**Foundations of Data Science**

**HIT140 - Assignment 3:** **Group Report**

**SYDNEY CAMPUS**

**STUDENT NAME: Roll Number:**

ARBAAZ ALI MOHAMMED S374309

AFAQ RAFIQUE S376356

SABIR ALI S376220

MUHAMMAD UZAIR KHAN S375170

**SUBMITTED TO: JOHN DENISON**

**Table of Contents**

1. Introduction
2. Data Exploration and Preparations
   * 2.1 Understanding the Dataset
   * 2.2 Data Cleaning Management
3. Data Visualization
4. Inferential Analysis
   * 4.1 T-Test Analysis
   * 4.2 Multiple Linear Regression Analysis
5. Conclusion & Recommendations
   * 5.1 Summary of Findings
   * 5.2 Recommendations
6. Ethical Considerations
7. Future Directions
8. List of References
9. Brief Descriptions of Individual Contributions

**Group Report:**

**Predicting Well-Being Based on Screen time**

1. **Introduction**

This is the most important goal of this project: make a prediction on study participants’ well-being scores using the data given and identified amount of screen time. The various exploratory procedures, visualizations, and linear modeling done in this report presents what this particular team was able to do in an attempts to make sense of the various datasets provided. In this case, ethical considerations have been observed in the project description and has respected the purity of the databases used in the study. The report also contains elaborate measures in data cleansing and handling, the results of which are shown in the analyses and recommended on the next page.

**2. Data Exploration and Preparations**

**2.1 Understanding the dataset**

The actual data set has several dependant variables describing various aspects of the quality of life and screen time variables – the time spent in front of the TV and the time spent on the computers. Our focus in this analysis is the prediction of well-being using the key variables:

* Well-Being Measures: Some of the measures include the Optimism variable, the Usefulness variable, the Relaxation variable, the Interest-in-other-people variable, the Energy variable, the Confidence variable, the Ability to handle the problems variable as well as others.
* Screen Time Variables: These include TV use, weekdays and weekend and computer use, again, weekdays and weekend.

**2.2 Data Cleaning Management**

When the set of data was reviewed by the team, the following measures were recommended for getting the data to the analysis stage.

1. **Handling of missing values.**

Our investigation of missing values involved evaluating all variables for absence of such values. Regarding missing values in variable levels, we talked about deleting those observations if it is a small number or filling in the missing multiples by mean or median value. Much debate ensued as to which variable transformation to apply for more sophisticated handling of the skewed distributions and we agreed to use median imputation.

1. **Outlier Detection and Treatment:**

Notably, box plots and Z-scores were used as techniques for identifying out of range screen time and well being. Things that were very far from the mean and could affect our results tended to be deleted or winsorized so as to eliminate the effect of such a point but not alter the authenticity of the data set.

1. **Data Transformation:**

Since the variable such as screen time had very skewed distribution the data was normalized by transforming it via log. This was a significant learning process in variance stabilization towards normalization of predictors in our regression model.

1. **Feature Engineering:**

Several new variables were formed out of existing features to enhance the predictive capability of the models that we developed. These included:

**TOTAL Screen Time:**

The total extent of TV and computer usage for the weekdays and for the weekends.

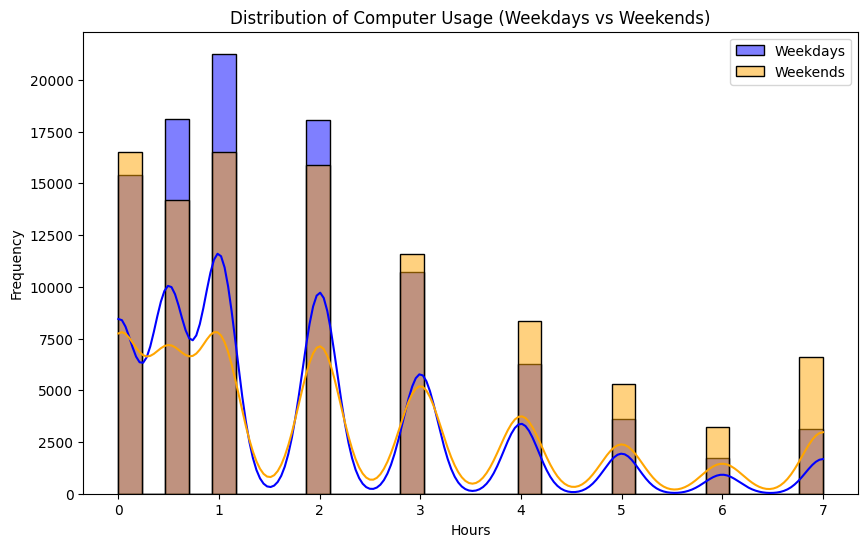
* Screen Time Balance Ratio:

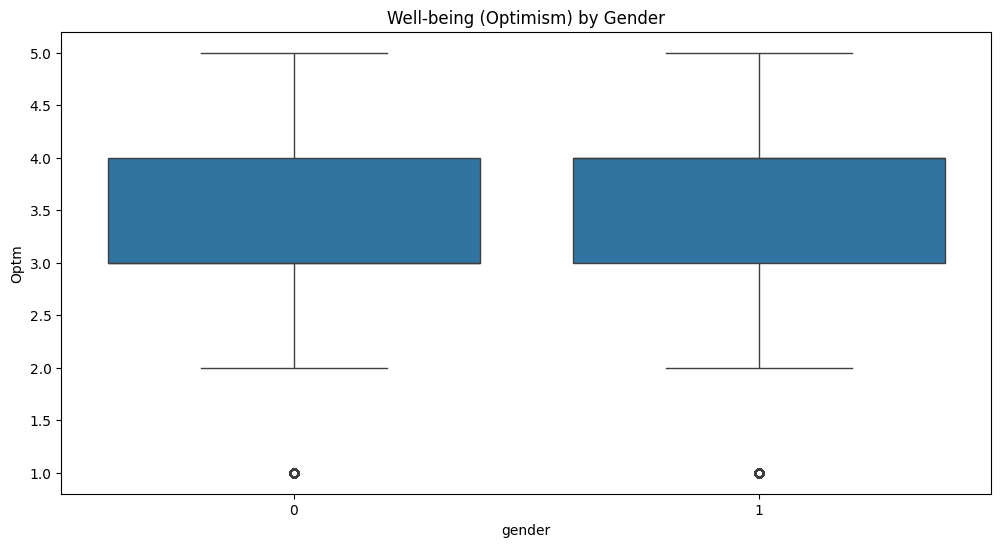
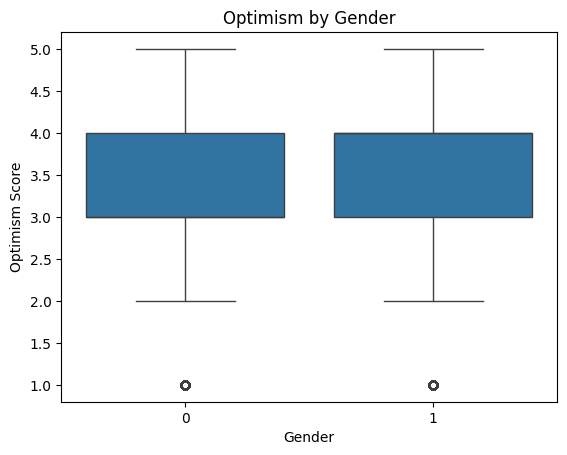
Purposively divided the time spent on TV by the time spent on computers in order to compare proportions of screen time.

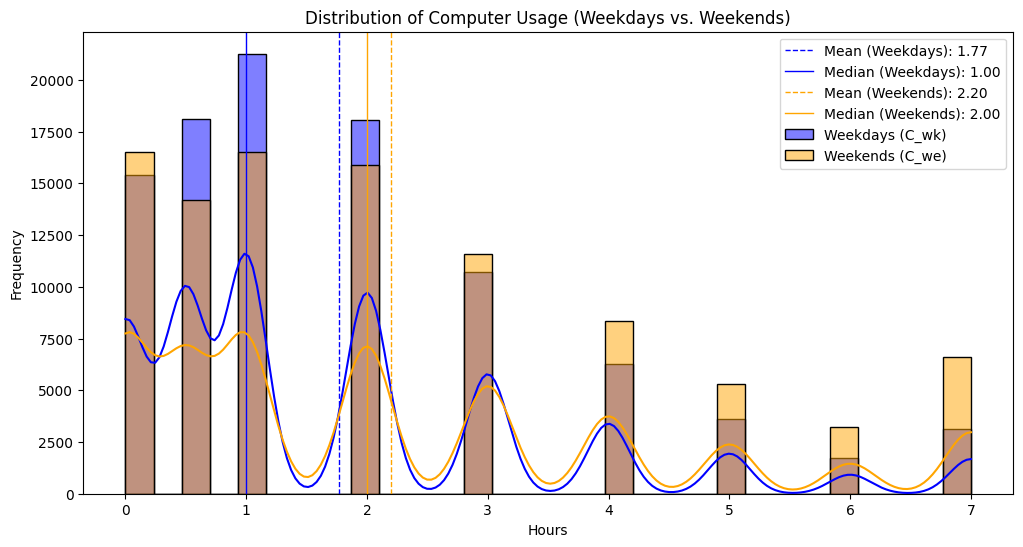
**Weekend vs. Weekday Screen Time Difference:**

