**Predicting Well-being Scores Based on Screen Time**

**Assignment – 03**

**HIT140- Foundations of Data Science**

**Team Members**

Mohammed Nurul Hasan (S376146) (Data Wrangling, Descriptive Statistics)

Praggawn Sara (S377784) (Exploratory Data Analysis)

Raianul Islam (S377227) (Inferential Statistics)

Bidur Chautariya (S374277) (Regression Model)

Date: 18/10/2024

**Table of Contents**

1. Introduction
2. Data Preparation (Data Wrangling)
3. Descriptive Statistics
4. Exploratory Data Analysis (EDA)
5. Inferential Statistics
6. Regression Model
7. Discussion and Limitations
8. Conclusion
9. Individual Contributions
10. References

**List of Figures**

1. Figure 1: Boxplot - Computer Use on Weekends by Gender
2. Figure 2: Scatter Plot - Computer Use on Weekends vs Optimism
3. Figure 3: Correlation Heatmap - Screen Time and Well-being Indicators
4. Figure 4: Actual vs Predicted Optimism Scores
5. **Introduction**

In this project, participants’ well-being scores (i.e. optimism, relaxation, cheerfulness) are explored and predicted using their screen time habits gathered from surveys. The main aim is to explore how various types of screen time (computer screen, smartphone, television) relate to well-being measures, and develop a regression model to predict optimism by using these measures. The tasks include data wrangling, descriptive and inferential statistics, exploratory data analysis (EDA) as well as building a regression model.

1. **Data Preparation (Data Wrangling)**

The dataset consists of three separate CSV files that were merged using a common identifier (ID). After merging, we addressed missing values by imputing the column means for numerical data, ensuring that no rows were removed due to missing values.

A screen shot of a computer program

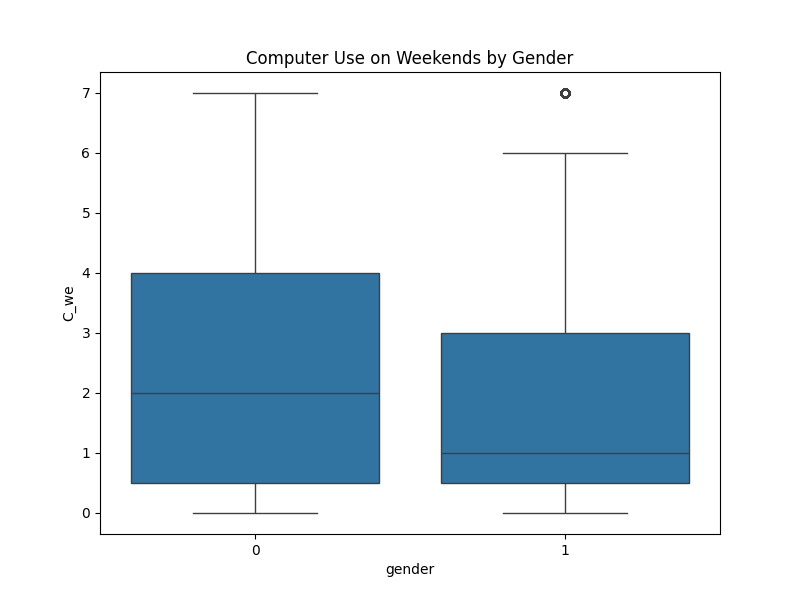
Description automatically generated

**Actions:**

* Datasets merged on ID.
* Missing values filled using column means to maintain data integrity.
* The cleaned dataset was used for subsequent analysis.

1. **Descriptive Statistics**

The central tendencies and variability of the dataset are summarized in descriptive statistics. We also classify screen time usage by gender, shown below. The difference in computer usage between male and female on the weekends was visualized using a boxplot.

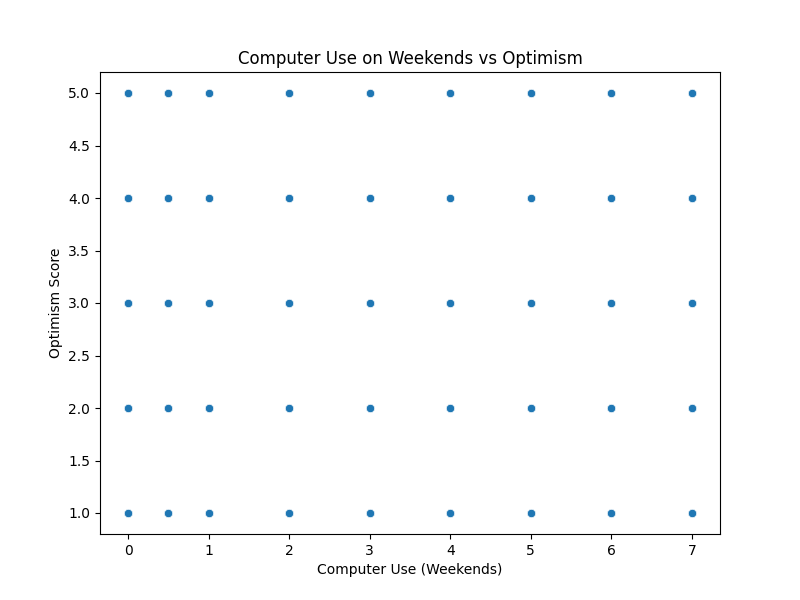


**Figure 1: Boxplot**

The boxplot shows that males (represented as 0) generally have a broader range of computer use on weekends compared to females (represented as 1). While the median computer use for males is slightly higher, the distribution varies more widely, with outliers indicating that some males spend significantly more time on the computer during weekends.

