean** | 519.030303 | 27.373737 | NaN | NaN | 87.929293 | 2.89899 | 33.868687 | 13.878788 | 20.40404 | NaN | 68.151515 | 25.220382 |
| **std** | 286.771560 | 3.652189 | NaN | NaN | 34.389798 | 1.75252 | 23.313588 | 8.189432 | 7.46819 | NaN | 38.028551 | 5.968245 |
| **min** | 16.000000 | 21.000000 | NaN | NaN | 40.000000 | 1.00000 | 5.000000 | 2.000000 | 10.00000 | NaN | 17.000000 | 12.000000 |
| **25%** | 275.000000 | 24.500000 | NaN | NaN | 60.000000 | 1.00000 | 15.000000 | 7.000000 | 13.50000 | NaN | 36.000000 | 21.500000 |
| **50%** | 526.000000 | 27.000000 | NaN | NaN | 75.000000 | 3.00000 | 28.000000 | 12.000000 | 20.00000 | NaN | 58.000000 | 25.000000 |
| **75%** | 774.500000 | 30.000000 | NaN | NaN | 105.000000 | 4.00000 | 42.500000 | 20.000000 | 25.00000 | NaN | 85.500000 | 29.000000 |
| **max** | 997.000000 | 35.000000 | NaN | NaN | 200.000000 | 8.00000 | 110.000000 | 40.000000 | 45.00000 | NaN | 195.000000 | 47.000000 |

*Figure 1. Description of the Cleaned Data Frame*

**EDA:** Furthermore, to better understand the data, I performed multiple exploratory analyses by grouping the variables. I grouped the data by gender and calculated the mean of their interactions (likes, comments, messages) they got. Then, I grouped them again by gender and calculated the sum of their interactions. For emotions, I grouped the data by dominant emotion and calculated the mean of their interactions. Then, I grouped the data by different platforms and calculated the mean of each interaction.

**Graphs:** To better visualize the data, I plotted several graphs, including bar, box, and scatter plots. I used a bar blot to see the frequency of posting between each emotion. I also used a box plot to see the distribution of daily usage time on social media by gender. Then, I plotted two scatter plots to see the relationship between age, daily usage time, and total reactions.

**ANOVA Test:** I want to test the relationship between the social media usage time and the user's emotions, so I performed an ANOVA one-way test for the mean usage time between different emotions. My null hypothesis for the test is that the mean daily usage time is the same across all dominant emotion groups. My alternative hypothesis is that at least one emotion group has a different mean daily usage time. If the p-value is smaller than 0.05, then we reject the null hypothesis; if not, then we do not reject the null hypothesis.

## Results

**EDA:** From the exploratory data analysis, the results show that, on average, the female has higher average interactions overall. For the summation of data, the male gets higher interactions through three sections. The results also show that (Figure 1) happiness has higher mean values over all interactions, and boredom has the lowest mean values. It suggests that interaction on social media might bring happiness. Lastly, some results show that most people use Instagram to engage online. LinkedIn has lower mean values since it is a more professional platform for people to engage with jobs and professional information.

| | mean_likes | mean_comments | mean_messages |
|---|---|---|---|
| Dominant_Emotion | | | |
| Anger | 35.666667 | 17.222222 | 21.666667 |
| Anxiety | 35.285714 | 14.619048 | 22.571429 |
| Boredom | 13.933333 | 6.600000 | 12.466667 |
| Happiness | 72.714286 | 25.785714 | 30.000000 |
| Neutral | 27.000000 | 10.923077 | 18.076923 |
| Sadness | 25.857143 | 12.000000 | 19.571429 |

*Figure 2. Table of Mean Interactions Between Each Emotion*

**Data visualization:** To find the relationship between gender and daily usage time of social media, a box plot was created (Figure 3). This box plot visualizes the distribution of daily social media usage time across different genders. For non-binary users, it has a narrow range, and the median usage time is around 60 minutes. For female users, it has the widest range, and the median usage time for females is also the highest. For male users, it has a narrow range, and the median usage time is lower than for females, but there are a few outliers with higher usage time.

