Mental health and technology usage Analysis

Final Project Documentation

Technical Instructor   Jackie Aboalaem   **Group ID**   ALX1-DAT1-G4D

Team Members:
Ahmed Farag
Basant Mostafa
Mohab Mohamed
Sameh Hasan
Shaimaa Ragab

Team Members Role

No	Member Name	MEMBER ROLE
1	Ahmad Farag Mostafa	Create preparatory visualizations (Python, Matplotlib), visualize preliminary insights (Python, Matplotlib), and prepare presentation slides (Microsoft PowerPoint).
2	Basant Mostafa	Record progress, organize notes, prepare Notebook comments, and perform project documentation.
3	Mohab Mohamed	Perform data transformation using Python Pandas.
4	Sameh Nabil Hasan	Brainstorm potential analysis questions, visualize potential results (Python, Matplotlib), prepare notebook markdown (Jupyter notebook), and build a Power BI visualization dashboard (Power BI).
5	Shaimaa Ragab	Conduct data exploration, and data cleaning (Python Pandas), and prepare the final report.

Contents

Executive Summary

Overview

This report explores the relationship between technology usage and mental health based on a dataset of individuals from various countries. The analysis focused on metrics such as technology usage hours, social media activity, sleep patterns, and mental health indicators like stress levels. The project aimed to discover trends and correlations between these factors and offer actionable insights to promote better mental well-being in the digital age.

Results

- **Technology Impact**: Increased Technology time, especially from social media usage, correlates with higher stress levels.
- **Sleep and Stress**: Lack of sleep significantly worsens stress, while participants who balanced sleep and technology usage reported better mental health.
- **Physical Activity**: Regular physical activity is a buffer against the negative impacts of excessive technology usage.
- **Support Systems**: Access to both online and offline support systems positively influence mental health outcomes.

About the Dataset

This dataset includes anonymized data from individuals in various countries, capturing demographic information (age, gender, country), daily technology usage, and self-reported mental health metrics.

Data Source

The data was sourced from a digital wellness survey to understand the relationship between technology usage and mental well-being.

Dataset Content and Features

The dataset consists of 15 columns capturing detailed information about technology usage, mental health, and lifestyle habits. Key features include:

1. **User_ID**: Unique identifier for each user.
2. **Age**: Participant's age.
3. **Gender**: Participant's gender.
4. **Country**: Country of residence.
5. **Technology_Usage_Hours**: Daily technology usage in hours.
6. **Social_Media_Usage_Hours**: Time spent on social media.
7. **Gaming_Hours**: Time spent gaming.
8. **Screen_Time**: Time spent on screen (TV, Cinema, etc.)
9. **Mental_Health_Status**: Self-reported mental health (Excellent, Good, Fair, Poor).
10. **Stress_Level**: Reported stress level (Low, Medium, High).
11. **Physical_Activity_Hours**: Daily physical activity duration in hours.
12. **Sleep_Hours**: Daily hours of sleep.
13. **Work_Environment_Impact**: Perceived impact of work environment on mental health.
14. **Support_Systems_Access**: Whether the individual has access to mental health support systems (Yes/No).
15. **Online_Support_Usage**: Whether the individual uses online mental health support systems (Yes/No).

Data Preprocessing

Exploratory Data Analysis (EDA)

