****

**HIT140 Foundation of Data Science**

**Final Report**

**Submitted by**

**Group: 12**

**Group members:**

Md Sydur Rahman – S374204

Rashed Mahmud - S383842

Rakibul Hassan Rimon - S383373

Ranjeeta Khatri – S375374

**Casuarina campus**

**Charles Darwin University**

Table of Contents

[1. Introduction 3](#_Toc180107816)

[2. Exploratory Data Analysis and Visualisation 3](#_Toc180107817)

[2.2 Descriptive Statistics 4](#_Toc180107818)

[2.3 Data Visualisation 5](#_Toc180107819)

[3. Predictive Modelling Approach 6](#_Toc180107820)

[3.1 Model Selection 6](#_Toc180107821)

[3.2 Feature Engineering 7](#_Toc180107822)

[3.3 Regression Modelling 8](#_Toc180107823)

[4. Results and Interpretation 9](#_Toc180107824)

[4.1 The finding of the Descriptive Analysis 9](#_Toc180107825)

[4.2 Results of the Analysis 10](#_Toc180107826)

[4.3 How accurate are predictions. 10](#_Toc180107827)

[5. Discussion and Limitations 10](#_Toc180107828)

[6. Conclusion 11](#_Toc180107829)

[7. Reference 13](#_Toc180107830)

[8. Individual Contributions 16](#_Toc180107831)

# Introduction

Recently, people’s concern on the impact of screen time on the mental health of the people particularly teenagers have been rising. As you know, most individuals have access to the smartphones, computers, gaming consoles and other TVs and therefore, it is poised that there is need to determine how various types os screen time affect psychological wellbeing. There is evidence from investigations including experiments on individuals to the effect that reduced use of the screens may assist to enhance conduct and psychological well-being among the young. The relationship between the time spent on screens and the health of individuals remains relatively unstudied. Such things as what the people do on the screens, how long they engage in the activity or their antecedents may all have variations.

This group project wants to investigate the effects of screen time in teenagers on the amount happiness teens report. We will utilise big data that contains details about individuals, their past, screen exposure duration, and some clues about the well-being. The datasets consist of over 100000 adolescents and reveal how teens employ computer, mobile, games, and television during weekdays and weekends. Furthermore, the survey focusing on how people feel about specific topics namely, hope, relaxation, confidence and energy levels, is used to measure well-being.

The goal of this project has two parts: first, to make an analysis by trying to see through numbers and identify many meaningful correlations. Second, to develop good models that can address well-being in terms of screen times and individual details about them. This project employs such data science techniques as; inspecting data, visualizing data, and modeling or conducting regression analysis to show how screen time influences teenagers’ lives for good or for worse. It is aimed at the provision of relevant information that can be used maybe in developing public health policies and concerning the utilization of the digital media.

The following report describes where we looked, what we learned, and what it means. Thus, it contributes to the topic of peer impact that analyzing spending time in front of digital devices decreases young people’s mental health.

# 2. Exploratory Data Analysis and Visualisation

**2.1** Dataset Overview: This work analysed three data sets to understand the impact of screen time on teenagers in terms of their well-being. Every dataset shows different information about the people in the study: their sedentary activities, their screen exposure history and some indicators of their well-being. Here’s a summary of these datasets:

**1. Demographic Data (dataset1. csv):** About 120000 youths are recorded in this file. Instead, each person who answers is given a special ID. The main demographic factors are gender, ethnicity and level of deprivation; the gender information is self-reported, with options: ‘male’=1 ‘female and others’ =0, Ethnicity ‘majority’: =0 ‘minority’ =1, The participation in ‘living in places with high poverty’=1, : ‘does not live in place with high poverty’ = 0.

**2. Screen Time Data (dataset2. csv):** The second data set contains detail on how many minutes about 113 000 people spend with screens daily. It records the time that the people spend on computers, playing game, on cellphone, and television on weekends, and weekdays. Each screen activity is presented as a variable enabling us to measure the time young people spend using various digital screens.

