Assignment 3: Report - The Unspoken Epidemic - Analysis to Combat the Rise of 'Brain Rot'

Adrian Leong (Student ID: 27030768)

Contents

Feedback from Assignment 1

This section details the incorporation of Assignment 1 feedback regarding the project's novelty.

Feedback: "Novelty: where is your methodology?"

Incorporation: This feedback highlighted the need to articulate the unique methodological approach. The project's novelty lies not just in the problem but in its **integrated, multi-data source, and advanced analytical methodology** for comprehensively tackling 'brain rot' beyond simple correlations. The current demonstration serves as a foundational step for this broader, novel approach, strengthening the project's unique value proposition through a clear methodological framework.

1 Introduction

The world's population is increasingly being online today, especially in the recent post-pandemic years. A concerning trend is arising: short-form, low quality but addictive content, usually on social media, is exploding in popularity (Ortiz-Ospina, 2019). For instance, TikTok has exploded in user base to 1.925 billion users since its launch in 2018 (How Many Users on TikTok? Statistics & Facts (2025), n.d.), Instagram Reels has similarly soared to 2 billion users, 726.8 million of which use Reels, despite Instagram Reels only having launched in August 2020 (Connell, 2025), and Youtube Short's has ballooned to 90 billion viewed videos in 2024 compared to 30 billion in 2021 (Ch, 2025). Moreover, video content has also been shortening in length in general. Social media, where brain rot content is typically found, has grown by an absurd 2.52x in the last 10 years, putting it at 5.24 billion people now - a solid majority of the world (Team, 2025). This phenomenon has come to be dubbed as "brain rot" (Heaton, 2024). While the pandemic lockdowns certainly exacerbated this trend, with increased time spent online globally, it is crucial to recognize that the rise of 'brain rot' content was already underway prior to this period. With how our behaviours shape us, it is vital to understand its impact, as research already shows a multitude of negative effects: shortened attention spans, reinforcement of existing viewpoints, dampened critical thinking, worsened academic anxiety, academic engagement and mindfulness, and causes depression - and this list is non-exhaustive. These are societal-level problems, that left unchecked, will have deep-running implications arising in the future. This project aims to have the joint goals of raising public awareness of the dangers of brain rot addiction, and discovering what works to mitigate this to help people live healthier lives.

2 Related Work

Naturally, as this is a very recent phenomenon, long-term effects of brain rot content cannot be studied just yet. Research shows that brain rot consumption shorten people's attention spans, reinforce existing viewpoints and dampens critical thinking (Kim, 2024). Another study on students (Li et al., 2024) found that brain rot content significantly affects student academic anxiety, academic engagement, and mindfulness for the worse. Li et al. (2024) also notes that practicing mindfulness mediates the effects on academic anxiety, thus showing an answer as to dampen the negative effects of brain rot addiction. Yet another study (Qu et al., 2023) shows that brain rot content addiction causes depression, which is known to cause many other negative effects. The combined results of these research highlights the significant negative impact of the consumption of brain rot content, and what to do about it. While there are many blogs and articles offering advice on how to deal with "brain rot", for example Curtis (2025) or Boys & Girls Clubs of America (2025), there is a lack of hard research to validate these. This project distinguishes itself by quantifying how bad the effects are and what helps with data.

3 Business Model

The project evaluates how ‘brain rot’ addiction impacts academic performance, quantifies its negative effects, and identifies effective mitigation strategies. It will have the joint goals of raising public awareness of the dangers of brain rot addiction, and discovering what works to mitigate this to help people live healthier lives. This project will be put best to use in educational and mental health institutions, and is also of great benefit to governments and society as a whole, to help implement mitigation strategies. The primary stakeholders who will benefit from this project are:

1. Educators
Brain rot affects **educational outcomes** through **reduced focus** and **increased anxiety** in students (Li et al., 2024). By understanding its roots, educators can tailor interventions to improve student performance at its fundamental cause.

2. Mental health practitioners
Brain rot excarbates **depression**, alongside other mental health problems (Qu et al., 2023), making it more difficult for clients to engage in healthier living. Through a better understanding of this phenomenon, counsellors and therapists will be able to better guide their clients towards healthier lives with improved well-being.

3. Government regulators
Brain rot poses a significant detriment to society, contributing to **reduced productivity** (Kim,2024) and **increased mental health costs** (Qu et al., 2023) at a societal level, thereby hindering both economic and technological progress. Providing concrete evidence of these impacts is crucial for driving government action. With compelling data, governments globally can be urged to recognize the urgency of combating brain rot and implement effective measures, such as regulations on social media/content creators and public mental health campaigns to promote awareness.

4. Parents
The developmental impact of brain rot on children and adolescents is a serious concern (Kim,2024). This project will empower parents to take proactive steps to shield their children from the addictive nature of this content, fostering healthier growth. Furthermore, it will provide guidance for parents seeking to mitigate the negative effects of brain rot in situations where children are already exposed (Boys & Girls Clubs of America, 2025).

5. Society as a whole
The general population will have better resources to better their well-being and productivity.

6. Research
This contributes to society’s knowledge of the impact of our behaviours on our lives.

4 Characterising and Analysing Data

4.1 Potential Data Sources and Characteristics

Addressing ‘brain rot’ comprehensively requires diverse data, from individual experiences to macro trends.

- **Primary Data (Surveys):** Self-reported data on social media usage, academic performance, sleep, and mental health (as used in this demonstration). Questions adapted from validated scales (e.g., Bergen Social Media Addiction Scale).
  - **Pros:** Directly addresses specific research questions; captures subjective experiences.
  - **Cons:** Prone to self-report bias, recall issues, social desirability bias; limited for objective behavior or long-term trends.
- **Secondary Data (for Future Expansion & Broader Trends):** Mobile engagement trends (Shorts/Reels), reading trends (average book length), Wikipedia dwell time data, neuroscience data, CommonCrawl/Google Books Ngrams Viewer, and academic score trends. These data types offer objective evidence and broader trend analysis.
- **Data Characteristics (The 4 V’s):** The characteristics of big data are commonly described by the “4 V’s”:
  - **Volume:** Current project is small. Future expansion involves **petabytes**, requiring scalable storage.
  - **Variety:** Current data is structured. Future expansion introduces **high variety**: structured, semi-structured, and unstructured (text, video, neuroimages).
  - **Velocity:** Current data is static. Future social media engagement data would be **high velocity**, necessitating real-time processing.
  - **Veracity:** Current self-reported data is subject to biases. Future data from APIs or web crawls may contain noise; rigorous validation and cleaning are critical (Marr, n.d.).
- **Platforms, Software, and Tools:**
  - **Current Project:** R (analysis, modeling), RStudio, local storage, CSV.
  - **Future Expansion:** Cloud object storage (AWS S3, Google Cloud Storage), NoSQL databases (MongoDB), Data Warehouses (Snowflake, BigQuery) for storage. Distributed computing frameworks (Apache Spark, Hadoop); cloud services (AWS Glue, Google Dataflow) for processing. Python (advanced ML, NLP, deep learning), specialized visualization (Tableau), workflow orchestration (Apache Airflow) for tools. These provide scalability, flexibility, and robust processing for diverse, large-scale, high-velocity data.

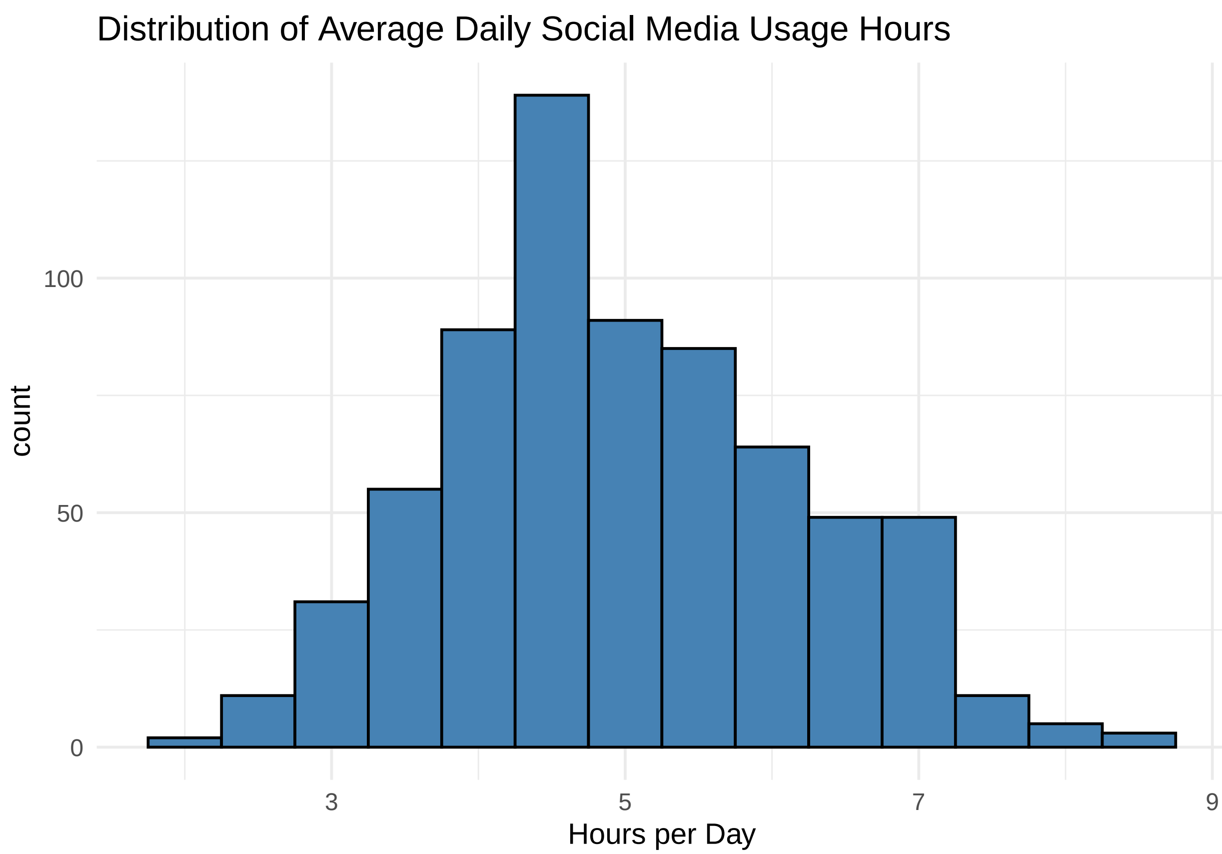
4.2 Data Analysis Techniques and Statistical Methods

