Exploring the Effects of Social Media Addiction on Student Well-Being

Amanda J. Reinhart

Western Governors University

Table of Contents

[A. Project Highlights 4](#_Toc1868160802)

[B. Project Execution 5](#_Toc2011082291)

[C. Data Collection Process 7](#_Toc1191971028)

[C.1 Advantages and Limitations of Data Set 7](#_Toc10980602)

[D. Data Extraction and Preparation 8](#_Toc512824111)

[E. Data Analysis Process 9](#_Toc1571068809)

[E.1 Data Analysis Methods 9](#_Toc1409685802)

[E.2 Advantages and Limitations of Tools and Techniques 9](#_Toc737998485)

[E.3 Application of Analytical Methods 10](#_Toc825242289)

[F Data Analysis Results 11](#_Toc1939531306)

[F.1 Statistical Significance 11](#_Toc109772385)

[F.2 Practical Significance 11](#_Toc1419440648)

[F.3 Overall Success 12](#_Toc1210557849)

[G. Conclusion 13](#_Toc1495356842)

[G.1 Summary of Conclusions 13](#_Toc613065263)

[G.2 Effective Storytelling 14](#_Toc16322062)

[G.3 Recommended Courses of Action 14](#_Toc660897025)

H Panopto Presentation 15

[References 16](#_Toc1410715514)

[Appendix A 17](#_Toc1695981861)

Dataset and Source Code [17](#_Toc1655579738)

# A. Project Highlights

**Research Question:** How do social media addiction levels vary across countries and academic levels, and how are higher addiction levels associated with students’ mental health, conflict frequency, and perceived academic impact?

**Project Scope:** This project analyzed the social media scores of students in relation to behavioral and well-being indicators that include: average hours of sleep per night, number of conflicts on social media, mental health score, perceived effect on academic performance, and addiction level. The analysis was split into two groups based on average daily usage to allow for comparative analysis between high-usage and low-usage groups.

The analysis was completed within JupyterLab using Python, Pandas, Matplotlib, Seaborn, and StatsModels. The dataset utilized will be manually downloaded from Kaggle.com and loaded into JupyterLab. The dataset will be cleaned and evaluated to complete the analysis. This report has been completed based on the final analysis. It will be presented to stakeholders as a summary of the project, its execution, utilized methodologies, results, and recommended actions, along with a video presentation.

**Solution Overview:** The project followed a structured analysis process to explore how social media addiction correlates with various well-being indicators among students. A waterfall methodology was used to guide each step of the project, from establishing requirements to gathering, preparing, and analyzing the data, and ultimately compiling and presenting findings. All work was completed within JupyterLab using Python, with Pandas used for data manipulation, Matplotlib and Seaborn for data visualization, and StatsModels for conducting Ordinary Least Squares (OLS) regression. The dataset was split into high-usage and low-usage groups to support comparative analysis, allowing for a more nuanced understanding of how usage patterns influence outcomes. The final solution consisted of scatterplots, regression summaries, and bar charts designed to communicate relationships across the variables studied clearly.

# B. Project Execution

**Project Plan:** This project was executed in close alignment with the plan laid out in Task 2. The two primary goals remained unchanged: (1) to examine how students’ social media addiction scores relate to behavioral and well-being indicators, and (2) to present the findings in a clear, actionable format for stakeholders. As originally outlined, the analysis included scatterplots with lines of best fit and OLS regressions to explore relationships between addiction scores and three continuous variables: average sleep per night, mental health score, and number of conflicts on social media. For the two categorical variables—academic level and perceived academic impact—average addiction scores were visualized through bar charts.

There were no significant variances from the original deliverables or intended analyses. The project proceeded as scoped, and all planned outputs were completed as described. This report, along with the final Jupyter Notebook (submitted as a PDF) and the Panopto video presentation, fulfills the original objective of communicating findings and next-step recommendations for stakeholder decision-making.

**Project Planning Methodology:** This project followed the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, providing a structured yet flexible framework for the analysis. The approach allowed for iterative refinement as new insights emerged throughout the project.

