Analyzing Students' Social Media Addiction and Academic Performance Using Big Data Techniques

Surya Medepalli, Uyen Nguyen

09/02/2025

Abstract

Social media addiction among students has become a critical concern affecting academic performance and mental health globally. This study examines the use of social media by students and its relationship to their well-being, based on data from 705 participants across 110 countries. A multi-layered big data pipeline was implemented, integrating Pandas for exploratory analysis, MapReduce for detecting usage patterns, Spark for distributed transformations and queries, graph theory algorithms for analyzing influence networks, and Apache Sedona for spatiotemporal modeling. We find that students who spend more time on social media tend to sleep less, report poorer mental health, experience stronger feelings of addiction, and encounter more conflicts. The level of self-reported addiction also depends on the platform, with short, visually focused applications showing higher reported effects. Graph analysis highlights the dominance of platforms that act as central hubs of influence. At the same time, spatial modeling reveals clusters of high-addiction regions that overlap with areas of lower reported well-being. These findings suggest that social media addiction is reinforced by both platform influence and geographic clustering, underscoring the importance of platform-specific and region-aware interventions. Since these results only show correlations rather than causation, they should be interpreted with caution. Future research could extend this work by applying predictive models, longitudinal tracking, and hypothesis-driven studies.

Keywords: social media, addiction, students, mental health, sleep, Spark, exploratory data analysis (EDA)

Introduction

Students frequently spend significant amounts of time on social media. While these platforms can foster connection and communication, they also carry the risk of promoting unhealthy usage patterns. This study examines platform references, time spent on social media, and how self-reported "addiction" scores relate to sleep quality, mental health, and interpersonal conflicts. The aim is to provide a clear, evidence-based overview that can guide future, more in-depth research. For this project, we use a data set obtained from the following site: https://www.kaggle.com/datasets. We encourage the readers of this paper to download the data set and explore the data.

Methodology

Data Source and Sample:

The data set contained 705 records and 15 variables capturing demographics, usage hours, most-used platform, self-reported academic impact, sleep, mental health score, conflicts over social media, and an addiction score.

Measures:

Key variables include:

- Avg_Daily_Usage_Hours (hours/day)
- Addicted_Score (0–10 scale),
- Sleep_Hours_Per_Night (hours)

- Mental_Health_Score (0–10 scale)
- Conflicts_Over_Social_Media (count-like)
- Most_Used_Platform (categorical),
- Affects_Academic_Performance (Yes/No).

Preprocessing:

Data preprocessing was performed using Python. Steps included:

1. Cleaning categorical text fields and standardizing values.
2. Converting relevant columns to numeric types.
3. Deriving a binary indicator for academic performance impact.
4. Creating a Risk_Bucket variable based on addiction scores: Low (0-3), Medium (4-6), High (7-10)

Analysis:

We conducted exploratory data analysis (EDA) using distributional plots and group summaries, followed by correlation analysis among numeric variables. We summarize per-platform averages and visualize relationships between usage and mental health. We then applied MapReduce (MRJob) to compute platform-level averages and correlations between usage and mental health. Next, Spark RDDs were used to derive new features, such as the usage-to-sleep ratio and a three-level Risk_Bucket, along with filtering and join operations. Spark DataFrames supported additional filtering, grouping, aggregation, and SQL queries. Finally, Graph Theory techniques (PageRank, triangle counting) and spatiotemporal analysis with Apache Sedona (convex hulls, buffers, unions, intersections) were employed to extend the analysis to networks and geography.