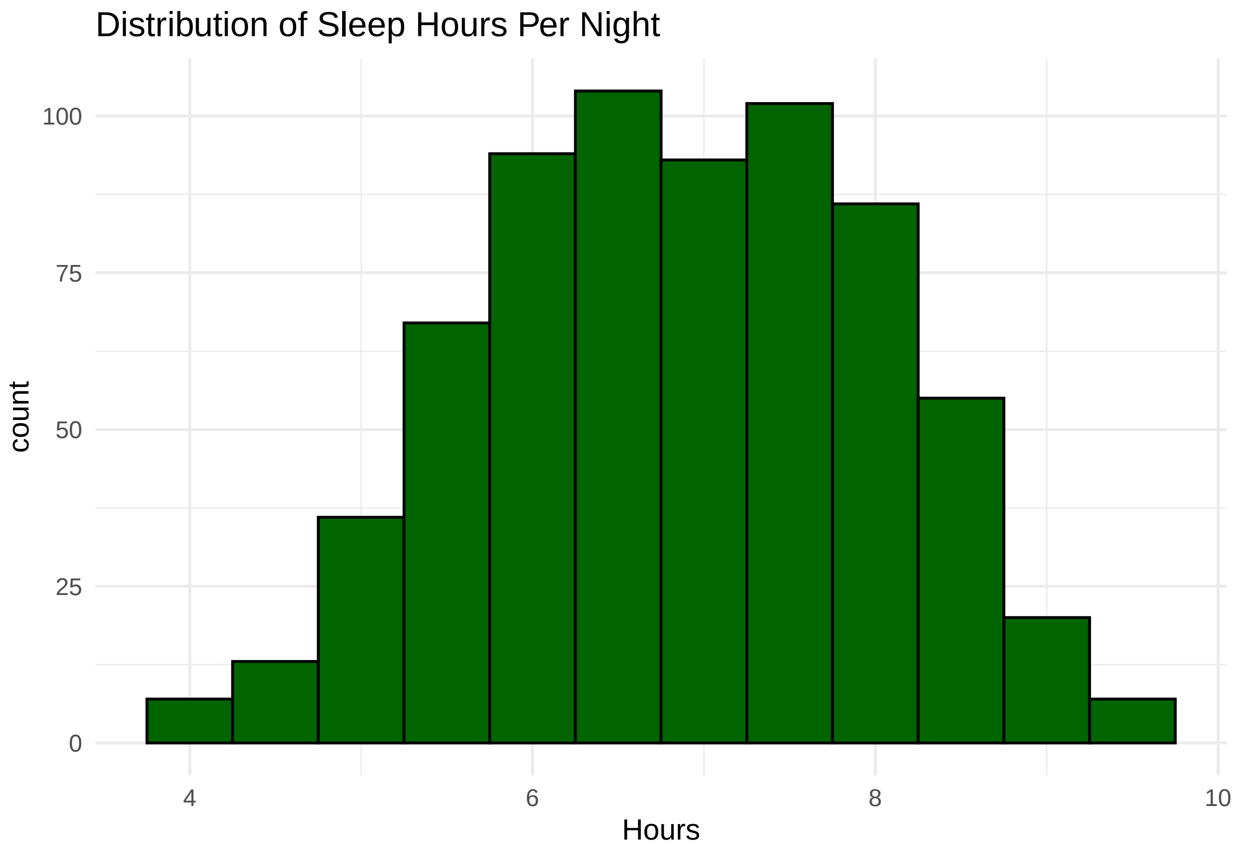
- **Descriptive Statistics & Visualization:** Methods include means, medians, distributions, time-series plots. Rationale: Summarize data, identify patterns and trends (e.g., usage, book length changes). Expected Outcomes: Baseline understanding of current patterns.
- **Inferential Statistics:** Methods such as T-tests, ANOVA (comparing group means), Chi-squared tests (categorical associations), Multiple Regression Analysis, Confidence Interval Tests, and Variance Inflation Factor (VIF) (Selection of Appropriate Statistical Methods for Data Analysis, n.d.). Rationale: Draw statistically sound conclusions, testing hypotheses. Multiple regression assesses simultaneous impact of multiple predictors, VIF diagnoses multicollinearity, and confidence intervals provide plausible ranges for population parameters. Expected Outcomes: Confirm significant differences or associations; understand independent effects of variables; ensure model robustness.
- **Time-Series Analysis:** Methods include ARIMA models, Prophet. Rationale: Analyze trends over time (e.g., usage, book lengths) for 'brain rot' onset and progression. Expected Outcomes: Detection of temporal patterns, forecasting.
- **Natural Language Processing (NLP):** Methods like N-gram analysis, readability scores (e.g., Flesch-Kincaid) (DataCamp, n.d.). Rationale: Quantitatively assess linguistic simplification in text corpora. Expected Outcomes: Statistical evidence of trends in linguistic complexity.
- **Machine Learning for Prediction:** Methods including Logistic Regression, Linear Regression, Decision Trees/Random Forests, Gradient Boosting Machines (Machine Learning for Social Science, n.d.). Rationale: Build predictive models to identify 'brain rot' drivers and outcomes. Expected Outcomes: Predict academic impact, sleep, mental health; identify influential factors.
- **Clustering (Unsupervised Learning):** Methods such as K-Means, Hierarchical Clustering (Machine Learning for Social Science, n.d.). Rationale: Identify natural groupings in student populations based on behavior. Expected Outcomes: Discover student profiles or 'brain rot' archetypes.