**3. Well-being Data (dataset3. csv):** The third data set consists of well-being data of 102580 individuals. These are anything but objective well-being indicators since base on people’s self-report answer to how hopeful they are, how useful, how relaxed, how energetic, how good they are in solving problems, and how emotionally connected they are to other people.

Each response is scored from 1 as an indication that a person never experienced that feeling in the two weeks preceding the survey, to 5 in as much as the person felt that way always in those two weeks.

These datasets enable us to see how screen time impacts numerous teenagers both in terms of their happiness and health. The information shared between the two allows the linking of screen time behaviour to well-being outcome based on demographic differences.

## 2.2 Descriptive Statistics

As a beginning step for our examinations, therefore, we systematically scrutinized the univariate summaries for all the data sets. This enabled us identify the skewness in the data distribution and identify any outlying or missing/incorrect values.

A majority of the people who responded to the survey were males, contributing to 52%, while females and others made up the remainders.

**Minority Status:** A sample of ethnic minority group comprised 15% of the people interviewed.

Deprivation: Precisely, 17.5% of people resided in very poor areas which might influence their use of the screens and general health.

**Screen Time Information:** Employed individuals used computers, smartphones, and video games for different durations in week days than during weekends. Despite the fact that the usage of computers is significantly lower than that of TV, people spend about 2. 5 days a week working on computers during weekdays and about 3. 5 days in weekends.

A current study shows that teens spend 3 hours a day using smart phones as opposed to about 2 hours spent on video games. For about 2. 5 hours each day, they also watch television.

**Health and Happiness Information:** The 4 most common responses included; I’m okay for most individuals. For instance, 60% of the teen interviewed reported that they sometimes or often felt positive about the future, with approximately 15% reporting to feel positive all the time.

Similar trends were observed on several other indicators of well-being, such as relaxed and confident = M = 3 and 4.

In fact, this analysis provides a first glance of the data that would help to define the right set of question for further analysis. It reveals aspects we might be interested in, such as are people in less affluent neighborhoods or who spend more time on some screen activities have lower well-being levels?

## 2.3 Data Visualisation

We also note that visualization heavily contributed to the comprehension of the data and the discovery of important trends. This makes it easier for me to noticing important trends when the data is presented without distraction. We made the connections different ways for example aging, time spent on technology and well-being.

Histograms: We created histograms to observe the distribution of the screen time more comprehensively embracing the computers, smartphones, video games, and TVs. From the histograms, it was demonstrated that majority of the sampled participants were engaged in digital activities for 1-3 hours within a day. But those who exceeded recommended screen time to 8 hours or more were identified to be using screens during weekends.

**Box Plots:** To compare the screen time in the scenario between boys and girls or in various level of poverty, we employed box plots. These visuals also demonstrated the fact that generally boys gamed on video games more and conversely girls frequently accessed smartphones. More often than not the children from low income neighborhoods spend more time in front of screens as compared to those from affluent neighborhoods.

**Correlation Matrix:** We thus generated a correlation matrix to examine the screen time and well-being relationship. This revealed how well correlated different variables are to one another. On weekends there was a slight positive correlation between computers and video game usage and positive mood, which is happiness, being optimistic and having confidence. But, the more the time spent watching TV, the small negative correlation was observed with the aspect of relaxation and clear headedness.

**Scatter Plots:** We employed scatter plots to examine how overall weekly screen time relates to some measure of well-being. The graphs demonstrated that screen time has an impact on wellbeing in different manners. For instance, for every 100 computer uses, well-being increased by 1, while for every 100 TV Viewings or uses well-being reduced by about 0.42.

These visuals enabled us to see how the data are distributed and provided some initial comparisons of screen time and well-being. These ideas so helped us in determination of further action in analyzing data and predicting.

