Analyzing Students' Social Media Addiction and Academic Performance

SURYA MEDEPALLI, UYEN NGUYEN

Project

- How students use social media & feel hooked
- Links to sleep, mental health, and conflicts
- Platforms with higher addiction risk
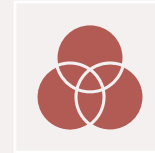- Insights for future research & solutions

Data selection

We used a data set from https://www.kaggle.com/datasets.

Dataset: Students Social Media Addiction.csv

Cleaning: fixed text, converted columns to numbers, added simple helper fields

Exploration (EDA): looked at distributions, by-platform comparisons, correlations, and two-variable charts

Scalable processing: used MrJob (MapReduce) and Apache Spark (RDDs, DataFrames, SQL)

Data Exploration & Preparation (Pandas)

Load data: pd.read_csv('Students Social Media Addiction.csv')

Quick look: df.head(), df.info(), df.describe()

Clean text: strip + title case; standardize Yes/No

Fix types: to_numeric() for hours/scores; handle missing with dropna()/fillna()

Engineer features: academic_impact (0/1), risk bucket (Low/Med/High)

Save outputs: cleaned CSV + model-ready (with get_dummies)

Spark Data Frames

Student ID	Age	Gender	Academic Level	Country	Avg Daily Usage Hours	Most Used Platform
1	19	Female	Undergraduate	Bangladesh	5.2	Instagram
2	22	Male	Graduate	India	2.1	Twitter
3	20	Female	Undergraduate	USA	6.0	TikTok
4	18	Male	Highschool	UK	3.0	YouTube

Spark RDDs

Platform	Avg Addicted Score
WhatsApp	7.463
Snapchat	7.462
TikTok	7.429
Instagram	6.554
YouTube	6.100
WeChat	6.067
Facebook	5.667
Twitter	5.500
LinkedIn	3.810

Graph Theory and Insights

Which platforms are the most influential among students?

- PageRank: Instagram (0.1604) and TikTok (0.0995) dominate; Facebook (0.0796), WhatsApp (0.0354) secondary.

- Triangle Counting: Students on Instagram/TikTok form dense cliques (~30k triangles each).

