Effects of Social Media (COMP3125 Individual Project)

Darlie Fils

Abstract—

The objective of my search was to gauge whether there is a correlation between the rise of social media- specifically Instagram- and mental health in the United States.
This research explores the multifaceted relationship between social media usage and mental health outcomes, focusing on both positive and negative effects. Social media platforms, including Instagram, and TikTok, have become central to modern communication and daily life, influencing individuals' self-perception, emotional well-being, and social dynamics. Through a combination of literature review and empirical analysis, this study investigates how excessive social media consumption correlates with mental health issues such as anxiety, depression, self-worth and concentration.

I. Introduction

Social media has become a very pervasive part of our society as more and more users are being created every day. With the rise of social media, it is to be expected that many other fabrics of our society would be affected such as mental health. In the past two decades, social media has fundamentally altered the way people communicate, share information, and engage with others, becoming an integral part of daily life. Platforms like Facebook, Instagram, Twitter, and more recently, TikTok, have revolutionized how individuals interact, create identities, and experience social connectedness. While these platforms offer numerous advantages, such as fostering relationships, providing opportunities for self-expression, and enabling the dissemination of information, there has been growing concern about their potential negative impact on mental health. With apps such as Instagram that are all about how many likes one receives on a post, gaining so much popularity over the years, there is reasonable concern that it might have a negative effect on feeling of self-worth for many Americans. Going into this research, I had 4 questions in mind: 1. How has the number of people on social media increased in the last few years?

2.Does this increase have any effect on other sectors of society like mental health?

3.How have sentiments regarding the internet shifted with its growing popularity?

4. Is there a link between the rise of social media usage and attention span/ decrease of collective intelligence.

II. Datasets

A.) Mental Health Social Media Dataset.csv obtained from Github. Link : [Mental-Health-and-Social-Media-Analysis/Mental Health Social Media Dataset.csv at main · randhir-05/Mental-Health-and-Social-Media-Analysis](#). This source is a credible and the dataset was created on 4/18/2022 through a survey of University Students.

B.) This data set is made up of data collected from surveyed university students. They are prompted to answer the questions such as: "how often do you use social media?" and "how often to you feel depressed?". They answer these questions with a numerical answer 1-5 where 1 is never and 5 is always. It is comprised of 481 rows and 22 columns. This dataset is comprised of mostly numeric values except for the column pertaining to how many hours spent on social media daily, which consisted of categorical data. I converted the values of the column to float values ie. 'Less than an Hour" became ".5". I also cleaned the data, getting rid of irrelevant rows and columns leaving me with a file dataset size of 478 rows x 8 columns.

III. Methodology

My methodology for finding this assumed link between mental health and social media usage throughout the years was to employ Jupyter Notebook and functions such as Pandas to clean and organize the data I have obtained. The dataset that I utilized on Jupyter was the 'Mental Health Social Media Dataset.csv' and it was a very large dataset with 481 rows and 22 columns.

A.Method A

As not all the columns were useful and relevant to my research, I started off by first employing the df1.drop() function to remove the columns that were not of importance, such as 'What_is_you_age" and "Relationship status". After removing all the columns, I did not need I was then left with 8 columns as opposed to the original 22.

B. Method B

I then converted the column that asks about the average time spent on social media from categorical data such as 'More than 5 hours' to numeric values like '6'. To do this I used the 'replace' function in Python to replace the values, 'Between 1 and 2 hours' , "Between 2 and 3 hours' etc to '1.5' , '2.5' , etc respectively.
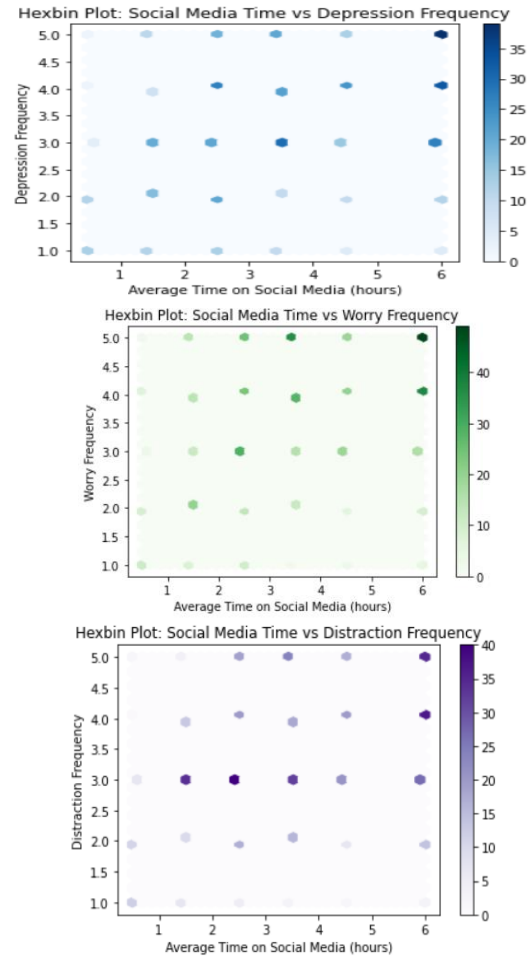
C. Method C

