Social Media Use and its Effects on Students

Student Name: Anuj Subedi
Student ID: 004009141

Objective

The objective of this project was to analyze the relationship between social media usage and the well-being of students across various demographics. Using a dataset of 705 student responses, this report employs data cleaning, statistical analysis, and visualization techniques to explore trends in academic performance, mental health, and sleep patterns.

Introduction

This dataset explores the relationship between students' social media usage, academic performance, sleep, and mental health. It contains 705 records with 13 attributes describing demographics, social media habits, and well-being scores. The attributes include Student_ID, Age, Gender, Academic_Level, Country, Avg_Daily_Usage_Hours, Most_Used_Platform, Affects_Academic_Performance, Sleep_Hours_Per_Night, Mental_Health_Score, Relationship_Status, Conflicts_Over_Social_Media, and Addicted_Score. Each attribute captures a specific part of a student's life to allow for a comprehensive analysis of how online habits correlate with offline well-being.

The dataset used for this analysis was obtained from Kaggle and is titled "Students' Social Media Addiction."

Method

The analysis was conducted following a structured methodology to ensure data integrity and the clarity of findings. The process involved data preparation and a visualization-based analytical strategy.

The initial step involved assessing the quality of the dataset. An investigation for missing values and duplicate entries was performed, and it was confirmed that the dataset contained no missing values or duplicated rows. This high level of data quality ensured the reliability of subsequent analyses. As a preparatory step, the Student_ID column was removed from the dataset, as it serves only as a unique identifier and holds no analytical value for this study.

The analysis was conducted using the Python programming language with the Pandas library for data manipulation, and the Matplotlib and Seaborn libraries for data visualization. The methodology included the generation of descriptive statistics to summarize the data, correlation analysis via a heatmap to identify high-level relationships, and a series of univariate, bivariate, and multivariate visualizations. These visualizations, including histograms, bar charts, box plots, and scatter plots, were employed to explore data distributions and investigate the relationships between social media habits and well-being metrics across different demographic groups.

Storytelling: Data Visualization and Interpretation

a) Establishing the Baseline of Student Behavior
To understand the impact of social media, it is first necessary to establish a baseline of typical student behavior. The following visualizations define the extent and nature of social media usage within the sample.