1. **Exploratory Data Analysis (EDA)**

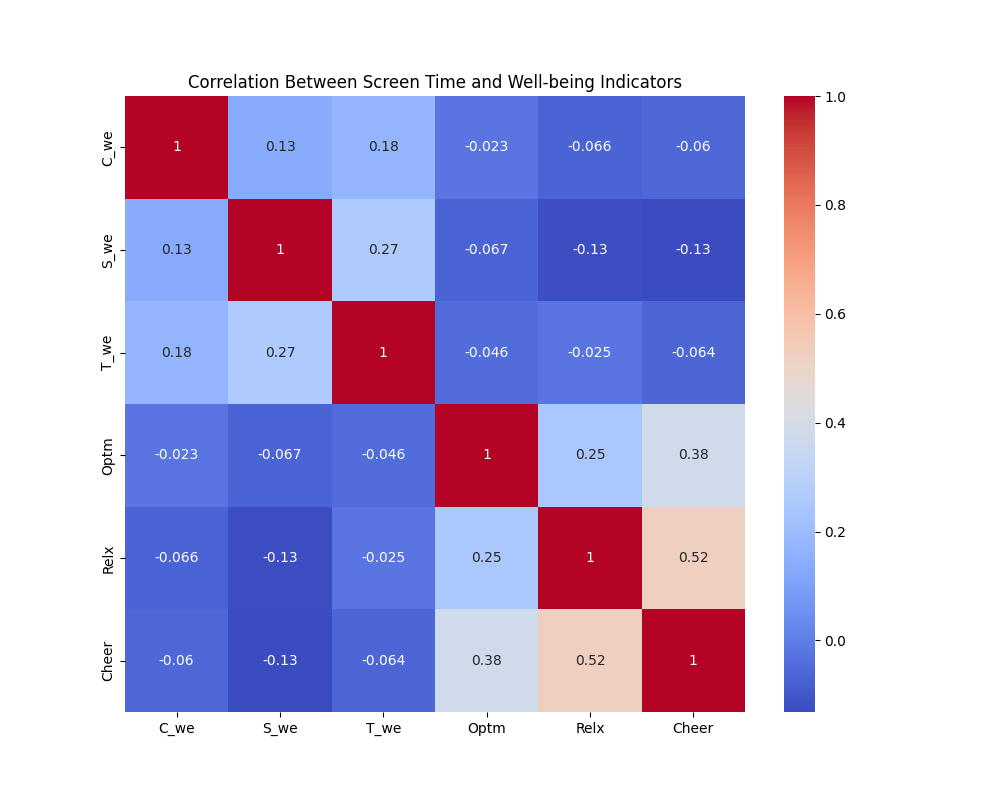
**Scatter Plot:**

****

**Figure 2: Scatter Plot**

The relationship between the score of computers use on weekends and optimism scores is visualized in scatter plot. Optimism shows no clear trend, and computer use does not seem to enhance optimism.

**Correlation Heatmap**:

****

**Figure 3: Heatmap**

The heatmap shows correlations between different screen time variables and well-being indicators:

* Smartphone use (S\_we) has the strongest correlation with well-being indicators like **relaxation** and **cheerfulness**.
* Computer use (C\_we) and TV use (T\_we) have weaker correlations with optimism, relaxation, and cheerfulness.
* This suggests that smartphone use may have a more substantial impact on overall well-being compared to other forms of screen time.

1. **Inferential Statistics**

A **t-test** was conducted to determine whether there is a statistically significant difference in optimism scores between males and females.