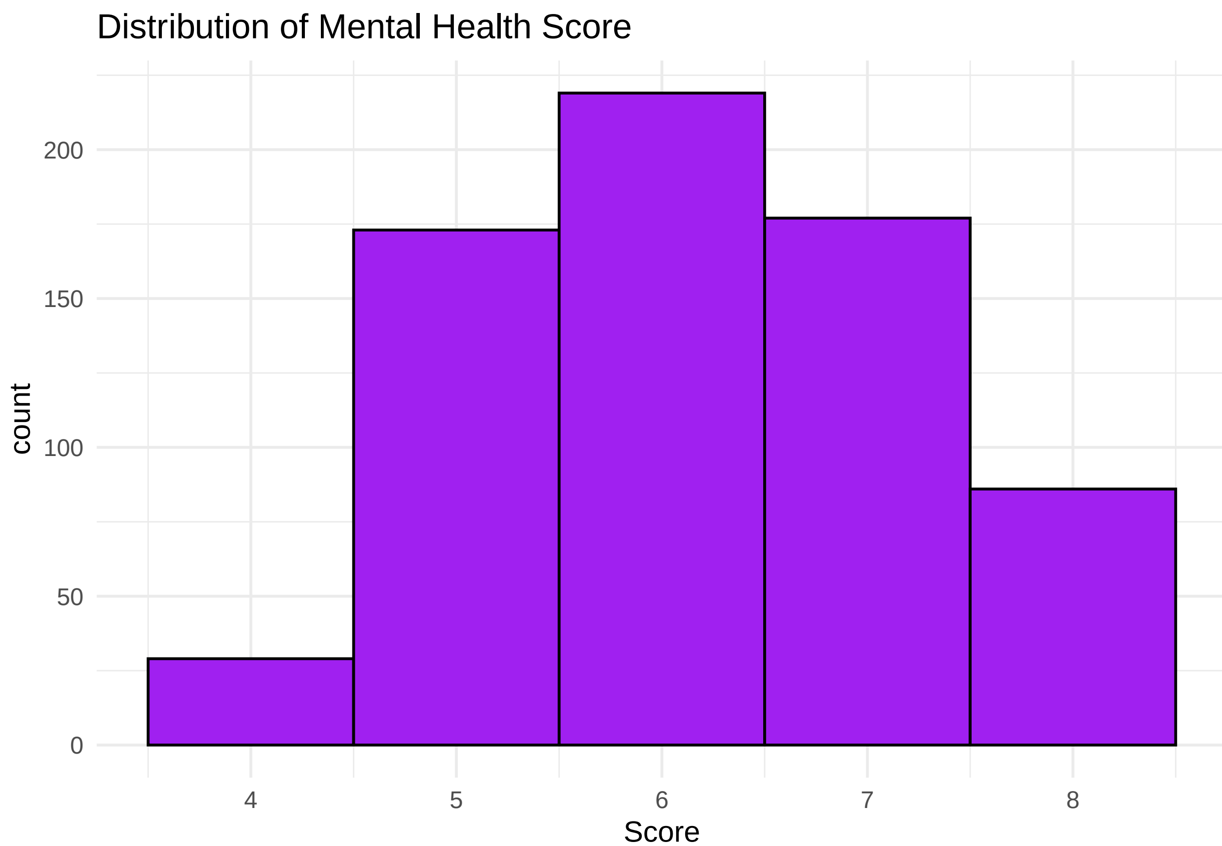
4.3 Demonstration

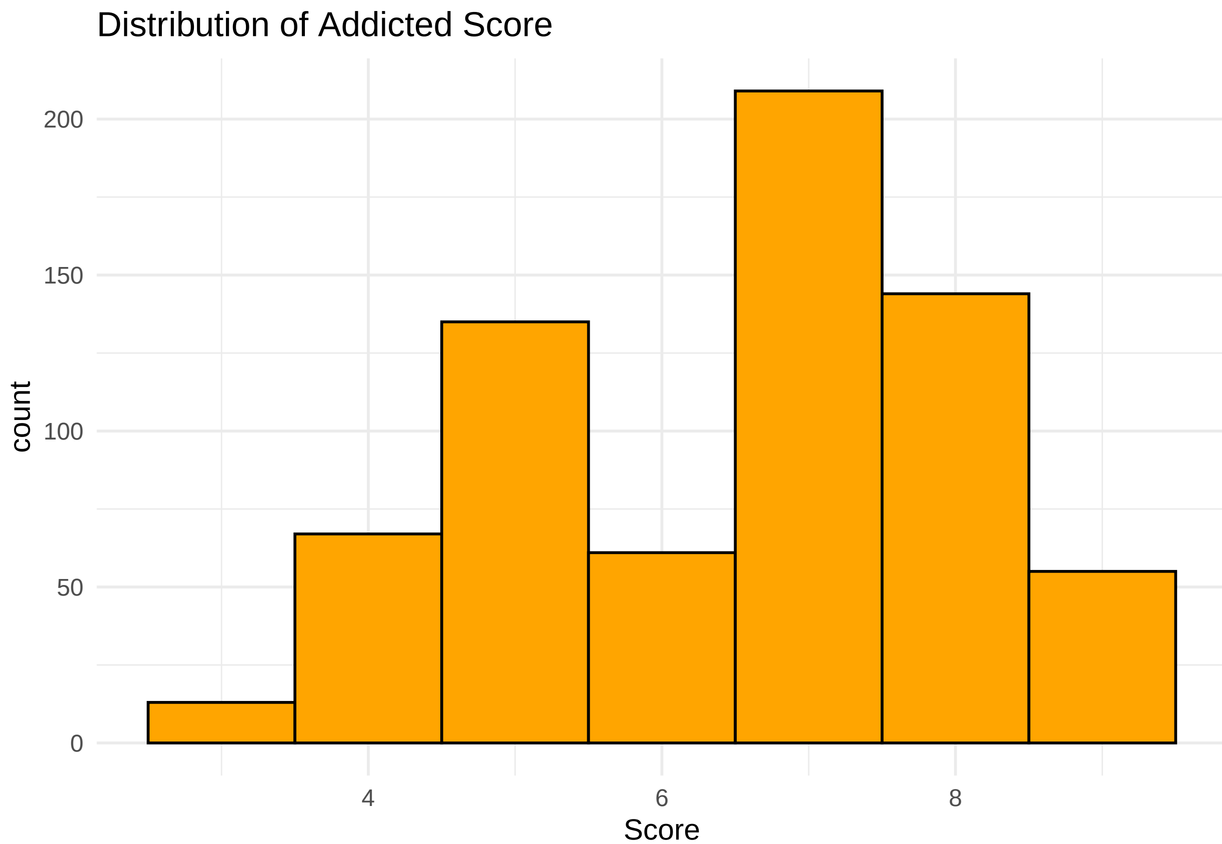
- **Dataset Identification:** A "Social Media Addiction and Mental Health" dataset from Kaggle was used. It was collected via surveys (university mailing lists, social media) with validation, de-duplication, and anonymization.
  - **Download Link:** https://www.kaggle.com/datasets/adilshamim8/social-media-addiction-vs-relationships
- **Data Description & Features:** 645 observations, 15 variables. Key features include `Age`, `Gender`, `Academic_Level`, `Avg_Daily_Usage_Hours`, `Most_Used_Platform`, `Sleep_Hours_Per_Night`, `Mental_Health_Score`, `Affects_Academic_Performance`, `Conflicts_Over_Social_Media`, `Addicted_Score`, `Relationship_Status`, `Country`.

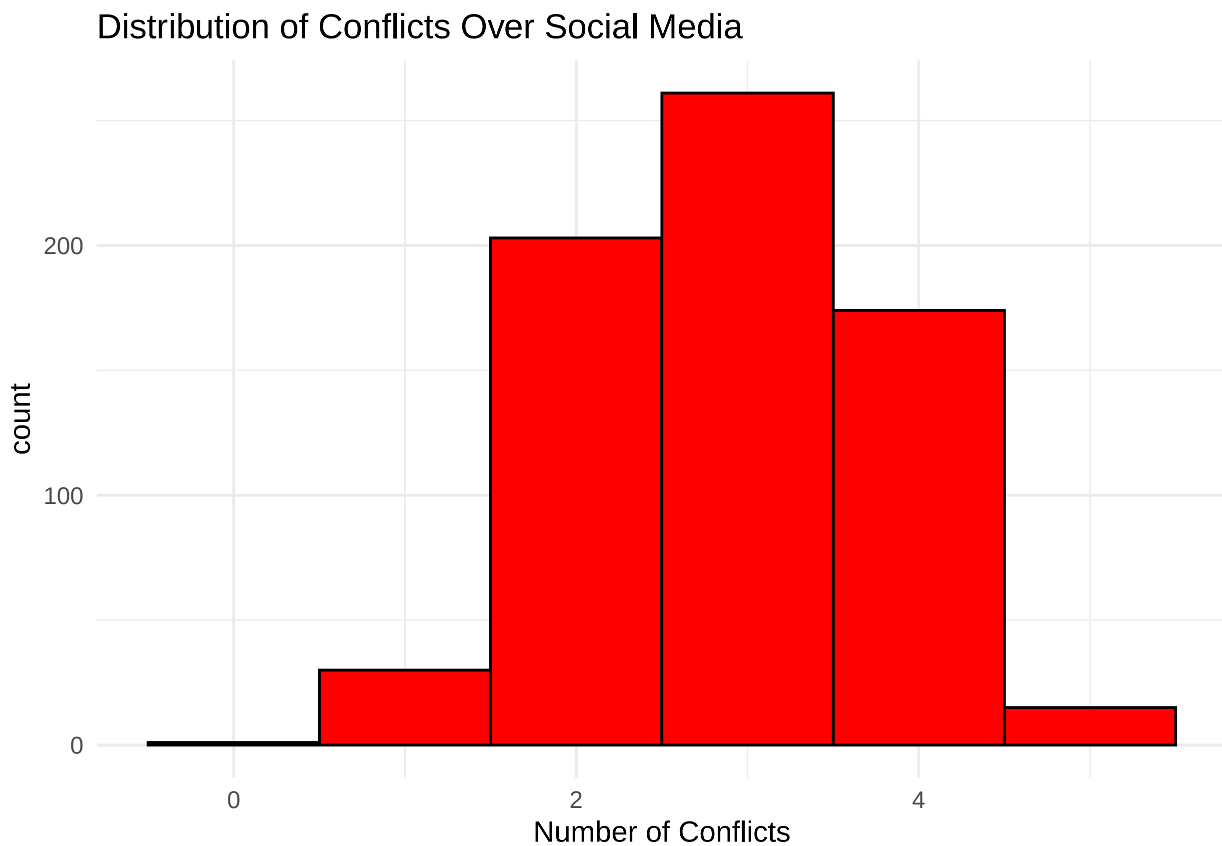
Descriptive Statistics and Distribution Visualizations




