**Connected, Yet Disconnected! A Deep Dive into the Social Media Paradox**

By Erik R. Lopez

**Introduction**

**Purpose of the Project:**

This project investigates the complex relationship between social media consumption and well-being, focusing on how different usage patterns can impact mental health, productivity, and overall satisfaction. Using a dataset of 1000 social media users, this study employs a combination of descriptive statistics, correlation analysis, ANOVA, and regression modeling to uncover potential consequences associated with various social media behaviors. As social media becomes increasingly integrated into the daily lives of individuals across diverse platforms like Facebook, TikTok, Instagram, and YouTube, this research aims to raise awareness and provide data-driven insights that can empower users to make informed choices about their online engagement.

**Background/Context**

Social media has become deeply ingrained in modern life, with millions of individuals of all ages and backgrounds worldwide engaging with various platforms. While offering undeniable benefits like global connectivity, diverse entertainment, and vast information resources, social media consumption has also raised concerns about its potential negative impacts. Studies have linked excessive use to mental health issues such as depression, anxiety, and low self-esteem, as well as to sleep disturbances and addictive behaviors (Andreassen et al., 2016; Lin et al., 2016; Primack et al., 2017; Woods & Scott, 2016). Furthermore, exposure to social media has been associated with the spread of misinformation and the occurrence of cyberbullying, leading to serious consequences like depression and even suicide (Kowalski et al., 2014; Van Cleemput et al., 2010; Vosoughi et al., 2018). Given these potential risks, this project undertakes a deep dive into user behavior and the associated consequences of social media consumption, utilizing a data-driven approach to uncover meaningful insights.

**Research Questions**

**Demographic Factors**

1. How does the Total Time Spent on social media platforms differ across demographic factors (Age Groups, Gender, Location)?
2. How does the Number of Sessions on social media platforms differ across demographic factors (Age Groups, Gender, Location)?
3. How does the Number of Videos Watched on social media platforms differ across demographic factors (Age Groups, Gender, Location)?

**Behavioral Factors**

1. How are platform, content types (Video Category), and usage frequency (Frequency, Watch Time) related to average Total Time Spent on social media?
2. How are the Platform, content types (Video Category), and usage frequency (Frequency & Watch Reason) related to the Number of Videos Watched?
3. Does Device Type influence social media consumption patterns (Total Time Spent, Number of Videos Watched, Number of Sessions), and does this vary across different platforms or content categories?
4. Do different motivational factors (Watch Reason) lead to variations in social media usage patterns (Number of Sessions, Number of Videos Watched, Total Time Spent)?

**Wellness Factors**

1. Do high amounts of social media consumption (Total Time Spent, Number of Sessions, Number of Videos Watched) impact individuals' mental wellness (Addiction Level, Self Control, Satisfaction, Productivity Loss, Importance Score)?
2. How do "Addiction Level," "Self Control," "Satisfaction," and "Productivity Loss" relate to each other, and are these relationships influenced by demographic factors or behavioral patterns?

**Temporal Factors**

1. What are the peak times throughout the day that individuals consume the most amount of time on social media platforms?
2. Do peak hours (Watch Time) and times of day (Frequency) differ between demographics (Age Group, Gender, Location)?

**Content Preferences**

1. How do content preferences (Video Category) vary across different demographic groups (Age Group, Gender, Location)?
2. Are there associations between content preferences (Video Category) and motivational factors (Watch Reason) or well-being outcomes (Addiction Level, Satisfaction, etc.)?

**Hypotheses**

**Demographic Factors**

1. Younger adults (18-34) will exhibit higher levels of social media consumption (measured by total time spent, number of sessions, and number of videos watched) compared to older age groups (35+). Fail to Reject the Null.
2. Males will exhibit higher levels of social media consumption (total time spent, number of sessions, number of videos watched) compared to females and individuals identifying with other genders. Fail to Reject the Null.
3. Individuals residing in the United States will demonstrate higher levels of social media consumption (total time spent, number of sessions, number of videos watched) compared to individuals residing in other countries. Fail to Reject the Null.
4. Social media consumption (total time spent, number of sessions, number of videos watched) will be highest during the evening hours compared to morning, afternoon, and night. Fail to Reject the Null.