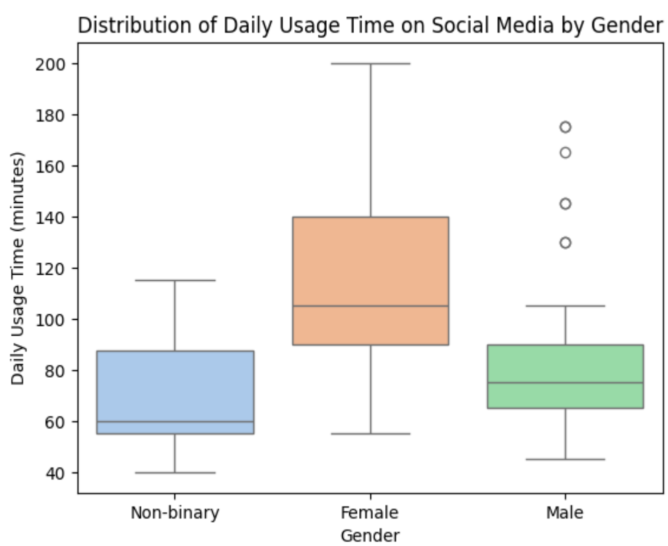


*Figure 3. Distribution of Daily Usage Time on Social Media by Gender*

Next, to examine the frequency of posting between different emotions, I visualized it using a bar graph (Figure 4). This bar chart illustrates the social media posts made per day by users based on the dominant emotion. It shows that users with happiness tend to post the most frequently, around 5-6 posts. Users with anger emotion also post relatively high. Users with boredom emotion tend to post less.

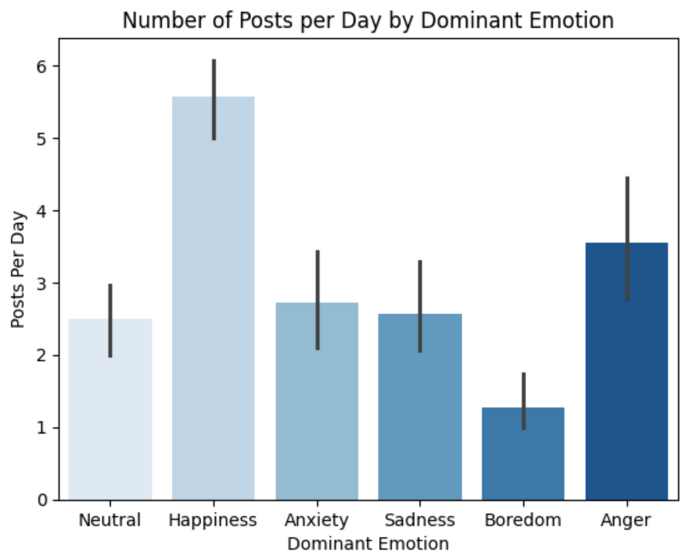


*Figure 4. Number of Posts per Day by Dominant Emotion*

In order to check the relationship between age and daily usage time, I used a scatter plot to see if there is a pattern in the data (Figure 5). This scatter plot shows a wide range of social media engagement. It does not imply any linear relationship, but it shows that around age 26-30, users have a wide variability, and the highest usage time can be 200 minutes. Users below 26 or older than 30 tend to have less usage time.

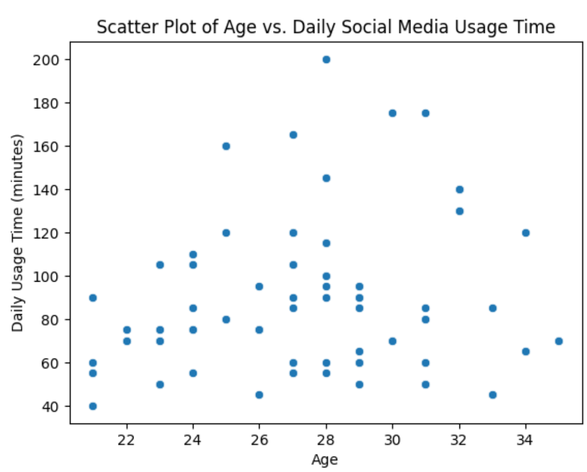


*Figure 5. Scatter Plot of Age vs. Daily Social Media Usage Time*

I also plotted another scatter plot to illustrate the relationship between daily usage time and total reactions received on social media (Figure 6). It shows a positive linear relationship that with more usage time, users tend to get more interactions online.