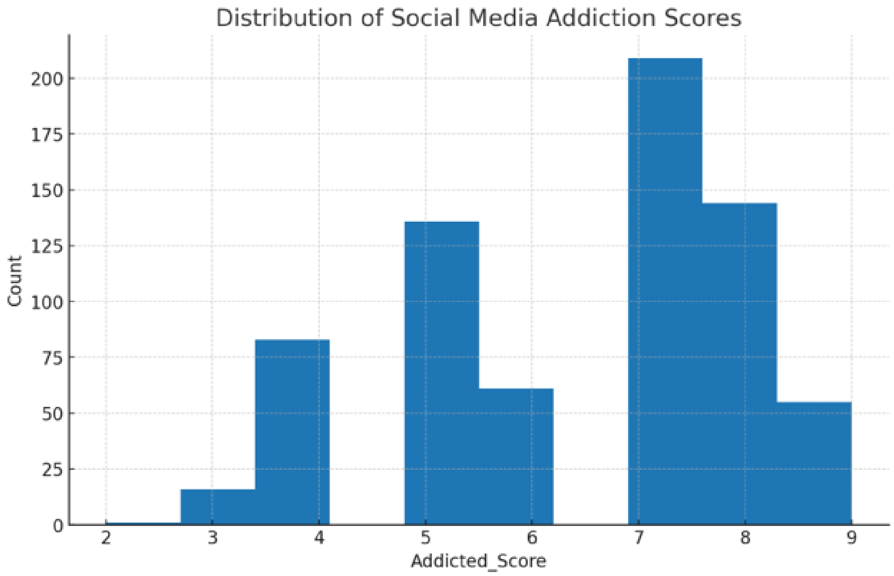
Results and Visualizations

Our analysis produced consistent patterns across statistical summaries, correlations, network graphs, and spatial models. We present the results progressively, starting with descriptive distributions, then moving to platform and demographic patterns, correlations, network structures, and finally spatiotemporal insights.

Descriptive Distribution

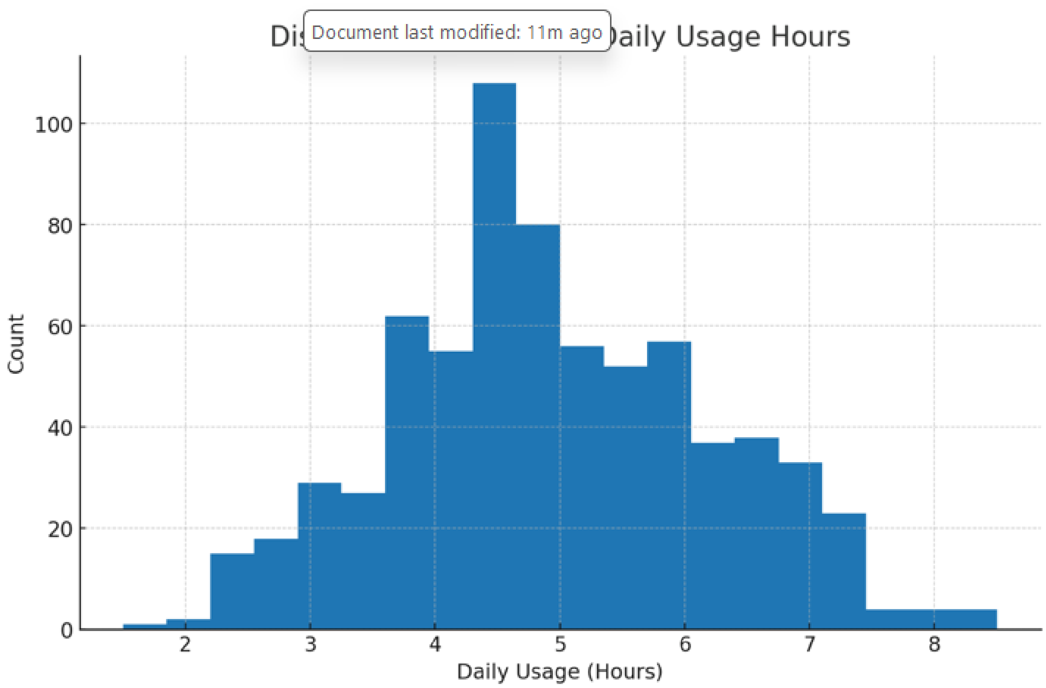
Addiction scores showed a multi-modal distribution, with peaks around scores of 4, 5, and 7–8 (Figure 1). This indicates that students tend to cluster into distinct groups of lower, medium, and higher self-reported addiction.