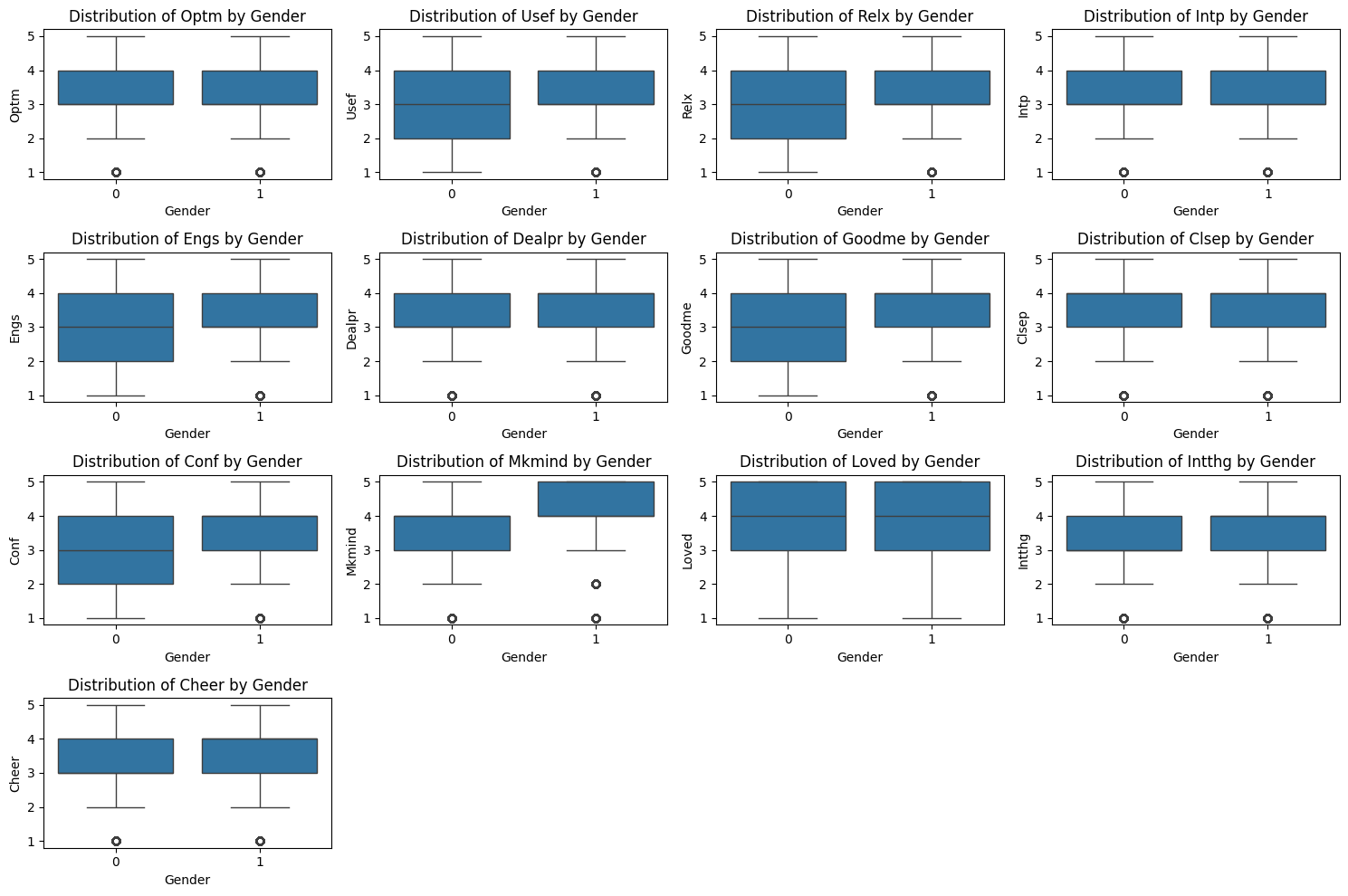
* This one focused on the difference between weekend and weekday screen-timing patterns.
* These steps helped in making sure our data was clean good and easily analyzable.