# 3. Predictive Modelling Approach

## 3.1 Model Selection

The purpose of this project is to compare the impact of screen time and young people’s background information on their happiness level. In order to achieve this, we had to distinguish between various forms of regression models like Linear, Ridge and decision tree. Our primary model was linear regression because it is a straightforward model, and it directly predicts the numbers such as the well-being scores.

The first and about the simplest method of showing how screen-time and well-being are connected is a linear regression. This way we can simply estimate how much each single component (such as the number of hours spent beforehand on a screen or age) might increase or decrease the levels of the outcome measures. Again, since we prominently target screen time, linear regression enables us to observe and analyse these relations without overcomplication.

This gave linear regression because this is fairly easy to explain after a brief discussion of the approach we also briefly considered the use of ridge regression. Ridge regression is a non-normally distributed linear regression analysis that finds use in preventing very large coefficients and thus acts as an enhanced technique of the linear regression model. This could be useful because the dataset in our study is vast and complex. Finally, we decided to use ridge regression together with our linear regression models for the augmentation to increase the accuracy of the outcomes and minimize overfitting.

During the initial stage of selection of models, decision trees have been considered as it retains the idea of more complex linkages and inter dependencies between many factors. As this project is to straightforward to explain precisely how screen time impacts well-being, we decided not to employ decision trees.

## 3.2 Feature Engineering

To improve the regression models and make the such models more effective we created new variables added few modifications to some of them. This has a name – feature engineering – and it allows the models to predict things much better.

**Total Screen Time:** Previously, data about the time that children spent with different devices (computers, video games, smartphones, and TV) were analyzed individually, while all these measures were combined into one new variable called the “Total Weekly Screen Time. This feature reveals how long each individual is spending gazing at screens – whether it’s on a computer, phone, tablet or TV.

**Weekend vs Weekday Screen Time:** It is reasonable to assume that people’s screen use varies from weekdays to weekend days; therefore, we distinguished between two variables: “Total Weekday Screen Time” and “Total Weekend Screen Time”. This difference inform the model whether adopting screens during weekdays or weekends has different impact on well-being.

**Demographic Interaction Terms:** We included interaction terms that define how self screen time is associated with individual attributes such as ‘gender’, ‘minority status’, and ‘deprivation index’. These interaction terms assist the model to pos it facilitates that utilization of screen time could affect people in various ways. For instance, learning the amount of time boys as well as girls spent playing video games may help explain if they perceive the pleasures or the worth of their health differently.

**Screen Time Distribution:** The other designed feature was duration that people spent on various devices and SOPs. For example, we introduced new options which reveal how long a person spends on smartphone or playing video games. It might aid us to know if specified screen activities such as employment of smartphones have different effects on well-being as compared to other forms of screen activities.

These special features enabled us to focus more on how screen time impacts on health. It also enhanced soundness results of our model to predict well-being and elucidate the outcomes that we have obtained.

## 3.3 Regression Modelling

Once we developed all of the features, we applied several regression models to estimate the well-being scores of the respondents. Every aspect such as optimistic view, confidence, and relaxation was considered as a separate outcome of well-being. With regard to the independent variables, these were the specific features which were developed earlier about.

**Linear Regression:** A method of determining the way two variables are related, specifically by comparing their data points on a straight line that best fit it. It helps forecast one thing depending on the other thing.

As for each of the well-being measures we generate multiple linear regression models. The basic structure of the model is like this:

Well-being indicator = starting point plus time spent on screen plus information about the person like age, gender and so on, plus a number to do with the sum of the amount of screen time, plus a random factor.

Where:

The starting coefficient, beta\_0 indicates the least developed well-being value, simplified as w0.

**The parameters:** (beta\_1), (beta\_2), and (beta\_3) are representatives of what in a way affects the well-being score through the independent factors such as time spend in front of screen and other personal details.

This final part of the model represents the error that allows the variant part of the changes that we cannot quantify.