Figure 1. Distribution of Social Media Addiction Scores.



Daily usage illustrates a unimodal distribution peaking between 4 and 5 hours per day, with most students in the 3–6 hour range (Figure 2). A smaller subset reported 7–8+ hours, indicating that while moderate use is typical, heavy use is also present.

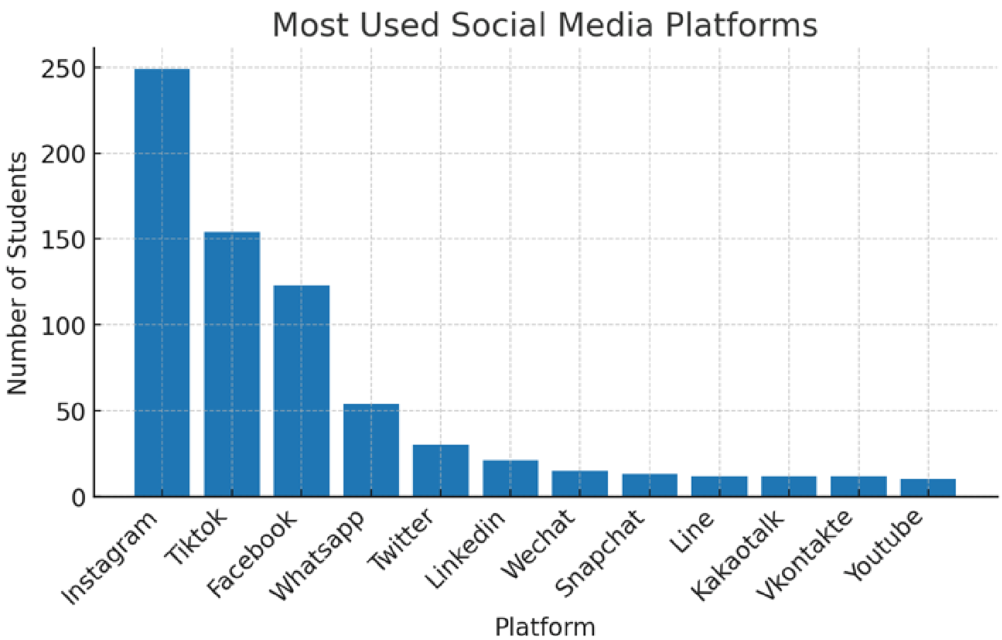
Figure 2. Distribution of Average Daily Usage Hours.



Platform Patterns

The chart shows that Instagram, TikTok, and Facebook were the dominant platforms (Figure 3), together accounting for the majority of students' reported 'most used' category. In contrast, students rarely selected platforms such as Line, WeChat, and VKontakte.

Figure 3. Most Used Social Media Platforms.



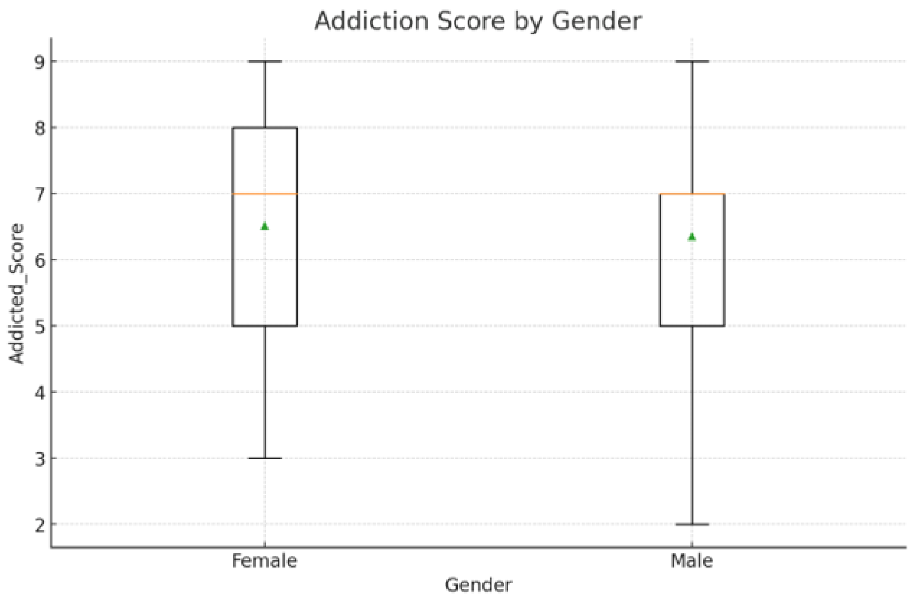
Students who mainly used WhatsApp, Snapchat, or TikTok reported the highest addiction scores (~7.4/10), showing these fast, social apps are especially engaging. Instagram was also elevated (6.6). In contrast, students on LinkedIn (3.8) or Line (3.0) scored much lower, reflecting the less addictive nature of professional or text-based platforms. Mental health scores followed the same pattern: lower on high-addiction apps, higher on low-addiction ones.