**Behavior Factors**

1. Entertainment and habit will be more frequently reported as primary motivations for social media use compared to boredom and procrastination. Fail to Reject the Null.
2. Comedy, jokes/memes, and entertainment video categories will be associated with higher levels of social media consumption (total time spent, number of videos watched, number of sessions) compared to other video categories. Fail to Reject the Null.
3. YouTube and TikTok will exhibit higher levels of user engagement (total time spent, number of videos watched, number of sessions) compared to Facebook and Instagram.

**Wellness Factors**

1. Higher levels of social media consumption (total time spent, number of sessions, number of videos watched) will be associated with increased productivity loss, lower self-control, lower satisfaction, and a higher importance score. Fail to Reject the Null.
2. Higher levels of social media consumption (total time spent, number of sessions, number of videos watched) will be positively correlated with higher addiction levels. Fail to Reject the Null.

**Temporal Factors**

1. Social media consumption (total time spent, number of sessions, number of videos watched) will be highest during the night hours compared to other times of the day. Fail to Reject the Null.
2. Higher levels of social media consumption during the morning and afternoon hours will be associated with higher addiction levels, increased productivity loss, lower self-control, and lower satisfaction. Fail to Reject the Null.

**Content Preferences**

1. Preference for life hacks and vlogs will be associated with lower productivity loss, higher self-control, and higher satisfaction compared to the preference for other video categories (jokes/memes, gaming, pranks, entertainment, ASMR, and trends). Fail to Reject the Null.
2. Preference for jokes/memes, pranks, and trends will be associated with higher addiction levels, lower self-control, lower satisfaction, and a higher importance score. Fail to Reject the Null.

**Data & Methodology**

**Data Source**: The dataset used in the project is from an open-source website called Kaggle, it was uploaded by Muhammed Roshan Riaz and is called Dark Side of Social Media (Dark Side Of Social Media). The sample size of the dataset is 1000. The key variables of the dataset are:

* UserID: unique identifier, primary key.
* Age: The age of the user.
* Gender: The gender of the user.
* Location: The geographical location of the user.
* Income: The annual income of the user.
* Debt: Tells If the is in Debt or Not.
* Owns Property: Indicates whether the user owns any property (Yes/No).
* Profession: The profession or job title of the user.
* Demographics: Additional demographic information about the user (Rural or Urban Life).
* Platform: The social media platform used by the user (e.g., Facebook, Instagram, TikTok).
* Total Time Spent: The total time the user has spent on the platform.
* Number of Sessions: The number of sessions the user has had on the platform.
* Video ID: A unique identifier for each video watched.
* Video Category: The category of the video watched (e.g., Entertainment, Gaming, Pranks, Vlog).
* Video Length: The length of the video watched.
* Engagement: The engagement level of the user with the video (e.g., Likes, Comments).
* Importance Score: A score representing the perceived importance of the video to the user.
* Time Spent On Video: The amount of time the user spent watching the video.
* Number of Videos Watched: The total number of videos watched by the user.
* Scroll Rate: The rate at which the user scrolls through content.
* Frequency: How frequently the user logs into the platform.
* Productivity Loss: The amount of productivity lost due to time spent on social media.
* Satisfaction: The satisfaction level of the user with the content consumed.
* Watch Reason: The reason why the user watched the video (e.g., Entertainment, Information).
* Device Type: The type of device used to access the platform (e.g., Mobile, Desktop).
* OS: The operating system of the device used.
* Watch Time: The specific time of day when the user watched the video.
* Self-Control: The user's self-assessed level of self-control while using the platform.
* Addiction Level: The user's self-assessed level of addiction to social media.
* Current Activity: The activity the user was engaged in before using the platform.
* Connection Type: The type of internet connection used by the user (e.g., Wi-Fi, Mobile Data).

