Project 1 Group 2 Written Analysis

We are a group of data analysts exploring the data on the usage of social media and emotional well-being. Social media has grown to become part of our modern day-to-day lives having an influence over various aspects of our emotional well-being. We will be analyzing demographic factors (age, gender), social media platform statistical factors (platform app, usage time, activity type), and emotional state factors (Happiness, Sadness, Anger, Anxiety, Boredom, Neutral). We seek to understand the relationship, if any, that may exist between these variables while also acknowledging that we are not experts and can only draw conclusions about the information in this dataset. This analysis will explore the dataset over platform activity and users’ emotional state, discuss our thought process of exploration, and include any conclusions found.

The data was sourced from Kaggle (see Citations). Temporal coverage dates were from February 15th to May 23rd of 2024 and the geospatial coverage was worldwide.

During the data retrieval and cleaning phase, we ran into an obstacle. There was a subsection of data from the ‘Age’ column that appeared to be switched with the values from the ‘Gender’ column. We tried to isolate this subsection of rows by using Booleans and using functions to store the values in a list before swapping them. We tried to use a ‘for loop’ to capture all rows with strings in the ‘Age’ column. After multiple attempts on our own, we then asked for guidance from our TA’s and Instructor. From this point, we were able to use the ‘.isin()’ method to filter data based on the 3 gender categories to specifically capture the rows with genders in them that are in the ‘Age’ column. For the rows captured in the ‘.isin()’ method, we renamed the columns to their corresponding values and ‘pd.concat()’ the data back in before we dropped the null values. We also had to ask the Learning Assistant the same question multiple times but wording it differently and being more specific each time. This took a team effort with brainstorming and deep critical thinking.

Questions:

* Is there a correlation between age and time spent on social media?
* What is the distribution of users by average daily time spent on social media?
* Which platform is used by each gender?
* What is the average age of platform users?
* Which platform is preferred based on gender and the average age?
* What are the most preferred platforms based on gender and average age?
* Are emotion and age connected when it comes to social media usage?
* What emotion affects time spent on social media the most?
* Which gender spends the most amount of time on social media?
* Which gender is affected the most by their emotions when it comes to using social media?
* Does emotion have an effect in how social media is used ie. Likes, comments, posts?

For us to find out if there is a correlative relationship between age and time spent on social media, we first established our x variable to be the data in the ‘Age’ column and our y variable to be the data in the ‘Daily Usage Time’ column. We used the plt.plot() function to display the chart as a scatter plot in order to visually see if there is any direction that the data is pointing to. The scatter plot doesn’t appear to show any obvious relationships or heavy clusters in certain areas. The areas that appear slightly clustered would be between the ages of 27 and 29 which implies that most users are within the ages of 27 - 29 . On the y axis, the data appears slightly clustered within the 40 to 100 minutes time frame. This indicates that most users spend between 40 minutes to 1 hour and 20 minutes on average, per day using social media. A supplemental line regression was also created to further portray the data between Age and Time Spent. It appears that the line is not as narrow and is more flat. This implies that there is possibly very little relationship between the two variables and the slope is a very small amount. Based on the regression line and the scatter plot, it can be concluded that there is not a correlative relationship between Age and Time Spent on social media. Finally, we ran a correlation test to find the P-value and R-value of the data. The R-value of 0.081 suggest a weak positive correlation between age and usage time meaning that as age increases, there is little to no change in usage time. The P-value of 0.014 indicates that the correlation has some statistical significance meaning that there is sufficient evidence to reject the null hypothesis. In other words, the correlation of the data is not by chance. In conclusion, there is weak positive correlation between these two variable but the data suggests statistical significance.

To get the distribution of users by average time spent on social media (in minutes), we plotted the data using .plot(kind=’box’) as well as a histogram by using plt.hist(). We calculated for the mean, median, mode, least time spent, most time spent, quartiles as well as outlier thresholds. The overall average time spent on social media is approximately 96 minutes per day, the most middle amount of time is 85 minutes per day, and the most common/frequent amount of times spent is 60, 70, 75, and 90 minutes per day. This implies that the data average is being skewed or pulled to the right towards 200 minutes per day. Based on this, we believe that the median of 85 minutes per day best fits the frequency of the data. The standard deviation is approximately 39 minutes which implies that 68% of the data falls within 46 and 124 minutes per day. The least amount of time spent is 40 minutes per day and the most amount of time spent is 200 minutes per day. To find any existing outliers, we established quartiles by using .quantile() established the lower and upper quartiles and calculating for the upper and lower bounds. Based on the thresholds, the outlier would be considered any time spent below -17 minutes per day which is impossible so we conclude that the outlier would be 1 minute or less spent per day on social media. The outlier would also be considered 202 minutes per day or more spent using a social media platform. In conclusion, the distribution of users by average daily time spent on social media (in minutes) appears to be skewed to the left. The data is being pulled towards a few outliers at 200 minutes per day spent on social media.