Most_Used_Platform	n	avg_addicted	avg_usage	avg_mental	sd_addicted
Whatsapp	54	7.463	6.48	5.54	0.503
Snapchat	13	7.462	5.09	5.54	0.776
Tiktok	154	7.429	5.35	5.71	1.041
Instagram	249	6.554	4.87	6.12	1.537
Youtube	10	6.1	4.08	6.6	1.37
Wechat	15	6.067	4.96	6.47	1.033
Kakaotalk	12	6.0	4.73	6.0	0.0
Facebook	123	5.667	4.51	6.72	1.435
Twitter	30	5.5	4.87	6.83	1.548
Vkontakte	12	5.0	4.25	7.0	0.0
Linkedin	21	3.81	2.52	8.0	0.602
Line	12	3.0	3.25	8.0	0.0

Demographic Differences

Addiction scores by gender showed minor differences (Figure 4). Both male and female students had median scores of about 7, showing that heavy social media use is a common issue for all genders.

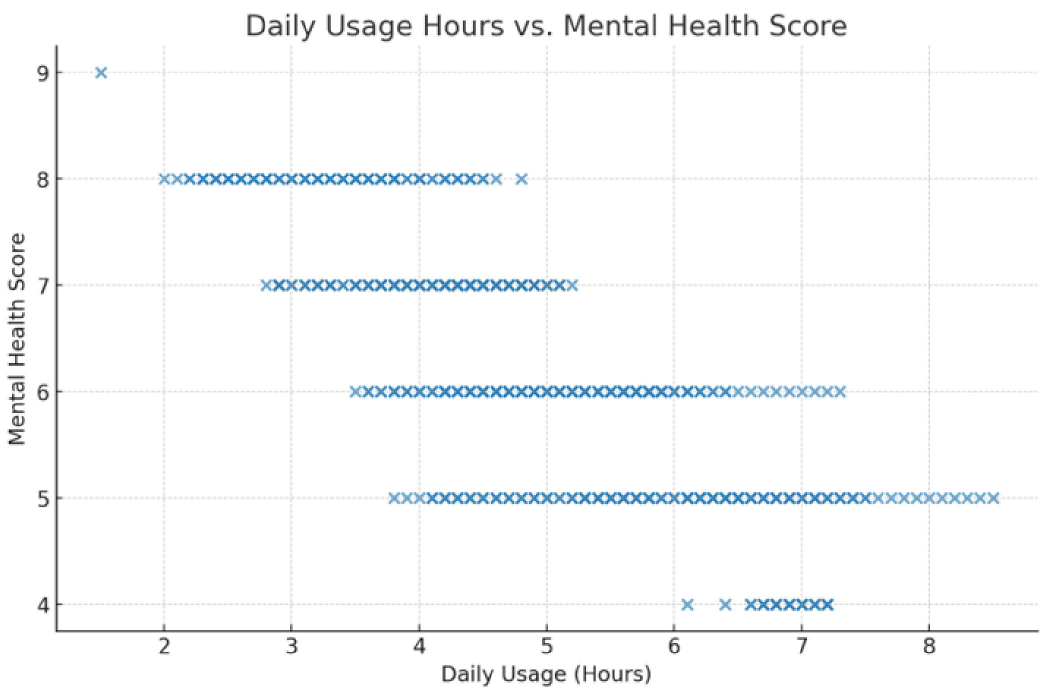
Figure 4. Addiction Score by Gender.



