Instituto Tecnológico y de Estudios Superiores de Monterrey – Campus Monterrey
School of Science and Engineering
Assignment Dashboard
2 Person - Team Assignment

M5: Data Viz Project

Professor: Sabut Butt

Jan Pascal Ludwig (A01830816)
Dominik Frank (A01830648)

November 28th, 2025
Monterrey, Nuevo León

Digital Environments & Mental Well–Being: A Data Story*

1 Context and Purpose

The audience for this project includes three groups: public health officials, parents, and young adults. Each group interacts with digital environments in different ways and needs slightly different guidance on how social media habits influence mental well–being. Concerns about screen time, online toxicity, and youth mental health have grown in recent years, so we wanted to move from opinions to data-based evidence. Acting as analysts within a national health department, our goal is to translate a complex dataset into a clear, understandable narrative.

We aim to:

- help **parents** establish healthier digital routines for their children,
- encourage **young adults** to reflect on how their online habits relate to stress, sleep, and mood,
- support **public health officials** by highlighting behavioural patterns worth targeting in future interventions.

We identify:

- **Who:** families, young users, and health authorities.
- **What:** an interpretation of the relationship between digital behaviour and mental health.
- **How:** through a short sequence of visuals that focus attention and avoid unnecessary complexity.
- **Intended action:** adjust routines, focus on preventive behaviours, and recognise high-risk digital patterns.

2 Big Idea

Healthy digital routines—moderate screen time, positive online interactions, and sufficient sleep and physical activity—are linked to lower stress and anxiety, suggesting that mental well–being depends as much on how *we use social media as on how much.*

*Course: Data Visualization and Storytelling

3 Dataset Description

We use the *mental_health_social_media_dataset*, a synthetic dataset approved for this project. It contains 5,000 rows and 15 columns of observations for individuals aged 13–69, with no missing values. For each person, the dataset records:

- demographic variables (age, gender),
- digital behaviour (daily social media time, platform, interaction quality),
- lifestyle factors (sleep, physical activity),
- mental health outcomes (anxiety, stress, mood, and a categorical mental state label).

The dataset is synthetic but designed to mimic patterns described in public health literature. Its clean structure makes visual exploration straightforward. A limitation is that categories such as *mental_state* are generated and may exaggerate trends, so we treat the results as illustrative rather than predictive. As can be seen in the `Data_Exploration.ipynb` notebook, the dataset contains no missing values. This is likely due to its synthetic nature, so no data cleaning was required before running the analysis.

4 Data Exploration and Visual Design

We distinguish between **exploratory** visuals, which helped us understand the dataset, and **explanatory** visuals, which we selected for the final story. We started with distributions to see who is in the data and how they behave. Then we looked at correlations and scatterplots to identify variables that seem related to mental health outcomes. Only visuals that supported the Big Idea and could be explained in one sentence were kept for the final dashboard.

In the design, we tried to apply:

- basic **Gestalt principles** (similarity, proximity, alignment),
- **preattentive attributes** (color, position, slope),
- simple **visual hierarchy** so that the main message stands out,
- enough **white space** to avoid clutter.

4.1 Exploratory Distribution of Mental States

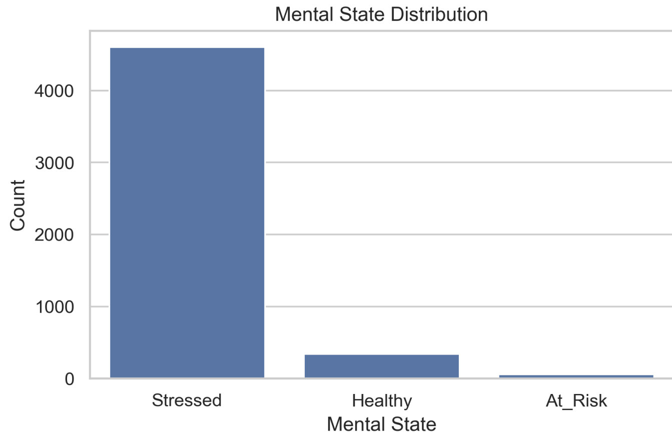
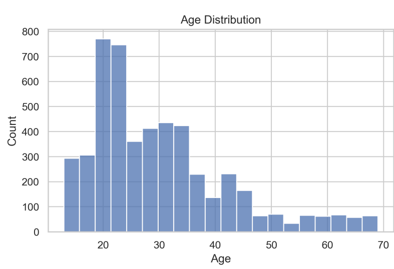