**Gender**

We examined the data in terms of gender, categorizing users as female, male, and non-binary. We ranked the platforms preferred by users based on their gender. For this analysis, we used the 'gender' and 'platform' columns. Our findings show that the platform preferred by women is 'Twitter,' the platform preferred by men is 'Instagram,' and the platform preferred by non-binary individuals is 'Facebook.'

**Age**

We used the 'age' and 'platform' columns to calculate the average age of social media users and found that the average age was 27.5. To improve the accuracy of the data, and because 27.5 is not an exact number, we first evaluated users aged 27, followed by users aged 28.

**Gender and the Average Age**

To increase the reliability of the study, we first examined platform users by dividing them into two age categories: 27 and 28**.** The results showed that the platform used by 27-year-old women, men, and non-binary individuals was 'Twitter.' Meanwhile, the platform used by 28-year-old women, men, and non-binary individuals was 'Instagram.' Thus, the age factor revealed that different genders preferred the same platform.

**Most preferred platforms based on gender and average age**

\*To increase the reliability of the study and to observe how the results may change when analyzing multiple factors, we focused on the most preferred platform by users based on gender and average age (27 and 28). Surprisingly, the most preferred platforms were:

* 27-year-old women/men: 'Twitter'
* 28-year-old women/men: 'Instagram'
* 27-year-old non-binary individuals: 'LinkedIn'
* 28-year-old non-binary individuals: 'Facebook'

One of the first analyses we did was comparing the user's age with their dominant emotion. The emotion that we are looking at is what the person is feeling while using whatever their dominant social media platform. The largest sample group was Happiness, while the smallest was anger. So, it made for a good comparison since both emotions are considered opposites of each other. There were 186 people who reported feeling happy while scrolling through their social media accounts, while we had 122 people who reported feeling angry while scrolling through their accounts. When looking at our graphs, people between the ages of 27-30 were the groups reporting feeling angry while they used social media. Even looking at the median age group which in this case was twenty-seven-year-olds, the most dominant emotion was anger. While happiness had one large group of people aged twenty-five who reported feeling happy while using social media. Even looking at the median age group which in this case was twenty-seven-year-olds, the most dominant emotion was anger. While looking at age and emotion there is no conclusive evidence that age has any factoring into what the person’s emotion was at the time of using social media.

One of the trends that was noticed amongst the data was that happiness does lead to a much higher amount of time spent on social media. The data showed that when happy the average person spends upwards of 140 minutes on social media. The lowest amount of time spent on social media came from people who were experiencing boredom who averaged about 60 minutes. When it came to looking at the three different genders, females spent on average 112 minutes on social media at a time, with males following at 94 minutes and non-binary individuals spending the least at 76 minutes. From here we decided to split the data set by each gender and see how emotions compared to time spent. Amongst all three genders if happiness was the main emotion each group spent upwards of 140 minutes on average on social media. While the lowest for each group was still boredom at 60 minutes on average. The second most dominant emotion was anger for males with it averaging 89 minutes. Females and non-binary individuals' second most dominant emotion was anxiety, with females averaging 117 minutes while non-binary was around 81 minutes. With all the data present we can conclude that the female gender spends the most time on social media, they are also the most affected by their emotions when it comes to social media usage. While happiness remains the most dominant emotion when it comes to social media usage.

In this data set we had to set our data frame to take the correct columns into account. "Posts Per Day", "Likes Received Per Day", "Comments Received Per Day", and "Messages Sent Per Day". From here the data was split for each column to compare it to each distinct emotion that is in the original data sheet. Using a pie chart we split the data for each column based upon the emotion that the person was feeling while using their social media. Happiness remained as the driving force behind interaction on social media. We noticed that within each of the four columns mentioned above, happiness is having at least 29% of the interaction and in the case of “Likes per day” it jumps to 38%. Boredom on the other hand shows that it does cause the least amount of interaction averaging between 5-8%. So there seems to be at least some correlation between our emotional state and how much we actually interact on social media.

Jenn’s Questions :

1) Is there a correlation between genders and emotional well-being on social media?

2) What age group spends the most time on social media and what is their emotional well-being while on social media?

3) Comparing the emotional state of individuals using different platforms?