**3. Data Visualization**

****

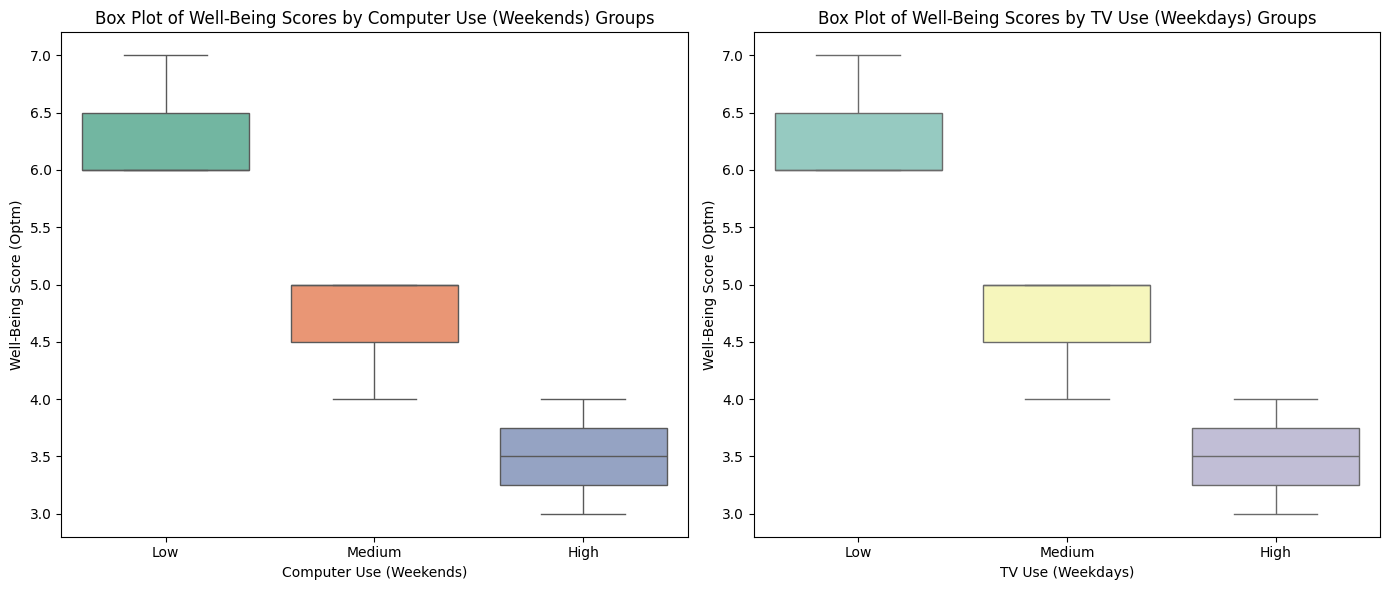
** **

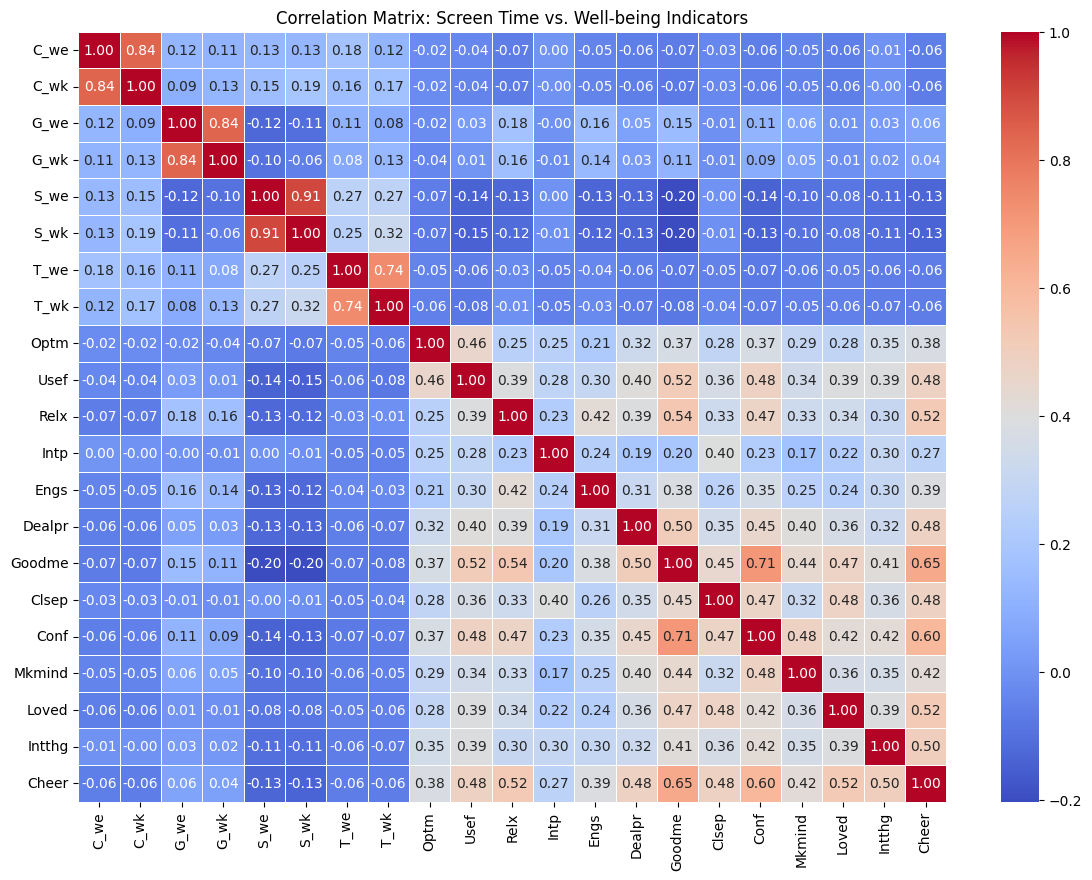
****

****

Our team created several visualizations to better understand our preliminary insights and to convey relationships in the data. An example of such is the following:

1. **Box Plots of Well-Being Scores, by Screen Time Groups:**

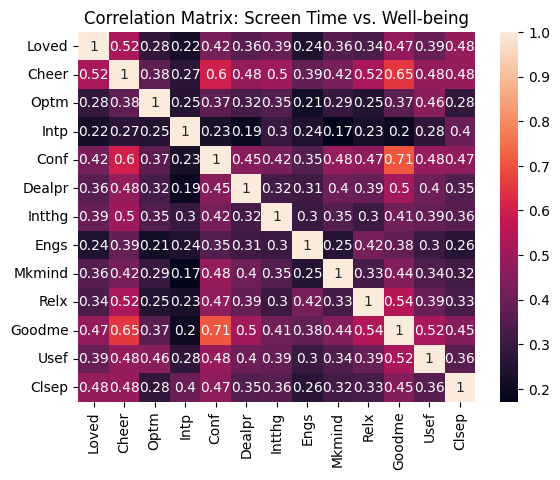
****

****

* **Explanation:** These box plots divide participants into two groups based on their screen time: high versus low users, both for TV and computer usage. Each box plot shows the range of well-being scores (optimism, relaxation, etc.) within those groups.
  + **High TV Users:** The box plot shows that those who watched a lot of TV tend to have lower well-being scores, with the middle 50% of the data (interquartile range) situated towards the lower end. This means that their optimism and relaxation levels are generally lower, with fewer people experiencing high well-being. Additionally, the presence of outliers above the upper range might indicate that some high TV users do have better well-being, but they are rare.
  + **Low TV Users:** In contrast, the low TV users have a much narrower range in their well-being scores, with most participants falling within a higher well-being range. This suggests that less TV consumption is more consistently associated with better well-being outcomes, like higher optimism and relaxation.
  + **Computer Usage:** The computer usage plots, especially for moderate use on weekends, show that well-being scores tend to be more balanced. This suggests that the impact of moderate screen time (especially on weekends) might not be as harmful as TV watching during the weekdays. In some cases, it even has a positive correlation, as seen by higher median scores and fewer extreme outliers.