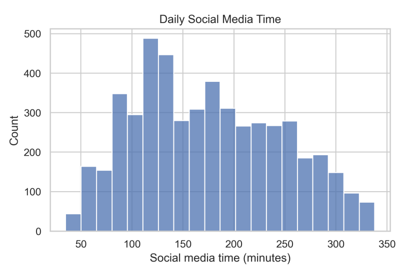
Figure 1: Distribution of mental states in the dataset.

Insight: The dataset is heavily skewed toward the *Stressed* category. This explains why average stress and anxiety scores are relatively high and suggests that we are looking at a population already under notable emotional load.

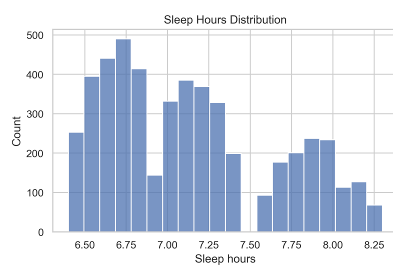
4.2 Key Behavioural Distributions



(a) Age distribution



(b) Daily social media time



(c) Sleep hours per night

Figure 2: Key behavioural and demographic distributions for the sample.

Insight: The histograms show a mostly young user base, moderate variation in sleep, and generally high daily social media usage. This gives a basic picture of the group we are analysing.

4.3 Correlation Structure

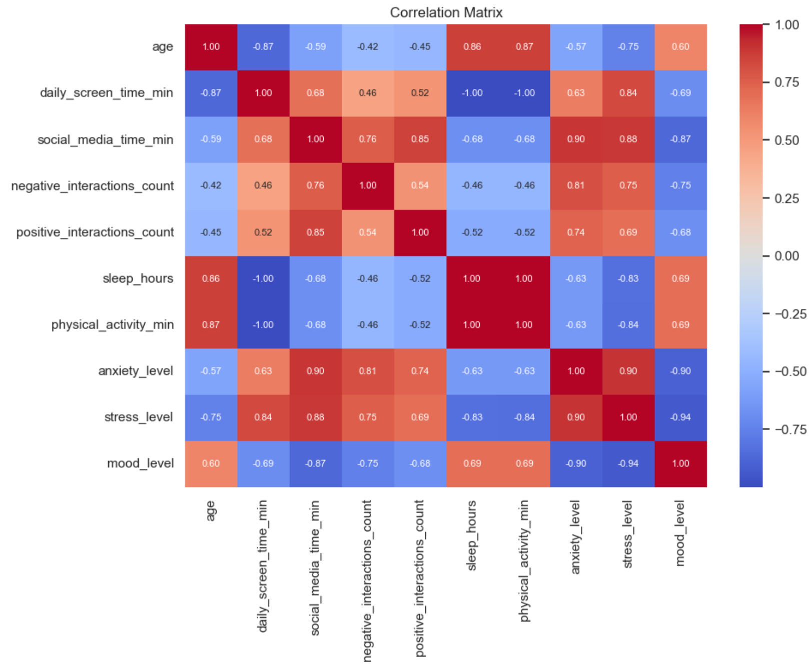
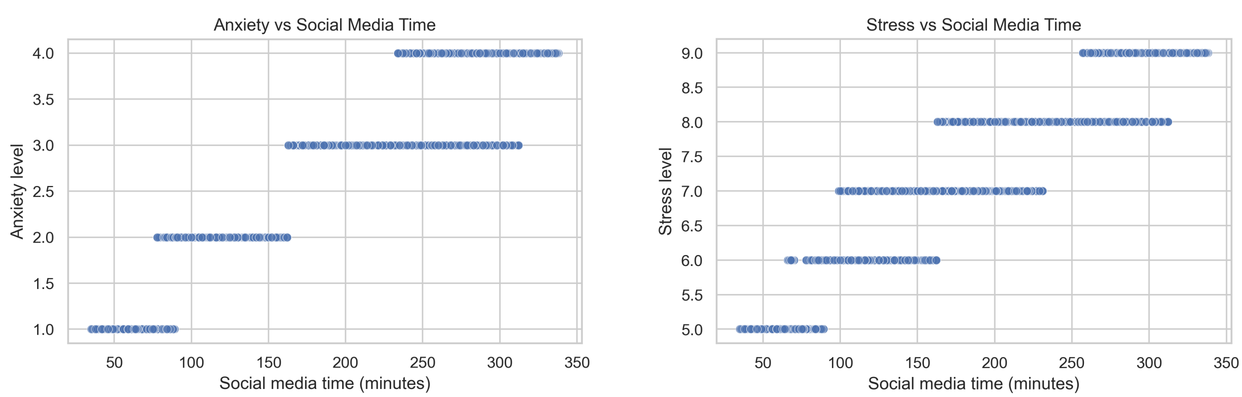


