**Overview**

**Research Question:**

How does digital screen time affect adolescents' well-being?

**Purpose of the Study:**

* To perform two inferential statistical analyses to investigate the relationship between digital screen time and self-reported well-being among a large cohort of adolescents.
* To provide actionable insights into how different types and amounts of digital screen time impact the psychological health and overall well-being of adolescents.

### **Justification of Analytical Approach:**

This objective aims to understand how digital screen time affects the well-being of adolescents. We also want to find any differences in well-being between across different demographic groups, such as gender. To do this, we used two main methods: regression analysis and an independent two-sample T-test. We chose these methods because they help us answer our research questions and meet the project’s goals.

We used **Regression analysis** to measure the link between total screen time and well-being scores. This method helps us find if there is a significant relationship between the independent variable (total screen time) and the dependent variable (average well-being score. The Ordinary Least Squares (OLS) regression shows how strong this relationship is. It also shows if the relationship is positive or negative. It tells us how much of the change in well-being scores is explained by screen time. This helps us answer our research question. It shows if more screen time is linked to changes in well-being, and if so, by how much.

For the second investigation, we used an **independent two-sample T-test**. This test compares the well-being scores between two groups: males and females. We chose this method because it is good for finding if there is a significant difference between the means of the two groups. The T-test checks whether the difference in average well-being scores is real or due to chance. It also considers the variability within each group. This method helps us meet our project objectives. It helps us find any differences in well-being between demographic groups. This can help guide targeted interventions.

Overall, we chose these analytical methods because they help us understand the impact of screen time on well-being. They also help us see the differences in well-being among different groups. Regression analysis helps us measure the relationship between screen time and well-being. The T-test helps us compare well-being scores between groups. Together, these methods address our research questions and meet the project’s goals.

**Data Preprocessing for Investigation**

**1. Data Loading**

* Below are the three datasets used in the analysis:
  + dataset1.csv: Contains basic demographic information for over 120,000 respondents.
  + dataset2.csv: Provides digital screen time data for approximately 113,000 respondents (a subset of dataset1.csv).
* dataset3.csv: Contains self-reported well-being indicators for about 102,580 respondents (a subset of dataset2.csv).

……..downloaded from Learnline

**2. Data Merging**

* All three datasets were merged on the common key ID, representing the unique identifier for each respondent to create a single comprehensive dataset that combines demographic data, screen time, and well-being indicators for analysis.
* Result:A merged dataset with complete information for respondents who appear in all three datasets (final sample size: ~102,580 respondents).

**3. Calculating Total Screen Time**

* Calculation of the Total\_Screen\_Time for each respondent:
* Variables Used: Combined the daily hours spent on various digital devices: computers (C\_we, C\_wk), video games (G\_we, G\_wk), smartphones (S\_we, S\_wk), and TV (T\_we, T\_wk).
* Method: Sum of all hours per day spent on different devices on both weekdays and weekends.
* Purpose:To quantify the overall screen time for each respondent, providing a consolidated measure to analyze its impact on well-being.

**4. Calculating Average Well-Being Score**

* Calculation of the Avg\_Well\_Being\_Score for each respondent:
* Variables Used: Self-reported scores on 14 well-being indicators (Optm, Usef, Relx, Intp, Engs, Dealpr, Thkclr, Goodme, Clsep, Conf, Mkmind, Loved, Intthg, Cheer).
* Method: Compute the mean score across all 14 indicators for each respondent.
* Purpose: To create a single measure of overall well-being. This makes the analysis simpler while keeping important information from the different aspects of well-being.

**5. Handling Missing Values**

* Method: Removed any rows with missing values (NaN) to ensure data quality and consistency.
* Purpose:To prevent incomplete data from biasing the results of the statistical analyses.
* Result:A cleaned dataset ready for statistical analysis, ensuring all respondents have complete data for both screen time and well-being indicators.

**6. Saving the Preprocessed Data**

* + - Saved the processed data to a CSV file (preprocessed\_data.csv) for use in further analyses. (e.g., T-test, ANOVA, regression).
    - Purpose:To enable efficient and consistent access to the cleaned and preprocessed data for all investigations.

