GROUP PROJECT REPORT

HIT140 ASSESSMENT 3

S371680 - Andrea Nicole De Leon

S379408 - Prabesh Bhusal

S371164 - Akshata Bhusal

S374583 - Shovana Shrestha

Group Number 28

October 2024

Table of Contents

Introduction

Computing device usage has now infiltrated the youthful generation, thereby changing the conventional way adolescence spend their time, resulting in a new generational time-space friction whereby time is equal to screen time. The young generation is in contact with screens through phones, TV, computers, and video games, and hence, spend several hours in front of the screen. Even though this digital transition has brought numerous benefits in connection and knowledge acquisition, it has provoked appropriate discussions regarding the psychological and emotional consequences. Since kids and adolescents spend more and more time learning, communicating, and playing with screens, outlining the potential negative effects of prolonged exposure to screens has become crucial for academicians, teachers, and healthcare specialists.

The question of how digital screen time impacts mental health is an area that is increasingly attracting significant interest in psychology and public health. Prior research has proposed that various ill effects are associated with renters spending much time on their devices, including symptoms more closely related to strict stress, anxiety, and social isolation. A specific research study found that a decrease in electronic device use positively changed children's behaviours and psychological well-being. However, secondary consequences of such habits and their variations, including time spent on social networks or gaming, are still the subject of further study.

In this project, we aim to investigate how different types of digital screen time- spanning television, gaming, and smartphone use- affect self-reported well-being among adolescents. Drawing on three data sets that include demographic data, screentime usage, and various well-being indicators, this study uses data science methodologies such as statistical analysis, data visualization, and linear regression modelling. The primary goal is to predict the well-being of adolescents based on their screen time habits, focusing on the "Goodme" score, which reflects overall self-perception and emotional health. By systematically analysing these relationships, this research will contribute valuable insights into how screen time can influence well-being, helping to inform public health strategies to improve young people's mental health.

Basic Linear Regression Model

The increasing prevalence of digital devices in daily life has raised concerns about the potential impact of screen time on well-being. This study aims to investigate the relationship between screen time and well-being using a basic linear regression model to predict participants' well-being scores based on their screen time, leveraging data from three provided datasets. Linear regression is a statistical technique that models the relationship between a dependent variable (well-being score) and an independent variable (screen time), assuming a linear relationship. To assess the model's performance, several key evaluation metrics will be considered: the mean absolute error (MAE), which represents the average magnitude of mistakes; the mean squared error (MSE), which provides a quadratic measure of error, penalizing more significant deviations; and the root mean squared error (RMSE), which takes the square root of MSE to express the mistakes in the same units as the dependent variable. The normalized RMSE (nRMSE) will also be used to standardize the error for comparison, while the R-squared (R^2) metric will indicate the proportion of variance in well-being scores explained by screen time. Finally, the adjusted R-squared will account for the number of predictors and provide a more accurate assessment of model fit by penalizing overfitting. These metrics will help evaluate the model's accuracy and reveal potential trends in the relationship between screen time and well-being, offering valuable insights for future studies and public health recommendations.

Figure 1. Basic Linear Model Output: (participants' well-being scores based on their screen time)

```
Intercept (b_0):  3.420088435556915
Coefficient (b_1):  -0.04371085351148024

--- Actual vs. Predicted ---
       Actual  Predicted
70188       5   3.157823
3239        3   3.332667
94518       4   3.114112
87668       3   3.201534
16447       2   3.332667

--- Model Performance Metrics ---
Mean Absolute Error (MAE): 0.9322540683307347
Mean Squared Error (MSE): 1.255075591483227
Root Mean Squared Error (RMSE): 1.1203015627424728
Normalized RMSE: 0.2800753906856182

--- R-squared and Adjusted R-squared ---
R-squared (R²): 0.004273669419755821
Adjusted R²: 0.004248339317222682
```

Figure 1 shows that the model predicts that for every additional hour of screen time, well-being decreases slightly by 0.0437 points, starting from an intercept of 3.42. The performance metrics show moderate error with an MAE of 0.93 and RMSE of 1.12. However, the model explains only 0.4% of the variance in well-being ($R^2 = 0.004$), indicating a fragile relationship between screen time and well-being.