Figure 3: Correlation matrix with numeric coefficients for all numeric variables.

Insight: Social media time and negative interactions correlate positively with stress and anxiety, while sleep and physical activity show negative correlations with these outcomes and positive correlations with mood. Based on this, we decided to focus more on screen time, sleep, and interaction quality.

4.4 Key Relationships

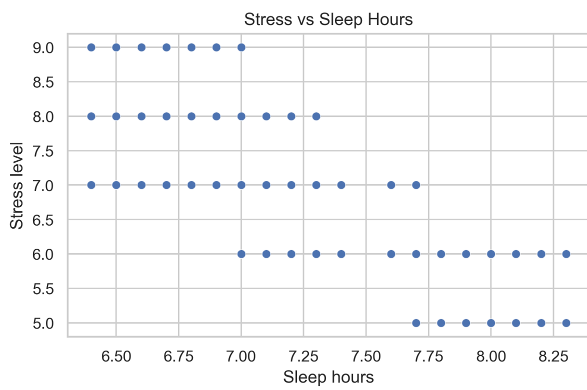


(a) Anxiety vs daily social media time

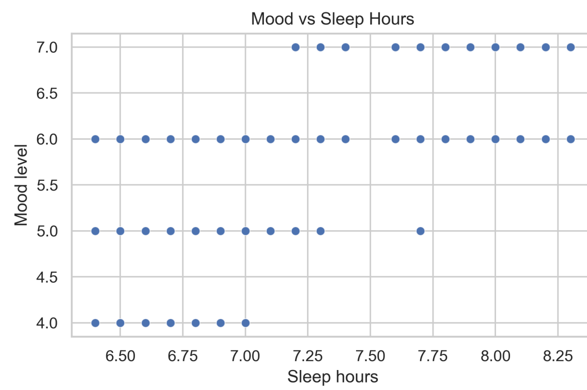
(b) Stress vs daily social media time

Figure 4: Relationships between social media use and mental health scores.

Insight: Both anxiety and stress increase with social media time. The horizontal bands reflect the discrete scale of the scores, but the overall upward trend is visible.



(a) Stress vs sleep hours



(b) Mood vs sleep hours

Figure 5: Sleep-related relationships with stress and mood.

Insight: Participants who sleep more tend to report lower stress and higher mood. This points to sleep as a possible buffer against digital stressors.

5 Data Storyboard

Our storyboard uses **horizontal logic** (from context to possible mechanisms) and **vertical logic** (headline, visual, takeaway). Each slide is shown below as it appears in the dashboard, followed by its role in the story and the main insight it is meant to convey.