Independent Variables: Gender, Age, Location,

**Data Cleaning/Preparation:** The dataset was cleaned in Microsoft Excel. Pivot tables, filters, and sorting were used to ensure there were no outliers, blank cells, misspellings, or typos. The issues found were that spaces needed to be added in a few cells with the variable title, for instance, ProductivityLoss to Productivity Loss or DeviceType to Device Type. Lastly, when it comes to the variable Location, one of the categories was misspelled from Barzil to Brazil.

**Analytical Techniques:** This project uses descriptive statistics, correlation analysis, regression analysis, and ANOVA analysis as statistical techniques.

**Program/Software:** Python (Jupyter Notebook), Microsoft SQL Server, Microsoft Excel, Microsoft Word, and Tableau Desktop.

**Analysis & Findings**

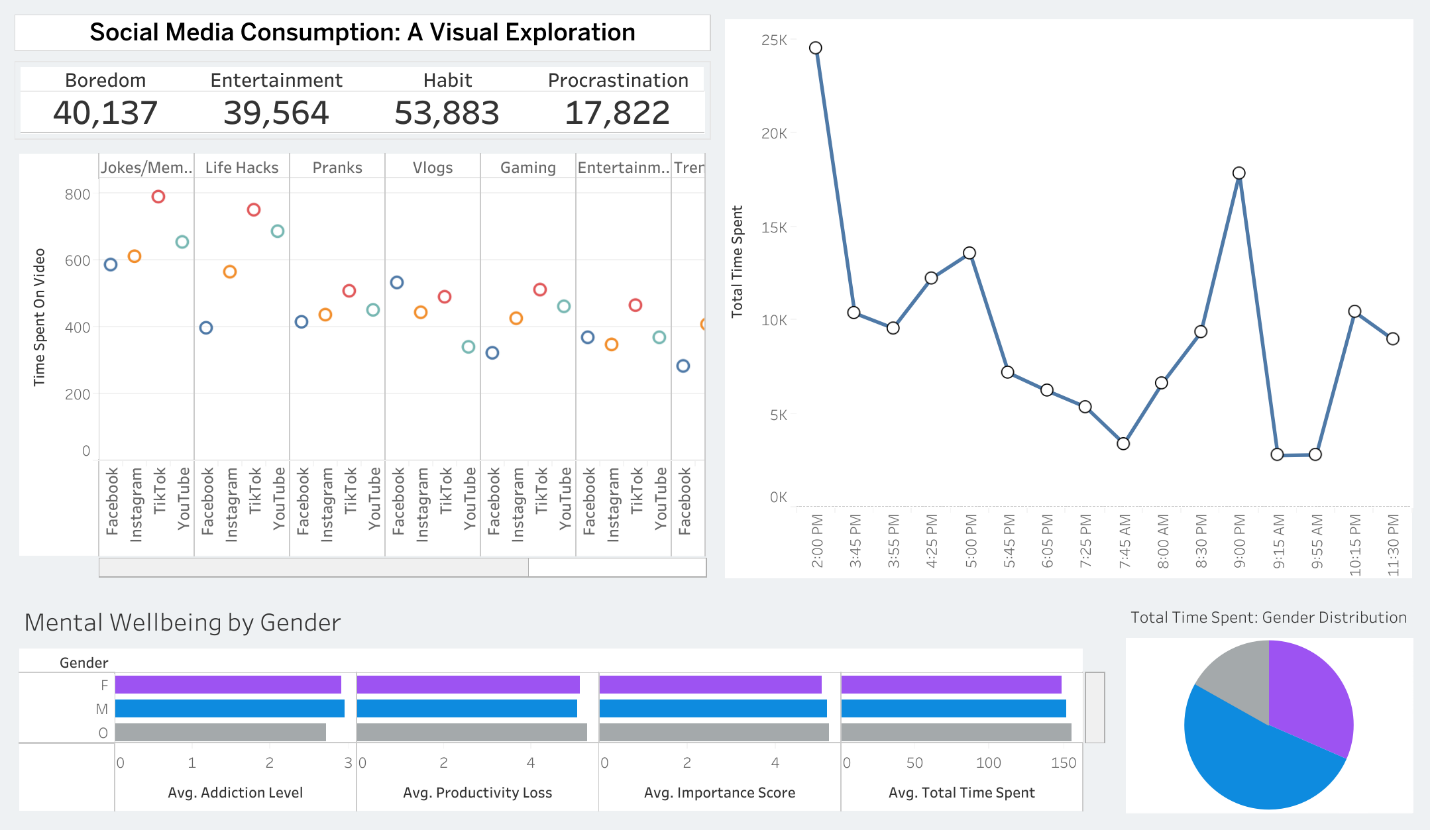
**Correlation Analysis:** Strong correlations were found between several well-being-related variables. Notably, "Productivity Loss" and "Satisfaction" showed a perfect negative correlation, while "Addiction Level" and "Self Control" were also perfectly negatively correlated. These findings highlight the complex interplay between social media use, self-control, productivity, satisfaction, and addiction.