Correlations

We found a strong negative correlation between daily usage and mental health ($r = -0.80$). Students who used social media more hours each day consistently reported lower mental health scores, less sleep, and more frequent conflicts.

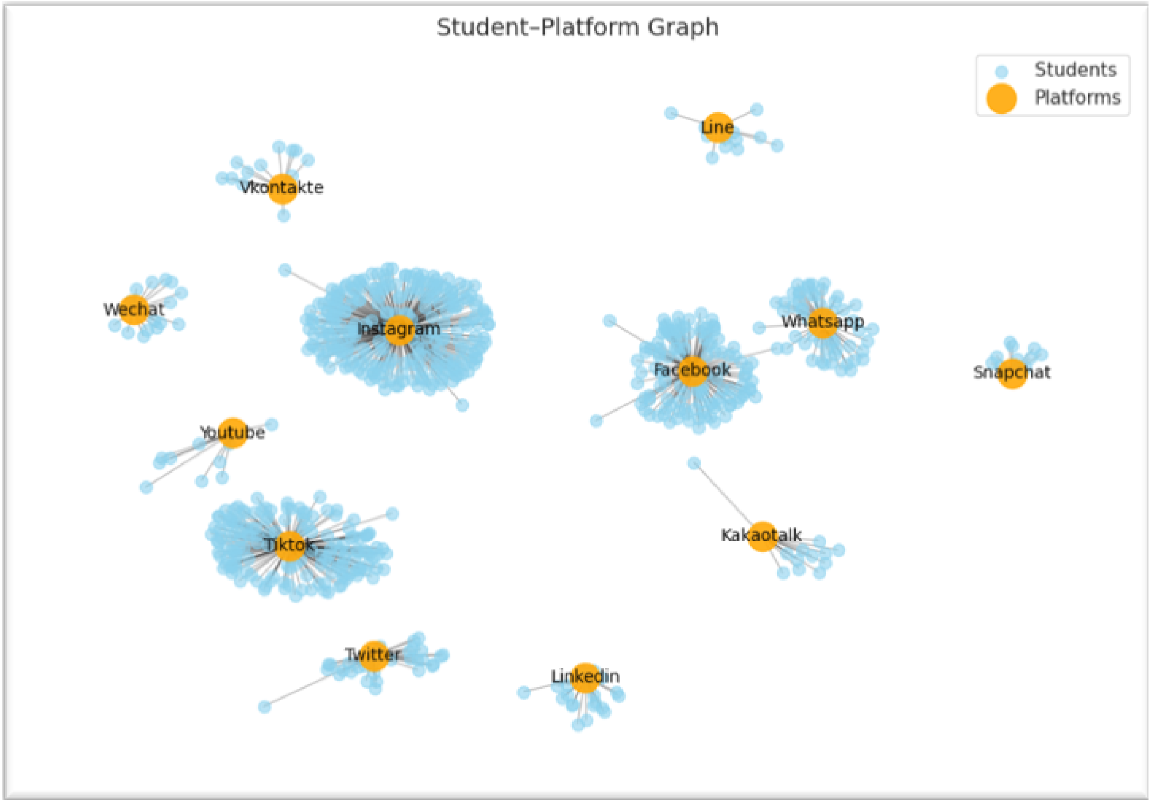
Figure 5. Daily usage hours vs. mental health score.



Graph Theory Results

The student–platform network contained 717 nodes and 705 edges. Instagram and TikTok emerged as the most influential hubs based on PageRank, while Facebook and WhatsApp were secondary.

Figure 6. Student–Platform Graph (Orange = Platforms, Blue = Students, Edges = Usage Connections).