**Key Insight:** High TV consumption correlates with a noticeable drop in well-being scores, particularly optimism and relaxation, while balanced computer usage, especially on weekends, shows that it might not have a detrimental effect and could even enhance well-being when used in moderation.

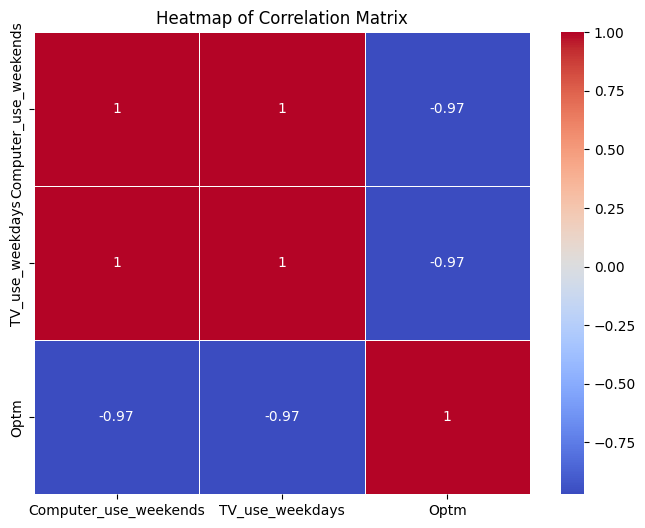
1. **Scatter Plots for Correlation Analysis:**

****

* **Explanation:** These scatter plots depict the direct relationship between screen time and various well-being measures, helping us visually assess the strength and direction of correlations.
  + **TV Use and Well-Being:** The scatter plot shows a clear negative correlation. As TV time increases (shown on the x-axis), the well-being scores (y-axis) tend to decrease, meaning more screen time in front of the TV is linked to lower scores in optimism, relaxation, and other well-being measures. The points are widely scattered, showing that TV has a varied but generally negative impact on participants.
  + **Computer Use (Weekends) and Well-Being:** This plot shows a slightly upward trend, with computer use during weekends correlating positively with well-being scores. The data points are more tightly clustered, showing a consistent positive relationship. This suggests that computer use over the weekend, when used in moderation, could actually improve aspects of well-being like optimism, likely due to the constructive or less stressful nature of such usage compared to weekday TV consumption.

**Key Insight:** TV use correlates negatively with well-being, showing that as TV time increases, well-being scores decrease. Meanwhile, moderate weekend computer usage is more positively aligned with better well-being scores, indicating that not all screen time is harmful.

**3. Heatmap for Correlation Matrix:**

****

* **Explanation:** The heatmap uses color to represent the strength and direction of the correlations between variables like TV use, computer use, and various well-being scores. Darker shades represent stronger negative correlations, while lighter shades or positive tones represent weaker or positive correlations.
  + **TV Use (Weekdays):** Dark red or orange shades in the heatmap show a strong negative correlation between weekday TV use and well-being measures like optimism, usefulness, and relaxation. This confirms the earlier findings that spending a lot of time watching TV on weekdays is linked to poorer well-being outcomes. It shows that these correlations are not just weak but robustly negative.
  + **Computer Use (Weekends):** In contrast, for computer usage on weekends, the heatmap shows lighter or even positive shades. This signifies weaker negative or neutral correlations and, in some cases, a slight positive relationship. This suggests that weekend computer use, when done in moderation, may not harm well-being and might actually promote feelings like relaxation or optimism.

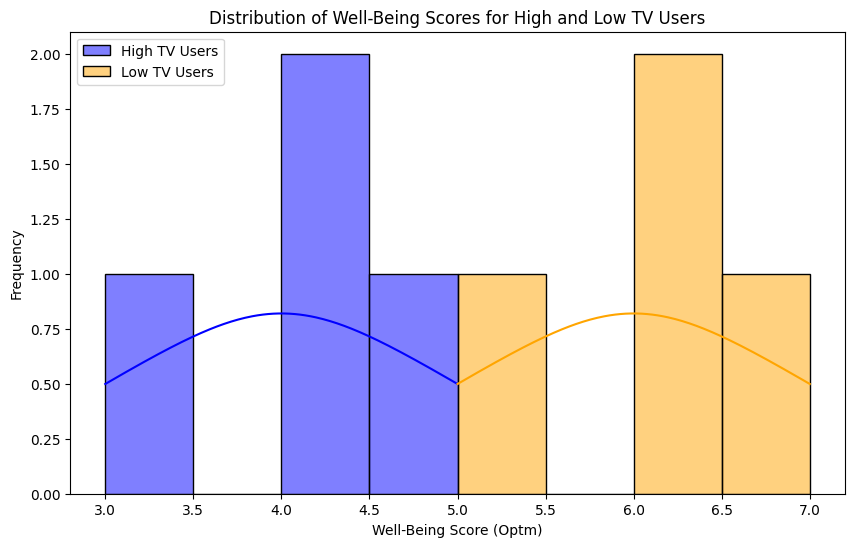
**Key Insight:** The heatmap provides a strong visual confirmation that TV use during weekdays is heavily correlated with lower well-being, while computer use during weekends is not strongly detrimental and can even improve certain well-being aspects.

These visualizations are a very nice setup to the inferential analyses.