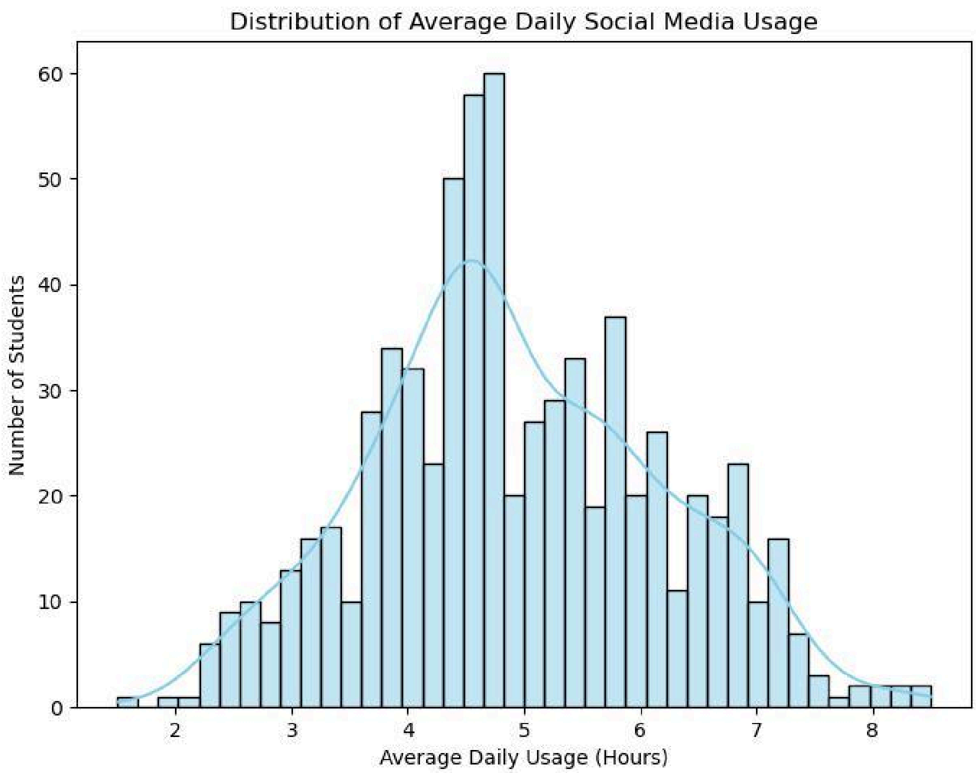

Figure 1: Distribution of Average Daily Social Media Usage
The histogram shows a unimodal distribution with a clear, single peak. The shape is asymmetrical with a slight right skew. The most frequent usage level is centered around 4.5 hours per day, and the bulk of the student population is concentrated in the 3 to 7-hour range. The right skew indicates that while the most common usage is high, there is a "tail" of students who engage in extremely high levels of daily use (7+ hours), pulling the average usage higher than the median.
It establishes a critical baseline for our story that heavy social media use is the norm for this student population, not the exception. The most common behavior is spending nearly 5 hours a day on social media. This finding gives significant weight to the subsequent analysis.

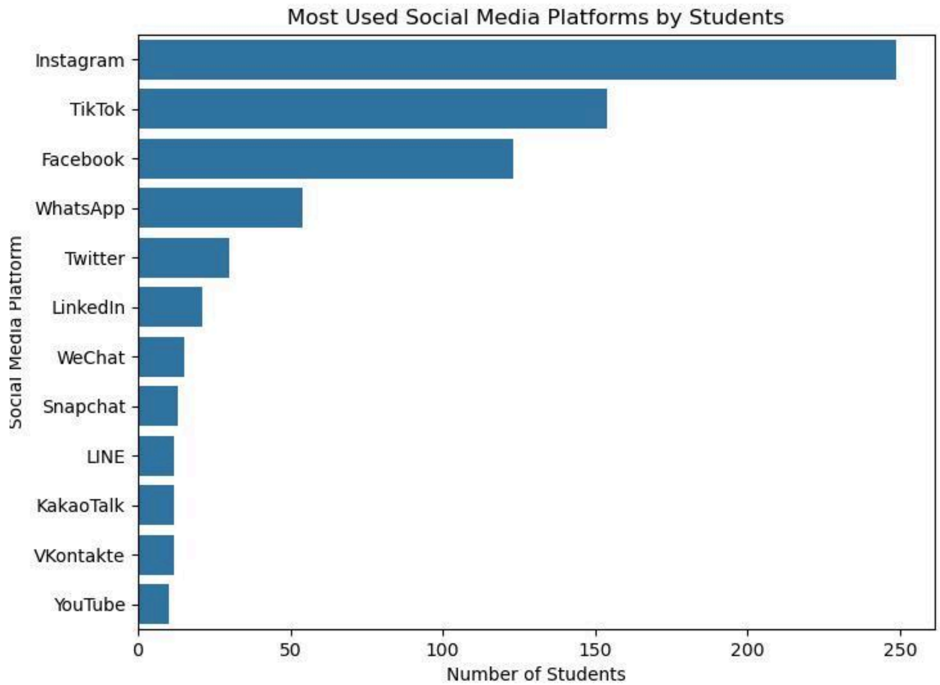

Figure 2: Most Used Social Media Platforms by Students
The chart clearly displays a ranked hierarchy of platform popularity. Instagram is the dominant platform, used by the most students. TikTok and Facebook are the second and third most used platforms, respectively. Together, these top three platforms represent the vast majority of primary student engagement. There is a substantial drop in the number of primary users after the top three, with platforms like WhatsApp, Twitter, and LinkedIn having a much smaller share.
This bar chart is essential for framing our narrative. It tells us that when we analyze the "effects of social media," we are primarily investigating the impact of Instagram, TikTok, and Facebook. This allows our story to be more focused and specific, linking the well-being metrics we explore later directly to the platforms where students spend the most time and energy.

b) Identifying the Core Problem
After establishing baseline behaviors, the next step is to explore the relationships between social media habits and key well-being indicators.