Triangle counting revealed dense peer clusters, with over 30,000 closest triplets around these hubs.

```
Top 10 students by triangle count:
Student_1: 30628
Student_6: 30628
Student_10: 30628
Student_14: 30628
Student_18: 30628
Student_23: 30628
Student_26: 30628
Student_30: 30628
Student_35: 30628
Student_38: 30628
```

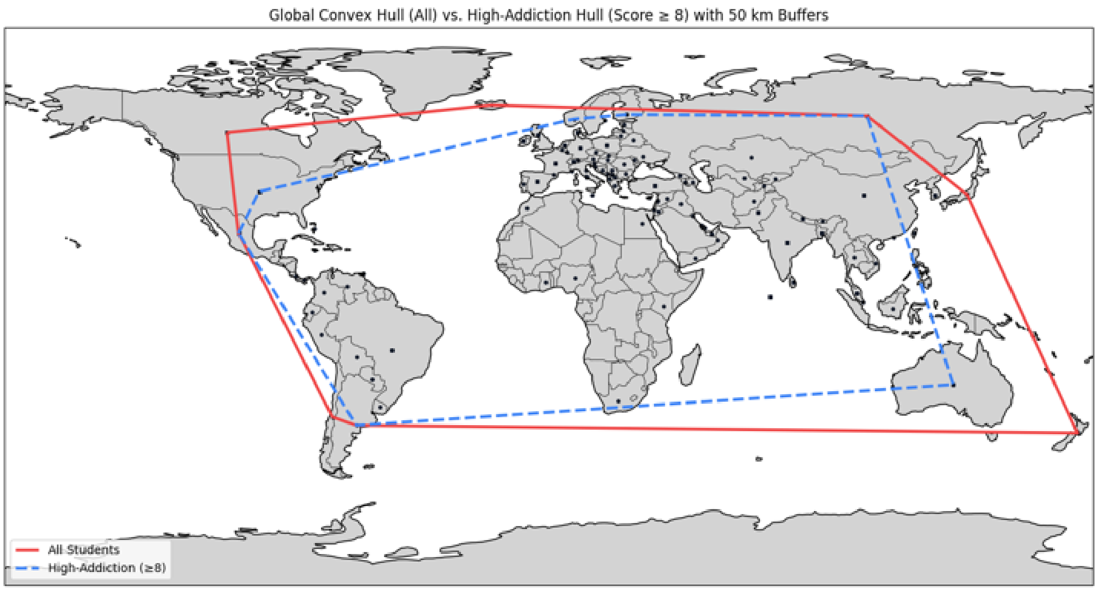
At the country level, Italy, Switzerland, the UK, India, and Mexico emerged as central connectors, showing that addictive use patterns are reinforced not only socially but also regionally.

```
Top 10 countries by triangle count (dense shared-platform clusters):
Country=Italy: 1920
Country=Switzerland: 1920
Country=UK: 1763
Country=India: 1736
Country=Mexico: 1736
Country=Bangladesh: 1730
Country=Canada: 1730
Country=Spain: 1730
Country=Finland: 1730
Country=Poland: 1730
```

Spatiotemporal Patterns and Enhanced Results

Figure 7 compares the global distribution of all students (red convex hull) with that of high-addiction students (blue convex hull, score ≥ 8). While the global hull spans nearly all major world regions—including the Americas, Europe, Asia, and Oceania—the high-addiction hull is much smaller, concentrated in North America, Europe, and Asia, and excluding much of South America, Africa, and Oceania. Black dots indicate individual student locations.

Figure 7. Global Convex Hull (All) vs. High-Addiction Hull (Score ≥ 8) with 50 km Buffers.



Red solid polygon = All Students (Global Convex Hull)
Blue dashed polygon = High-Addiction Students (Convex Hull, Score ≥ 8)
Black dots = Student locations

Quantitative Coverage Analysis

Apache Sedona buffer analysis revealed that the global student population covered approximately **851,936 km^2**, while high-addiction students (score ≥ 8) were concentrated in only **351,163 km^2**, representing **41.2%** of the total coverage. This **2.43:1** ratio confirms that addiction hotspots are geographically clustered rather than evenly distributed (Table 1).