I followed by removing the rows of people who answered 'no' in column 'Do_you_use_social_media' to ensure that when I took the averages of the responses from columns that inquire about aspects of mental wellbeing such as, 'How_easily_distracted_are_you' , it would be a true representation of the those who actually use social media. To do this I used the python function 'df1 = df1.drop([134, 9, 54,]' where 134, 9 and 54 are the indexes in which we find the only 3 people in the entire dataset who answered no to using social media, leaving me with my final data set comprised of 478 rows x 8 columns as opposed to the original 422 rows x 22 columns. This is my final df1 and I proceeded to find the averages and standard deviation of the responses for each column using the df1['Column'].mean() and df1['column'].std() functions respectively. I also used the df1.['…social_media_every_day]. mode() python function to decipher which response showed up the most for the column that asks about average time spent on social media.

D. Method D.

Now that I had df1 comprised of everyone who answered 'yes' to social media, I needed to create df2 comprised of the no answers for comparative reasons. I set df2 equal to the 'Mental Health Social Media Dataset.csv', utilized df2.drop() to get rid of the same columns I removed for df1 and then finally used the df2.loc[[134,9,54]] to be left with just those three values. I then proceeded to take the averages of their responses using the same python mean functions as I used for df1.

IV.   Results

For df1, when asked about the average time spent on social media daily, I found the mode to be '6' or 'More than 5 hours'. I found the average responses to the question pertaining to how easily distracted one is to be 3.36 with a standard deviation of 1.171. For the column relating to how bothered one is by worries, I found the mean to be 3.57 with a standard deviation of 1.28. For the column about seeking validation from social media, I found the mean to be 2.46 and standard deviation of 1.25. Finally for df1, I found the mean of the column that inquires about how often one feels depressed or down to be 3.28 with a standard deviation of 1.31. Conversely for df2(the table with the 'no' responses) I found the mean of the responses for the column relating to distraction, worries, validation and depression to be 1.67, 1.67. 1.0 and 1.33 respectively.

Hexbin Plot: Social Media Time vs Depression Frequency

Hexbin Plot: Social Media Time vs Worry Frequency

Hexbin Plot: Social Media Time vs Distraction Frequency

V.   Discussion

It was incredibly difficult to obtain the datasets for this project as many of them were private or only accessible through payment.

It took a very long time to find this dataset although I am very pleased with it. The caveat to this specific dataset is that although it consists of many inputs, it is not very varied in terms of the types of people who were surveyed. Most of the participants were around the same age, and university students, which might influence the variability of responses.

To improve this search, I would conduct my own research in which I survey a multitude of random people of a range of different ages and occupations and ask them similar questions to create a more varied and inclusive dataset.

VI.   CONCLUSION

In conclusion, based on the dataset, the results suggest that there is a connection to the amount of time spent on social media and feelings of worry, depression and distraction. Most students responded that they spend more than 5 hours a day. This highlights the pervasiveness of social media today. Only 3 students out of the 481 surveyed answered that they do not use social media, evincing just how

common it is for most people to be on the internet and at least one social media platform. There also exists a pretty big difference in the averages of each response between those you answered yes to using social media and those who don't. For instance, those who say that they do use social media had an average of 3.28 when questioned how often they feel depressed or down whereas, the average for the three who don't use social media was 1.33. This is a pretty big difference as it suggests that the less time spent on social media has a link to mental health and productivity. The hexabin plots also express this sentiment as visually we can see that the color of the points are darker in the areas that have a higher density of responses, For instance, for the 'Social Media Time Vs. Depression Frequency', we see the darkest point at (6, 5) where 6 is hours spent on social media and 5 is the response to how often one feels depressed. This pattern is also found in the hexabin plots, Social Media Time vs Worry Frequency' and 'Social Media Time vs Distraction Frequency'. These graphs further support 'the notion that the more time spent on social media takes a toll on mental health and focus. Based on this dataset, it is fair to conclude that there is a bit of a correlation and link between time spent on social media and the mental.

REFERENCES

[1] randhir-05,"Mental-Health-and-Social-Media-Analysis/Mental Health Social Media Dataset.csv at main · randhir-05/Mental-Health-and-Social-Media-Analysis," *GitHub*, 2024. https://github.com/randhir-05/Mental-Health-and-Social-Media-Analysis/blob/main/Mental%20Health%20Social%20Media%20Dataset.csv