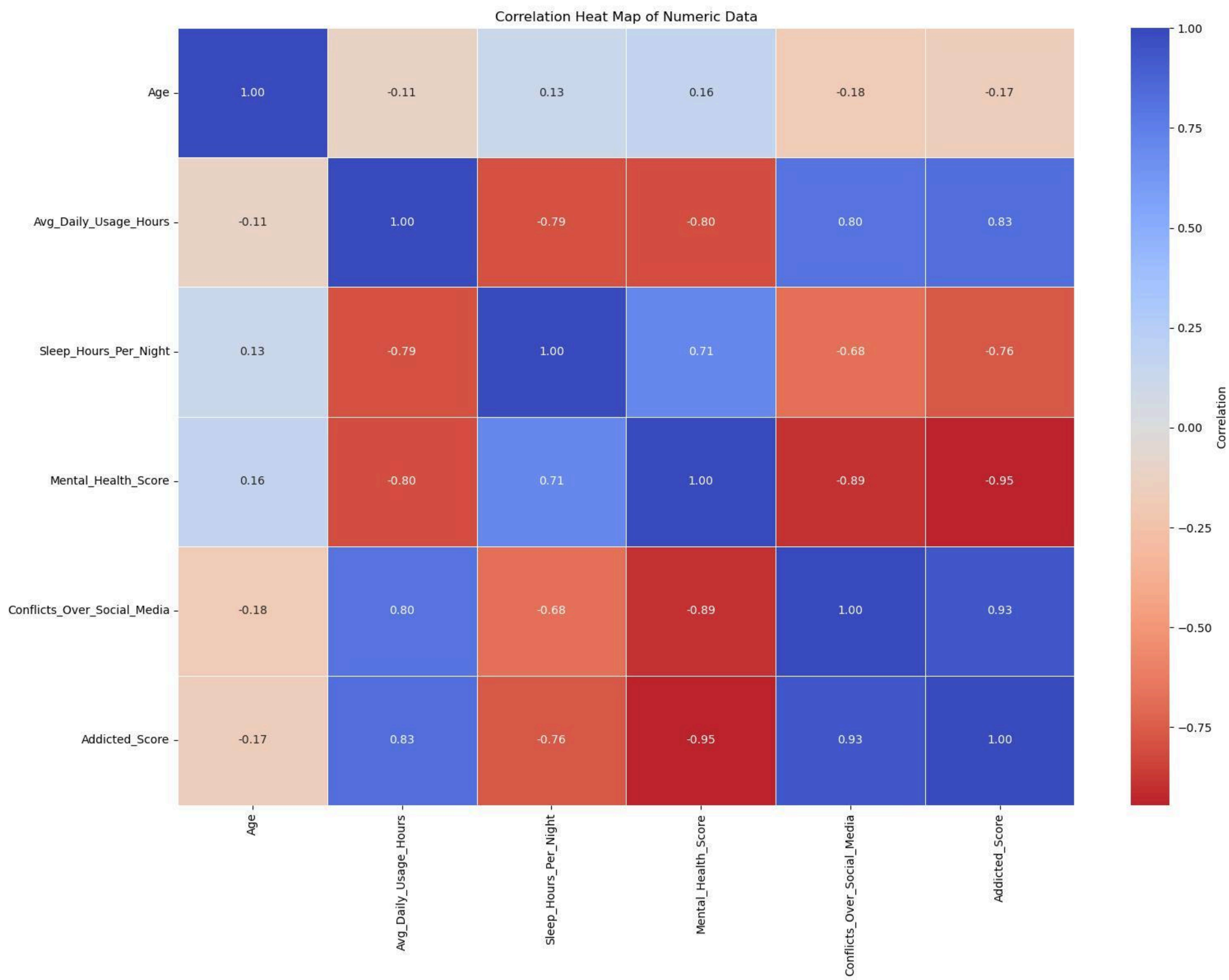

Figure 3: Correlation Heatmap of Numeric Variables

The heatmap reveals several strong and statistically significant correlations. Key insights include:

i) A strong negative correlation between Avg_Daily_Usage_Hours and both Mental_Health_Score (-0.80) and Sleep_Hours_Per_Night (-0.79).

ii) A strong positive correlation between Avg_Daily_Usage_Hours and both Addicted_Score (0.83) and Conflicts_Over_Social_Media (0.80).

iii) An extremely strong negative correlation (-0.95) between Addicted_Score and Mental_Health_Score.