Distribution of Sleep Hours Per Night
100
75
50
25
0
count
4
6
8
10
Hours

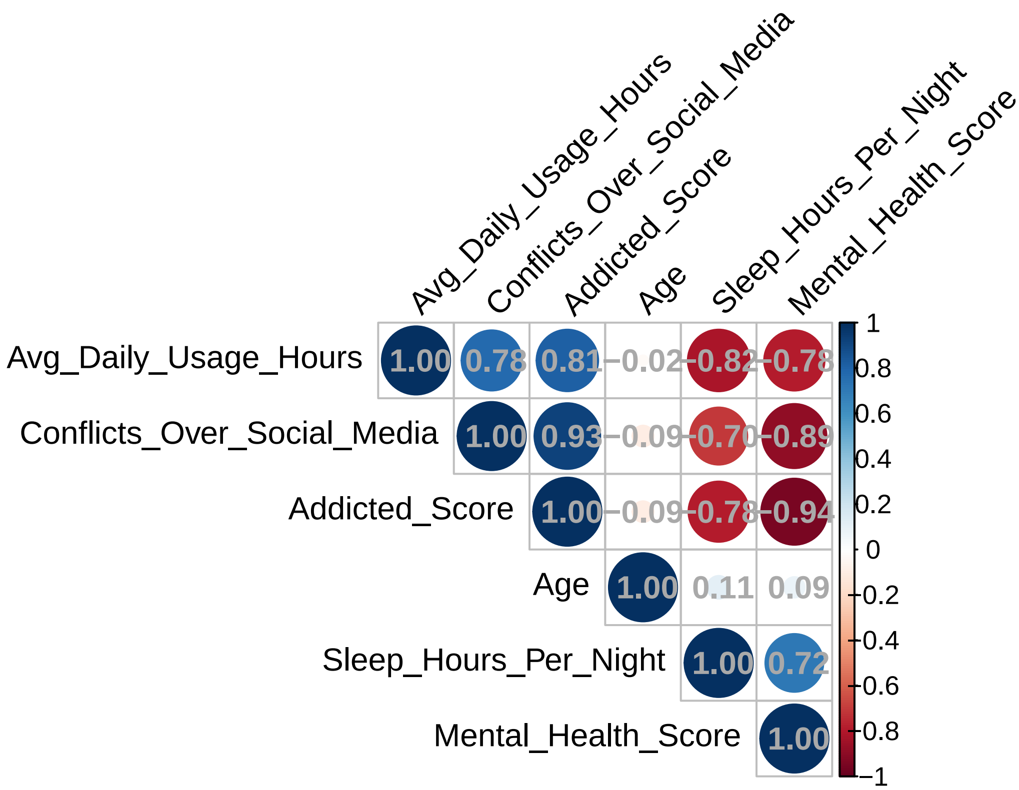

Distribution of Mental Health Score
200
150
100
50
0
count
4
5
6
7
8
Score


Distribution of Addicted Score
200
150
100
50
0
count
4
6
8
Score

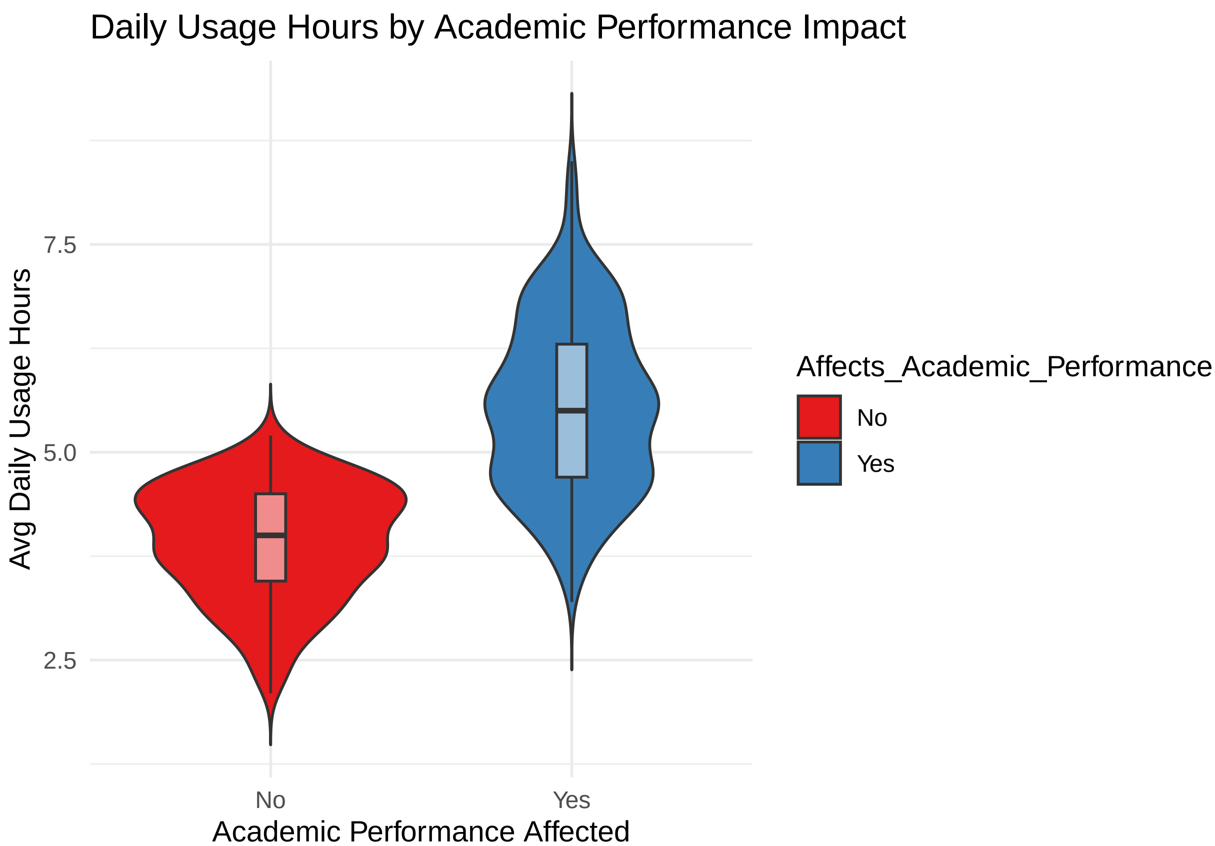

Distribution of Conflicts Over Social Media
200
100
0
count
0
2
4
Number of Conflicts

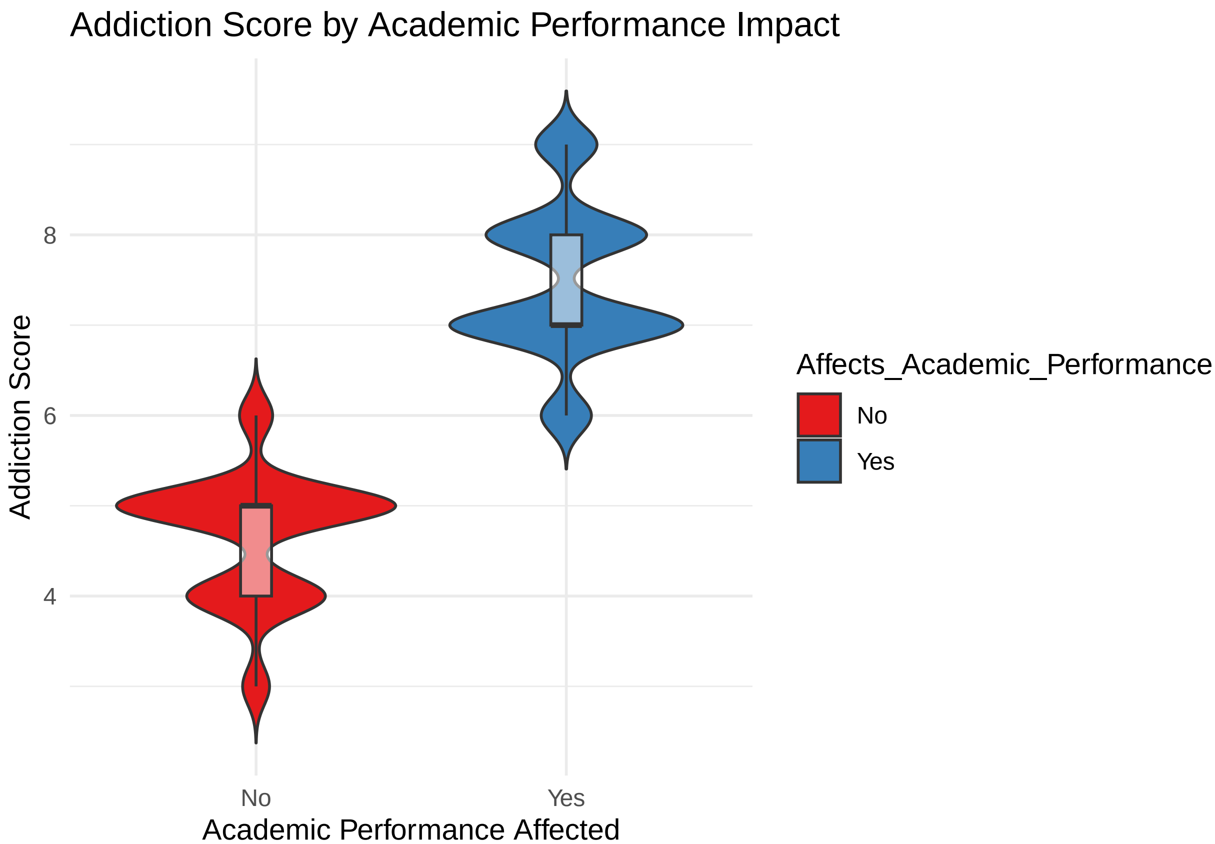
- **Analysis Process (Using R): Initial Inspection:** Loaded data (`str()`, `summary()`, `head()`). **Cleaning & Preprocessing:** `na.omit()` removed missing data. Variables converted to factors. “LinkedIn” users filtered out. `Most_Used_Platform` grouped, with “Facebook” as reference. **Exploratory Data Analysis (EDA):** Visualizations revealed initial patterns. These box plots are enhanced using hybrid box-violin plots for richer data distribution insights:

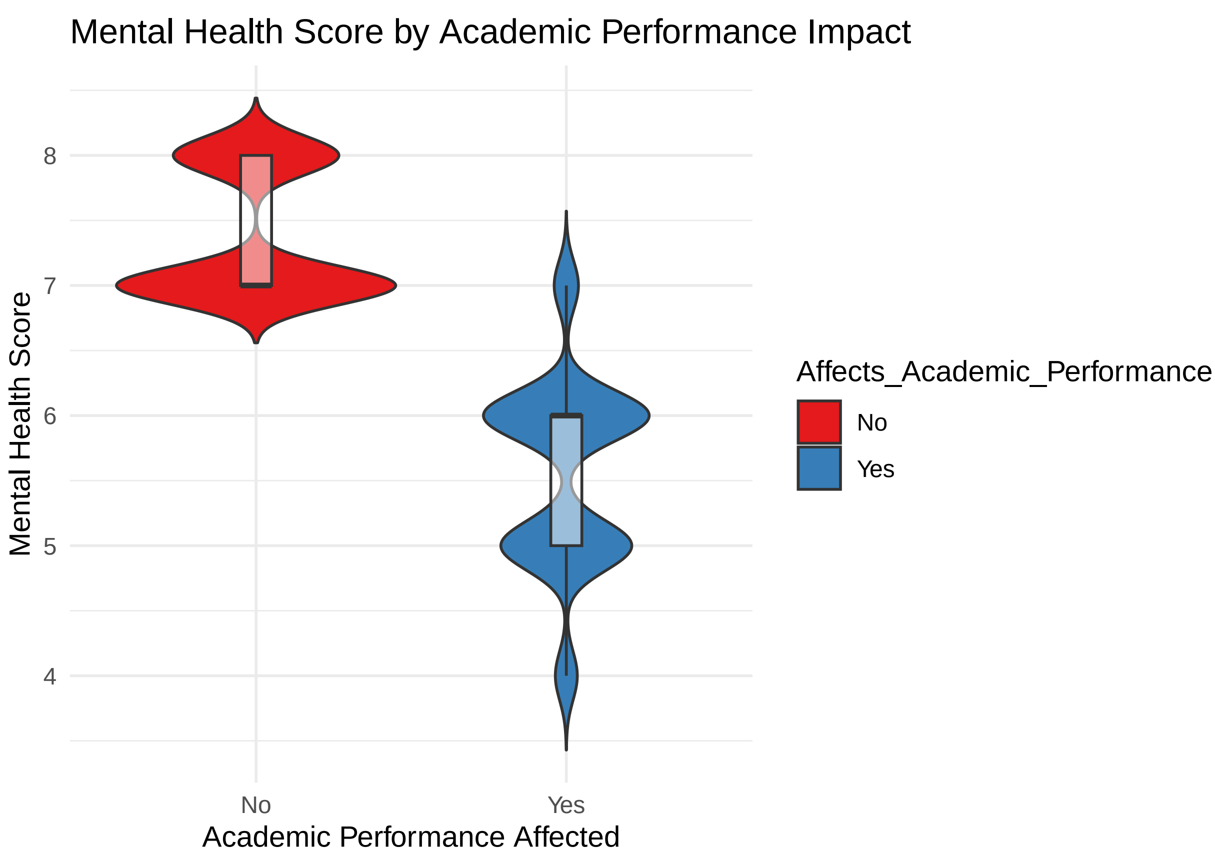
Correlation matrix

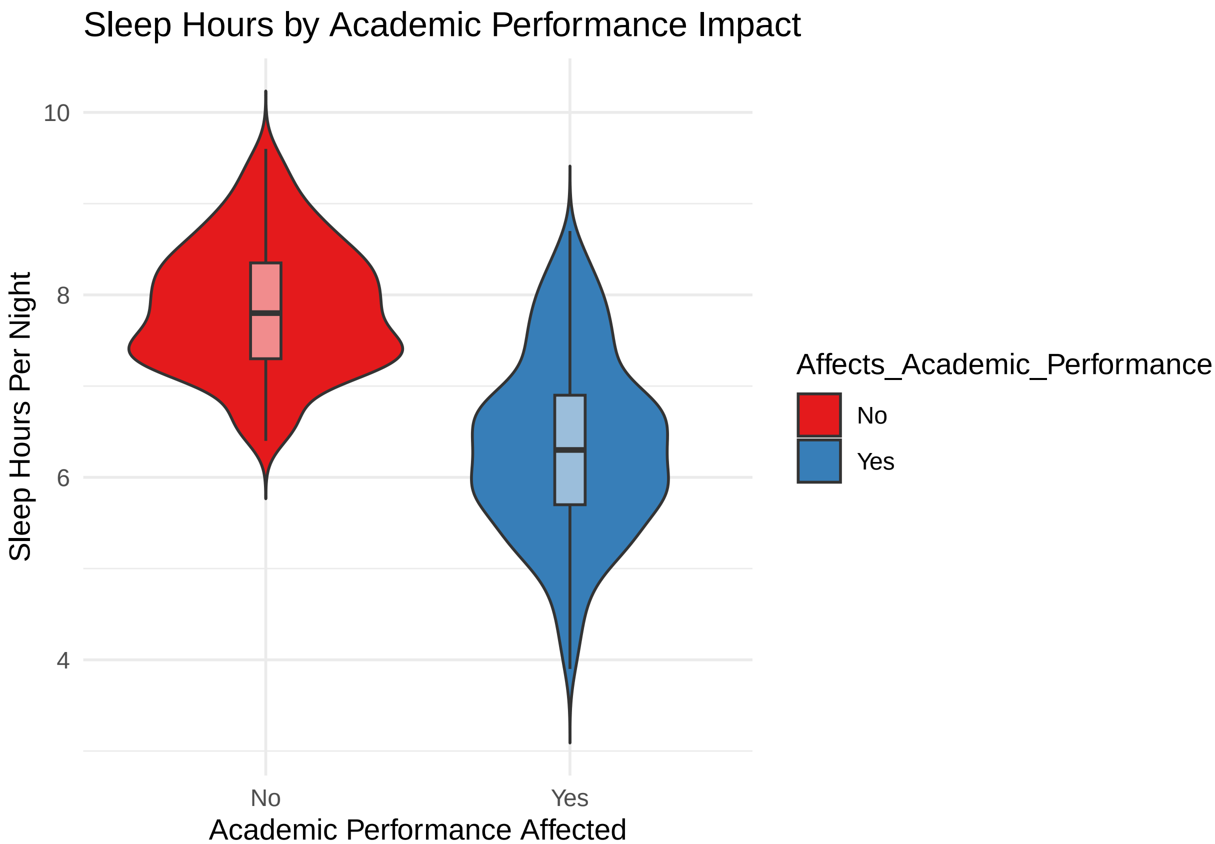


Hybrid box-violin plots for numeric variables by Affects_Academic_Performance

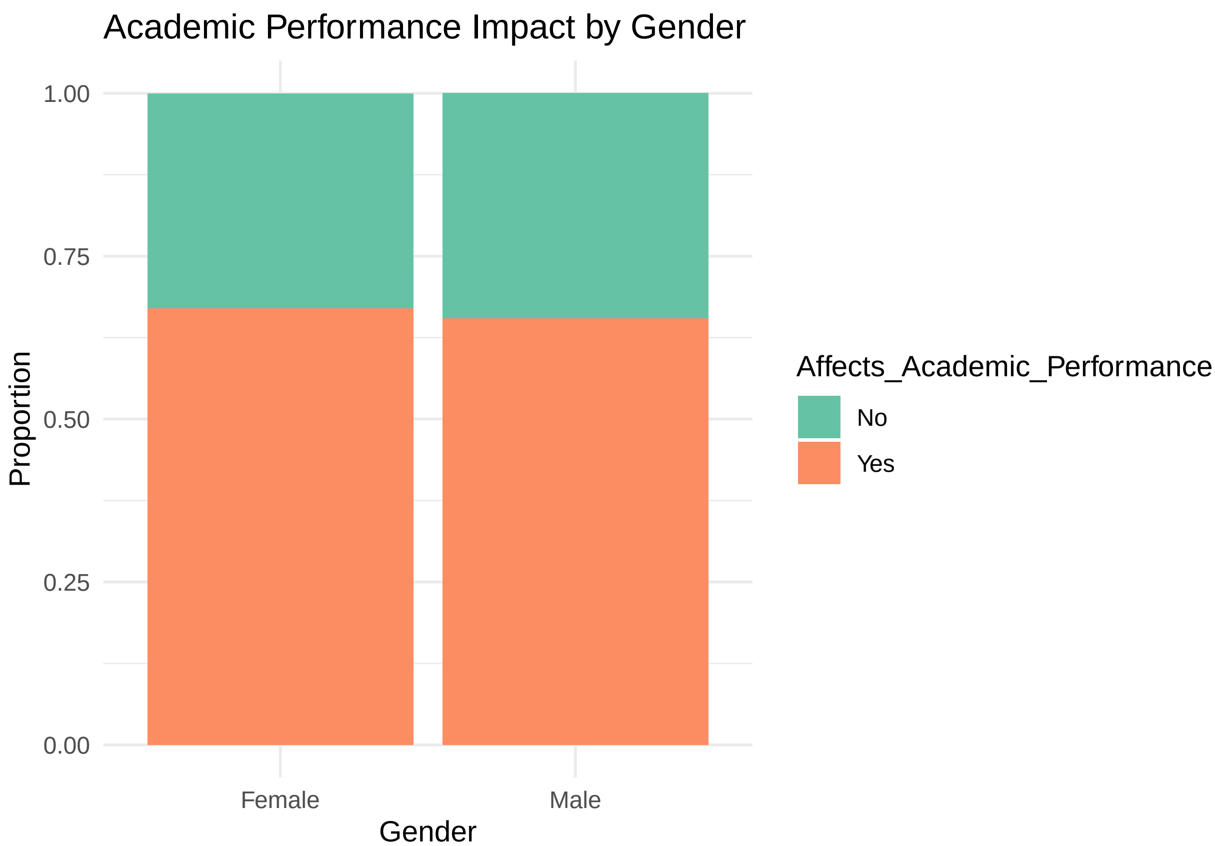


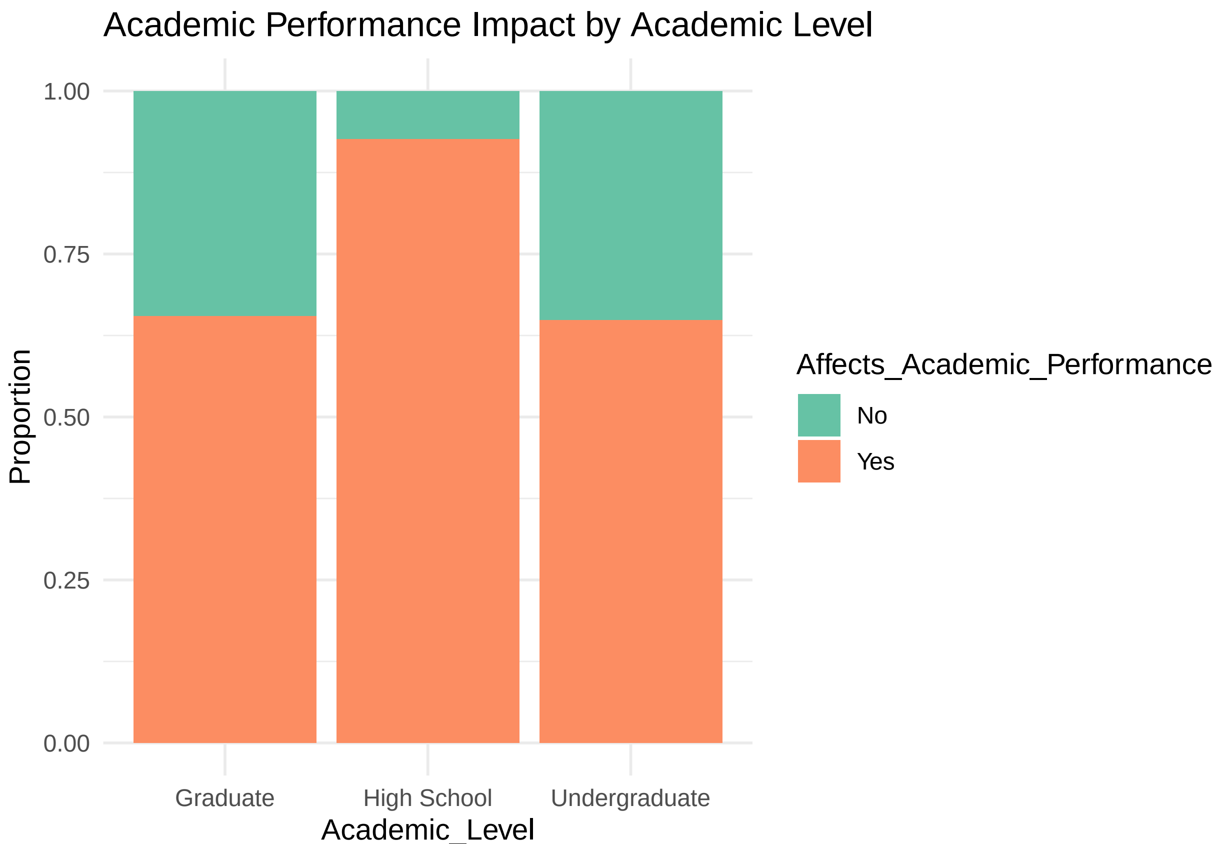

Addiction Score by Academic Performance Impact
Addiction Score
8
6
4
No
Yes
Academic Performance Affected
