Decoding the Digital Well-being Gap

A Multi-Tool Analysis of User Behavior and Mental Health Risk

1. Project Overview

This project explores the intersection of digital behavior and clinical mental health markers, GAD-7 (Generalized Anxiety Disorder) and PHQ-9 (Patient Health Questionnaire), using a dataset of 8,000 user records. By pairing SQL-driven data segmentation with Python visualizations, I identified the specific digital habits that escalate psychological risk.

2. Dataset & Methodology

The analysis is based on 8,000 diverse records representing a range of social media engagement styles.

- Data Origin- Generated via a Behavioral Simulation Engine (Python), enforcing logic based on established psychological literature, such as the correlation between late-night usage and circadian disruption.
- Psychometric Scales- Standardized GAD-7 (Anxiety, 0-21) and PHQ-9 (Depression, 0-27) scores.
- Key Metrics- Daily screen time, user archetypes, sleep duration, and activity types (Active vs. Passive).

Columns Descriptor:

- User_ID: Unique identifier.
- Age: User age (18-60).
- Gender: Biological sex (Male/Female).
- User_Archetype: Persona category (e.g., 'Digital Minimalist', 'Hyper-Connected').
- Primary_Platform: The app where the user spends the most time.
- Daily_Screen_Time_Hours: Total active screen time per day.
- Dominant_Content_Type: Main genre consumed (e.g., Gaming, Lifestyle).
- Activity_Type: 'Active' (posting) vs 'Passive' (scrolling).
- Late_Night_Usage: 1 if user is active after 12:00 AM, 0 otherwise.
- Social_Comparison_Trigger: 1 if content consumed typically induces envy/insecurity.
- Sleep_Duration_Hours: Average nightly sleep.
- GAD_7_Score: Anxiety score (0-21).
- GAD_7_Severity: Categorical interpretation (Minimal to Severe).
- PHQ_9_Score: Depression score (0-27).
- PHQ_9_Severity: Categorical interpretation (None to Severe).

Executive Summary

The analysis reveals that mental health risk is not merely a product of total screen time, but is heavily influenced by timing, intentionality, and engagement quality.

- Users active after 12:00 AM experience a 131% increase in average anxiety scores (4.91 to 11.35). This is strongly linked to a significant sleep deficit, with late-night users averaging only 5.08 hours of rest.
- Engagement style is a critical predictor of mood. Users primarily engaging in "Passive Consumption" (browsing without interacting) report 32% higher depression scores than those who engage actively.
- Threshold of Escalation: Risk does not scale linearly; it accelerates once a user crosses 6 hours of daily screen time. High-usage users report depression scores (8.97) nearly 4.5x higher than those under 3 hours (2.00).

- Social Comparison Triggers: Content that induces envy or social comparison acts as a primary driver for acute anxiety, raising GAD-7 scores by approximately 43% (7.04 to 10.06).

---

3. Phase 1: SQL Data Analysis & Segmentation

MySQL Script:
https://drive.google.com/file/d/1Cdxtw8pL57IDeiK7gziU54yH7J3pVD89/view?usp=sharing

Using SQL, I performed targeted aggregations to uncover how specific behaviors impact mental health scores.

A. User Archetype Risk Assessment

Analysis of user personas reveals a clear "health-to-risk" spectrum.

- Passive Scrollers and Hyper-Connected users exhibit the highest average anxiety (10.26) and depression (8.72) scores, respectively.
- Digital Minimalists maintain the healthiest baseline, with an anxiety score of only 3.33 and depression at 2.01.

B. Temporal Analysis (Late-Night Usage)

Late-night digital engagement correlates strongly with physiological and psychological strain.

- Users active late at night lose over an hour of sleep, averaging 5.08 hours compared to 6.24 hours for regular users.
- This behavior correlates with a 131% increase in average anxiety scores (4.91 \rightarrow 11.35).

C. Volume & Trigger Analysis

- Mental health symptoms scale aggressively with time. Users with High Screen Time (>6 hrs) report average depression scores nearly 4.5x higher than Low Usage users (8.97 vs 2.00).
- Exposure to envy-inducing content is a primary driver for anxiety, raising GAD-7 scores from 7.04 to 10.06.

---

4. Phase 2: Python Data Visualizations (Matplotlib/Seaborn)

Python Visualizations:
https://drive.google.com/file/d/12csrm-n64CF0J_pB9kB52xpekRdT5KcI/view?usp=sharing

To validate the SQL findings and visualize the distribution of risk, I utilized Python to create the following visualizations.

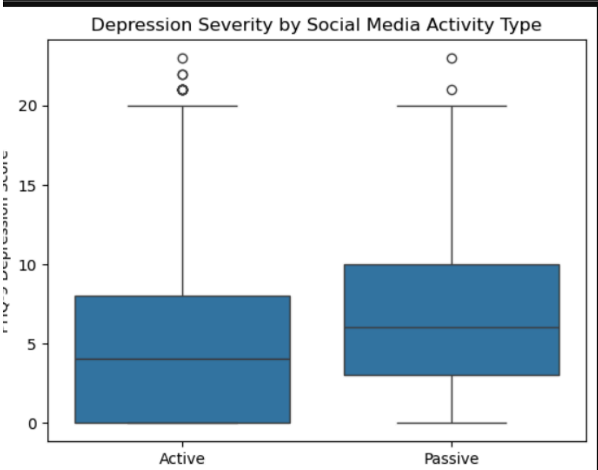
I. Distribution of Depression by Activity Type

Using a box plot, I compared depression severity between active and passive users.