- **Exploring Data**: Display the head of the data (first 3 rows), the tail of the data (last 3 rows), and a random sample of the data.
- **Descriptive Statistics**: Summary statistics (mean, median, standard deviation) were computed for all numerical variables. For instance, the average daily screen time was 6.6 hours.
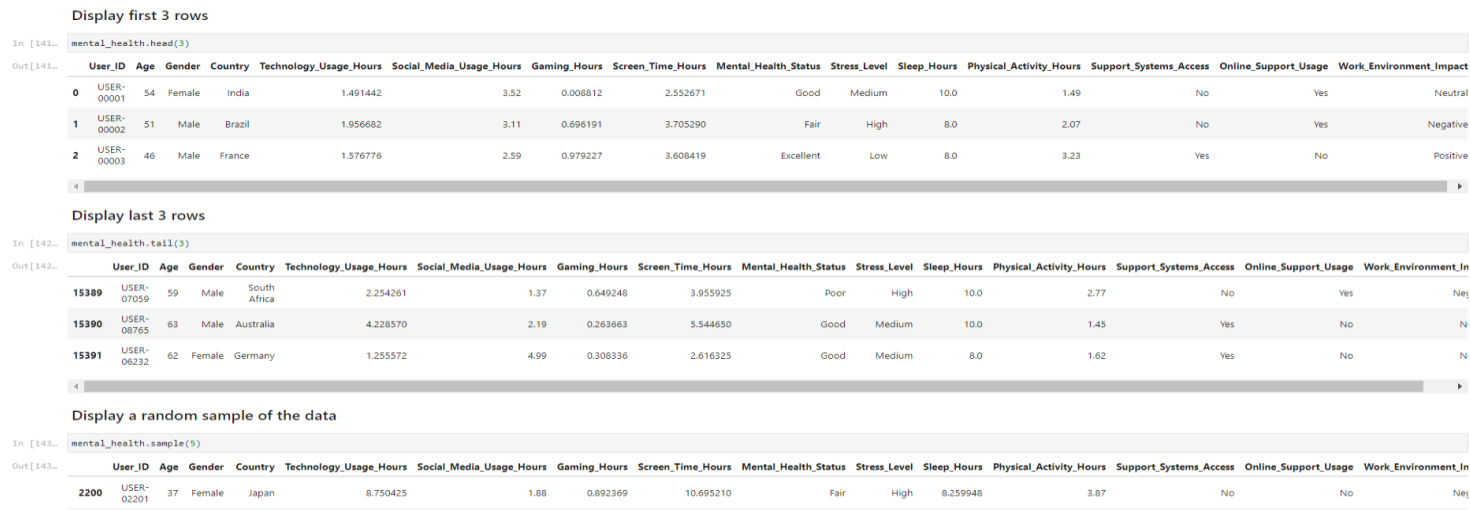
Display first 3 rows

In [141]: `mental_health.head(3)`

	User_ID	Age	Gender	Country	Technology_Usage_Hours	Social_Media_Usage_Hours	Gaming_Hours	Screen_Time_Hours	Mental_Health_Status	Stress_Level	Sleep_Hours	Physical_Activity_Hours	Support_Systems_Access	Online_Support_Usage	Work_Environment_Impact
0	USER-00001	54	Female	India	1.491442	3.52	0.008812	2.552671	Good	Medium	10.0	1.49	No	Yes	Neutral
1	USER-00002	51	Male	Brazil	1.956682	3.11	0.696191	3.705290	Fair	High	8.0	2.07	No	Yes	Negative
2	USER-00003	46	Male	France	1.576776	2.59	0.979227	3.608419	Excellent	Low	8.0	3.23	Yes	No	Positive

Display last 3 rows

In [142]: `mental_health.tail(3)`

	User_ID	Age	Gender	Country	Technology_Usage_Hours	Social_Media_Usage_Hours	Gaming_Hours	Screen_Time_Hours	Mental_Health_Status	Stress_Level	Sleep_Hours	Physical_Activity_Hours	Support_Systems_Access	Online_Support_Usage	Work_Environment_In
15389	USER-07059	59	Male	South Africa	2.254261	1.37	0.649248	3.955925	Poor	High	10.0	2.77	No	Yes	Neg
15390	USER-08765	63	Male	Australia	4.228570	2.19	0.263663	5.544650	Good	Medium	10.0	1.45	Yes	No	N
15391	USER-06232	62	Female	Germany	1.255572	4.99	0.308336	2.616325	Good	Medium	8.0	1.62	Yes	No	N

Display a random sample of the data

In [143]: `mental_health.sample(5)`

	User_ID	Age	Gender	Country	Technology_Usage_Hours	Social_Media_Usage_Hours	Gaming_Hours	Screen_Time_Hours	Mental_Health_Status	Stress_Level	Sleep_Hours	Physical_Activity_Hours	Support_Systems_Access	Online_Support_Usage	Work_Environment_In
2200	USER-02201	37	Female	Japan	8.750425	1.88	0.892369	10.695210	Fair	High	8.259948	3.87	No	No	Neg

Figure 1

Get summary statistics

In [16]: `mental_health.describe()`

Out[16]:

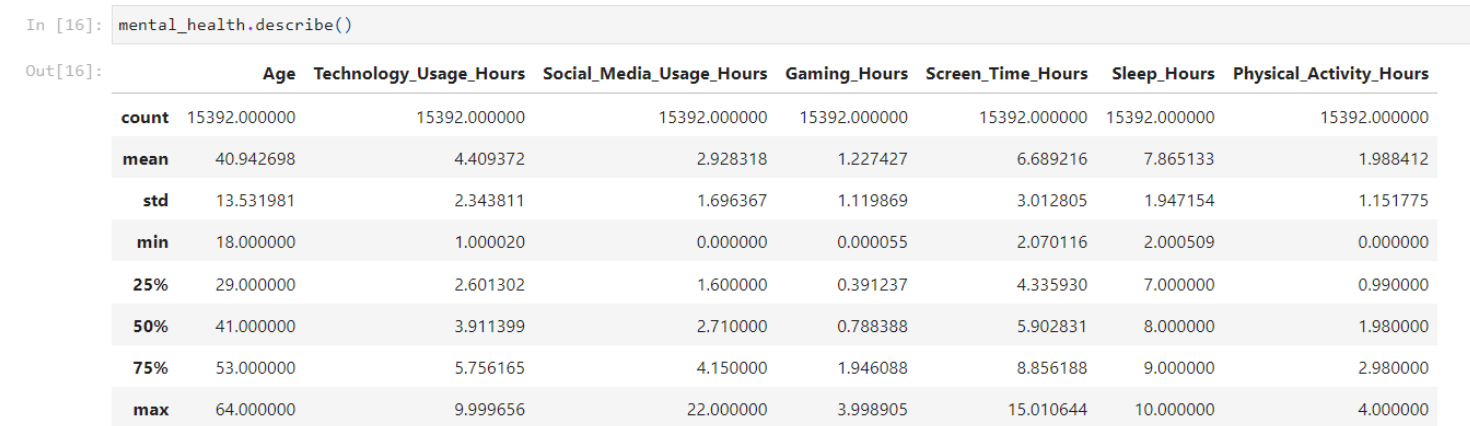|	Age	Technology_Usage_Hours	Social_Media_Usage_Hours	Gaming_Hours	Screen_Time_Hours	Sleep_Hours	Physical_Activity_Hours
count	15392.000000	15392.000000	15392.000000	15392.000000	15392.000000	15392.000000	15392.000000
mean	40.942698	4.409372	2.928318	1.227427	6.689216	7.865133	1.988412
std	13.531981	2.343811	1.696367	1.119869	3.012805	1.947154	1.151775
min	18.000000	1.000020	0.000000	0.000055	2.070116	2.000509	0.000000
25%	29.000000	2.601302	1.600000	0.391237	4.335930	7.000000	0.990000
50%	41.000000	3.911399	2.710000	0.788388	5.902831	8.000000	1.980000
75%	53.000000	5.756165	4.150000	1.946088	8.856188	9.000000	2.980000
max	64.000000	9.999656	22.000000	3.998905	15.010644	10.000000	4.000000

Figure 2

Managing Missing Data

- **Removal of Incomplete Records**: Rows with too many missing fields were removed to ensure data consistency.
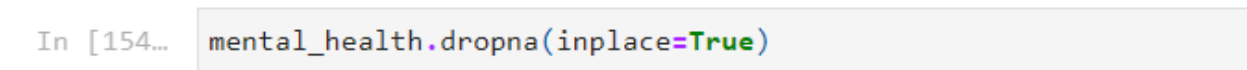
Dropping null values

In [154...] `mental_health.dropna(inplace=True)`

Figure 3

- **Removal of duplicated records:** Rows that are duplicated more than once were removed to avoid biased and misleading records.

Checking duplicated values

```
In [185…   mental_health.duplicated('User_ID').value_counts()

Out[185…   False    15283
           True        35
           Name: count, dtype: int64
```

> 35 duplicated values were found and need to be removed

Dropping duplicated values

```
In [186…   mental_health = mental_health.drop_duplicates('User_ID', keep='first')

In [187…   mental_health['User_ID'].duplicated().value_counts()

Out[187…   User_ID
           False    15283
           Name: count, dtype: int64
```

> Duplicated Values were deleted

Figure 4

Data Transformation

- **Normalization**: Continuous variables like **Technology_Usage_Hours** and **Screen_Time_Hours** were normalized to bring them onto the same scale.
- **Categorical Encoding**: Categorical columns like **Online_Support_Usage** and **Support_Systems_Access** were converted into Boolean data type.