Table 1: Coverage Areas by Population and Platform (EPSG:3857, buffer unions, approximate)

Category	**Coverage (km^2)**
Global student coverage	851,936
High-addiction coverage (score ≥ 8)	351,163
Instagram	343,359
TikTok	304,341
Facebook	218,501
LinkedIn	163,876
Snapchat	101,447
YouTube	78,036
Twitter	62,429
WhatsApp	31,214
Line	7,804
WeChat	7,804
VKontakte	7,804
KakaoTalk	7,804

Platform-Specific Geographic Patterns

Coverage varied substantially by platform. Instagram dominated at **343,359 km^2** (40.3%), followed by TikTok at **304,341 km^2** (35.7%) and Facebook at **218,501 km^2** (25.6%). Combined, these three platforms exceeded 100% of global coverage due to substantial overlaps, indicating that most addiction risk occurs within interconnected ecosystems.

Strategic Overlap Analysis

Key intervention zones were identified in overlapping regions: Instagram–TikTok (**132,661 km^2**), Facebook–Instagram (**117,054 km^2**), and Facebook–TikTok (**93,643 km^2**). These concentrated intersections represent optimal targets for coordinated, multi-platform interventions (Table 2).

Table 2: Pairwise Platform Overlaps (km^2, EPSG:3857, buffer intersections)

Platform A	**Platform B**	**Overlap (km^2)**
Instagram	TikTok	132,661
Facebook	Instagram	117,054
Facebook	TikTok	93,643
Instagram	Twitter	54,625
TikTok	Twitter	46,822
Facebook	Twitter	39,018
Facebook	LinkedIn	31,214
Instagram	YouTube	29,872
Facebook	YouTube	29,872
Instagram	LinkedIn	23,411

Convex Hull Extent Comparison

Convex hull areas reinforce these findings: the global student distribution spanned **365.4 million km^2**, while high-addiction students occupied **268.8 million km^2** (73.6%). This 26.4% reduction demonstrates that severe addiction cases are clustered within a smaller geographic extent, consistent with buffer-based coverage results (Table 3).

Table 3: Convex Hull Areas for Global, High-Addiction, and Platforms (EPSG:3857, visualization reference only)

Category	**Hull Area (km^2)**
Global student hull	365,430,267
High-addiction hull (score ≥ 8)	268,795,992
Instagram	305,376,000
TikTok	246,205,000
YouTube	236,330,000
Facebook	209,828,000
LinkedIn	175,852,000
Snapchat	112,994,000
Twitter	69,370,000
WhatsApp	61,224,000
Line	0
WeChat	0
VKontakte	0
KakaoTalk	0

Regional Development Correlation

High-addiction concentrations in developed regions—such as North America, Europe, and East Asia—align with previous research linking cultural and economic factors to social media dependence. For example, Cheng et al. (2021) found that addiction rates were significantly higher in collectivist countries (31%) compared to individualist ones (14%), with a global average of 17.4%. In contrast, Africa, South America, and much of Oceania generally fall outside high-addiction zones, likely due to differences in infrastructure, digital access, and cultural engagement with social media.

Note: All area calculations use the EPSG:3857 projection and are approximate, intended for comparative analysis rather than precise measurement.

Discussion

People who use social media often score higher on measures of addiction, which fits with the idea that apps keep us hooked by offering unpredictable rewards that encourage constant checking. At the same time, heavier use is linked with poorer mental health and worse sleep, though we cannot say for sure which causes which. Higher addiction scores were also tied to more arguments or conflicts about social media, suggesting that the effects go far beyond the individual and can strain relationships, too.

Several reasons could explain the observed patterns: (i) displacement, where screen time interferes with sleep and offline restorative activities; (ii) arousal and vigilance, where late-night engagement affects the onset and quality of sleep; (iii) social comparison and feedback loops, which can exacerbate stress or mood volatility; and (iv) design elements, such as endless feeds and intermittent notifications that promote extended sessions. These mechanisms may interact differently across platforms and user segments, and are not mutually exclusive.