**ANOVA Analysis:** Multiple ANOVA analyses were done using Python and found that no matter what independent and dependent variables were used in the study, findings demonstrated no statistical significance between them (p values were greater than 0.05).

1. **Demographic Factors:**
   1. There was no statistical significance between Total Time Spent and Age Groups (p = 0.66 > 0.05), Gender (p = 0.68 > 0.05), and Location (p = 0.32 > 0.05).
   2. There was no statistical significance between the Number of Videos Watched and Age Groups (p = 0.79 > 0.05), Gender (p = 0.18 > 0.05), and Location (p = 0.07 > 0.05).
   3. There was no statistical significance between the Number of Sessions and Age Groups (p = 0.15 > 0.05), Gender (p = 0.17 > 0.05), and Location (p = 0.21 > 0.05).
2. **Behavioral Factors:**
   1. There was no statistical significance between Engagement and Watch Reason (p = 0.99 > 0.05).
   2. There was no statistical significance between Total Time Spent and Watch Reason (p = 0.20 > 0.05).
   3. There was no statistical significance between the Number of Sessions and Watch Reason (p = 0.37 > 0.05).
   4. There was no statistical significance between the Number of Videos Watched and Watch Reason (p = 0.53 > 0.05).
3. **Content & Platform Preferences**
   1. There was no statistical significance between the Total Time Spent and the Video Category (p = 0.99 > 0.05).
   2. There was no statistical significance between the Number of Sessions and the Video Category (p = 0.31 > 0.05).
   3. There was no statistical significance between the Number of Videos Watched and the Video Category (p = 0.98 > 0.05).
   4. There was no statistical significance between the Total Time Spent and the Platforms (p = 0.74 > 0.05).
   5. There was no statistical significance between the Number of Sessions and the Platforms (p = 0.98 > 0.05).
   6. There was no statistical significance between the Number of Videos Watched and the Platforms (p = 0.26 > 0.05).
   7. There was no statistical significance between the Total Time Spent and the Device Type (p = 0.93 > 0.05).
   8. There was no statistical significance between the Number of Sessions and the Device Type (p = 0.61 > 0.05).
   9. There was no statistical significance between the Number of Videos Watched and the Device Type (p = 0.92 > 0.05).
4. Temporal Factors:
   1. There was no statistical significance between the Total Time Spent and the Watch Time (p = 0.52 > 0.05).
   2. There was no statistical significance between the Number of Sessions and the Watch Time (p = 0.09 > 0.05).
   3. There was no statistical significance between the Number of Videos Watched and the Watch Time (p = 0.65 > 0.05).
   4. There was no statistical significance between the Total Time Spent and the Frequency (p = 0.51 > 0.05).
   5. There was no statistical significance between the Number of Sessions and the Frequency (p = 0.07 > 0.05).
   6. There was no statistical significance between the Number of Videos Watched and the Frequency (p = 0.83 > 0.05).

**Regression Analysis:** Numerous regression models were tested, but individual predictors, such as age, gender, device type, platform, and watch reason, generally did not have strong or statistically significant effects on social media usage patterns or their associated outcomes. However, the "Frequency" of social media use emerged as a significant predictor of both productivity loss and addiction level. Individuals who use social media most frequently in the morning or evening tend to report higher productivity loss, while those who use it most frequently at night tend to report higher addiction levels.