Transforming data types

> Data type correction

```
In [160…   mental_health['Age'] = mental_health['Age'].astype('int')

In [161…   mental_health['Support_Systems_Access'] = mental_health['Support_Systems_Access'].astype('bool')
           mental_health['Online_Support_Usage'] = mental_health['Online_Support_Usage'].astype('bool')

In [162…   mental_health['Gaming_Hours'] = mental_health['Gaming_Hours'].round(2)
           mental_health['Physical_Activity_Hours'] = mental_health['Physical_Activity_Hours'].round(2)
           mental_health['Screen_Time_Hours'] = mental_health['Screen_Time_Hours'].round(2)
           mental_health['Sleep_Hours'] = mental_health['Sleep_Hours'].astype('int')
           mental_health['Social_Media_Usage_Hours'] = mental_health['Social_Media_Usage_Hours'].round(2)
           mental_health['Technology_Usage_Hours'] = mental_health['Technology_Usage_Hours'].round(2)
```

Figure 5

Handling Outliers

- Regarding summary statistics outliers were noticed, so the IQR method was followed to handle the outliers.

Handling Outliers

```
In [28]:  def outliers_values(mental_health, columns = ['Technology_Usage_Hours', 'Social_Media_Usage_Hours',
                                                        'Gaming_Hours', 'Screen_Time_Hours',
                                                        'Physical_Activity_Hours']):
              for column in columns:
                  Qr1 = mental_health[column].quantile(0.25)
                  Qr3 = mental_health[column].quantile(0.75)
                  IQR = Qr3 - Qr1
                  upper_extreme = Qr3 + 1.5 * IQR
                  lower_extreme = Qr1 - 1.5 * IQR
                  mental_health = mental_health[(mental_health[column] >= lower_extreme) & (mental_health[column] <= upper_extreme)]
              return mental_health
          mental_health_cleaned = outliers_values(mental_health)
```

Code explanation

First step: the function was defined with the data frame and a list of numeric columns that contain outliers to remove outliers

Second step: The for loop was made to iterate over each column in columns.

Qr1 is the calculation of the first quartile in each column in columns.

Qr3 is the calculation of the third quartile in each column in columns.

IQR is inter quartile range and it calculates the range between third quartile and first quartile.

Upper_extreme define the above outliers.

lower_extreme define the below outliers.

Third step filter out the data frame to exclude the outliers.

Fourth step return the cleaned data frame.

Final step function was called and stored in a new variable (mental_health_cleaned).

Feature Engineering

- **Categorizing Ages**: A new feature was created by categorizing ages into groups.
  - Function Definition: The function categorize_age is defined to categorize ages into three groups:
  - Old Adult: for ages 46 and above.
  - Adult: for ages 31 and 45.
  - Young Adult: for ages 30 and below.

Categorizing Age to groups

```
In [46]:  def categorize_age(age):
              if age >= 46:
                  return 'Old Adult'
              elif age >= 31:
                  return 'Adult'
              else:
                  return 'Young Adult'

          mental_health_cleaned['Age_Group'] = mental_health_cleaned['Age'].apply(categorize_age)
```

Figure 6

- **Sleep Hours Categorization:** A new feature was created by categorizing Sleeping hours into groups.

- function Definition: The function categorize_sleep_h(sleep) categorizes sleep hours into three statuses:
- "Well Sleeping": for sleep hours greater than or equal to 6.
- "Fair Sleeping": for sleep hours between 4 and 6.
- "Bad Sleeping": for sleep hours less than 4.

Sleep hours categorization

```python
In [50]: def categorize_sleep_h(sleep):
    if sleep >= 6:
        return 'Well Sleeping'
    elif sleep >= 4:
        return 'Fair Sleeping'
    else:
        return 'Bad Sleeping'

mental_health_cleaned['Sleep_Status'] = mental_health_cleaned['Sleep_Hours'].apply(categorize_sleep_h)
```

Figure 7

Key Insights from Analysis

Strong Positive Correlations:

Technology Usage and Screen Time: There are strong positive correlations between Technology_Usage_Hours and Screen Time Hours, indicating that individuals who spend more time on technology tend to have higher overall screen time.

Gaming Hours and Screen Time: A strong positive correlation exists between Gaming Hours and Screen Time Hours, suggesting that individuals who play more video games also have higher screen time.

Negative Correlation with Sleep:

Technology Usage and Sleep: There's a moderate negative correlation between Technology_Usage_Hours and Sleep Hours, suggesting that increased technology usage might be associated with reduced sleep duration.

Screen Time and Sleep: Similarly, a negative correlation exists between Screen Time Hours and Sleep Hours, reinforcing the potential link between excessive screen time and sleep disturbances.

Weak or No Correlations:

Social media and Mental Health: The correlations between Social_Media_Usage_Hours and the mental health indicators (stress_level_score, Sleep Hours) are weak or non-existent, suggesting that social media usage alone might not be a primary driver of these outcomes.