Age shows only weak correlations with all other variables, suggesting these patterns are consistent across different age groups in the sample.

It suggests that heavy social media use is a central factor shaping well-being, driving poorer mental health, reduced sleep, increased addiction, and more frequent conflicts. These conflict-related behaviors appear to reinforce each other, creating a feedback loop.

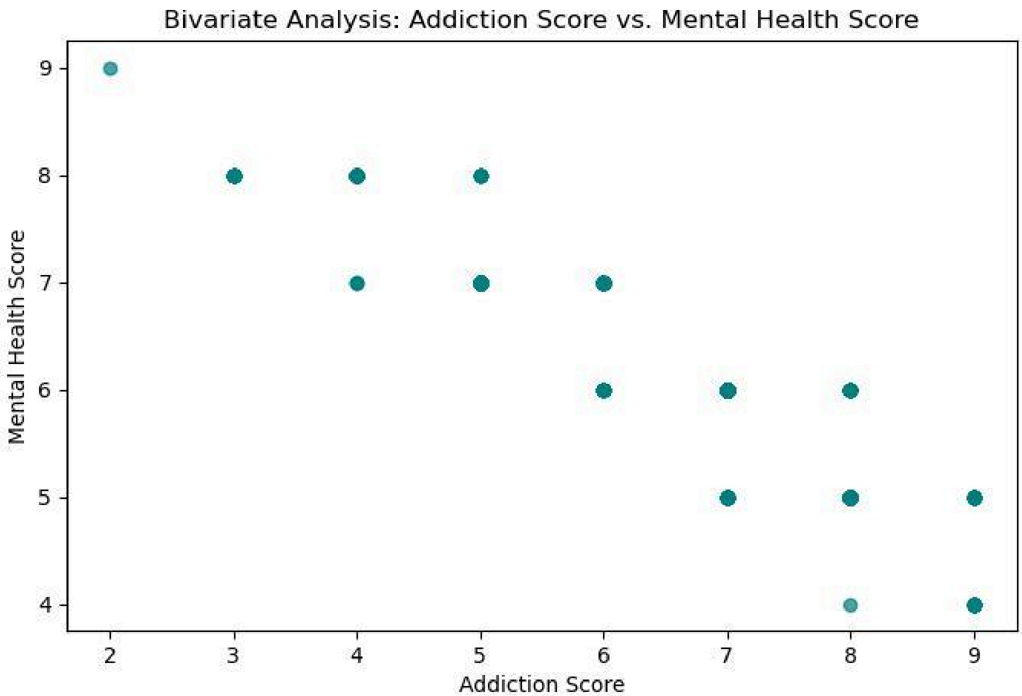

Figure 4: Bivariate Analysis: Addiction Score vs. Mental Health Score

The diagram reveals a clear, strong negative correlation between the two variables, forming a distinct downward-sloping pattern from the top-left to the bottom-right. The points are tightly clustered, indicating a very strong relationship with little deviation from the main trend.

This scatter plot is the most powerful piece of evidence in the analysis. It visually confirms the extremely strong correlation (-0.95) found in the heatmap and serves as the "smoking gun" for the project's central theme. The tight clustering makes it clear that for this group of students, higher levels of social media addiction are directly and powerfully associated with poorer mental well-being.

c) Exploring Influencing Factors and Demographics

Next, we discuss how different demographic and situational factors influence social media habits and their outcomes.