**Visualizations & Interpretations**



<https://public.tableau.com/views/SocialMediaConsumptionAVisualExploration/SocialMediaConsumptionAVisualExplorationDashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>

A screenshot of a computer

Description automatically generated

<https://public.tableau.com/views/SocialMediaConsumptionDemographicInsight/SocialMediaConsumptionDemographicDashboard?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>

**Interpretation:**

Total Time Spent by Watch Reason

* Social media users tend to spend a significant amount of time on these platforms due to habit, followed by boredom, entertainment, and procrastination. This finding indicates that habit plays a major role in driving social media consumption, raising concerns about the potential for developing unhealthy usage patterns if not consciously managed.

Total Time Spent by Watch Category

* Across all platforms, the most popular video categories are Jokes/Memes and Life Hacks. Interestingly, Facebook stands out as having relatively low consumption for these two categories compared to other platforms. This suggests that users on different platforms might have varying content preferences, with Facebook users potentially less interested in Jokes/Memes and Life Hacks compared to other types of content

Total Time Spent by Watch Time

* Social media users are most active in the late afternoon and evening, with noticeable spikes in total time spent around 2:00 PM and 9:00 PM. Activity remains relatively high between these two peaks, indicating consistent engagement throughout the afternoon and evening. Notably, evenings exhibit higher activity levels compared to mornings, suggesting that social media use might be more prevalent during leisure hours.

Average Addiction Level, Productivity Loss, Importance Score, and Total Time Spent by Gender

* This visualization reveals some intriguing gender differences in how social media impacts well-being. Individuals identifying as "Other" report higher productivity loss, place higher importance on social media, and spend more time on these platforms compared to males and females. However, they also report the lowest average addiction levels. Males and females show similar patterns across most metrics, with males slightly higher on addiction level and females slightly higher on productivity loss. These findings highlight the complex relationship between gender, social media usage, and well-being, suggesting that different genders might experience the impact of social media in unique ways.

Total Time Spent by Gender Distribution

* Males tend to spend more time on social media than females and those identifying as "Other." This finding suggests that gender might affect the overall time commitment to social media platforms.

Critical Thinking:

* Habit Formation: The high time spent due to habit could be related to social media's addictive nature. Platforms are designed to keep users engaged, and over time, this can lead to habitual checking and scrolling even without a legitimate purpose.
* Boredom & Procrastination: Users tend to turn to social media as a means of escapism or to avoid tasks they find wasteful. This can be because of the ease of access or the constant stimulus spikes on these platforms.
* Users who engage with social media platforms because of habit or boredom, most likely lack or refuse to engage in other activities in their lives or lack hobbies.
* The popularity of jokes/memes and life hacks might be because of the prevalence of short-term form, making the content entertaining and informative in an easily consumable way.
* Facebook’s lower consumption can be due to its being an older platform, being only popular with older users. The only issue with this conclusion is that YouTube is an old platform as well but constantly contains high traffic on the platform. One reason is that YouTube contains longer forms of content. Well, Facebook focuses on short-form content, not being able to compete with TikTok or Instagram.
* High user login occurs during the afternoon and evening. This falls in line with times of the day when people take lunch breaks and get off of work, causing them to be more social with friends and loved ones.
* Females' higher productivity loss can be caused by the fact that women spend more time on social media tailoring their profile to look at a certain image.
* Jokes/Memes provide a short form of entertainment that provides instant stimulus, making them highly consumed by all three genders. Women tend to lean and care more towards trends, well males lean toward life hacks and gaming. Gaming does make sense for males since there are more male gamers than there are females. A limitation would be the type of life hacks and jokes/memes that females and males consume because there may be differences there.
* Social Norms and Expectations: Gender differences in time spent could be related to social norms and expectations around technology use, leisure activities, and social interaction.