5.1 Slide 1 — Who Is Online & How Much Time Do They Spend?

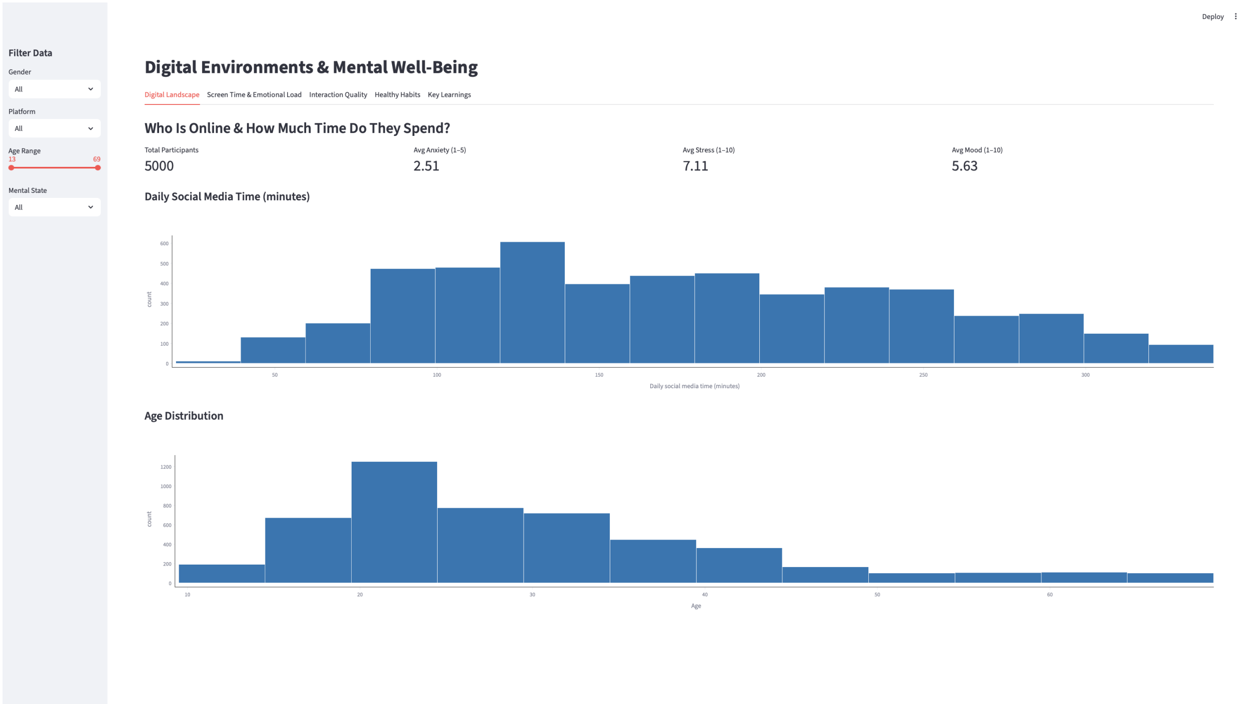


Figure 6: Slide 1 — Digital landscape overview.

Slide 1 introduces the demographic and behavioural baseline. The histograms show that most participants are young adults and that one to three hours of daily social media use is completely normal in this group. This starting point helps the audience see that later results do not come from a small or extreme subgroup but from a fairly typical high-usage population.

Key message: high screen-time is common, especially among younger users, and is the starting point for our analysis rather than an exception.