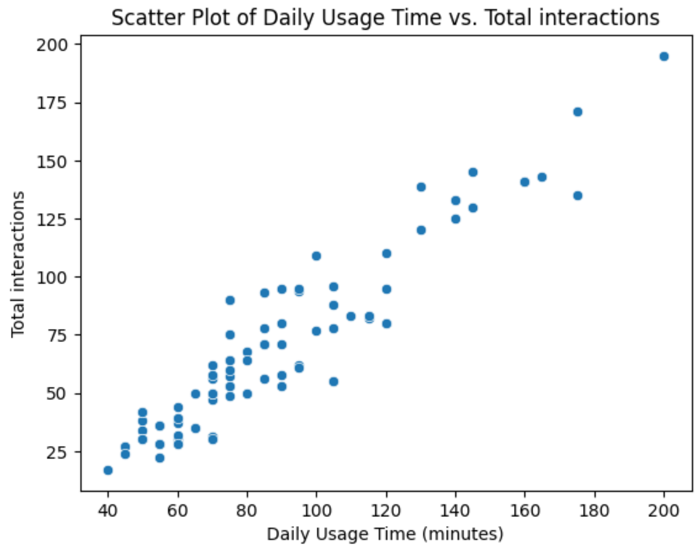


*Figure 6. Scatter Plot of Daily Usage Time vs. Total Interactions*

Based on the four graphs, we can find the social media usage patterns with distinct characteristics. From all the platforms, users experiencing happiness post most frequently, while those feeling boredom post the least. Female users tend to spend more time on social media and receive the most interactions overall. Individuals with anxiety or happiness tend to have a high variability in usage time. While age shows no clear relationship with usage time, individuals around 25 to early 30s tend to engage more on social media. Lastly, a strong positive linear relationship between daily usage time and total interactions illustrates that increased screen time often results in greater engagement. Therefore, these insights emphasize that social media behavior is shaped by a combination of factors, including emotions, age, gender, and platform.

**Statistical Analysis:** Based on the ANOVA test result, we see that the F-statistic is around 17.99 and the p-value is around 1.93*10^ (-12), which is smaller than 0.05. Therefore, we need to reject the null hypothesis, and there is a statistically significant difference in average daily usage time between at least two emotion groups. I made a box plot to see the distribution of daily usage time by emotion (Figure 7). The plot shows the result from the test that people with different emotions spend different amounts of time on social media. For example, those experiencing happiness tend to have higher usage times, while people feeling neutral, anxiety, sadness, and anger spend less time. This suggests that daily usage time may be related to users' emotional responses to social media.

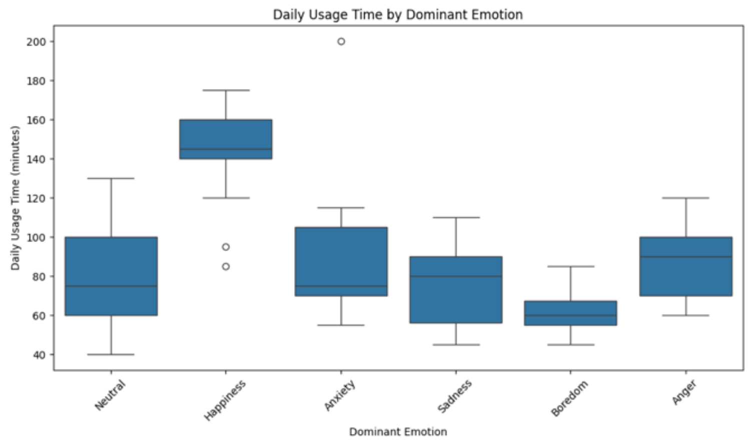


*Figure 7. Distribution of Daily Usage Time by Dominant Emotion*

## Conclusion

This study explored the relationship between social media usage patterns and emotional responses. By analyzing data collected from surveys, I found some factors associated with users' dominant emotions while using digital platforms. The results showed that people experiencing emotions such as happiness or anxiety tend to spend more time on social media. The one-way ANOVA test confirmed a statistically significant difference in average daily usage time across different emotion groups, suggesting that time spent online may influence users' emotional experiences. Exploratory analysis and visualizations further supported this conclusion. We see that people who showed happiness also showed higher posting frequency with more social interactions. Additionally, female users tended to spend more time online and received more interactions, while age showed less relationship with usage time.

However, there are some limitations in the study. First, the dataset is relatively small, which restricts the generalizability of the findings. Also, the data is self-reported, which means that there could be some bias in users' own reporting. Another limitation is that the dominant emotion could be changed over time, so it is not a stable variable. We should do some future research with larger datasets and more features to provide more precise insights.

In conclusion, this study demonstrates that social media behavior is linked to emotional patterns with some factors such as gender, usage time, interactions, and age. By understanding these relationships, we can improve digital social media well-being.

**References**

Saiphoo, Alyssa G., and Tyler M. Vahedi. 2020. “Social Media Use and Its Connection to Mental Health: A Systematic Review.” *Cureus* 12 (6): e8627. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7364393/.

Bulut, Emirhan. n.d. *Social Media Usage and Emotional Well-Being*. Kaggle. https://www.kaggle.com/datasets/emirhanai/social-media-usage-and-emotional-well-being.