A screenshot of a graph

Description automatically generated

<https://public.tableau.com/views/SocialMediaConsumptionMentalWellnessInsight/SocialMediaConsumptionMentalWellnessInsightDashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>

**Interpretation:**

Number of Sessions by Watch Reason

* On average, users have a significantly higher number of sessions when using social media for boredom (X sessions) and procrastination (Y sessions) compared to habit (Z sessions) and entertainment (W sessions). This could indicate that users turn to social media more frequently as a way to escape boredom or avoid tasks, highlighting the potential for social media to become a distraction or a means of procrastination.

Avg Number of Videos Watched by Platform & Age Group

* Instagram: Usage is highest among the 18-24 and 35-44 age groups, suggesting that this platform might be particularly appealing to these demographics. This could be related to Instagram's focus on visual content and its popularity among younger adults and millennials.
* TikTok: The 25-34 and 45-54 age groups show the highest average number of videos watched on TikTok. This might reflect the platform's appeal to a wider range of age groups with its diverse content offerings, including short-form videos, music, and comedy.
* YouTube: The 18-24 age group dominates YouTube usage, likely due to the platform's extensive library of long-form videos, including educational content, vlogs, and entertainment.

Scroll Rate by Video Category

* Pranks, ASMR, and Trends have the highest scroll rates, while Vlogs, Entertainment, and Life Hacks have the lowest. This suggests that users tend to engage more deeply with Vlogs, Entertainment, and Life Hacks, possibly due to their longer format, informative nature, or personal connection with creators. The higher scroll rates for Pranks, ASMR, and Trends might indicate that users are quickly browsing or skipping through these categories, potentially due to their shorter length or less engaging content. However, it's important to consider that content length might also play a role in scroll rates.

Avg Mental Wellness Visualization

* Avg Self-Control: Across all genders, users report high self-control scores, with an average score of 7.0 out of 10. This suggests that users generally feel in control of their social media usage. However, it's important to consider this in relation to other well-being metrics, such as addiction levels, to get a complete picture.
* Avg Productivity Loss: The average productivity loss across all genders is 5.0, indicating a moderate level of productivity loss due to social media use.
* Avg Importance Score: With an average importance score of 5.0, social media seems to hold a moderate level of importance in users' lives.
* Avg Satisfaction: Across all genders, satisfaction falls below average at around 4.0 out of 10. This suggests that users might not be fully satisfied with their social media experiences. Further research could explore the reasons for this dissatisfaction, such as negative social comparison, content quality, or platform features. This lower satisfaction might be related to the high prevalence of habit as a motivation for social media use, as seen in the 'Number of Sessions by Watch Reason' visualization.

**Critical Thinking:**

* Negative Social Comparison: Social media often presents idealized versions of people's lives, which can lead to negative social comparison and feelings of inadequacy.
* Content Quality: Users might be dissatisfied with the quality of content they encounter on social media, finding it repetitive, shallow, or unfulfilling.
* Platform Features: Certain platform features, such as algorithms that promote divisive content or addictive design elements, might contribute to dissatisfaction.
* Habitual Use: As you mentioned, the high prevalence of habit as a motivation for social media use might lead to less intentional and less satisfying experiences. Users might be scrolling out of habit rather than actively seeking out meaningful content or connections.

**Relationships:**

* Self-Control and Addiction: The high self-control scores might seem contradictory to the potential for addiction. This could suggest that users are aware of the addictive nature of social media and actively trying to control their usage, even if they haven't fully overcome addictive tendencies.
* Productivity Loss and Satisfaction: The moderate productivity loss and below-average satisfaction might be linked. If users feel like they're wasting time on social media without getting much enjoyment or fulfillment, it could lead to feelings of guilt and decreased productivity.
* Importance and Satisfaction: The moderate importance score and below-average satisfaction could indicate that while social media plays a role in users' lives, it might not be fulfilling its potential to provide meaningful connections or experiences.