5.2 Slide 2 — Does More Screen Time Mean Higher Emotional Load?

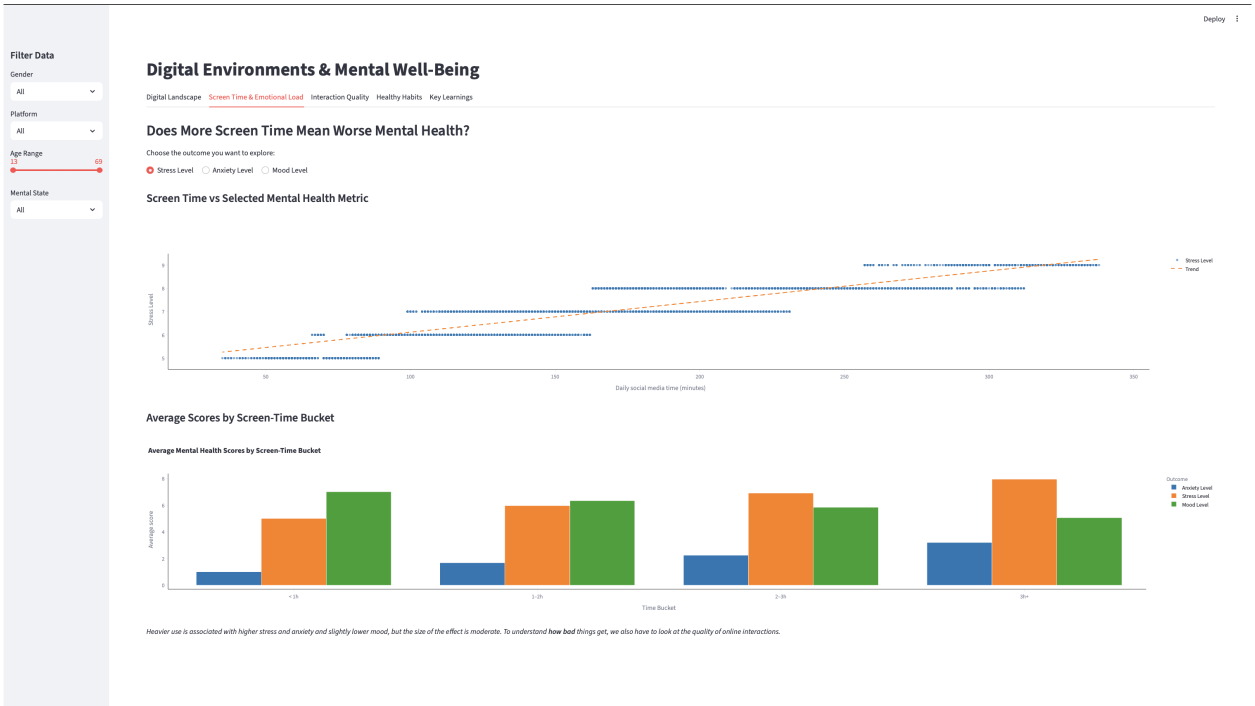


Figure 7: Slide 2 — Screen time vs stress and mood.

Slide 2 moves from context to impact. A scatterplot with a trend line shows the positive relationship between daily social media time and stress levels and the slight decline in mood scores. Here the focus is on the direction of the pattern rather than exact values: as the dots move to the right (more minutes online), the fitted line for stress clearly moves up and the one for mood moves down.

Key message: more time online is associated with higher stress and slightly lower mood across the whole sample.