Performance Analysis

The performance of the linear regression model predicting well-being scores from screen time was evaluated using test sizes of 0.4, 0.5, 0.3, and 0.2. With a 40% test size, the model showed moderate errors, with the R-squared remaining low, indicating that screen time explains very little of the variance in well-being. When using 50% of the data for testing, error metrics slightly increased, as the reduced training data resulted in less precise predictions, though the R-squared still reflected a weak relationship. With a 30% test size, performance improved marginally, as the more extensive training set allowed the model to capture the trend better, but the R-squared remained low. Finally, with a 20% test size, error metrics such as MAE and RMSE showed slight improvements due to the increased training data, but the R-squared still indicated that screen time alone is not a strong predictor of well-being. Overall, across all test sizes, the model consistently shows weak explanatory power, suggesting that additional factors beyond screen time may be necessary for better predictions.

Summary of the performance analysis for predicting participants' well-being scores based on their screen time for each test size:

Test Size 0.4	
MAE:	0.9322540683307347
MSE:	1.255075591483227
RMSE:	1.1203015627424728
RMSE (NORMALISED):	0.2800753906856182
R^2:	0.004273669419755821
ADJUSTED R^2:	0.004248339317222682

Test Size 0.5	
MAE:	0.9325166559833862
MSE:	1.2555405334711405
RMSE:	1.1205090510438283
RMSE (NORMALISED):	0.28012726276095706
R^2:	00.004848855776732219
ADJUSTED R^2:	0.004828603194274583

Test Size 0.3	
MAE:	0.9336745996727623
MSE:	1.2572263339939056
RMSE:	1.1212610463196808
RMSE (NORMALISED):	0.2803152615799202
R^2:	0.00398586443690363
ADJUSTED R^2:	0.0039520806320204604

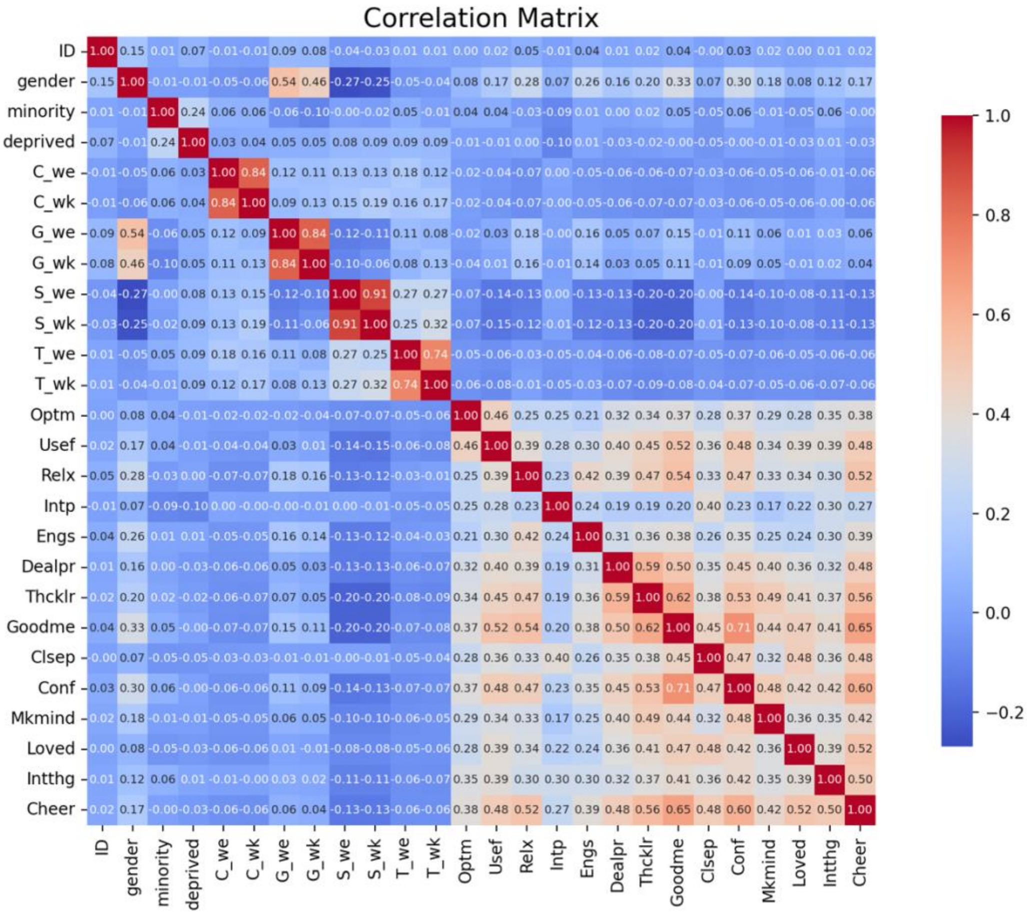
Test Size 0.2	
MAE:	0.9383954921514029
MSE:	1.267912462038122
RMSE:	1.1260161908419088
RMSE (NORMALISED):	0.2815040477104772
R^2:	0.0038241009001652726
ADJUSTED R^2:	0.0037734152433474