Analysis:

The relationship between gender and emotional well-being on social media reveals some interesting patterns. We did separate pie charts showing the distribution of emotions across females, males, and non-binary individuals. Just by looking at the charts there are noticeable differences in how these groups engage emotionally with social media.**Happiness** is the most prominent emotion among females, making up **30%** of their emotional distribution, significantly higher than males **20%** and non-binary individuals **7%**. This suggests that females in the dataset are more likely to experience or express positive emotions on social media. In contrast, non-binary individuals exhibit the highest proportion of **Neutral** emotions **33%**, indicating a more emotionally indifferent or balanced state compared to males **14%** and females **16%**. When examining negative emotions, males show a broad spread with **Anger 17%**, **Anxiety 17%**, **Sadness 16%**, and **Boredom 16%**, collectively accounting for **66%** of their emotional profile. Females, while expressing higher happiness, still display a significant portion of negative emotions **53%**, dominated by **Anger 16%** and **Anxiety 16%**. Non-binary individuals demonstrate a different profile, with minimal **Anger (4%)**, but higher levels of **Sadness 19%** and **Boredom 18%**, highlighting potential emotional challenges or unique stressors in their social media experience. The visualization process involved creating individual pie charts for each gender using a contingency table of emotions and genders. By looping through each gender and plotting the respective emotion percentages, the charts clearly depict the variation in emotional states. Custom colors for each emotion further enhance interpretability. This approach not only highlights differences but also facilitates a deeper understanding of how emotions vary across genders in social media contexts.

When exploring the connection between age, time spent on social media, and emotional well-being, the data reveals that individuals in their late twenties, particularly those aged 27 to 30, spend the most time on social media, often exceeding 150 minutes per day. The scatter plot shows the connection between age, time spent on social media, and emotional well-being, represented by dominant emotions. Each dot on the graph represents a specific emotion for an age group, with the size of the dot indicating how often that emotion was recorded. This allows us to see patterns of social media use and emotional states at different ages. The average age of participants is **27.5**, with the youngest being **21** and the oldest **35**. People aged **27 and 28** are the most common in the dataset, followed by those aged **29 and 22**. Individuals in their late twenties, especially those aged **27 to 30**, spend the most time on social media, with many exceeding **150 minutes per day**. When it comes to emotions, **Happiness** and **Neutral** feelings are the most common across all ages and are associated with higher daily usage times. Younger individuals, especially those aged 21 to 24, tend to report more frequent negative emotions like Sadness and Boredom, although these are present across all age groups. Anxiety is also noted in several age ranges, though it is less frequent overall. Interestingly, Anger is the least common emotion recorded. The scatter plot was created by grouping the data by age and dominant emotion, counting how often each emotion occurred, and plotting it. Different colors represent different emotions, and the size of each dot shows how many people in that age group experienced a particular emotion. This method helps highlight the relationship between time spent on social media, age, and emotional well-being. Overall, the visualization shows that people in their late twenties tend to spend the most time on social media and are more likely to experience positive emotions like happiness during their usage.

To analyze the emotional states of individuals using different platforms, we first reviewed the distribution of platforms and dominant emotions in the dataset. The platform distribution showed that **Instagram** had the highest number of users **236**, followed by **Twitter** **188**, **Facebook** **178,** and **Linkedin 118**. Platforms like WhatsApp, Telegram, and Snapchat had an equal but smaller number of users, 68 users each. Regarding emotions, **Happiness** was the most frequently reported emotion **186** occurrences, closely followed by **Neutral** **184** occurrences. Negative emotions like Anxiety, Sadness, Boredom, and Anger appeared less frequently, with **Anxiety** being the most common negative emotion **156** occurrences. To visualize the relationship between platforms and emotional states, we created pie charts showing the emotional breakdown for each platform. The code used to generate these charts utilized a subplot layout with two rows and four columns to display the data for all platforms in one figure. Each platform’s emotional distribution was represented by a pie chart, with slices indicating the proportion of each emotion. Colors corresponding to specific emotions were applied for better clarity. Since there were only seven platforms, one subplot was left blank. The results provided interesting insights. For instance, Instagram stood out as a predominantly positive platform, with Happiness accounting for an overwhelming 70% of its users' emotional states. In contrast, Twitter exhibited a more balanced emotional distribution, with significant portions of Anxiety 41% and Sadness 24%, highlighting the platform's tendency to evoke or reflect negative emotions. LinkedIn users reported high levels of Boredom 53%, likely tied to its professional focus and job hunting. Snapchat, on the other hand, showed a more neutral emotional tone, with Neutral emotions 38% being the most dominant. Platforms like Facebook, WhatsApp, and Telegram demonstrated varied emotional distributions, with Neutral and Negative emotions such as Anxiety and Sadness being more prominent. This analysis helps illustrate the emotional impact of different platforms on their users, highlighting Instagram’s positivity and Twitter’s emotional diversity. The use of pie charts provided a clear, visual comparison of emotions across platforms, making it easier to identify trends and differences.

**Citations**

Emirhan BULUT. (2024). Social Media Usage and Emotional Well-Being [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/8460631