5.3 Slide 3 — When Does It Become Harmful? Screen-Time Buckets

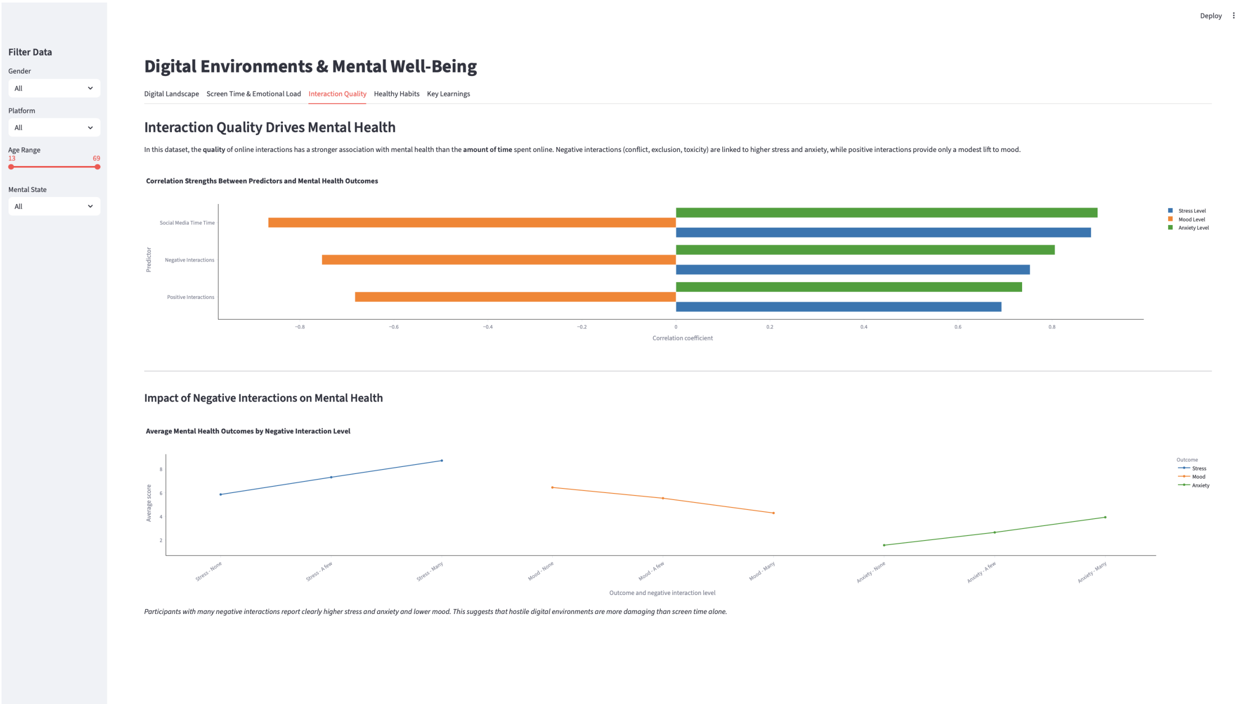


Figure 8: Slide 3 — Average mental health scores by screen-time bucket.

Slide 3 turns the continuous pattern into something more actionable. By grouping users into four screen-time buckets (under 1h, 1–2h, 2–3h, 3h+), we can compare average scores directly. The bars show that stress and anxiety are clearly higher in the 2–3h and 3h+ groups, while mood drops, which makes it easier to talk about approximate thresholds rather than only "more or less".

Key message: negative effects become visible once daily use is above roughly two hours, which is a useful rule-of-thumb for recommendations.