**Limitations**

1. **Limited Scope of Variables**
   1. Watch Reason: The current study included a limited set of categories for Watch Reason. Future analysis could benefit from expanding these categories to include motivations such as:
      1. Avoidance/Escapism: Using social media to avoid negative emotions, stressful situations, or responsibilities.
      2. Professional Networking: Using social media for career advancement, job searching, or building professional connections.
      3. Social Interaction: Using social media to connect with friends and family, maintain relationships, or engage in online communities.
   2. Platform: The study focused on a subset of popular social media platforms. Future research could incorporate a wider range of platforms, such as:
      1. Twitch: A live-streaming platform popular for gaming and creative content.
      2. X (formerly Twitter): A microblogging platform used for news, discussions, and social commentary.
      3. Snapchat: A multimedia messaging app known for its disappearing messages and filters.
   3. Unused Variables:
      1. Income: How does income level relate to social media usage patterns and well-being outcomes?
      2. Profession: Are certain professions more likely to use social media extensively or experience specific impacts on their well-being?
      3. Debt: Does debt level influence social media use or relate to escapism or financial anxieties?
      4. Owns Property: Does homeownership relate to social media usage patterns or reflect different life stages or priorities?
      5. Current Activity: How does the activity users were engaged in before using social media affect their usage patterns and experiences?
      6. Connection Type: Does the type of internet connection (Wi-Fi vs. mobile data) influence usage habits or access to certain platforms?
      7. Demographics: Explore more detailed demographic variables, such as race, ethnicity, education level, and relationship status, to understand how social media use varies across different population groups.
      8. Video Length: How does the length of videos consumed relate to user engagement, satisfaction, and potential addiction?
2. **Additional Variables for Future Consideration**
   1. Future research could benefit from incorporating additional variables that might provide a more comprehensive understanding of social media use and its impact on mental wellness:
   2. Sleep: Assess the relationship between social media use and sleep quality, duration, and habits. This could involve collecting data on sleep patterns (e.g., bedtime, wake-up time, sleep disturbances) or using objective measures like sleep trackers.
   3. Personality Traits: Incorporate more detailed personality assessments, such as the Big Five Inventory, to explore how specific personality traits relate to social media usage patterns, preferences, and well-being outcomes.
   4. Self-Esteem: Include measures of self-esteem to investigate how social media use influences self-perception, body image, and social comparison.
   5. Education Level: Examine how education level relates to digital literacy, critical thinking skills, and the ability to navigate the complexities of online information and social interactions.

**References**

Andreassen, C. S., Torsheim, T., Brunborg, G. S., & Pallesen, S. (2016). The Bergen Facebook Addiction Scale: Construct validity and psychometric properties. *Psychological Reports, 118*(1), 107–117.

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). *Cyberbullying: Bullying in the digital age*. John Wiley & Sons.

Lin, L. Y., Sidani, J. E., Shensa, A., Radovic, A., Miller, E., Colditz, J. B., Hoffman, B. L., Giles, L. M., & Primack, B. A. (2016). Association between social media use and depression among U.S. young adults. *Depression and Anxiety, 33*(4), 323–331. [https://doi.org/10.1002/da.22466](https://www.google.com/url?sa=E&source=gmail&q=https://doi.org/10.1002/da.22466)

Primack, B. A., Shensa, A., Sidani, J. E., Whaite, E. O., Lin, L. Y., Rosen, D., Miller, E., Colditz, J. B., Hoffman, B. L., & Miller, E. (2017). Social media use and perceived social isolation among young adults in the U.S. *American Journal of Preventive Medicine, 53*(1), 1–8. [https://doi.org/10.1016/j.amepre.2017.01.010](https://www.google.com/url?sa=E&sourc%3C2%3Ee=gmail&q=https://doi.org/10.1016/j.amepre.2017.01.010)

Van Cleemput, K., Vandebosch, H., & Pabian, S. (2010). Cyberbullying: Prevalence, stability, and risk factors during adolescence. *Journal of Child Psychology and Psychiatry, 51*(4), 432-440.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151.

Woods, H. C., & Scott, H. (2016). #Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of Adolescence, 51*, 41–49. [https://doi.org/10.1016/j.adolescence.2016.05.008](https://www.google.com/url?sa=E&source=gmail&q=https://doi.org/10.1016/j.adolescence.2016.05.008)