Milestone 3: Near-Final Draft Analysis Report

Project Title: Analyzing the Impact of Daily Screen Time on Mental Health
(This is an individual project.)

GitHub : https://github.com/Chandana3940/mental-health-screen-time.git

1. Data Overview

The dataset used is the **"Mental Health and Technology Usage Dataset"** obtained from Kaggle. It contains **10,000 entries** and **14 variables**, including demographic information, technology usage behavior, and mental health indicators. No missing or duplicate values were identified, so the data was ready for analysis without imputation.

For modeling, the **Mental_Health_Status** variable was encoded into a numeric score (**MH_Score**) from 1 to 4:

- 1 – Poor
- 2 – Fair
- 3 – Good
- 4 – Excellent

This allowed the use of statistical tests and regression models.

2. Methods and Additional Analysis

Milestone 3 extends the exploratory work from Milestone 2 by adding **group comparisons**, **formal hypothesis testing**, and a **multiple regression model**.

1. **Group Means by Mental Health Status**
    - Average values of screen time, sleep, physical activity, gaming, and social media hours were calculated for each mental-health category (Poor, Fair, Good, Excellent).
2. **Kruskal–Wallis Non-Parametric Test**
    - Screen_Time_Hours for each mental-health group was compared using the Kruskal–Wallis test to check if the distributions differ significantly between groups.
3. **Multiple Linear Regression**
    - A regression model was built to predict **MH_Score** using:
        - Screen_Time_Hours, Sleep_Hours, Physical_Activity_Hours, Gaming_Hours, Social_Media_Usage_Hours
    - Model performance was evaluated using R^2 and Mean Squared Error (MSE).
4. **Residual and Prediction Diagnostics**

- A **Predicted vs Actual** scatterplot and a **Residuals vs Predicted** plot were created to inspect model fit and linearity.

5. **Stress-Level Analysis**
    - The distribution of **Stress_Level** (Low, Medium, High) was visualized with a count plot.
    - This provides context for how balanced the stress categories are across participants.

Mental_Health_Status	Screen_Time_Hours	Sleep_Hours	Physical_Activity_Hours	Gaming_Hours	Social_Media_Usage_Hours
Excellent	7.95	6.48	5.05	2.54	3.93
Fair	7.92	6.51	4.98	2.49	4.02
Good	8.09	6.50	4.96	2.50	3.99
Poor	7.94	6.52	5.02	2.52	3.95

3. Figures and Visualizations

Figure 1. Average screen time, sleep, physical activity, gaming hours, and social media usage by mental-health status