The Business Understanding and Data Understanding phases were executed together, grounding the analysis in a clearly defined research question and an exploratory review of the dataset's available variables. Data Preparation followed, involving cleaning, standardizing column names, handling missing values, and splitting the dataset based on reported daily social media use.

During the modeling phase, visual and statistical tools were used to explore relationships between social media addiction scores and other behavioral indicators. Scatterplots with lines of best fit, bar charts, and OLS regression models were developed as appropriate for the variable types. The Evaluation phase was integrated into the modeling step, using visual patterns and statistical significance to assess the strength and validity of observed relationships.

The final Deployment phase involved compiling results into a written report and recording a video presentation to summarize the methodology, findings, and recommended courses of action for stakeholders.

**Project Timeline and Milestones**

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestone or deliverable** | **Duration**  **(hours or days)** | **Projected start date** | **Anticipated end date** |
| Establish Requirements for Analysis Project | 2 days | 5/20/2025 | 5/21/2025 |
| Gather and Clean Data | 1 day | 5/26/2025 | 6/26/2025 |
| Evaluate and Analyze Data | 1 day | 5/26/2025 | 5/26/2025 |
| Compile Written Analysis Report | 1 day | 5/28/2025 | 5/28/2025 |
| Record Analysis Presentation | 1 day | 5/31/2025 | 5/31/2025 |

Overall, the project timeline saw no significant variances from the proposed timeline. The primary difference is that evaluating and analyzing the data and compiling the written analysis report took less time than initially anticipated, at one (1) day each instead of two (2). The rest of the project stuck to the proposed project timeline.

# C. Data Collection Process

**Data Selection and Collection:** There was no variance in the data selection and collection process from the one originally proposed. The dataset was downloaded directly from Kaggle.com in CSV format. The CSV file was manually loaded into the Jupyter Notebook via .read\_csv() for cleaning, exploration, and analysis.

**Obstacles Encountered:** Initially, this project aimed to utilize the Kaggle.com API to pull our dataset and ensure it would stay up-to-date whenever the notebook was run. However, due to TLS Certificate issues on my laptop and the extensive nature of the workaround to get the Kaggle API to function properly, I opted for manual download and upload for simplicity's sake. Since the dataset was based on a one-time survey, data freshness was not a concern, and this adjustment had no material impact on the project.

**Unplanned Governance Issues:** No unplanned data governance issues were encountered during this project. The dataset was publicly available and freely licensed through Kaggle, with no personally identifiable information or restricted data requiring special handling. All usage remained within the bounds of the dataset’s original terms, and no additional governance measures were needed beyond what was outlined in the project proposal.

## C.1 Advantages and Limitations of Data Set

The dataset used for this project had several clear advantages. It was complete and appropriately sized from the start, allowing us to move directly into analysis without needing significant time to handle missing or null values. All relevant information was contained within a single table, so we could avoid complex joins and still conduct a thorough analysis. The dataset was also freely available for public and educational use, eliminating the need to secure special licensing or permissions before proceeding.

The main limitation of the dataset is that it was generated from a one-time survey. Because of this, any relationships identified in the analysis are correlational rather than causal. While the dataset's size provides a degree of robustness, the lack of repeated or longitudinal data prevents us from making broader claims about trends over time or across different student populations.

# D. Data Extraction and Preparation

The dataset was manually downloaded from Kaggle.com in CSV format and loaded into the Jupyter Notebook using the .read\_csv() function from the pandas library. From there, the data was reviewed for consistency and cleaned to ensure it was suitable for analysis. Since the dataset had no null values, minimal cleaning was required. Basic exploratory techniques were used to confirm column types and ranges and verify that each variable aligned with the intended analysis goals. Columns that were unnecessary for the analysis were dropped to streamline the dataset and reduce clutter.

To compare different social media usage levels, two separate DataFrames were created based on reported average daily time spent on social media. One DataFrame included students in countries with higher daily usage, while the other DataFrame comprised students with lower daily usage. This separation allowed for more targeted analysis within each group without modifying the original dataset.