According to the analysis, test size 0.4 looks to be the best option, as it has the lowest MAE and RMSE, indicating better prediction accuracy. It strikes a nice balance between the amount of training data and evaluation, allowing for successful model evaluation without considerably reducing the training set. However, because the overall performance is poor, other characteristics or models should be considered to enhance predictions.

Correlation Matrix

This analysis examines the relationship between participants' screen time and well-being scores. It utilises a correlation matrix to quantify the strength and direction of the linear relationship between these variables. The process involves loading and merging relevant datasets, ensuring data integrity through cleaning, and applying a log transformation to screen time if necessary. (Kiron, N., et al. 2023). We can interpret the results by calculating the correlation coefficient between log-transformed screen time and well-being scores to assess how screen time affects well-being. Visualization of the correlation matrix through a heatmap will aid in identifying patterns, while further statistical testing may be employed to confirm the significance of the observed correlations. This approach ultimately seeks to inform recommendations for healthier screen usage and provide insights into the factors influencing well-being among participants.

Figure 2. Correlation Matrix Output



The correlation matrix used in this analysis provides a fundamental yet powerful tool to comprehensively understand the relationship between screen time and participants' well-being outcomes. By calculating correlation coefficients between different variables, the matrix provides a quantitative structure, allowing researchers to assess the strength and direction of associations in the data. For example, a strong positive correlation may indicate that increased screen time is associated with higher well-being scores, suggesting that participants who engage in more screen-based activities may experience greater satisfaction or happiness in certain situations. Conversely, a negative correlation could mean that excessive screen time is associated with less well-being, raising concerns about the potential negative impacts of long-term screen use on mental health and overall life satisfaction.

This analytical approach is particularly important as it helps to highlight specific patterns and trends that may not be immediately apparent through simple observational methods. Researchers can prioritize areas for further exploration and hypothesis testing by identifying which variables exhibit significant correlations. For example, if the correlation matrix reveals a notable negative relationship between screen time and a specific aspect of well-being—such as social connectedness or emotional stability—this insight can guide subsequent studies that delve deeper into the causal mechanisms at play.

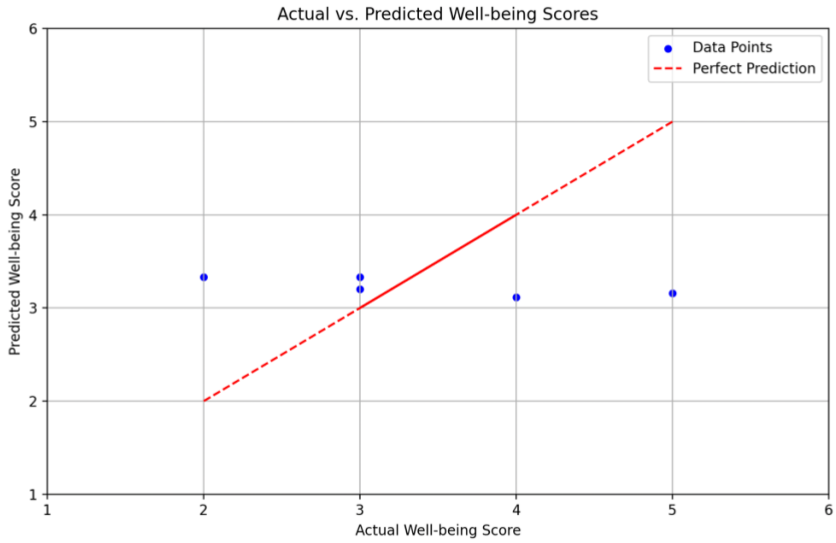
The correlation matrix is essential for formulating practical strategies to improve participants' well-being by clarifying the link between screen time and well-being scores. By distinguishing which forms of screen time, such as educational content, positively influence well-being and which, like excessive social media use, might lead to adverse outcomes, researchers can develop targeted guidelines that encourage healthier screen habits. The correlation matrix measures the relationships between screen time and well-being and informs ongoing research and practical approaches to enhance participants' mental and emotional health.