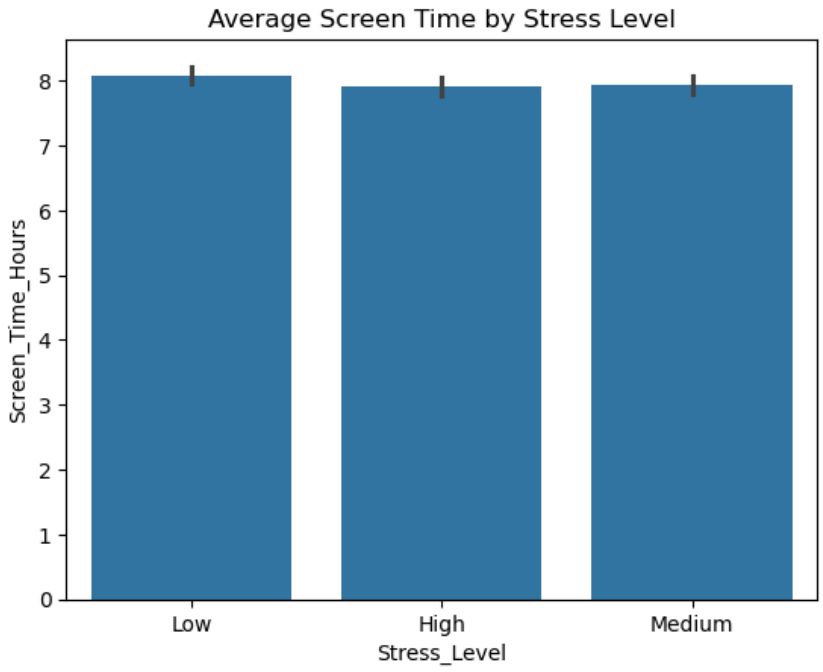

Figure 2. Distribution of stress levels

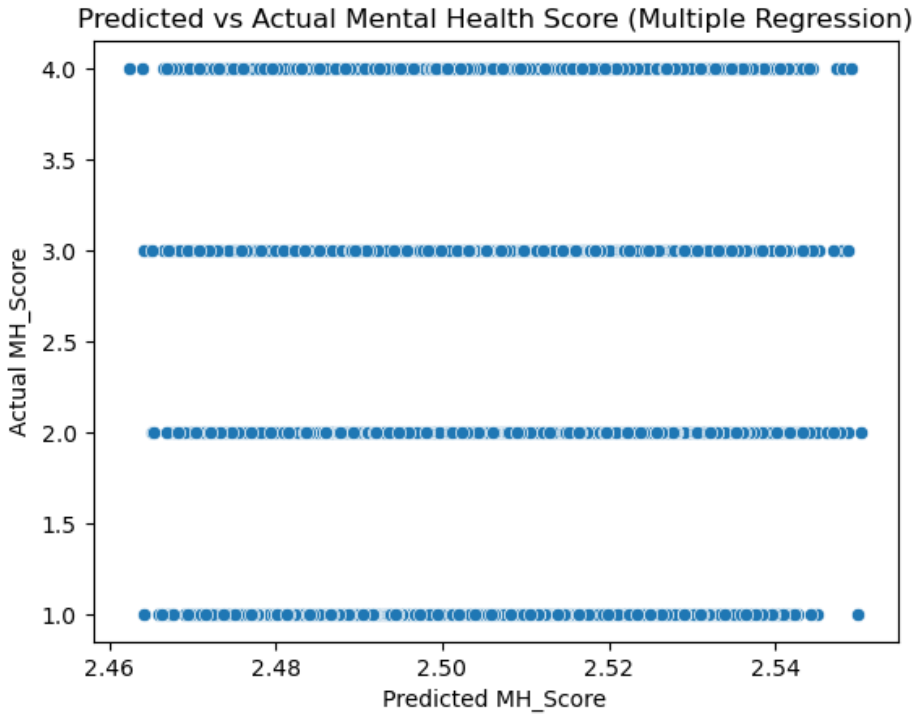

Figure 3. Predicted vs actual mental health score (MH_Score) from the multiple regression model.

4. Key Findings

- **Group Means:**
  The group means table shows that **screen time, sleep, physical activity, gaming hours, and social media usage are very similar** across all four mental-health categories. There is no strong visual separation between "Poor", "Fair", "Good", and "Excellent" groups.

- **Kruskal–Wallis Test:**
  The Kruskal–Wallis test for Screen_Time_Hours across the four mental-health groups produced a **high p-value**, meaning we **fail to reject the null hypothesis**. In simple terms, there is **no statistically significant difference** in screen-time distributions between the mental-health groups.

- **Multiple Regression Performance:**
  The multiple linear regression model using all behavioral predictors (screen time, sleep, activity, gaming, social media) yielded a **very low R^2** and relatively high error. This indicates that the model explains **almost none of the variance** in MH_Score.

- **Predicted vs Actual & Residuals:**
  The predicted vs actual plot shows that predicted values are clustered in a narrow range and do not track the four MH_Score levels well.
  The residuals vs predicted plot shows **horizontal bands with no clear trend**, confirming that the linear model is not capturing meaningful structure in the data.

- **Stress Levels:**
  The stress-level distribution is fairly balanced across Low, Medium, and High categories, so there is no severe class imbalance. However, differences in sleep and physical activity by stress level are **small**, and they still do not translate into strong predictive power for mental health status.

5. Interpretation

The analysis confirms that:

- **Daily screen time is not a strong predictor of self-reported mental health** in this dataset.
- Even when additional behavioral variables (sleep, physical activity, gaming, social media usage) are included, the model still performs poorly.
- Both the statistical test (Kruskal–Wallis) and the regression diagnostics show that the relationship between these behaviors and mental health is **weak or non-existent** in the available data.

This suggests that mental health is likely driven by **many other factors**—such as personal history, life events, social support, and clinical conditions—that are not captured here. Therefore, this dataset alone is not sufficient to build an accurate predictive model of mental-health status.

6. Early Conclusions and Next Steps

For the final project stage, the plan is to:

- Experiment with **non-linear models** (e.g., Random Forests, Gradient Boosting) to see if they capture small, non-linear patterns that linear regression misses.
- Explore **interaction effects**, such as combinations of high screen time and low sleep, to check whether they relate more strongly to mental health.
- Consider **feature engineering** or incorporating **additional datasets** with richer psychological or clinical variables.

These steps will help refine the analysis and move from exploratory modeling to more interpretable and robust conclusions.