Student–Platform Graph
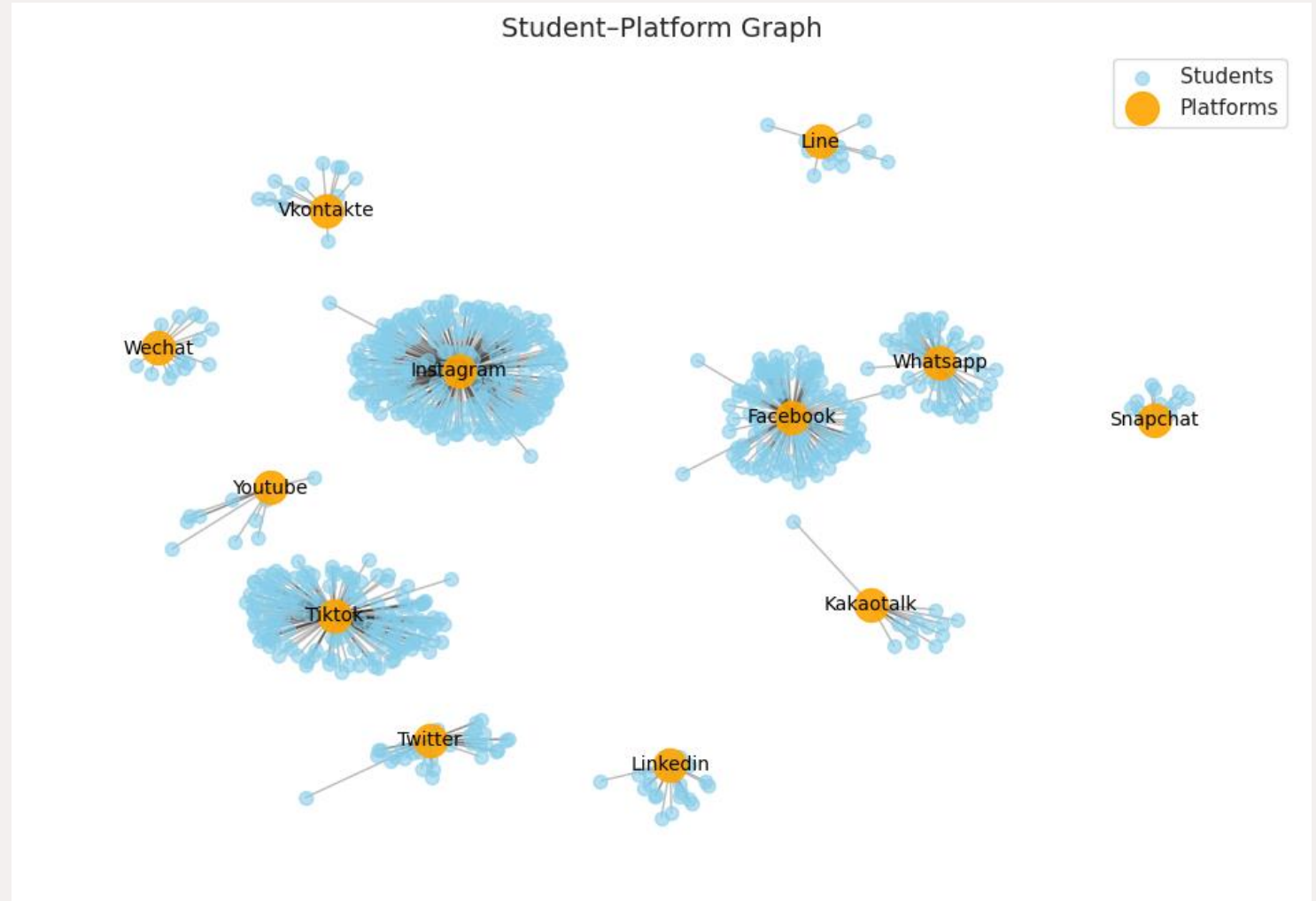
Graph Theory and Insights