A screen shot of a computer code

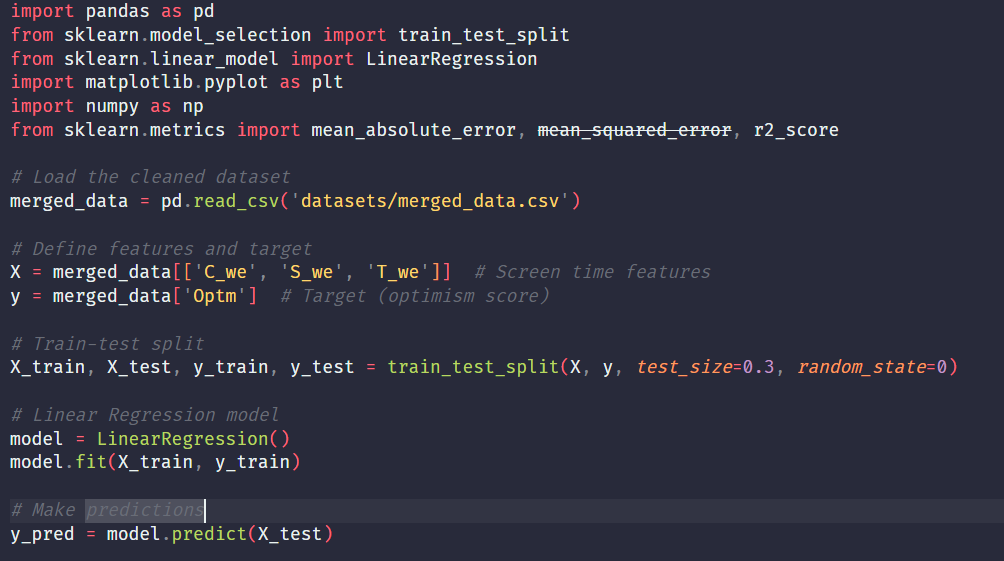
Description automatically generated

**Result:**

* T-statistic: 25.37
* **P-value**: ~0.000 (extremely small)
* There is a statistically significant difference between the optimism scores of males and females. Females generally score higher in optimism, as indicated by the histogram and t-test results.

1. **Regression Model**

A linear regression model was built to predict optimism based on screen time variables (computer, smartphone, and TV usage). The model’s performance was evaluated using **MAE**, **MSE**, **RMSE**, and **R²**.

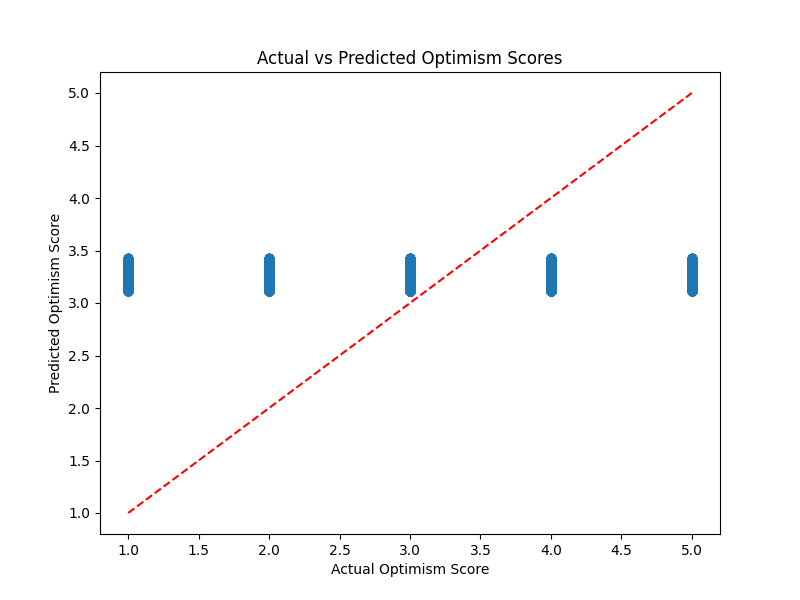


A screen shot of a computer program

Description automatically generated

**Model Evaluation:**

* **MAE**: 0.81
* **MSE**: 0.99
* **RMSE**: 0.99
* **R²**: 0.005 (Very low)
* Actual vs Predicted (Figure 4):



**Figure 4: Actual vs Predicted**

When mimicking the relationship between screen time on optimism, the regression model explains a mere 0.5%, suggesting screen time alone is not a strong predictor of optimism. R² value is low because the predicted values don’t closely align with actual optimism scores in Actual vs Predicted plot.

1. **Discussion and Limitations**

By analysing the screen time with the same well-being indicators, we can see some interesting trends, namely that smartphone use correlated more with well-being indicators like relaxation and cheerfulness. However, there are several limitations to this analysis:

* **Small Predictive Power**: The regression model had a very low R² value (0.005), indicating that screen time alone is insufficient to predict well-being outcomes like optimism.
* **Missing Factors**: Other factors influencing well-being, such as social interactions, sleep quality, or mental health status, were not included in the dataset.
* **Data Limitations**: The dataset may have sample biases, and missing data was imputed with means, which may not reflect true values.

1. **Conclusion**

In this project I explored screen time and well-being. Screen time was a weak predictor of optimism, but analysis finding significant gender differences in optimism. The findings that smartphone use is more strongly related to well-being indicators of relaxation and cheerfulness than television or print media use implies that there may be qualitative differences in the relationship between different types of screen time and well-being, although this needs further investigation.

1. **Individual Contributions**

* **Nurul Hasan**: Conducted data wrangling and performed descriptive statistics.
* **Praggawn Sara**: Conducted exploratory data analysis (EDA), including visualizations like scatter plots and the correlation heatmap.
* **Raianul Islam**: Performed inferential statistics, including t-tests comparing optimism between genders.
* **Bidur**: Developed the regression model and evaluated its performance.