Average addiction scores vary by platform, indicating that social graph structures, pacing, and content modality (e.g., short-form video versus text) may be important moderators. Subgroup analyses may reveal different profiles of high usage, with some students maintaining relatively preserved well-being and others experiencing greater strain. Demographic and situational factors, such as academic workload, may also increase risk, and gender differences were minimal, suggesting that heavy social media use is a challenge shared across both male and female students.

Beyond platform-level differences, patterns of use reflect social and regional factors. Popular platforms act as central hubs where students spend much of their time, and close friend groups often make heavy use feel typical and even expected. In some countries, social and cultural trends push specific platforms to become especially dominant. Overall, these patterns show that addiction risks concentrate in particular communities and regions, creating greater challenges for some groups of students and highlighting the need for solutions that address both peer influence and the broader environment where students live.

For educators and student support teams, prevention efforts might focus on sleep hygiene (e.g., evening wind-down routines), attention management (e.g., batching notifications, timeboxing), and reflective prompts that help students monitor urges to check feeds. At the platform level, choice-preserving nudges such as optional session reminders or default 'quiet hours' could reduce late-night engagement without restricting access.

Coordinated strategies across multiple platforms are likely more effective than isolated platform-specific measures, since students often engage across overlapping ecosystems such as Instagram, TikTok, and Facebook. Regionally tailored approaches may also help, with intensive interventions in high-risk areas and prevention-focused strategies in emerging regions where patterns are still developing.

Since this study only looked at one point in time, it is hard to know which way the relationship goes (for example, students who feel down might end up using social media more). The results may appear firmer than they are because they rely on self-reports. Other factors could include personality or how much stress someone is under. It might shape both social media use and well-being. Finally, the averages across different platforms could be skewed depending on who was sampled or how much variety there was in the data.

Future research could focus on the following items: (1) use more advanced statistics to rule out other possible explanations; (2) check whether the effects change at different levels of use, such as when benefits taper off or risks suddenly increase; (3) look at whether certain factors, like reduced sleep, help explain the connection between social media use and mental health; and (4) follow people over time to see which comes first. When possible, small experiments—like testing scheduled "quiet hours"—could also help show cause and effect more directly.

Network and spatial analysis also offer promising directions. Future work could investigate how peer clustering shapes addictive behaviors, how hotspots change as digital access expands, and how cultural or economic contexts influence where risks are highest. It would also be valuable to study whether successful interventions in one region spread to neighboring areas, helping shift from a static snapshot toward a more dynamic and predictive framework.

This study focused on students; however, we might see similar patterns in other groups who similarly use social media. Different groups have different ways of using social media, which include platform habits, social norms, and daily responsibilities—like work—which means it is important to repeat the research in other settings to see if the results hold up.

Conclusion

Our discussion shows how the links we observed can be understood through everyday patterns of behavior, while also pointing to practical, ethical, and research directions for the future. Overall, the findings remind us that the relationship between social media and well-being is complicated, and we still need better studies to sort out cause and effect. By looking at more diverse groups, testing small changes in real life, and paying attention to context, future work can help us figure out not just whether social media use matters, but how it matters—and what can be done to support healthier use for individuals and communities.

References

Kaggle. (2025). *Students Social Media Addiction*. https://www.kaggle.com/datasets

Cheng, C., Lau, Y., Luk, J. W., Choi, J., Chio, F. W., Lam, L., & Cheng, S. H. (2021). *Prevalence of social media addiction across 32 nations: Meta-analysis*. Addictive Behaviors, 117, 106845. https://doi.org/10.1016/j.addbeh.2021.106845

Sahay, I., Chouhan, A., Sunilkumar, S., & Raval, R. (2025, July 31). *Data analysis on impact of social media bias on the mental health of individuals*. SSRN. https://ssrn.com/abstract=5389364