**7. Key Outcomes of Data Processing**

* + - Produced a comprehensive dataset with both screen time and well-being information for approximately 102,580 respondents.
    - The processed data is now ready for inferential statistical analyses, such as regression and hypothesis testing.

**Code for data\_preprocessing.py**

* This script handles data loading, cleaning, merging, and preprocessing

# data\_preprocessing.py

import pandas as pd

def load\_data():

# Load datasets

demographics = pd.read\_csv('dataset1.csv')

screen\_time = pd.read\_csv('dataset2.csv')

well\_being = pd.read\_csv('dataset3.csv')

return demographics, screen\_time, well\_being

def merge\_data(demographics, screen\_time, well\_being):

# Merge datasets on ID

merged\_df = demographics.merge(screen\_time, on='ID').merge(well\_being, on='ID')

return merged\_df

def preprocess\_data(df):

# Calculate total screen time for each respondent

df['Total\_Screen\_Time'] = df[['C\_we', 'C\_wk', 'G\_we', 'G\_wk', 'S\_we', 'S\_wk', 'T\_we', 'T\_wk']].sum(axis=1)

# Calculate average well-being score for each respondent

df['Avg\_Well\_Being\_Score'] = df[['Optm', 'Usef', 'Relx', 'Intp', 'Engs', 'Dealpr', 'Thkclr', 'Goodme',

'Clsep', 'Conf', 'Mkmind', 'Loved', 'Intthg', 'Cheer']].mean(axis=1)

# Handle missing values

df.dropna(inplace=True)

return df

if \_\_name\_\_ == "\_\_main\_\_":

# Load data

demographics, screen\_time, well\_being = load\_data()

# Merge datasets

merged\_df = merge\_data(demographics, screen\_time, well\_being)

# Preprocess data

preprocessed\_df = preprocess\_data(merged\_df)

# Save the preprocessed data to a CSV file for further analysis

preprocessed\_df.to\_csv('preprocessed\_data.csv', index=False)

I**nvestigation 1 - Effect of Total Screen Time on Well-being**

**0bjective**

The goal of this investigation was to explore the link between total digital screen time and self-reported well-being in adolescents. Specifically, we aimed to test the null hypothesis that there is no significant relationship between total screen time and well-being scores, against the alternative hypothesis that such a relationship exists. This investigation helps us understand if, and how, the amount of time spent on digital screens affects the overall well-being of young people.

**Code for investigation1\_analysis.py**

# investigation1\_analysis.py

import pandas as pd

import statsmodels.api as sm

def perform\_regression\_analysis(data):

# Define dependent and independent variables

X = data['Total\_Screen\_Time']

y = data['Avg\_Well\_Being\_Score']

# Add a constant term for the intercept

X = sm.add\_constant(X)

# Fit the regression model

model = sm.OLS(y, X).fit()

return model

if \_\_name\_\_ == "\_\_main\_\_":

# Load preprocessed data

preprocessed\_data = pd.read\_csv('preprocessed\_data.csv')

# Perform regression analysis

model = perform\_regression\_analysis(preprocessed\_data)

# Display results

print(model.summary())

**Methodology-Regression Analysis**

To study this relationship, we used Ordinary Least Squares **(OLS)** regression analysis. This is a statistical method that measures the effect of one variable on another. In our study, the dependent variable was the average well-being score**(Avg\_Well\_Being\_Score**). This score was calculated as the mean of 14 self-reported well-being indicators. The independent variable was total screen time**(Total\_Screen\_Time)**. This was defined as the total hours spent on digital devices each day. We used Python's **statsmodels** library to create the regression model. We added a constant term to account for the intercept in the equation.

### **Data Preparation for Regression**

Before doing the regression analysis, we prepared the data carefully to ensure it was accurate and reliable. We merged three datasets using a common key, the respondent ID. This created a complete dataset that included demographic information, screen time data, and well-being scores. The final dataset had about 102,580 respondents. This is a strong sample size for the analysis. We calculated the well-being score and total screen time based on the available data. We also removed any missing values to make sure the data was clean and ready for analysis.