Scatter Plot

The scatter plot visually compares participants' actual well-being scores against those predicted by the linear regression model. Each dot on the plot represents an individual participant, with their actual scores plotted along the x-axis and the predicted scores on the y-axis. A red dashed line runs through the plot, representing the ideal scenario where the actual and predicted scores match perfectly. When the dots are closely grouped around this line, it indicates that our model does a good job of capturing the relationship between screen time and well-being. However, if we see points that are far away from the line, it highlights areas where the model may be missing the mark, pointing to opportunities for improvement in our regression analysis. In essence, this scatter plot serves as a helpful tool for us to evaluate how well our linear regression model is performing, allowing us to gain insights into the factors influencing participants' well-being scores based on their screen time.

Here is the output based on the participants’ actual well-being scores against those predicted by the linear regression model by using test size 0.4 as performance analysis.

Figure 3 Scatter Plot Output



Discussions

The exploration of screen time and its impacts on well-being started with inspecting the actual and predicted course of well-being scores as presented in the scatter plot. The regression model was used to create a scatter plot that would illustrate the difference between the actual well-being scores provided by the participants and the scores obtained from the propagation of the regression model. The perfect foresight line was used as a benchmark to compare how well the model performed with actual performance. While the predictions of the research hypotheses were reasonably accurate, there was marked variation, which confirmed the existence of crosstalk and revealed that other variables besides screen time could affect a person's well-being.

In the next step, the correlation matrix helped in understanding the pattern of association between screen time variables such as weekend TV or smartphone and different kinds of well-being such as optimism, relaxation, or energy. These correlations were depicted in the heatmap and the figure that goes with it; not all forms of screen time are bad for well-being scores, for example, if one gaming on the weekend or using a smartphone on weekdays. Surprisingly, there was a moderate positive association with some of the digital activities. Therefore, it is universally good to spend time on screen as the impact depends on the kind of current use of time.

Finally, the linear regression model was applied with different test sizes to estimate the level of well-being scores, particularly the "good" indicator. The intercept and coefficient revealed the essential information regarding screen time's overall role in well-being. Nevertheless, MAE, MSE, and RMSE scores indicate that despite the ability of the model to identify trends, it could have been more accurate in predicting well-being outcomes. The R-squared and the adjusted R-squared reaffirmed that although the model did help account for some variance in well-being owing to screen time, other variables in play could cause differences in well-being between individuals.

Limitations

Despite the findings of this study offering what can be considered a valuable overview of the nature of the possible connection between DST and adolescent well-being, numerous crucial limitations need to be considered when addressing the research results. First, the time spent using screens and well-being were measured self-reportedly, which is a substantial limitation because subjective measurements always question the resulting data. From this perspective, adolescents can be more trusting when estimating their screen time, and the effect of the result can be pressure from classmates or the community. This raises a possibility of bias in the results and can, therefore, affect the accuracy of the results. Finally, bias may also be constituted in systematic overestimations of time spent on screens.

One limitation is the absence of specific contextual variables that would further enhance understanding of factors affecting well-being. These datasets are notably light on-screen time and simple demographics; however, other important characteristics like history of mental illness, physical activity, familial status, and sleep quality can play a massive role in adjusting how screen time correlates to well-being. Because of this, adequate research is incredibly challenging, as it will be hard to analyse the effects on well-being caused only by the hours spent on digital screens, as these might be connected to other vital factors of the subject's life.