**Ways to Measure Success:**

Through R-squared, the ability of the various input factors to explain changes in the output factor was evaluated as demonstrated in the next section. We also relied on RMSE to see just how correct the forecasts were that we had made. The cross-validation was also applied to check the model’s ability to generalize on the other data and not to over-fit.

**Understanding the Results:**

There were several interesting outcomes with the line of analysis that we used. For instance, screen time predicted lower relaxation and higher energy price, as well as being worse in the worst-off poor neighborhoods. On the other hand, when people used computers, especially on weekends, they felt hopeful and self confident. The findings indicated that boys who used more time playing video games felt better than girls if they did the same thing. This means that form of screen matters for feeling in people.

Performing feature engineering and multiple regression analysis we herein provide a model which estimates the well-being scores using screen time and provides the detailed breakdown of the effect of given type of screen activities on various subgroups of population.

# 4. Results and Interpretation

## 4.1 The finding of the Descriptive Analysis

The study indicates that there are differences in the screen-time gender analyses for boy and for girl. According to the survey, boys spent longer hours in playing video games than those used by girls, and on the other hand, while girls spent longer hours in using smartphones than boys. This indicates that boys and girls use screens in dissimilar manner indicating that it influences their feelings and health. All the respondents pointed out that they classify positive effects as ‘’sometimes or ‘’often’’ through affective statements such as hope, relaxed, and confident. This means that, on the balance, they have normal well-being.

## 4.2 Results of the Analysis

In particular, the correlation with the time we spend on screens and happiness was clearly outlined and revealed in the study. Those who spent much time on computers indicated higher level of positive affective states such as optimism and energy probably because computers are used for work or learning activities. On the other hand self reported watching television for more than forty minutes was associated with perceived lesser relaxation and perceived ability to think clearly. This implies that sitting down while using screens is not healthy for our mind. Furthermore, the analysis indicated that being occupied with the screens on the weekends was higher by detrimental impacts to human well-being compared to on weekdays. It might be as such because the notion of a weekend is hardly rigid in terms of the time table that follows and therefore it is possible that ways employed to screen time and the repercussions are different on the two days.

## 4.3 How accurate are predictions.

Overall, our linear regression models had moderate prediction power, with values of R-squared ranging from 0. 3 to 0. 5 for most well-being indicators were observed. Feature engineering was crucial for Section 4.3 in making the model improve. These consisted of total hours and the interconnectivity, between devices and it facilitated in capturing more complex phenomena in the data by the model. Althrough, the models assist us in comprehending how the time spent in front of screens as a factor impacts teenagers’ happiness.

# 5. Discussion and Limitations

**Discussion**

It reveals that positive interaction with the screens led to increased happiness levels whereas negative interaction with screens led to decreased happiness levels. Children who spent more time texting felt less happy and had less energy than other children who read, used a computer, or watched TV showing that some types of screen time were beneficial. This may be because through a computer you can learn something new or get more done. On the other hand passive activities such as watching television were associated with poor mental health status. This implies that excessive time in front of a screen is aggressive and was found to negatively affect relaxation and cognitive mental health.

We also found that screen time differently impacts distinct categories of a population. Time spent on screens was more crucial to persons from a poor household and those of a minority population. This means that time spent on screens could increase the severity of issues affecting well-being for some people the situation worsens for those already dealing with it.

**Limitations:**

A main limitation is that there is self-reported data related to time spent watching screens and perceptions of one’s health. Self-reporting is based on one’s opinion which is prone to mistakes, for instance, most people might claim to use their screen for more or less time than they actually do, this can also include exaggerated or minimized accounts of their well-being This may alter the accuracy of the findings.

Furthermore, the data employed in this research were found to be only one in a particular period. So, at the present moment, it is impossible to have complete certainty regarding the impact of screen exposure on an individual’s well-being. It would be ideal for the research to be carried out on long-term basis, which will help in establishing how screen time, if at all, influences the well-being of an average person.