### **Regression Model Fitting**

The regression model was created to study the effect of total screen time on well-being scores. The regression equation is: Avg\_Well\_Being\_Score = β0 + β1 \* Total\_Screen\_Time + ε. Here, β0 is the intercept, which is the starting value of well-being when screen time is zero. β1 is the coefficient that shows the effect of screen time on well-being. ε represents the error term or the unexplained part of the model. We used OLS regression to estimate these coefficients and find the best fit for the observed data.

**Key Results of the Regression Analysis**

The regression analysis shows a statistically significant but weak negative relationship between total screen time and well-being scores of adolescents. The **R-squared value of 0.015** suggests that only **1.5% of the variation** in well-being scores is explained by total screen time, indicating that other factors likely play a more substantial role in influencing well-being. While the relationship is statistically significant **(p-values are effectively zero)**, the effect size is very small **(coefficient of -0.0092),** meaning that additional screen time is associated with only a minimal decrease in well-being scores. These findings suggest that just reducing screen time may not greatly improve well-being. More research is needed to understand other factors that affect well-being.

#### Investigation 2: Differences in Well-being Across Demographic Groups

#### ****Objective of the Investigation****

The second investigation aimed to find out if there are significant differences in well-being scores between different demographic groups. We focused specifically on male and female adolescents. The null hypothesis states that there is no significant difference in average well-being scores between these two groups. The alternative hypothesis suggests that there is a significant difference. This analysis helps to find any gaps in well-being between the groups. It can also help in planning specific interventions to address these differences.

**Code for investigation2\_analysis.py**

# investigation2\_analysis.py

import pandas as pd

from scipy.stats import ttest\_ind

def perform\_ttest(data):

# Group data by gender

group1 = data[data['gender'] == 1]['Avg\_Well\_Being\_Score']

group2 = data[data['gender'] == 0]['Avg\_Well\_Being\_Score']

# Perform t-test

t\_stat, p\_val = ttest\_ind(group1, group2)

return t\_stat, p\_val

if \_\_name\_\_ == "\_\_main\_\_":

# Load preprocessed data

preprocessed\_data = pd.read\_csv('preprocessed\_data.csv')

# Perform T-test

t\_stat, p\_val = perform\_ttest(preprocessed\_data)

# Display results

print(f'T-statistic: {t\_stat}, P-value: {p\_val}')

#### ****Methodology: T-test****

We used an independent two-sample **T-test** to check for differences between the two groups. The T-test is a statistical method that compares the averages of two independent groups to see if there is a significant difference between them. In this investigation, the two groups are males (coded as 1) and females (coded as 0). The dependent variable is the average well-being score(Avg\_Well\_Being\_Score), which is the mean of 14 self-reported well-being indicators. The T-test helps us see if the average well-being scores are significantly different between these two gender groups.

#### ****Data Preparation for T-test****

The data for this T-test came from the preprocessed dataset. This dataset included well-being scores and demographic information for about 102,580 respondents. The dataset was filtered to create two groups based on gender: one for males (gender = 1) and another for females (gender = 0). This grouping made sure that the T-test could accurately compare the well-being scores between the two gender groups.

#### ****Execution of the T-test****

The T-test was done using Python's scipy.stats.ttest\_ind function. This function calculates the T-statistic and the corresponding p-value. The T-statistic shows how big the difference is between the two group averages compared to the variation within each group. The p-value tells us how likely it is that the observed difference happened by chance. A low p-value (**usually less than 0.05**) means the difference is statistically significant.

Top of Form

Bottom of Form

#### ****Key Results of the T-test****

The T-test results showed a T-statistic of 87.91, which indicates there is a large difference between the mean well-being scores of the two groups. The p-value was almost 0.0, suggesting that this difference is very unlikely to have happened by chance. Therefore, we reject the null hypothesis. We conclude that there is a statistically significant difference in well-being scores between male and female adolescents. These findings show that gender has an important effect on well-being among adolescents. In conclusion, the second investigation confirms a significant difference in well-being scores between males and females. This suggests that gender-specific factors may affect well-being and should be considered when creating strategies to improve adolescent health.