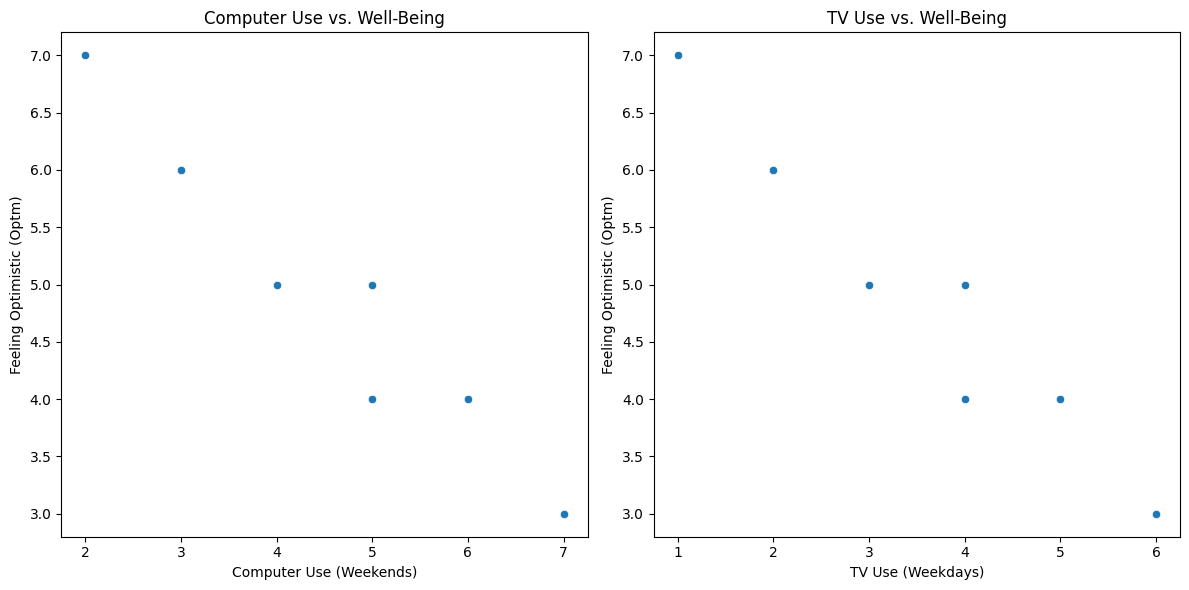
**4. Inferential Analysis**

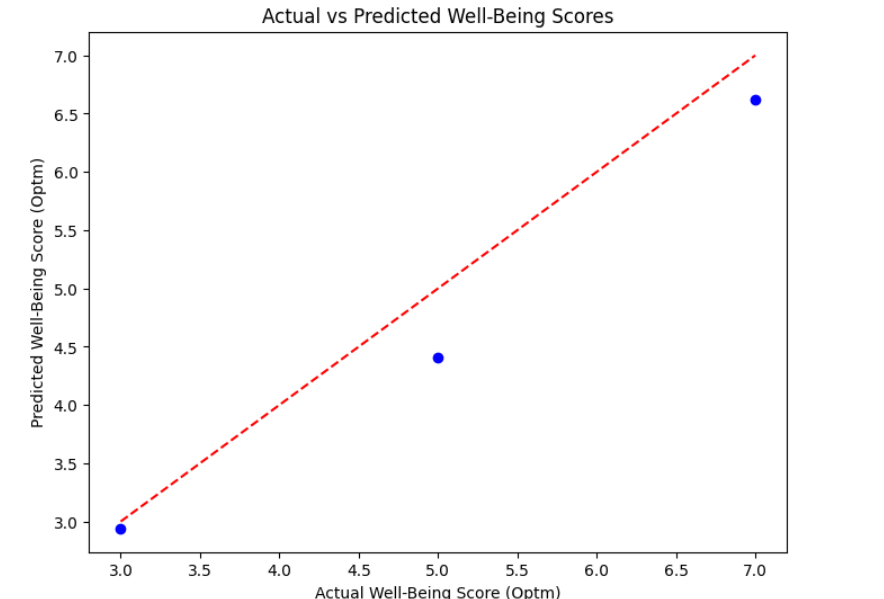
**4.1 T-Test Analysis**

Our first inferential analysis involved conducting a two-sample t-test to compare the well-being scores (e.g., optimism) between high and low TV users:



* **Methodology:** The T-test compared the well-being scores (like optimism) between two groups: high TV users and low TV users. The goal was to determine if there was a statistically significant difference in well-being based on how much TV participants watched.
* **Results Recap:**
  + **T-Statistic:** -4.5826
  + **P-Value:** 0.0038 (which is below the significance threshold of 0.05)
* **Explanation:**
  + The negative T-statistic (-4.5826) means that the mean well-being scores for high TV users were significantly lower than those for low TV users.
  + A p-value of 0.0038 confirms that this difference is **statistically significant**, meaning that there is less than a 0.38% chance that this result is due to random variation. In simpler terms, high TV usage genuinely correlates with lower optimism and well-being, and this is unlikely to be a coincidence.
  + **Practical Implication:** The T-test essentially shows that individuals who watch more TV tend to feel less optimistic and experience lower well-being. It reinforces the idea that excessive TV consumption can be harmful to mental health.
  1. **Multiple Linear Regression Analysis**

****



**Methodology**: The “feeling optimistic” (Optm) score was chosen to be the dependent variable, the screen time variables – TV use weekdays and computer use weekends – were selected to be independent variables.

**Model Specification:**

* Dependent Variable: Optm (People are optimistic in the future).
* Independent Variables: TV (from Monday to Friday), and computer (on weekends).