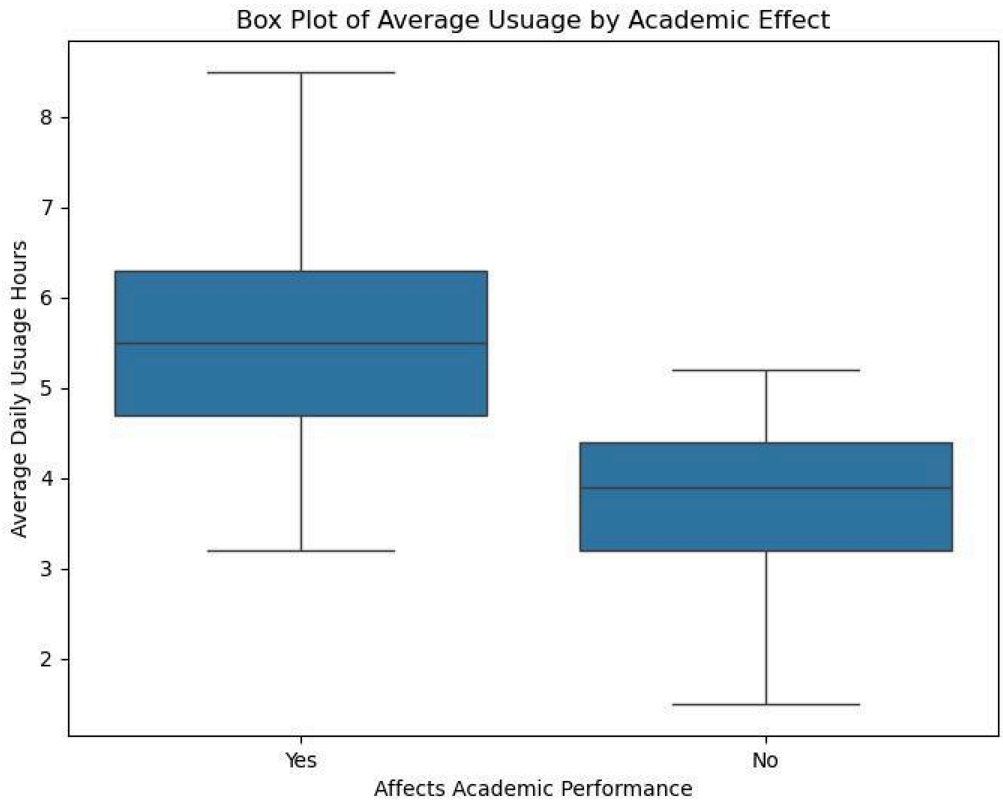

Figure 5: Box Plot of Average Usage by Academic Effect

The median daily social media usage is higher for students who report that social media affects their academic performance ("Yes") compared to those who report no effect ("No"). The IQR for the "Yes" group is larger, meaning their usage levels vary more widely. The "No" group has a narrower IQR, indicating their usage is more consistent and clustered at lower levels. The whiskers for the "Yes" group extend further upward, showing they have a higher overall maximum usage, with some students regularly exceeding 7-8 hours of use.

This box plot suggests a clear pattern: students who report academic harm from social media tend to use it significantly more and with greater variability. The wider spread and higher values in the "Yes" group highlight that excessive or irregular social media use may play a meaningful role in disrupting academic focus and performance.