**Visualization**

**Code for visualization of both investigation;**

# visualization.py

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

def plot\_screen\_time\_vs\_wellbeing(data):

plt.figure(figsize=(8, 6))

sns.regplot(x='Total\_Screen\_Time', y='Avg\_Well\_Being\_Score', data=data, scatter\_kws={'alpha':0.3}, line\_kws={'color': 'red'})

plt.title('Total Screen Time vs. Average Well-being Score')

plt.xlabel('Total Screen Time (hours per day)')

plt.ylabel('Average Well-being Score')

plt.show()

def plot\_wellbeing\_by\_gender(data):

plt.figure(figsize=(8, 6))

sns.boxplot(x='gender', y='Avg\_Well\_Being\_Score', data=data)

plt.title('Well-being Scores by Gender')

plt.xlabel('Gender (1 = Male, 0 = Female)')

plt.ylabel('Average Well-being Score')

plt.show()

if \_\_name\_\_ == "\_\_main\_\_":

# Load preprocessed data

preprocessed\_data = pd.read\_csv('preprocessed\_data.csv')

# Plot total screen time vs well-being

plot\_screen\_time\_vs\_wellbeing(preprocessed\_data)

# Plot well-being scores by gender

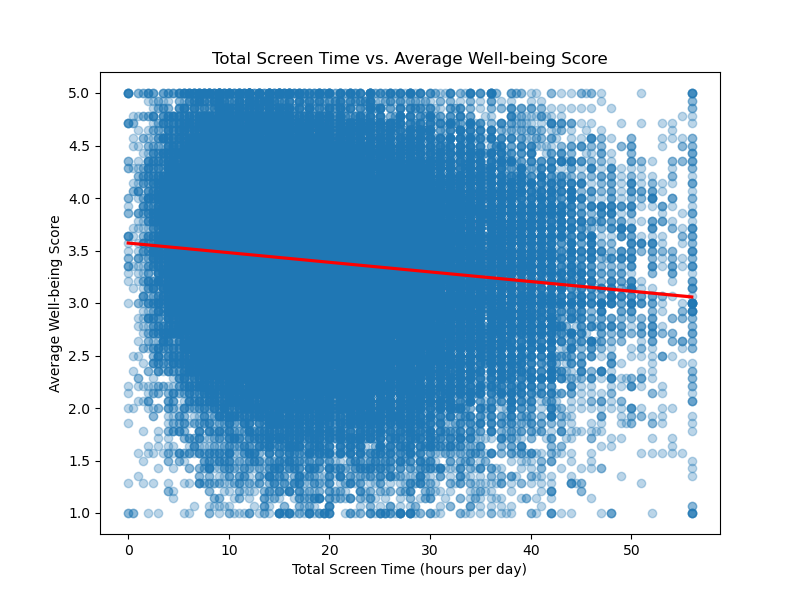
plot\_wellbeing\_by\_gender(preprocessed\_data)

### **Visualization for Investigation 1: Effect of Total Screen Time on Well-being**

#### ****Objective of the Visualization****

The first visualization aims to illustrate the relationship between total screen time and the self-reported well-being scores of adolescents. This scatter plot with a regression line helps to visually validate the findings from the regression analysis by showing the overall trend and distribution of data points.

#### ****Description of the Plot****

The scatter plot shows individual data points representing each respondent's total screen time (in hours per day) on the x-axis and their average well-being score on the y-axis. A red regression line is overlaid to indicate the overall trend between the two variables. The plot also includes transparency settings for the scatter points (alpha=0.3) to highlight data density and minimize visual clutter.