Affects_Academic_Performance
No
Yes

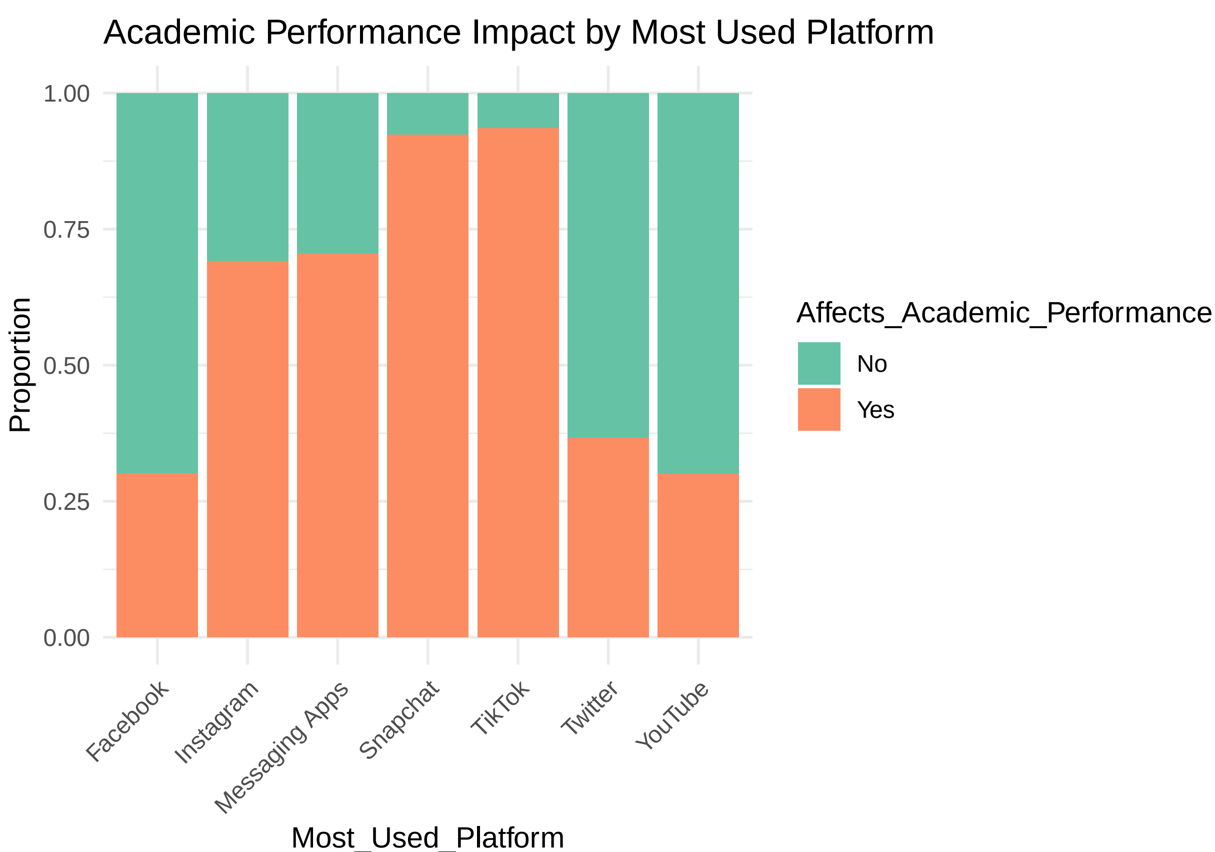

Mental Health Score by Academic Performance Impact
Mental Health Score
8
7
6
5
4
No
Yes
Academic Performance Affected
Affects_Academic_Performance
No
Yes


Sleep Hours by Academic Performance Impact
Sleep Hours Per Night
10
8
6
4
No
Yes
Academic Performance Affected
Affects_Academic_Performance
No
Yes