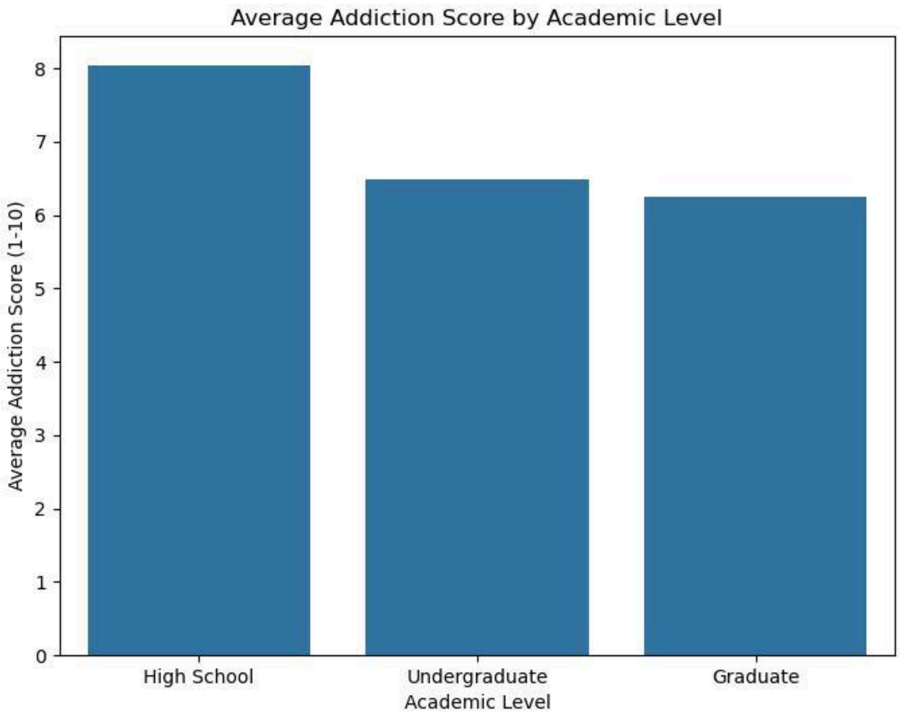

Figure 6: Average Addiction Score by Academic Level
High School students in this sample show the highest average addiction score. However, as this group is very small (n=27), this result may not be representative. The more statistically reliable trend is the slight decrease in average addiction scores from the Undergraduate level to the Graduate level.
This chart suggests that the undergraduate experience correlates with slightly higher levels of social media dependence compared to graduate studies. The lower scores for graduate students might indicate that increased academic rigor or maturity lead to more controlled usage patterns, though the difference is not significant enough to draw rigid conclusions.

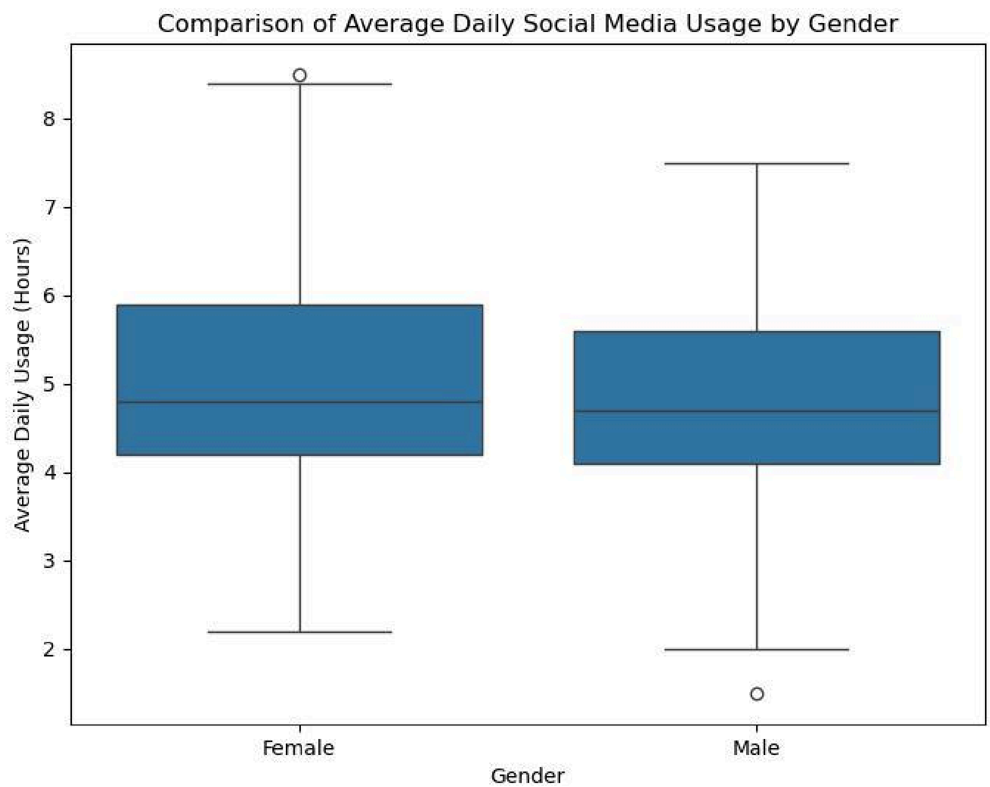

Figure 7: Comparison of Average Daily Social Media Usage by Gender

The median usage is nearly identical for both male and female students, sitting just below 5 hours per day. The IQRs are also of a similar size, and the whiskers extend to similar ranges, indicating that the overall distribution of usage is comparable across genders.