How do student-level behaviors aggregate into country-level clusters of shared platform adoption that may reinforce social media addiction?
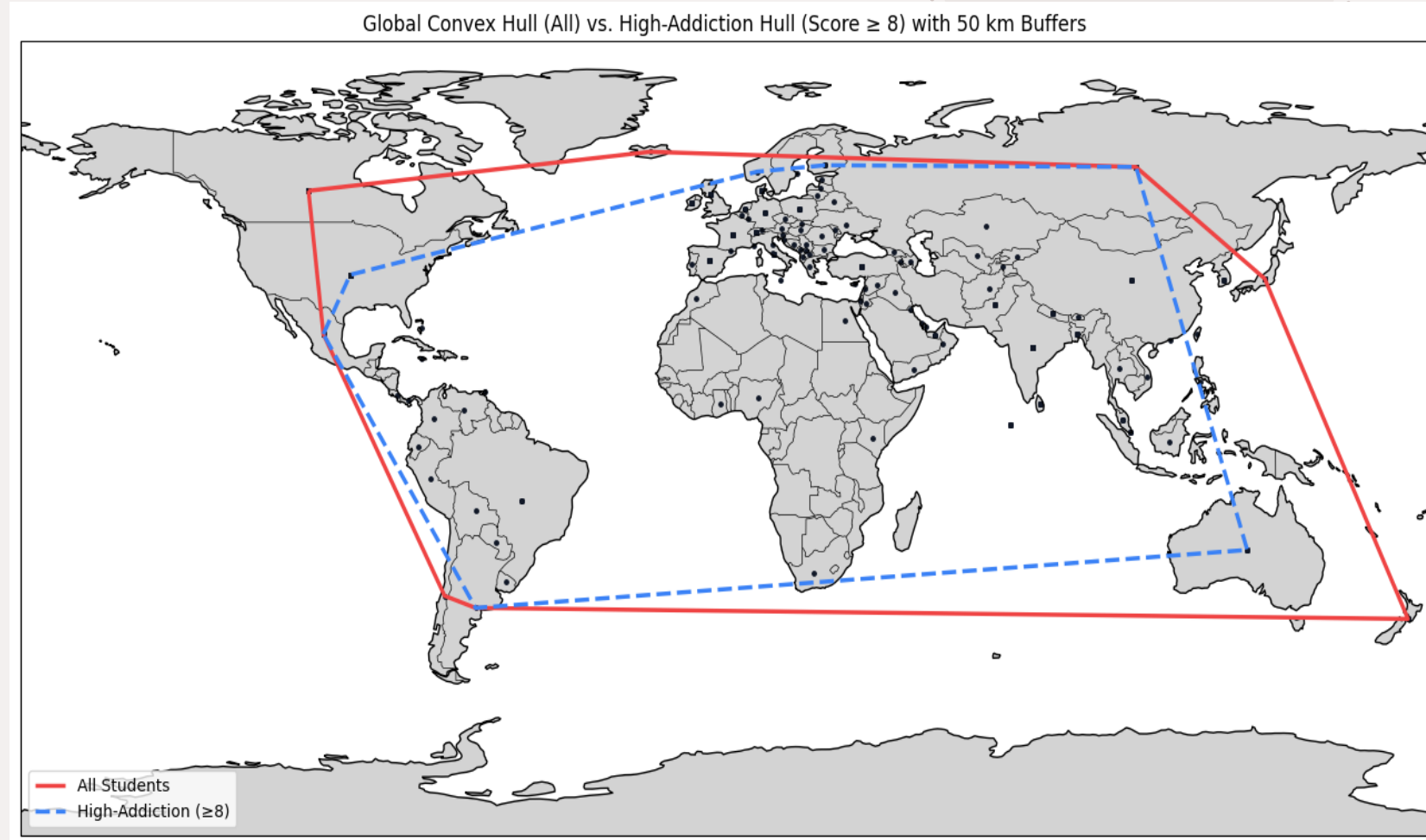
- PageRank: Instagram (0.1142) and TikTok (0.1017) still dominate.

- Italy (0.0124), UK (0.0123), India (0.0119) central connectors.

- Triangle counts: Italy & Switzerland ≈ 1,920; UK/India/Mexico > 1,700

Country	Triangles
Italy	1920
Switzerland	1920
UK	1763
India	1736
Mexico	1736

Spatiotemporal Analysis and Insights

How is student social media addiction distributed geographically?

- This is the world map with convex hull (red = all students, blue = high-addiction)

- Global hull covers all regions (Americans, Europe, Asia, Oceania)

- High- addiction hull smaller, denser around North America, Europe, Asia

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

All Students
High-Addiction (≥8)

Key Results

- Higher daily usage associates with higher addiction and lower mental health

- Addiction correlates negatively to sleep and positively to conflicts

- Platform averages differ, engagement design and content modality may matter

- Instagram and Tiktok dominate; Italy, UK and Indian are key cross-country hubs

- Addiction is global but densest in North America, Europe, and Asia; high and low – addiction zones overlap, with biggest overlaps between Instagram and Tiktok.

Conclusion

Global issue, reinforced socially & geographically

Strongest: TikTok, Instagram, Snapchat; weaker: LinkedIn, Twitter

Impacts: less sleep, lower mental health, more conflicts

Link to wellbeing is complex; causality unclear

Addiction spreads via peers & regions

Future work: diverse groups, real-life tests, context

Need awareness, balance, platform-specific solutions

References

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

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

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