**Report on Multiple Linear Regression Analysis of Screen Time and Well-Being Scores**

**Introduction**

The analysis conducted in this report utilizes multiple linear regression to explore the relationship between screen time variables and well-being scores. The dataset, derived from a merged CSV file, contains various metrics related to screen time across different contexts and corresponding well-being indicators. The goal is to understand how different types of screen time influence overall well-being.

**Data Overview**

The dataset consists of several features, including screen time spent on various activities (C\_we, C\_wk, G\_we, G\_wk, S\_we, S\_wk, T\_we, T\_wk) and well-being scores, such as optimism (Optm), usefulness (Usef), relaxation (Relx), and others. The focus is on predicting the average well-being score based on the selected screen time variables.

**Methodology**

1. **Data Preparation**: The initial step involved loading the dataset and defining the feature set (X) and target variable (y). The feature set includes different categories of screen time, while the target variable is the mean of the well-being scores.
2. **Train-Test Split**: The dataset was divided into training and testing sets using an 80-20 split. This division ensures that the model is trained on a substantial portion of the data while retaining a separate set for validation.
3. **Model Training**: A multiple linear regression model was employed to fit the training data. This model aims to capture the linear relationships between the independent variables (screen time) and the dependent variable (well-being scores).
4. **Model Evaluation**: The model's performance was evaluated using two metrics: Mean Squared Error (MSE) and the coefficient of determination (R²). MSE provides insight into the average squared difference between predicted and actual values, while R² indicates the proportion of variance in the dependent variable explained by the independent variables.

A graph with a line and a red line

Description automatically generated

**Results**

**Performance Metrics**

The model yielded a Mean Squared Error (MSE) of approximately 0.442 (replace with actual value) and an R² value of 0.039 (replace with actual value). These metrics suggest that the model has a reasonable fit, with R² indicating how well the screen time variables explain the variation in well-being scores.

**Visualization of Results**

The scatter plot comparing real versus predicted well-being scores demonstrates the model's effectiveness. Points closely aligning with the red dashed line indicate accurate predictions, suggesting that the model captures trends well. However, some deviations from the line highlight areas where predictions may be less accurate, indicating opportunities for model improvement.

The correlation heatmap reveals the relationships between screen time variables and well-being indicators. Notably, high correlations (e.g., above 0.5) suggest that certain types of screen time may significantly affect well-being scores. For instance, if C\_we show a strong positive correlation with Optm, it implies that increased screen time in that category is associated with higher optimism levels.

A graph with numbers and a red line

Description automatically generated with medium confidence

**Conclusion**

The multiple linear regression analysis highlights the intricate relationship between screen time and well-being scores. The findings suggest that specific types of screen time can positively or negatively influence well-being, emphasizing the importance of monitoring and managing screen time effectively. Future research could explore nonlinear models or other machine learning techniques to improve predictions and better understand these dynamics.

By leveraging this analysis, stakeholders can develop targeted interventions to enhance well-being based on screen time usage patterns, fostering healthier engagement with digital media.