Gender alone does not seem to be a strong predictor of usage in this sample. This suggests that social media habits may be shaped more by personal lifestyle, personality, or platform preference, rather than gender identity, making the issues identified in this report a universal student concern.

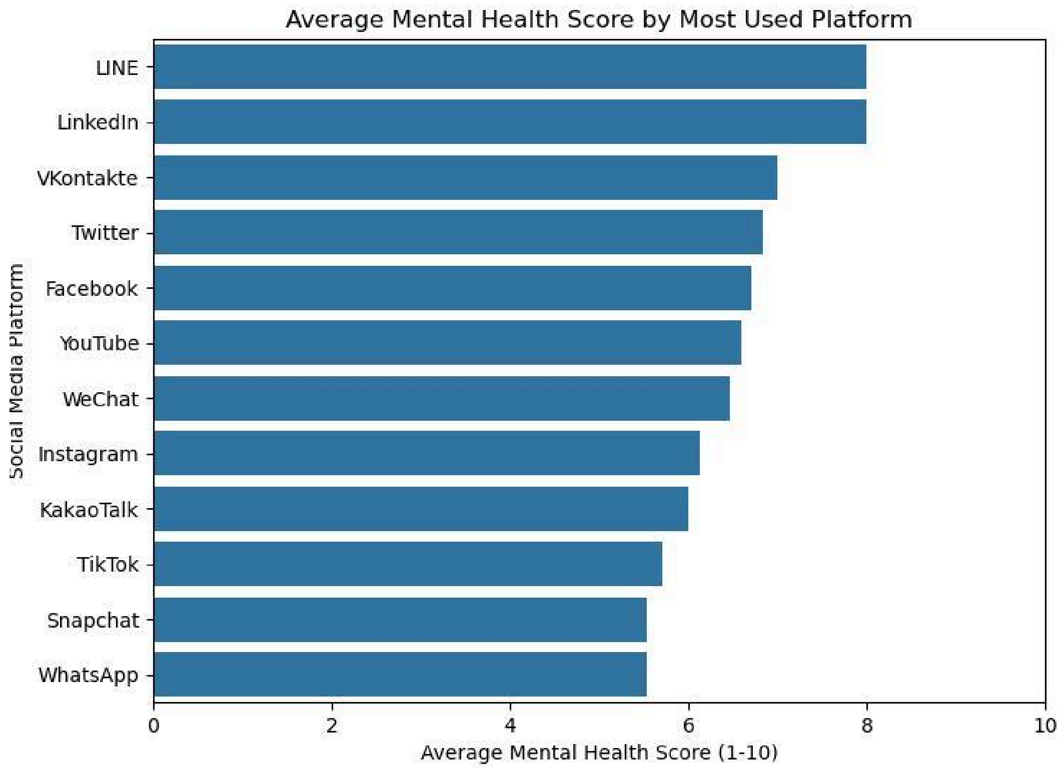

Figure 8: Average Mental Health Score by Most Used Platform
The chart ranks platforms by the average mental health score of their users. Platforms often used for specific functions or professional networking (e.g., LINE, LinkedIn) are associated with higher average mental health scores. Conversely, platforms that are heavily visual and entertainment-driven (e.g., WhatsApp, Snapchat, TikTok) are associated with the lowest average scores.
The type of social media platform seems to matter. It suggests that not all social media is created equal in its relationship with mental well-being. However, since not all platforms have a significant number of data points, these findings should be interpreted with caution.

d) A Multivariate Analysis
To conclude the analysis, a multivariate plot explores the interaction between several key variables simultaneously.