- Observation: The median PHQ-9 score is visibly higher for Passive users compared to Active users. Passive usage also shows a wider range of higher-severity scores, suggesting that "scrolling" is a significant risk factor for depressive symptoms.

```
highlight differences between passive scrolling and active engagement. Depression by Activity Type
lt.figure()
ns.boxplot(
   data=df,
   x="Activity_Type",
   y="PHQ_9_Score"

lt.title("Depression Severity by Social Media Activity Type")
lt.xlabel("Activity Type")
lt.ylabel("PHQ-9 Depression Score")
lt.show()
```



II. Correlation: Screen Time vs. Anxiety

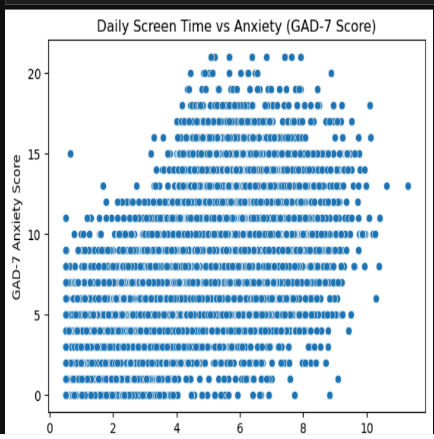
I visualized a scatterplot to observe the relationship between daily usage hours and anxiety levels.

- Observation: The dense cluster of points shows a positive upward trend; as screen time exceeds the 4-hour mark, we see a significant concentration of users reaching the highest possible GAD-7 scores (near 20+).

jupyter mental_heath analysis Last Checkpoint: yesterday

File Edit View Run Kernel Settings Help Trusted

```
[4]: #Show whether higher screen exposure aligns with increased anxiety symptoms. Screen-time vs Anxiety
plt.figure()
sns.scatterplot(
    data=df,
    x="Daily_Screen_Time_Hours",
    y="GAD_7_Score"
)
plt.title("Daily Screen Time vs Anxiety (GAD-7 Score)")
plt.xlabel("Daily Screen Time (Hours)")
plt.ylabel("GAD-7 Anxiety Score")
plt.show()
```

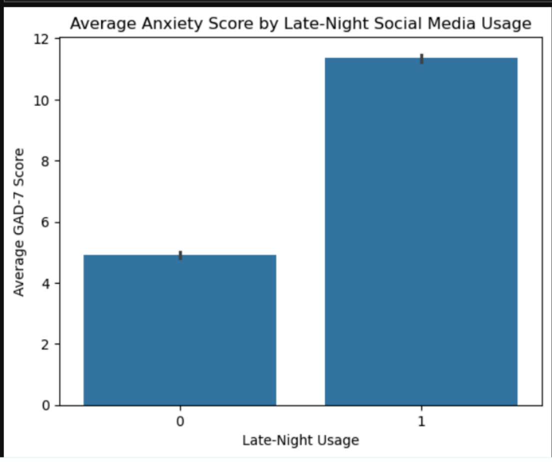


III. Visualizing the Anxiety Toll of Late-Night Use

A **bar plot** clearly illustrates the clinical impact of engagement timing.

- Observation: There is a stark visual contrast between Daytime users (0) and Late-Night users (1, defined as active after 12:00 AM). While daytime users remain in a manageable range, the late-night group clearly crosses the threshold into moderate-to-severe anxiety categories (averaging approximately 11.35).

```
plt.figure()
sns.barplot(
    data=df,
    x="Late_Night_Usage",
    y="GAD_7_Score"
)
plt.title("Average Anxiety Score by Late-Night Social Media Usage")
plt.xlabel("Late-Night Usage")
plt.ylabel("Average GAD-7 Score")
plt.show()
```



5. Recommendations for stakeholders, users, and social media creators.

- Introduce late-night usage alerts or optional “wind-down” nudges after a defined time threshold.
- Provide personalized usage summaries highlighting late-night behavior.
- Flag late-night usage as a risk signal for further well-being analysis.
- Design features that promote active participation (comment prompts, low-friction interactions).
- Reduce algorithmic overexposure to passive, endless-scroll content.
- Segment users by engagement style to tailor healthier usage experiences.
- Use multi-factor signals for risk monitoring instead of single metrics. For example, when screen time exceeds 6 hours, multifaceted factors, such as high daily screen time, late-night usage, and sleep duration below recommended levels, may indicate excessive digital overexposure, including heavy social media usage.

6. Conclusion

The synthesis of behavioral logs and psychometric data in this project confirms that the relationship between technology and mental health is defined by quality over quantity. While total screen time remains a significant indicator of risk, with symptoms more than tripling beyond the 6-hour threshold, the most profound clinical escalations are tied to specific behavioral choices, namely late-night usage and passive consumption.

The analysis of 8,000 records demonstrates that users are not a monolith; rather, they fall into archetypes with vastly different clinical profiles. The "Digital Minimalist" baseline (3.33 GAD-7 / 2.01 PHQ-9) serves as a proof of concept that social media can be utilized without incurring severe psychological costs. Conversely, the 131% increase in anxiety linked to late-night activity highlights a critical physiological vulnerability; the disruption of sleep directly fuels emotional instability.

Ultimately, this project highlights that social media platforms and health stakeholders must move beyond single-metric tracking. By monitoring "High-Risk Clusters," such as the convergence of high-volume usage, late-night activity, and social comparison triggers, we can transition from reactive treatment to proactive, data-driven digital wellness. This research provides the roadmap for designing digital environments that prioritize human well-being over raw engagement metrics.