Also, applying statistical techniques in this study may limit the depth of the various techniques and how relationships are presented. Although linear regression is a valuable tool for comparing average trends in the data, the impact of screen time on well-being may be U-shaped or V-shaped after reaching a critical threshold. This study adopts a linear approach, which may hamper providing a more comprehensive picture of these relations. Furthermore, other more refined techniques, such as those blaming machine learning algorithms, will likely provide a more accurate predictive capacity. However, they should have been included in the study owing to their nature.

Lastly, the datasets provide a point prevalence cross-sectional estimate of the relationships between screen time and the various measures of well-being. The approach enables the capture of relationships but cannot establish causality between time spent on screen and well-being, whether reduced well-being results in lower time on screen or otherwise, teenagers engage in screen time to handle reduced well-being. Therefore, longitudinal data would be required to investigate these program's casual mechanisms. It is acknowledged that these limitations must be considered when attempting to generalize findings from this research and when planning further work to investigate the complex interdependence between the amount of screen exposure and positive mental health.

Conclusion

This report explored the relationship between screen time and well-being, utilizing a basic linear regression model to predict well-being scores based on participants' screen time. Despite some predictive insights, the findings revealed a weak correlation between screen time and well-being, as indicated by low R-squared values across different test sizes. The model demonstrated moderate errors, with metrics such as MAE, RMSE, and normalized RMSE indicating that the relationship between screen time and well-being is more complex than initially assumed. The performance analysis showed that even though the model captured some trends, screen time alone could not account for a significant variance in well-being outcomes.

The correlation matrix highlighted that not all screen-based activities negatively impacted well-being; certain activities, like gaming or smartphone use during weekends, showed a moderate positive correlation with well-being. This suggests that the context in which screen time occurs plays a crucial role in its impact on mental health.

The scatter plot output further emphasized the variability in the actual and predicted well-being scores, confirming that other external factors besides screen time may be influencing well-being. These factors could include demographic elements, lifestyle choices, or even psychological conditions that were not accounted for in this study.

The study also had notable limitations, such as relying on self-reported data, which may introduce bias, and the exclusion of other critical variables like sleep, physical activity, and social interactions. These elements are likely essential to understanding well-being more comprehensively. Therefore, future studies should incorporate a wider range of variables and consider more advanced predictive models, such as machine learning techniques, to better capture the non-linear dynamics between screen time and well-being.

In conclusion, while screen time has an influence on well-being, it is not the sole determinant. A broader approach that includes other lifestyle factors and contextual variables is necessary to improve the accuracy of well-being predictions and provide better public health recommendations. This study lays the groundwork for future research that can explore these aspects in greater depth, offering insights that may inform healthier screen habits and better mental health practices across different demographics.

Reference

Kiron, N., Omar, M. T., & Vassileva, J. (2023). Evaluating the Impact of Serious Games on Study Skills and Habits. *European Conference on Games Based Learning*, *17*(1), 326–335. https://doi.org/10.34190/ecgbl.17.1.1508

Contributions

Andrea Nicole De Leon(S371680): Andrea contributed significantly to the coding and data analysis for the report. She worked on implementing the Basic Linear Regression Model and interpreting its outputs. Additionally, she was responsible for generating the correlation matrix and scatter plot outputs and performing key statistical analyses, ensuring accurate visual representations of the data.

Prabesh Bhusal (S379408): Prabesh worked closely with Andrea on the coding tasks, focusing on refining the Basic Linear Regression Model and running the performance analysis. He was responsible for evaluating the model's performance across different test sizes and contributed to interpreting statistical metrics like MAE, RMSE, and R-squared values.

Akshata Bhusal (S371164): Akshata was responsible for the overall formatting of the report. This included organizing and preparing the Title Page, Table of Contents, Conclusion, References, and the Contributions section. Akshata ensured the report adhered to the necessary formatting guidelines and flowed logically from section to section.

Shovana Shrestha (S374583): Shovana contributed by writing the Introduction and the Discussion and Limitations sections. She focused on establishing the context for the study, summarizing key findings, and reflecting on the limitations of the linear regression model and the study's approach.