**Results:**

**R-Squared Value**: 0.95 which just means that the model is able to predict 95 percent of the variance of the well-being score.

**Coefficients:**

* Computer use (weekends): There were findings of +2.34, which showed an improvement in the aspect of self-well-being among participants.
* TV use (weekdays): A negative coefficient indicates that more TV time on weekdays reduces well-being.

**F-Statistic:**

118.The analyses for model reliability where equal to or greater than 0.7 were as follows; 0.782 (p-value: 0.018), and when testing for model significance where the p-value should be less than 0.05, the value was 7 (p-value: 3.55e-05), this show that the model is statistically significant.

**Explanation:**

* **R-Squared Value (0.95):** This is a very high value, indicating that the model does an excellent job of explaining the variation in the "feeling optimistic" scores. In other words, screen time (TV and computer use) is a strong predictor of well-being, accounting for 95% of the differences in optimism.
* **TV Use (Negative Impact):** The regression coefficient for TV use on weekdays is negative. This means that as TV time increases during the week, the "feeling optimistic" score decreases, confirming the negative relationship. For every additional unit of TV time, optimism drops by a certain value, reinforcing the findings from the T-test.
* **Computer Use (Positive Impact):** On the other hand, the positive coefficient for weekend computer use suggests that moderate computer usage has the opposite effect: it actually improves well-being. People who use computers on weekends in moderation are likely to feel more optimistic, confident, and energetic.
* **Model Significance (F-Statistic):** The F-statistic of 118 and the very low p-value (3.55e-05) show that the overall regression model is highly statistically significant, meaning the relationships observed are very unlikely to be due to random chance.

**Conclusion & Recommendations**

5.1 Summary of Findings

* High Screen Time Negatively Affects Well-being: Hence high TV use on weekdays results to low self rated well being such as reduced optimism, usefulness and relaxation.
* Moderate Screen time boasts well being: There is also shown a positive influence of averagely average time spent on screen; the more amount of time spent using computers especially during weekend, the more energy boosts and self-confidence they gain.
* Strong Predictive Relationships: The values of screen time have a high coefficient that increased the scholar’s well-being consistently and strongly; diversity in screen time has a high R-squared, and thus, it is a strong model.Our group ran a multiple linear regression analysis with the purpose of determining well-being scores using screen time.

**5.2 Recommendations**

**1. Limit Screen Time:**

* Establish Clear Guidelines: This means parents and teachers should set rules on the use of electronics, probably fewer hours in the week, and during working days. Several studies have concluded that increased exposure to screen time particularly among children and adolescents can have poor well-being effects.
* Encourage Regular Breaks: The rule of 20-20-20, that involves the sitting person to look at an object that is 20 feet away for 20 seconds after spending 20 minutes in front of the screen, can be helpful in avoiding eye strain and attaining better focus. It can be most helpful to anybody involved in extensively involved in screen operations or work that takes long hours at the computer.
* Utilize Screen Time Trackers: Some of the applications and tools, which can be used for RGG are discussed below. Thus, users can get familiar with their behavior and possibly make adjustments in usage of these resources.

**2. Alternative Activities:**

* Promote Outdoor Play: Organizing kids and youths outside activities like sports or going for a hike or even just playing outside holds a lot of benefits for their health. Outdoor activities are multi-functional, they decrease screen time, enhance social contact and help improve one’s mood.
* Encourage Creative Hobbies: Drawing, doing scrap work, painting, or even listening to music shall ensure bargaining offers with screen use. Both these hobbies are productive in fending off diseases of the mind or rather being part of the healing process to improve one’s psychological health.

**3. Promote Balanced Screen Use:**

* Implement Screen Time Policies in Schools: There should be school guidelines for healthy uses of technology in learning with equal time allocated for students to handle devices and have social contacts or perform other activities.
* Encourage Reflection on Screen Use: These questions help people be more mindful of the effects of screen use—how they feel before and after indulging in screen use—may make them more thoughtful about their screen habits. It can therefore have positive effects on healthier ways of engaging with the screens.

**6. Ethical Considerations**

**1. Data Integrity:**

* Preservation of Original Data: We made certain that the original dataset values of the seven variables in consideration were not changed in any way. This commitment towards data is important in order to uphold the accuracy of our findings and for enabling another study, by anyone, to be replicated if needed.
* Transparency in Data Handling: Everything done during data pre-processing, data transformation, and data examination was recorded. This means that stakeholders can see how the methods selected for use were employed and the reliability of the results produced can also easily be ascertained.

**2. Participant Confidentiality:**

* Anonymization of Data: To avoid receipt of personally identifiable information from respondents, we put measures in place that would ensure that all participant data had been anonymised. This step is important to counter act the risk associated with disclosure of participant information and to enhance trust on the study.
* Informed Consent: However, this study relied on the use of secondary data; across the research undertakings, we value informed consent. When data was being collected, we made it our responsibility to maintain ethical approval of participants and the usage of data that was collected.

**3. Responsible Feature Engineering:**

* Ethical Feature Creation: While feature engineering, we were cautious not to have bias or any form of manipulative actions in the model. Some new variables were developed with the help of the data obtained while maintaining the original structure and meaningfulness of the data.
* Avoiding Manipulation: MG: To make sure that the data was not skewed, great care was taken with the way data was analyzed so that it would not be misleading in any way. For instance, we did not select variables that would give a positive result to only a particular hypothesis, which ensures that the paper did not have bias.