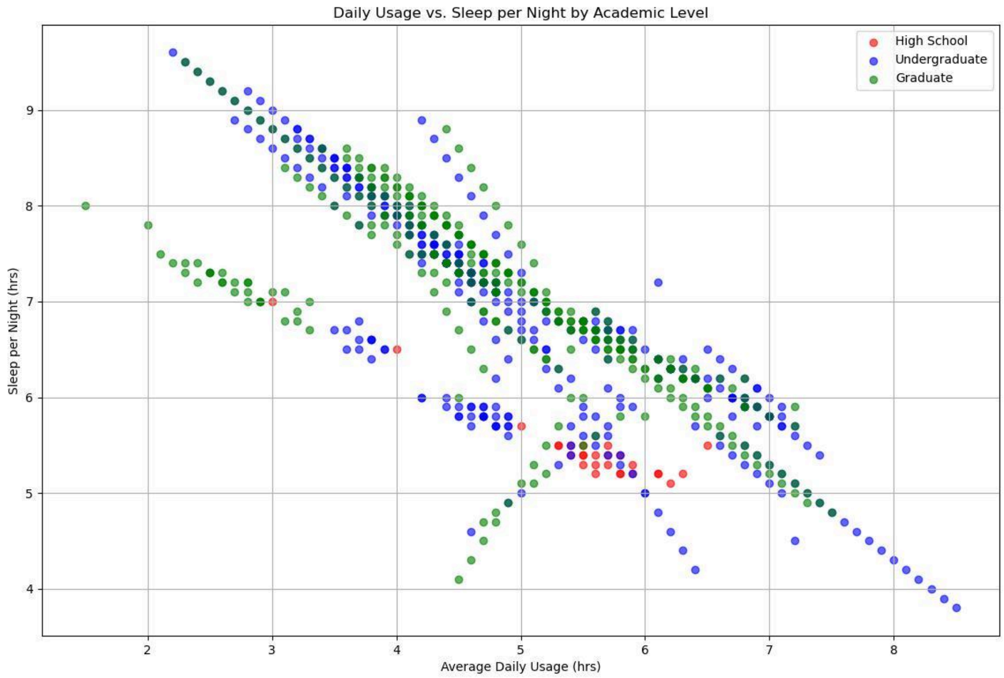

Figure 9: Daily Usage vs. Sleep per Night by Academic Level

This scatterplot demonstrates a strong negative correlation between daily usage hours and sleep duration across all academic levels. As usage increases from 2 to 8+ hours, sleep decreases from approximately 9-10 hours to as little as 4 hours. Undergraduates (blue) show the highest daily usage hours and greatest variability, with some extreme cases of 8+ hours usage and minimal sleep.

This multivariate plot demonstrates the universal nature of the trade-off between social media and sleep. The negative relationship holds true for high schoolers, undergraduates, and graduate students alike. It tells a compelling story that regardless of academic level, spending more time on social media consistently comes at the cost of sleep, a foundational component of both academic success and mental health.

Conclusion

This analysis of 705 students has demonstrated a clear and concerning link between heavy social media usage and negative well-being outcomes. Heavy social media use (averaging nearly 5 hours daily) is the norm for this population, with usage concentrated primarily on Instagram, TikTok, and Facebook. There is an undeniable negative correlation between addiction levels and mental health, and a direct trade-off between usage hours and sleep. Higher usage is also strongly linked to a greater number of interpersonal conflicts. These effects also appear to be universal, holding true across different academic levels and genders. Students who self-report academic harm are also the heaviest users, suggesting a strong connection between behavior and perceived outcomes. Lastly, the type of platform may play a role, with entertainment-focused platforms correlating with poorer self-reported mental health compared to utility-focused ones.

In summary, the data tells a compelling story that while social media is a ubiquitous part of student life, its intensive use presents significant risks to a student's success by impacting mental health, sleep, and academic focus.

References

Shamim, Adil. (2025). Students' Social Media Addiction [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/adilshamim8/social-media-addiction-vs-relationships

Acknowledgements

I would like to express my sincere gratitude to my professor, Dr. Meenakshi Nerolu, for her invaluable guidance and mentorship throughout this project. Her insightful feedback and structured framework were instrumental in shaping the direction and quality of this analysis.

I also wish to extend my appreciation to Adil Shamim for making the "Students' Social Media Addiction" dataset publicly available on Kaggle, without which this research would not have been possible. Furthermore, I am grateful to the Department of Mathematics at Howard University for providing the resources and academic environment necessary to complete this work.

Finally, I would like to thank my peers and classmates for the engaging discussions and collaborative learning environment. I am also grateful to my family and friends for their unwavering support and encouragement throughout my academic endeavors.