# 6. Conclusion

The analysis of the relationship between the level of technology use and happiness in adolescents, made it possible to ascertain that the issues of technology use and the impact on mental health of a person is not a simple one. There are positive effects, when one has to use a particular screen for work or study such as the energy and enthusiasm one feels from using a computer. On the contrary, if a person engages in passive consumption of content, such as television viewing, it can have a negative impact on overall well-being. It follows from these findings that not every moment that a person spends looking at the screen has same effects, therefore it depends on the activity that the person is doing on the screen.

The research showed that the effect that screen time is not the same across all people. The residents of low-income neighborhoods and minorities suffer from more serious negative impacts. Whenever social media is abused and is consumed for longer periods of time, it can exacerbate the social and health inequalities.

# 7. Reference

1. Anderson, M., & Jiang, J. (2018). Teens, social media & technology 2018. Pew Research Center. <https://www.pewresearch.org>

2. Atwood, S. A., & Strasburger, V. C. (2019). Screen time and adolescents: What do we know? Pediatric Clinics of North America, 66(5), 945-959. <https://doi.org/10.1016/j.pcl.2019.06.006>

3. Cain, N., & Gradisar, M. (2010). Electronic media use and sleep in school-aged children and adolescents: A review. Sleep Medicine, 11(8), 735-742. <https://doi.org/10.1016/j.sleep.2010.02.006>

4. Campbell, A. J., & Cumming, S. R. (2016). The impact of digital screen time on adolescent mental health. International Journal of Behavioral Development, 40(4), 311-316. <https://doi.org/10.1177/0165025415598222>

5. Ferguson, C. J. (2017). Everything in moderation: Moderate use of screens unassociated with child behavior problems. Psychiatric Quarterly, 88(4), 797-805. <https://doi.org/10.1007/s11126-016-9486-3>

6. George, M. J., Russell, M. A., Piontak, J. R., & Odgers, C. L. (2018). Concurrent and subsequent associations between daily digital technology use and high-risk adolescents’ mental health symptoms. Child Development, 89(1), 78-88. <https://doi.org/10.1111/cdev.12819>

7. Hale, L., & Guan, S. (2015). Screen time and sleep among school-aged children and adolescents: A systematic literature review. Sleep Medicine Reviews, 21, 50-58. <https://doi.org/10.1016/j.smrv.2014.07.007>

8. Houghton, S., Hunter, S. C., Rosenberg, M., Wood, L., & Zadow, C. (2015). Virtually impossible: Limiting Australian children and adolescents daily screen-based media use. BMC Public Health, 15, 5-15. <https://doi.org/10.1186/s12889-015-1445-4>

9. Kardefelt-Winther, D. (2017). How does the time children spend using digital technology impact their mental well-being, social relationships, and physical activity? An evidence-focused literature review. UNICEF Office of Research - Innocenti.

10. King, D. L., Delfabbro, P. H., Zwaans, T., & Kaptsis, D. (2014). Clinical features and axis I comorbidity of Australian adolescent pathological Internet and video game users. Australian & New Zealand Journal of Psychiatry, 48(12), 1058-1068. <https://doi.org/10.1177/0004867414538676>

11. Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., Shablack, H., Jonides, J., & Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. PLOS ONE, 8(8), e69841. <https://doi.org/10.1371/journal.pone.0069841>

12. Lemola, S., Perkinson-Gloor, N., Brand, S., Dewald-Kaufmann, J. F., & Grob, A. (2015). Adolescents’ electronic media use at night, sleep disturbance, and depressive symptoms in the smartphone age. Journal of Youth and Adolescence, 44(2), 405-418. <https://doi.org/10.1007/s10964-014-0176-x>

13. Przybylski, A. K., & Weinstein, N. (2017). A large-scale test of the Goldilocks hypothesis: Quantifying the relations between digital-screen use and the mental well-being of adolescents. Psychological Science, 28(2), 204-215. <https://doi.org/10.1177/0956797616678438>