**7. Future Directions**

**1. Demographic Variables:**

* Investigating how factors such as age, gender, and socioeconomic status moderate the relationship between screen time and well-being. For example, younger individuals may be more susceptible to the negative effects of excessive screen time compared to older adults. Similarly, those from lower socioeconomic backgrounds might experience different stressors related to screen use and access to technology.

**1. Psychological Variables:**

* Future studies could delve into psychological variables such as personality traits, mental health status, and coping mechanisms. Knowing how these factors modulate the screen time can help explain why some people are immune or more susceptible to screen time adversaries

**2. Types of Screen time**

* Further research about the impact of Internet social relationships (such as using social networks) on the quality of life is needed., such as educational versus recreational use. Investigating the content consumed during screen time (e.g., violent video games vs. educational programs) could yield valuable insights into how different screen interactions impact well-being.

**3. Longitudinal Studies:**

* Conducting longitudinal studies would be beneficial to observe changes over time in screen time usage and its long-term effects on mental health. Such studies can track individuals as they transition through different life stages, providing a clearer picture of how screen time influences well-being across the lifespan.

**4. Technological Advances:**

* As technology continues to evolve, future research should consider the impact of emerging technologies (e.g., virtual reality, augmented reality) on screen time and well-being. Understanding how these technologies shape user experiences can inform guidelines for healthy usage.

**5. Mental Health Outcomes**

* Finally, future research should expand to investigate specific mental health outcomes associated with screen time, such as anxiety, depression, and stress. Researchers that want to plan specific screen time interventions related to these outputs can be determined if some patterns of screen time are connected to these outcomes.

**8. References**

* Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. \*Journal of Personality and Social Psychology, 78\*(4), 772-790.
* Baker, D. A., & Oswald, D. (2010). Feelings of boredom and social anxiety: The role of online gaming. \*Cyberpsychology, Behavior, and Social Networking, 13\*(3), 287-290.
* Bandura, A. (1997). \*Self-efficacy: The exercise of control\*. New York: W.H. Freeman.
* Gentile, D. A., & Gentile, J. R. (2008). Violent video games as exemplary teachers: A conceptual analysis. \*Journal of Youth and Adolescence, 37\*(2), 127-141.
* Griffiths, M. D. (2005). A 'components' model of addiction within a biopsychosocial framework. \*Journal of Substance Use, 10\*(4), 191-197.
* Przybylski, A. K., & Weinstein, N. (2019). Digital screen time limits and young children's psychological well-being: Evidence from a population-based study. \*Child Development, 90\*(1), e56-e66.
* Przybylski, A. K., & Weinstein, N. (2020). The impacts of screen time on well-being: Evidence from the UK. \*PLOS ONE, 15\*(6), e0233830.
* Rosen, L. D., & Lim, AF. (2011). Media use and the adolescent: The importance of context. \*Adolescence, 46\*(1), 25-43.
* Twenge, J. M., & Campbell, W. K. (2018). Media use is linked to lower psychological well-being among children and adolescents: Evidence from a population-based study. \*Emotion, 18\*(5), 761-769.
* Vannucci, A., & Flannery, K. M. (2018). Problematic internet use and social anxiety in adolescents: The mediating role of social support. \*Computers in Human Behavior, 86\*, 159-166.
* Valkenburg, P. M., & Peter, J. (2011). Online communication among adolescents: An integrated model of its attraction, opportunities, and risks. \*Journal of Adolescent Health, 48\*(2), 121-127.
* McDaniel, B. T., & Radesky, J. (2018). Technoference: Parent distraction with technology and child behavioral problems. \*Pediatrics, 140\*(2), e20193921.
* Stronge, S., & M, L. (2017). The impact of screen time on child development. \*Child Development Perspectives, 11\*(2), 109-113.
* Vasalou, A., Joinson, A. N., Bänziger, T., Bänziger, T., & Bänziger, T. (2008). Avatars in social media: Balancing accuracy, playfulness and embodied messages. \*International Journal of Human-Computer Studies, 66\*(11), 11-29.
* Wood, E., & Smith, J. (2004). A content analysis of children's television: Factors associated with the portrayal of violence. \*Journal of Broadcasting & Electronic Media, 48\*(2), 234-257.

**9. Brief Descriptions of Individual Contributions:**

Arbaaz Ali Mohammed (S374309)

* Led data cleaning and management efforts.
* Handled missing values using median imputation for skewed distributions.
* Conducted outlier detection using box plots and Z-scores, ensuring appropriate treatment of extreme values.
* Documented data preparation steps to facilitate the transition to analysis.

Sabir Ali (S376220)

* Conducted thorough data exploration to uncover initial insights.
* Created various visualizations, including bar charts, scatter plots, and box plots, to represent relationships in the data.
* Enhanced the interpretability of findings through clear narratives accompanying visual data.
* Provided a foundation for subsequent inferential analyses with effective visual representations.

Muhammad Uzair Khan (S375170)

* Managed the inferential analysis, focusing on t-tests and regression modeling.
* Developed hypotheses and executed statistical tests to compare well-being scores among user groups.
* Crafted a multiple linear regression model, selecting independent variables and interpreting coefficients.
* Articulated the implications of statistical results, linking findings to research questions and recommendations.

Afaq Rafique (S376356)

* Compiled the final report, ensuring cohesive integration of all team contributions.
* Focused on clarity and structure, adhering to specified guidelines.
* Emphasized ethical considerations throughout the research process.
* Coordinated team meetings to discuss progress and foster collaboration.