#### ****Interpretation of the Plot****

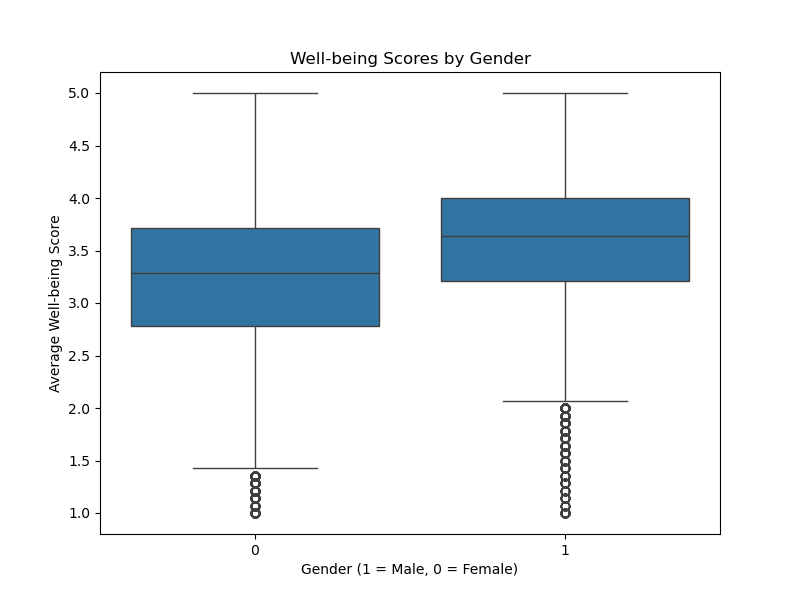
The regression line in the plot has a slight downward(negative) slope. This matches the regression analysis results, which showed a small but statistically significant negative relationship between screen time and well-being scores. The points in the plot are spread out widely. This means there is a lot of variation in well-being scores at different levels of screen time. However, the slight downward trend suggests that more screen time is generally linked to lower well-being scores. This visual supports the conclusion that while the relationship between screen time and well-being is statistically significant, it is practically small.

### **Visualization for Investigation 2: Differences in Well-being Across Demographic Groups**

#### ****Objective of the Visualization****

The second visualization is designed to compare the well-being scores between different demographic groups, specifically between males and females. This box plot provides a clear and concise way to visualize the distribution and central tendency of well-being scores for each gender, highlighting any differences between the groups.

#### ****Description of the Plot****

The box plot displays the distribution of average well-being scores for two groups: males (gender = 1) and females (gender = 0). The x-axis represents the gender, while the y-axis shows the well-being scores. Each box represents the interquartile range (IQR) of the well-being scores for that gender, with a line inside the box indicating the median score. The plot also shows any potential outliers beyond the whiskers.

#### ****Interpretation of the Plot****

The box plot shows a clear difference in well-being scores between males and females. It suggests that one gender has a higher median well-being score than the other. This matches the **T-test results**, which found a significant difference between the two groups' average well-being scores. The box plot also shows the range of scores within each group, helping us see how well-being scores are spread out for each gender.

### **Conclusion and Key Takeaways**

The visualizations give important insights that support the findings from both investigations. The scatter plot for Investigation 1 shows a weak negative link between screen time and well-being. This suggests that screen time has some impact, but other factors are probably more important. The box plot for Investigation 2 shows clear differences in well-being scores between genders. This highlights the need for gender-sensitive approaches when planning interventions to improve adolescent well-being.

### **Findings and Conclusions**

#### ****Summary of Findings:****

The investigations provide valuable insights into the relationship between digital screen time and adolescents' well-being, as well as differences in well-being across demographic groups.

* **Investigation 1:** The first investigation used regression analysis to examine the effect of total screen time on self-reported well-being scores among adolescents. The results revealed a statistically significant negative relationship, with each additional hour of screen time associated with a slight decrease in well-being scores by 0.0092 points. However, the effect size was minimal, as indicated by the low R-squared value of 0.015, suggesting that total screen time accounts for only 1.5% of the variance in well-being scores. This indicates that while screen time may have some impact, its practical significance is limited, and other factors are likely more influential in determining well-being.
* **Investigation 2:** The second investigation used a T-test to compare well-being scores between male and female adolescents. The results showed a statistically significant difference in well-being scores between the two groups, with a T-statistic of 87.91 and a p-value effectively at zero. This finding indicates that gender plays a significant role in influencing well-being, highlighting potential disparities that may need to be addressed in interventions targeting adolescent health.

The findings from both investigations show that screen time has a statistically significant negative effect on well-being, but the impact is small. This means that just reducing screen time may not greatly improve adolescent well-being. A better approach might include other factors, like the type of content, how screens are used, and individual differences among people. The differences in well-being between genders also suggest that specific approaches are needed to address the unique needs of different groups. To conclude, these findings highlight the need for a balanced view on screen time. While it can affect well-being, it is not the only factor. Broader strategies are needed to support adolescent health effectively.