14. Robinson, T. N., Banda, J. A., & Hale, L. (2017). Sleep and screen exposure across multiple platforms: Implications for adolescent well-being. Journal of Pediatrics, 185, 228-234. <https://doi.org/10.1016/j.jpeds.2017.02.037>

15. Radesky, J. S., & Christakis, D. A. (2016). Increased screen time: Implications for early childhood development and behavior. JAMA Pediatrics, 170(5), 480-481. <https://doi.org/10.1001/jamapediatrics.2016.0028>

16. Twenge, J. M., & Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. Preventive Medicine Reports, 12, 271-283. <https://doi.org/10.1016/j.pmedr.2018.10.003>

17. Odgers, C. L., & Jensen, M. R. (2020). Annual research review: Adolescent mental health in the digital age: Facts, fears, and future directions. Journal of Child Psychology and Psychiatry, 61(3), 336-348. <https://doi.org/10.1111/jcpp.13190>

18. 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-e65. <https://doi.org/10.1111/cdev.13007>

# 8. Individual Contributions

**Md Sydur Rahman – S374204**

Md Sydur Rahman was mainly involved in the first phase where he started coding with data preprocessing to prepare the data for analysis. His work included data pre-processing by writing scripts on Python aimed at handling missing values, cases of duplicate data, as well as other cases of inconsistency in the datasets. During the group coding sessions, and particularly selecting the codes for spadji and cqs sections, Sydur made sure that all the members of the group had similar ideas about the overall coding plan in their data. When it was possible to make mistakes in work with data he corrected some problems connected with wrong indexing and type of the data. Since Md Sydur was able to rectify these errors, it became even possible to go to the next levels of coding such as feature engineering. This laid strong coding foundation made his work to provide an organized and free-from-error data set to enable paving of team’s model development base.

**Rashed Mahmud – S383842**

Rashed Mahmud was mainly responsible for the coding on building and fine-tuning the prediction models. Work carried out involved the involvement in the incorporation of the linear regression models developed in party with the libraries such as the Scikit-learn and ensured that the linear regression models used by the team where properly incepted. During the coding of this model, Rashed also came across and fixed several problems with model convergence, mainly with the convergence of hyperparameters. While coding in groups, Rashed instilled the importance of group contribution, where he used to guide the team from learning debugging processes when the models generated unfavorable results. He also focused on tweaking the models through tuning the features as well as the models’ regularization so that they should execute appropriately. Rashed was helpful in identifying essential coding issues towards models’ performance and accuracy of the predictions the team was making.

**Rakibul Hassan Rimon – S383373**

Finally, coding the data visualization and management part of the project was facilitated by Rakibul Hassan Rimon. He used programming in Python to create almost any kind of plot such as scatter, histogram, or correlation matrix using libraries known as Matplotlib and Seaborn. As a part of assessment Rakibul discussed the problem with his colleagues regarding the axis, data points, and scaling problems in the graphical outputs during the coding sessions. He often enlisted several people’s help to make sure that data is encoded correctly, and the figures are correctly designed for further presentations. Rakibul was also instrumental to fix errors which were associated with the visualization code so that the plots created were easily understood. That freedom enabled him to resolve these problems and facilitate presentation of the team’s work in the most appealing and truthful way.

**Ranjeeta Khatri – S375374**

Ranjeita Khatri was much involved in twin activities of documenting the coding process and keeping the team code intelligible throughout the process. As to the project documentation, she had to write the comments and descriptions of the code sections, so the group members did not lose time on reading and comprehending what other members have done. Ranjeeta also organized a coding error tracker for the issues that are identified during the coding sessions, the bugs and solutions were written in a repository. This proved effective especially in tracking of repeat incidences and preventing reoccurrence within the subsequent phases. In particular, Ranjeeta also actively participated in coding by correcting syntax errors and logical mistakes which could lead to the wrong results in the program. Immersed herself in the documentation including tracking the errors that came up and helped to keep the team’s coding in check and free from interruptions due to outstanding bugs.