This preparation work was performed entirely in Python using pandas for manipulation, and the workflow remained consistent with what was outlined in Task 2. The use of JupyterLab for environment management and pandas for data wrangling was appropriate for this scale and type of analysis, offering flexibility and transparency throughout the process.

# E. Data Analysis Process

## E.1 Data Analysis Methods

This project used a combination of OLS regression, scatterplots, and bar charts to analyze the relationship between students’ social media addiction scores and key well-being indicators. OLS regression and scatterplots were used to evaluate the strength and direction of relationships between numeric variables. In contrast, bar charts were used to compare addiction scores across categorical variables, including perceived academic impact and academic level. These analyses were performed separately on both the high-usage and low-usage student groups to allow for comparison.

These methods were selected because they align well with the evaluated data types and the project’s research goals. OLS regression provided clarity on the relationship's strength and statistical significance, and bar charts allowed for clean visual comparisons between categorical groups.

**E.2 Advantages and Limitations of Tools and Techniques**

All the tools utilized for this project, JupyterLabs, Python, Pandas, Matplotlib, Seaborn, and Statsmodels, have one significant advantage: they are all free to the public and user-friendly. Each tool is incredibly flexible, making it ideal for this project. Additionally, extensive documentation can be easily found online, making these ideal tools for any analyst, regardless of their experience level.

JupyterLab and the Python libraries used in this project (Pandas, Matplotlib, Seaborn, and StatsModels) were well-suited for exploratory analysis but came with some limitations. JupyterLab isn’t designed for production workflows or real-time collaboration, and Python can be slower with large-scale data. Pandas and statsmodels are memory-intensive and can be complex to interpret without a strong statistical background, while matplotlib and seaborn lack interactivity. None of these limitations impacted this project, but they’re worth noting for potential future expansion.

**E.3 Application of Analytical Methods**

To apply the analytical methods described in E1, the dataset was first filtered into two separate dataframes: one for students with high social media usage and one for those with low usage, based on average reported screen time. From there, numeric variables such as average sleep per night, number of conflicts on social media, and mental health score were analyzed using OLS regression. For each pairing, a scatterplot was created to visualize the relationship, and a line of best fit was included to illustrate trend direction. Regression summaries were then generated using the statsmodels library to provide coefficients, R-squared values, and p-values.

Categorical variables, specifically perceived impact on academic performance and academic level, were visualized using bar charts that compared average social media addiction scores across different groupings. The bar plots provided a quick visual summary of how scores differed across categories, allowing for interpretation alongside the numeric regression results.

Because the dataset was complete and well-formatted, no imputation or transformation of variables was required before analysis. No violations of assumptions were encountered during analysis. The data was suitable for linear regression, and p-values were well below the alpha level of 0.05, supporting statistical significance where noted.

# F Data Analysis Results

## F.1 Statistical Significance

For each relationship explored in the analysis, the null hypothesis assumed no connection between a student’s social media addiction score and the well-being or behavioral indicator being evaluated. For the continuous variables (average sleep per night, mental health score, and number of conflicts on social media), OLS regression was used to determine whether a statistically significant relationship exists. Each regression produced a p-value, and an alpha level of 0.05 was used to assess significance. If the p-value was below 0.05, the null hypothesis was rejected and the relationship was considered statistically meaningful.

Each relationship we explored during this analysis was deemed to be statistically significant. Regardless of the strength or direction of the relationship being explored, each p-value came in below our designated alpha level of 0.05.

The bar chart comparisons on our categorical variables focused on visual differences and did not include visual testing. We could not determine statistical significance for either of these relationships (academic level and perceived impact on academic performance). However, we could still identify functional patterns worth noting in the context of our other findings.

**F.2 Practical Significance**

While each relationship evaluated in this project was statistically significant, it is equally important to consider whether these findings are meaningful in a real-world context. The practical significance of these results lies in the consistent patterns that emerged across both the high-usage and low-usage student groups. For example, as social media addiction scores increased, so did the number of reported conflicts on social media and the severity of reported mental health challenges. These are patterns thateducation professionals, counselors, and academic support staff can use to understand student behavior and well-being better.