5.4 Slide 4 — Interaction Quality Matters More Than Time

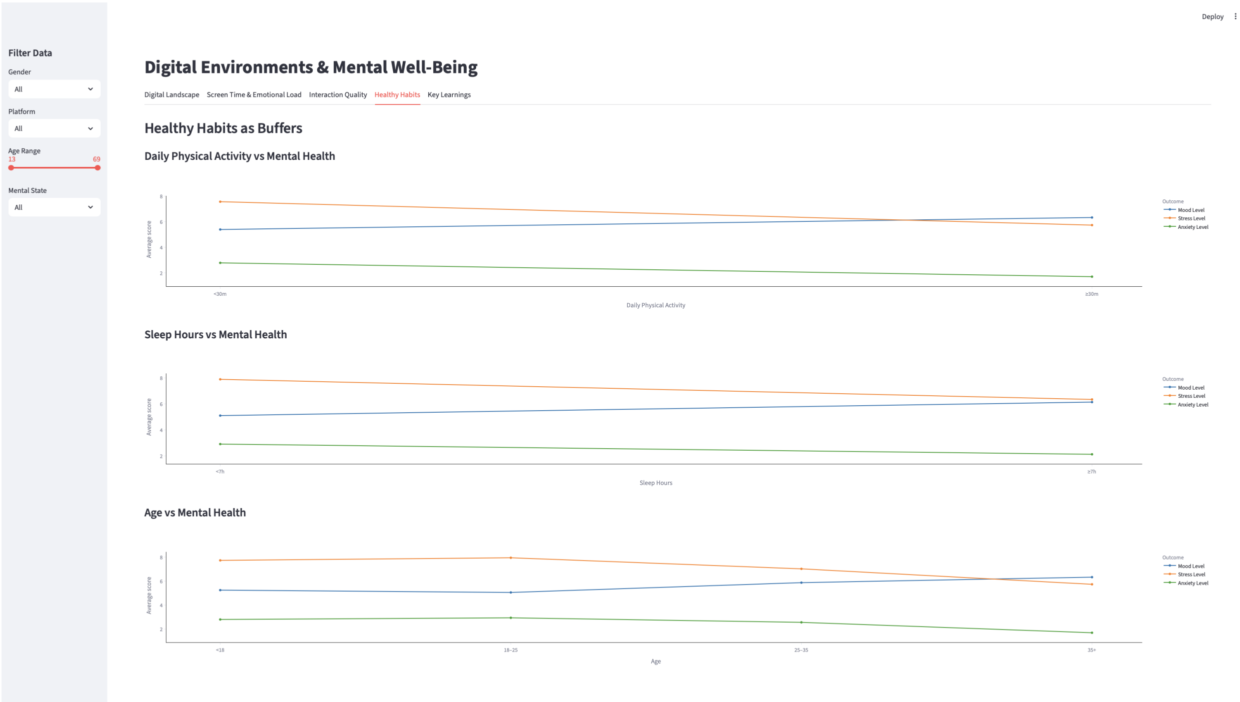


Figure 9: Slide 4 — Impact of interaction quality on mental health.

Slide 4 adds nuance to the story by focusing on what happens during online time. The correlation bars and line chart show that negative interactions (conflict, exclusion, toxicity) are strongly linked to higher stress and anxiety and lower mood, even when overall screen time is similar. Positive interactions help, but their effect is noticeably smaller, which suggests that simply "spending time with friends online" is not enough if the tone of those interactions is hostile.

Key message: the emotional tone of interactions is at least as important as total time online and explains why some heavy users are affected more than others.

5.5 Storyboard Summary

Taken together, the slides show a simple progression:

1. who is online and how much they use social media (Slide 1),
2. that higher use is linked to higher stress and lower mood (Slide 2),
3. that problems become clearly visible once use exceeds about two hours per day (Slide 3),
4. and that negative interactions are a key mechanism behind these outcomes (Slide 4).

This sequence connects basic descriptive facts with more specific patterns and ends with a concrete lever for action: reducing exposure to hostile interactions and keeping daily use within a reasonable range.

6 Narrative Section

In a live presentation we would proceed as follows.

Slide 1 — Digital Landscape. We start with who is in the sample and how much time they spend online. This gives the audience a reference point before any interpretation of mental health scores.

Slide 2 — Screen Time & Stress. Next, we show the scatterplot with the trend line. We briefly point out the slope and mention that these are associations, not proof of causality.

Slide 3 — Screen Time Buckets. We then move to the bar chart by screen-time bucket. The focus here is on the jump in stress and anxiety beyond two hours per day and the gradual decline in mood.

Slide 4 — Interaction Quality. Finally, we show the slide on interaction quality and explain that negative interactions seem to drive a large part of the emotional burden.

7 Interactive Filters and Dashboard Modes

The dashboard is not only a static story but also an interactive tool. A filter panel on the left lets users adapt the visuals:

- **Gender (dropdown):** compare male and female participants.
- **Platform (dropdown):** focus on a specific social media platform.
- **Age range (slider):** restrict the analysis to a narrower age band, for example teenagers.
- **Mental state (dropdown):** filter by overall mental state (*Healthy*, *At_Risk*, *Stressed*).

All filters default to *All*. This keeps the first view simple and consistent with the story described above. Afterwards, users can explore subgroups that matter to them.

To make the dashboard easier to use in different contexts, it is available in both **dark mode** and **light mode**. Dark mode is used in the screenshots because it works well on screens. Light mode offers higher contrast on white backgrounds and is more suitable for printing or use with projectors. Only the colour theme changes; the data and visuals stay the same.

8 Challenges and Lessons Learned

Working with a synthetic dataset was the first challenge. The data looks realistic, but we had to be careful not to oversell the findings and to phrase them as patterns in this particular dataset. Deciding which visuals to keep was another step. We produced more exploratory charts than we finally used and tried to keep only the ones that supported the main message.

On the design side we had to balance colour contrast and simplicity, and we adjusted layouts several times to improve readability. The project was also a reminder that correlations and trends are useful for communication, but they do not prove causality. For that reason, the dashboard focuses on associations and possible practical recommendations rather than strong causal claims.

Appendix

The source code used to clean the data, generate exploratory plots, and produce all figures in this report is contained in the file `Data_Exploration.ipynb`. The notebook is written in Python and uses `pandas`, `matplotlib`/`seaborn` for static graphics, and `plotly` together with `Streamlit` for the interactive dashboard. Running the notebook on the original CSV file reproduces the figures in a transparent way.