Univariate Analysis for Categorical Variables

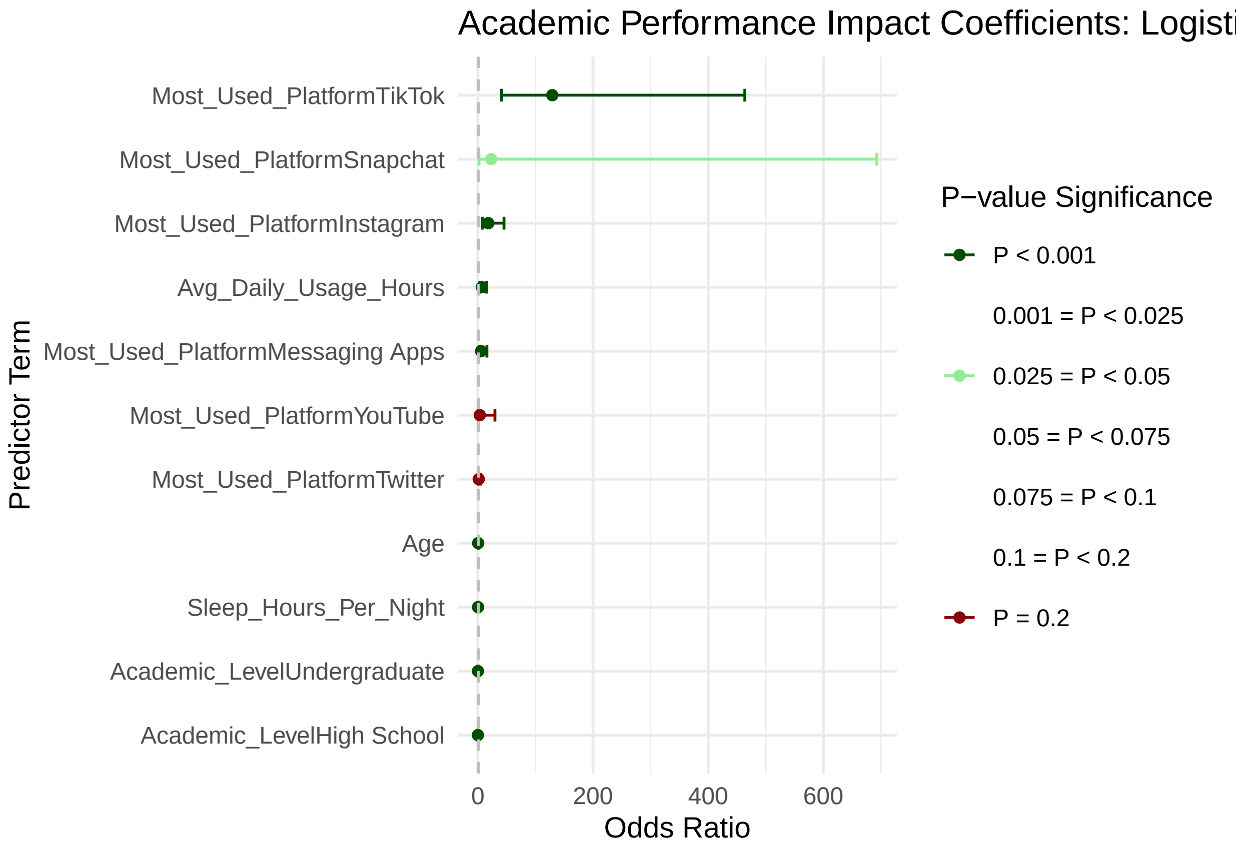


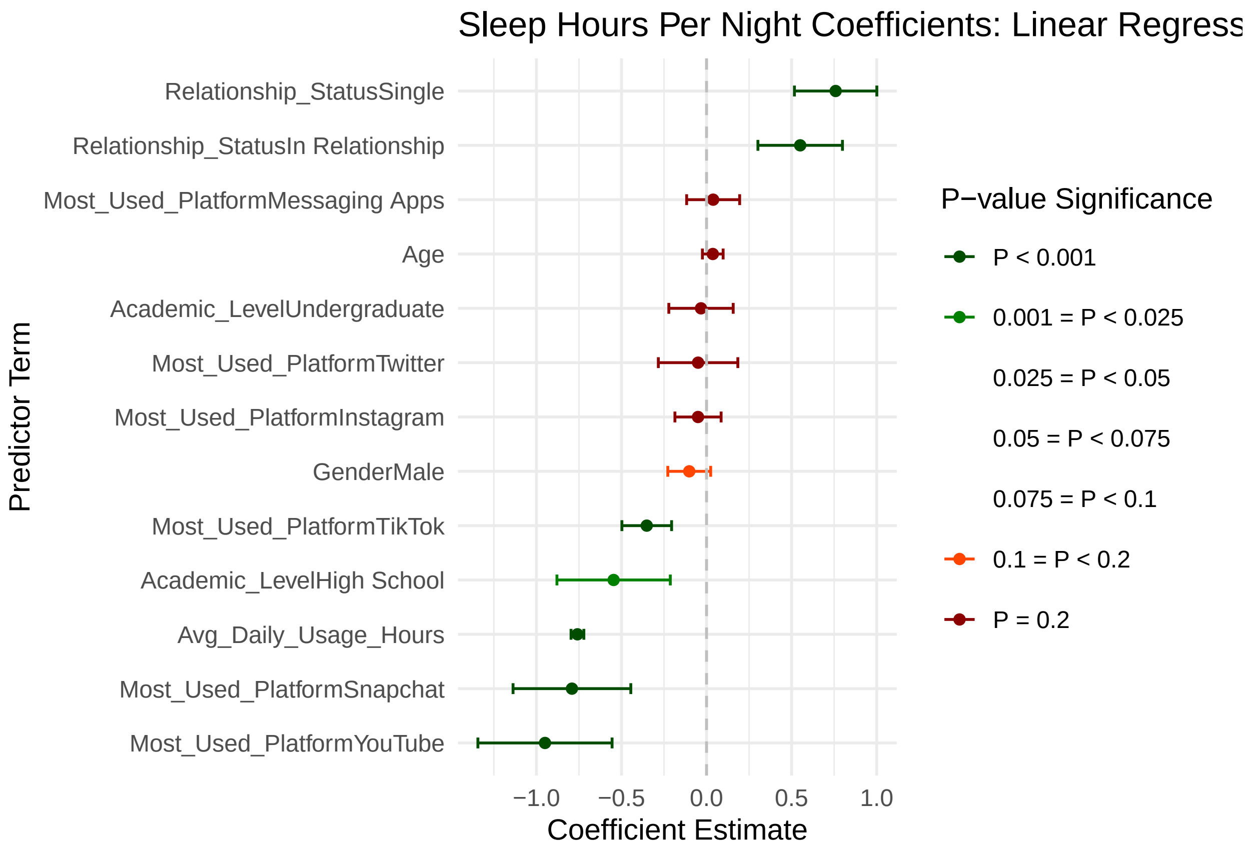

Academic Performance Impact by Academic Level
1.00
0.75
0.50
0.25
0.00
Proportion
Graduate
High School
Undergraduate
Academic_Level
Affects_Academic_Performance
No
Yes


Academic Performance Impact by Most Used Platform
1.00
0.75
0.50
0.25
0.00
Proportion
Facebook
Instagram
Messaging Apps
Snapchat
TikTok
Twitter
YouTube
Most_Used_Platform
Affects_Academic_Performance
No
Yes

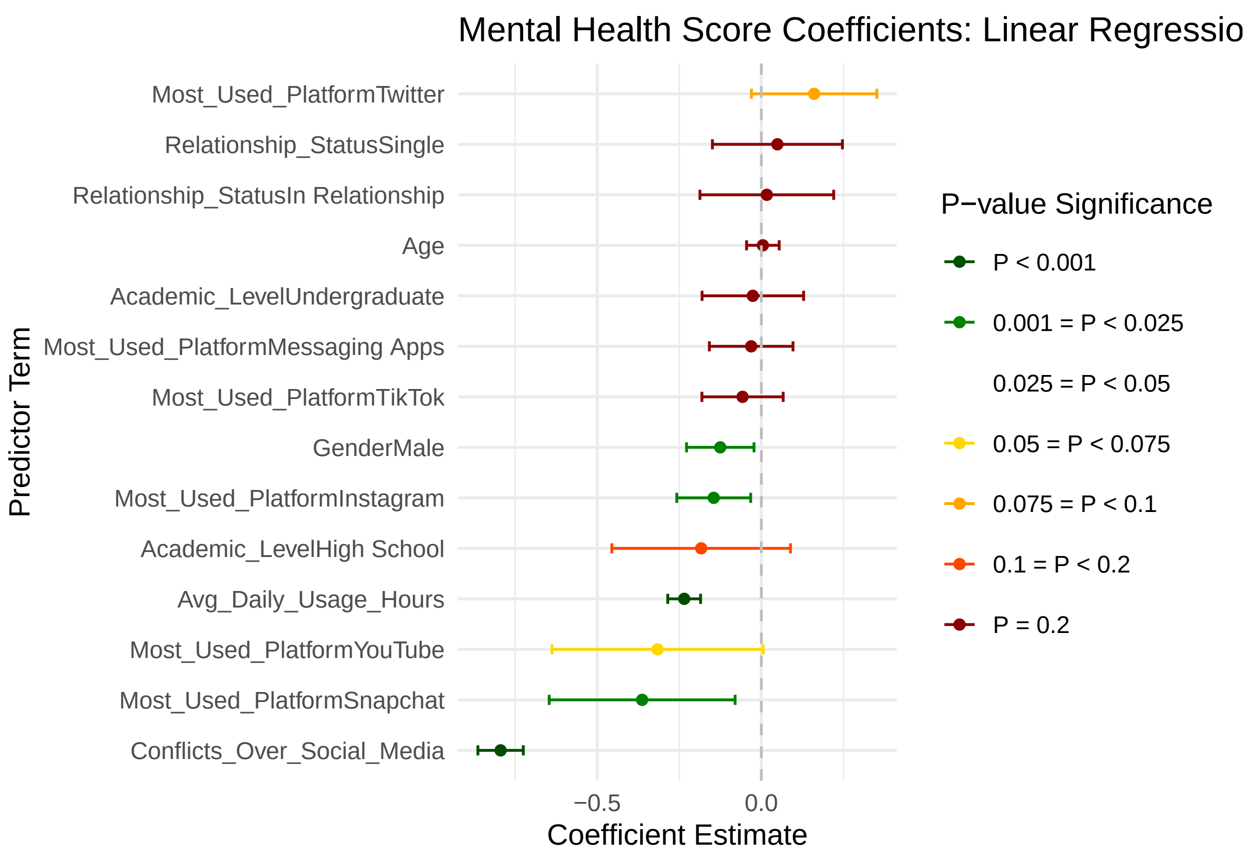
Statistical Modeling: Three regression models were built: Logistic Regression for `Affects_Academic_Performance`, Linear Regression for `Sleep_Hours_Per_Night` and `Mental_Health_Score`. Predictors included `Avg_Daily_Usage_Hours`, `Academic_Level`, `Most_Used_Platform_Grouped`, `Age`, `Gender`, `Relationship_Status`, plus `Conflicts_Over_Social_Media` for mental health. VIFs confirmed no problematic multicollinearity. The regression analysis included various predictors to predict academic performance impact, sleep hours, and mental health scores. The coefficients and their statistical significance for these models can be effectively visualized using coefficient plots with a sequential color scale based on p-value:

Binary Logistic/Linear Regression Analysis





Sleep Hours Per Night Coefficients: Linear Regress
Predictor Term
Relationship_StatusSingle
Relationship_StatusIn Relationship
Most_Used_PlatformMessaging Apps
Age
Academic_LevelUndergraduate
Most_Used_PlatformTwitter
Most_Used_PlatformInstagram
GenderMale
Most_Used_PlatformTikTok
Academic_LevelHigh School
Avg_Daily_Usage_Hours
Most_Used_PlatformSnapchat
Most_Used_PlatformYouTube
−1.0
−0.5
0.0
0.5
1.0
Coefficient Estimate
P−value Significance
P < 0.001
0.001 = P < 0.025
0.025 = P < 0.05
0.05 = P < 0.075
0.075 = P < 0.1
0.1 = P < 0.2
P = 0.2



Mental Health Score Coefficients: Linear Regressio
Predictor Term
Most_Used_PlatformTwitter
Relationship_StatusSingle
Relationship_StatusIn Relationship
Age
Academic_LevelUndergraduate
Most_Used_PlatformMessaging Apps
Most_Used_PlatformTikTok
GenderMale
Most_Used_PlatformInstagram
Academic_LevelHigh School
Avg_Daily_Usage_Hours
Most_Used_PlatformYouTube
Most_Used_PlatformSnapchat
Conflicts_Over_Social_Media
−0.5
0.0
Coefficient Estimate
P−value Significance
P < 0.001
0.001 = P < 0.025
0.025 = P < 0.05
0.05 = P < 0.075
0.075 = P < 0.1
0.1 = P < 0.2
P = 0.2

- **Analysis Results:** Models showed strong predictive power ($p < 2.2e\text{-}16$) and low VIFs (all $\text{VIF}^{1/(2*\text{Df})} < 2.0$).
- **Feasibility Conclusion:** The demonstration identified significant relationships between social media usage, platform choices, and student well-being. Findings highlight that while some platforms are strongly associated with negative outcomes, others present counter-trends. This confirms project feasibility and underscores the importance of detailed, platform-specific analysis.

5 Standard for Data Science Process, Data Governance and Management

5.1 Standard for Data Science Process

The **Cross-Industry Standard Process for Data Mining (CRISP-DM)** was adopted for its structured, iterative approach (Shimaoka et al., 2024).

- **Business Understanding:** Defined 'brain rot' problem and goals.
- **Data Understanding:** Acquired and explored dataset, identifying patterns and quality.
- **Data Preparation:** Cleaned data, handled missing values, converted types, filtered, and engineered features.
- **Modeling:** Selected and applied regression models, performed diagnostic checks (VIF).
- **Evaluation:** Assessed model performance and interpreted findings.
- **Deployment:** Disseminated findings via this report and R Markdown file.

CRISP-DM ensures robustness and reliability.

5.2 Data Governance and Management

- **Data Accessibility:**
  - **Current Project:** Uses a public, anonymized Kaggle dataset, promoting transparency.
  - **Future Expansion:** New primary/sensitive data would have strictly controlled access, limited to authorized personnel via secure platforms with ethical approvals.
- **Data Security & Confidentiality:**
  - **Current Project:** Minimal direct risk due to anonymized public data.
  - **Future Expansion:** For sensitive data, measures include: anonymization/pseudonymization, encryption (at rest/in transit), role-based access, secure compliant storage, and data minimization.
- **Ethical Concerns Related to Data Usage:**
  - **Self-Reported Bias:** Acknowledged limitation of the current dataset (cross-sectional, online recruitment bias).
  - **Privacy & Confidentiality (Future Data):** Crucial to obtain explicit informed consent for new individual-level data.
  - **Potential for Misinterpretation/Stigmatization:** Findings must be presented with nuance, emphasizing correlation over causation to avoid unfair generalizations. Statistical associations are not causal claims.
  - **Responsible Reporting:** Findings communicated responsibly, highlighting limitations and actionable insights.

Adherence ensures robust insights and responsible data handling.

6 References

Students' social media addiction. (2025, May 10). Kaggle. https://www.kaggle.com/datasets/adilshamim8/social-media-addiction-vs-relationships

Boys & Girls Clubs of America. (2025, February 26). *Supporting Digital Well-being: 12 Ways to Help Teens Unplug from Technology*. https://www.bgca.org/news-stories/2025/February/unplugging/

Ch, D. (2025, January 8). *YouTube shorts statistics*. SendShort - Create Viral Shorts Instantly With SendShort. https://sendshort.ai/statistics/shorts/

Connell, A. (2025, January 1). *32 Top Instagram Reels Statistics for 2025*. Adam Connell. https://adamconnell.me/instagram-reels-statistics/

Curtis, L. (2025, January 6). *12 habits to Prevent "Brain Rot" Health*. https://www.health.com/habits-to-prevent-brain-rot-8766150

DataCamp. (n.d.). *What is Natural Language Processing (NLP)? A Beginner's Guide*. Retrieved May 25, 2025, from https://www.datacamp.com/blog/what-is-natural-language-processing

Heaton, B. (2024, December 2). *'Brain rot' named Oxford Word of the Year 2024*. Oxford University Press. https://corp.oup.com/news/brain-rot-named-oxford-word-of-the-year-2024/

Kim, I. (2024). EXPLORING THE COGNITIVE AND SOCIAL EFFECTS OF TIKTOK ON ADOLESCENT MINDS: a STUDY OF SHORT-FORM VIDEO CONSUMPTION. *ierj.in*. https://doi.org/10.21276/IERJ24769489007345

Li, G., Geng, Y., & Wu, T. (2024). Effects of short-form video app addiction on academic anxiety and academic engagement: The mediating role of mindfulness. *Frontiers in Psychology, 15*. https://doi.org/10.3389/fpsyg.2024.1428813

Machine Learning for Social Science. (n.d.). *Researcher reasoning meets computational capacity: Machine learning for social science*. Retrieved May 25, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC10893965/

Marr, B. (n.d.). *What are the 4 Vs of Big Data?* Bernard Marr. Retrieved May 25, 2025, from https://bernardmarr.com/what-are-the-4-vs-of-big-data/

Ortiz-Ospina, E. (2019, September 18). *The rise of social media*. Our World in Data. https://ourworldindata.org/rise-of-social-media

Qu, D., Liu, B., Jia, L., Zhang, X., Chen, D., Zhang, Q., Feng, Y., & Chen, R. (2023). The longitudinal relationships between short video addiction and depressive symptoms: A cross-lagged panel network analysis. *Computers in Human Behavior, 152*, 108059. https://doi.org/10.1016/j.chb.2023.108059

Selection of Appropriate Statistical Methods for Data Analysis. (n.d.). *Selection of Appropriate Statistical Methods for Data Analysis*. Retrieved May 25, 2025, from https://pmc.ncbi.nlm.nih.gov/articles/PMC6639881/

Shimaoka, M., Ferreira, J., & Goldman, A. (2024). The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks. *Journals-Sol.SBC.Org.Br*. Retrieved May 25, 2025, from https://journals-sol.sbc.org.br/index.php/reviews/article/view/3757

Team, B. (2025, February 10). *Social Media Usage & Growth Statistics*. Backlinko. https://backlinko.com/social-media-users

How many users on TikTok? Statistics & Facts (2025). (n.d.). https://seo.ai/blog/how-many-users-on-tiktok

I used AI in the following ways:
I used Gemini to brainstorm ideas.
I used Gemini to help me elaborate my points into more verbose explanations.
I used Gemini to help me write the technical code used in the project.
I used Gemini and Quillbot to generate the citations.