## The strength of these relationships, particularly the high R-squared values in the OLS regressions, suggests that social media addiction levels are more than just passively associated with these indicators; they explain a substantial portion of the variation. This makes the findings actionable. Institutions may consider monitoring social media usage patterns or surveying students about perceived impact to identify those who may benefit from early intervention or additional support. These results provide a data-driven foundation for conversations around policy, wellness programming, or awareness campaigns targeting student populations.

## F.3 Overall Success

This analysis can be determined a success from various viewpoints. We saw success from the standpoint of completeness and accuracy of the proposed statistical analysis. The planned visualizations were all correctly executed for our datasets, including scatterplots, bar charts, and OLS regression models. Additionally, all results returned by the statistical tests could be interpreted using our established alpha level.

The dataset was complete and large enough for our analysis to be meaningful and demonstrate clear student social media use trends. Though the survey's individualistic nature limits analytical expansion on this topic using the same data and prevents us from drawing longitudinal success, it provided a large enough population for meaningful and statistically significant analysis.

# G. Conclusion

## G.1 Summary of Conclusions

The findings of this analysis directly support the research question by confirming that social media addiction scores are meaningfully related to various behavioral and well-being indicators in students. In both usage groups, students with higher addiction scores generally reported fewer hours of sleep, lower mental health scores, and more frequent conflicts on social media. These relationships were strongest in the high-usage group. While relationships in the low-usage group were weaker, they consistently reflected the same trends. For all of the explored relationships and both usage groups, the p-values were calculated below our alpha level, meaning that we rejected the null hypothesis.

Additionally, our results showed that students at higher academic levels tended to report lower addiction scores than their peers, and those who believed their social media use impacted their academic performance had higher addiction scores than their peers. These trends offer valuable insight into how social media use may intersect with academic experience and emotional well-being.

Put together, these findings provide strong evidence that social media addiction is associated with other patterns of student behavior and can serve as a meaningful indicator in conversations about student well-being.

## G.2 Effective Storytelling

## The tools and visualizations chosen for this project supported effective communication of findings by making complex relationships easier to understand. Scatterplots with lines of best fit allowed us to clearly show the direction and strength of relationships between addiction scores and continuous variables, helping to reinforce the trends identified in the regression models visually. Bar charts were an appropriate choice for our categorical variables, as they made it simple to compare average addiction scores across distinct groups. JupyterLab allowed for seamless integration of code, outputs, and visualizations in one environment, while libraries like Matplotlib and Seaborn provided flexible formatting to ensure the graphics were clean and accessible. Together, these choices ensured that the narrative built from the data was visually intuitive and analytically sound.

## G.3 Recommended Courses of Action

Based on the findings of this analysis, stakeholders in student well-being, including educators, mental health counselors, and parents, are encouraged to implement targeted interventions. Two primary strategies are recommended.

The first is encouraging students to engage in a daily mindfulness and journaling practice. These techniques are well-documented in improving emotional regulation, self-awareness, and anxiety reduction. A short 10–15 minute reflective routine each day can promote healthier mental states and encourage students to become more conscious of their social media habits, potentially lowering addiction scores over time.

The second recommendation is the promotion of screen-free sleeping environments. Removing access to phones or connected devices at bedtime reduces opportunities for late-night scrolling or compulsive checking. Limiting this behavior may lead to improved sleep quality and duration, a factor strongly associated with both mental health and addiction levels in this analysis.

While these interventions primarily target sleep and mental health indicators, the impact of improving these areas can extend further. Better rest and stronger mental well-being are closely linked with higher academic performance, healthier relationships, and increased confidence, making these interventions both impactful and broadly beneficial.

# H Panopto Presentation

# The video presentation that accompanies this analysis can be found at the link below:

# 

# References

No sources were cited.

# Appendix A

# Dataset and Source Code

The dataset for this project, ‘Students Social Media Addiction.csv’, can be found at <https://www.kaggle.com/datasets/adilshamim8/social-media-addiction-vs-relationships>.

The source code is presented with this project as a PDF file and will be highlighted during the video presentation.