Physical Activity and Mental Health: The correlations between Physical_Activity_Hours and the mental health indicators are also weak, indicating that physical activity might not have a strong direct impact on stress levels or sleep quality in this dataset.

```
In [90]: mental_health['stress_level_score'] = mental_health['Stress_Level'].map({'Low': 1, 'Medium': 2, 'High': 3})
         selected_columns2 = ['Technology_Usage_Hours', 'Social_Media_Usage_Hours', 'Gaming_Hours', 'Screen_Time_Hours', 'Sleep_Hours', 'Physical_Activity_Hours', 'stress_level_score']
         columns_corr = mental_health[selected_columns2].corr()
         columns_corr
```

Out[90]:

	Technology_Usage_Hours	Social_Media_Usage_Hours	Gaming_Hours	Screen_Time_Hours	Sleep_Hours	Physical_Activity_Hours	stress_level_score
Technology_Usage_Hours	1.000000	0.000756	0.443334	0.942867	-0.541570	0.007193	-0.016564
Social_Media_Usage_Hours	0.000756	1.000000	0.003524	0.001890	-0.014205	-0.001456	0.346965
Gaming_Hours	0.443334	0.003524	1.000000	0.716641	-0.437850	0.000747	-0.009151
Screen_Time_Hours	0.942867	0.001890	0.716641	1.000000	-0.584131	0.005859	-0.016286
Sleep_Hours	-0.541570	-0.014205	-0.437850	-0.584131	1.000000	-0.006659	0.008119
Physical_Activity_Hours	0.007193	-0.001456	0.000747	0.005859	-0.006659	1.000000	0.002947
stress_level_score	-0.016564	0.346965	-0.009151	-0.016286	0.008119	0.002947	1.000000

Figure 8

Visualizations

Correlation Heatmap

- A heatmap was created to show correlations between key variables like **Technology_Usage_Hours**, **Stress_Level**, and **Sleep_Hours**. Strong correlations were observed between high screen time and stress, while sleep showed an inverse relationship with stress.
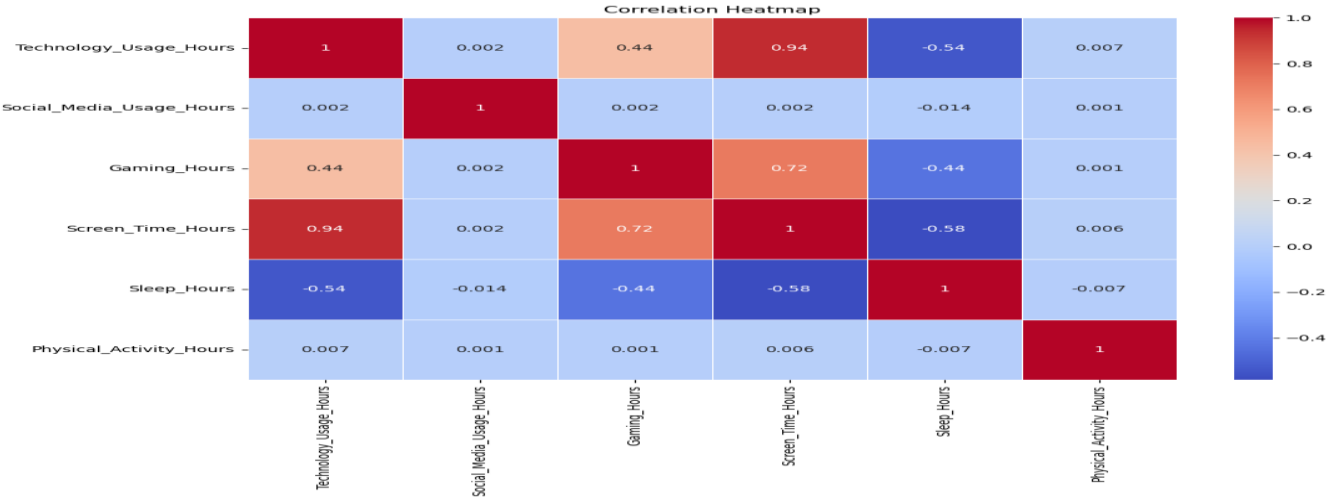
Figure 9

Box Plot of Continuous Variables

- The box plot visually represents the distribution of the continuous variables: **Technology_Usage_Hours, Social_Media_Usage_Hours, Gaming_Hours, Screen_Time_Hours, and Sleep_Hours.**

Distribution for Continuous Variables
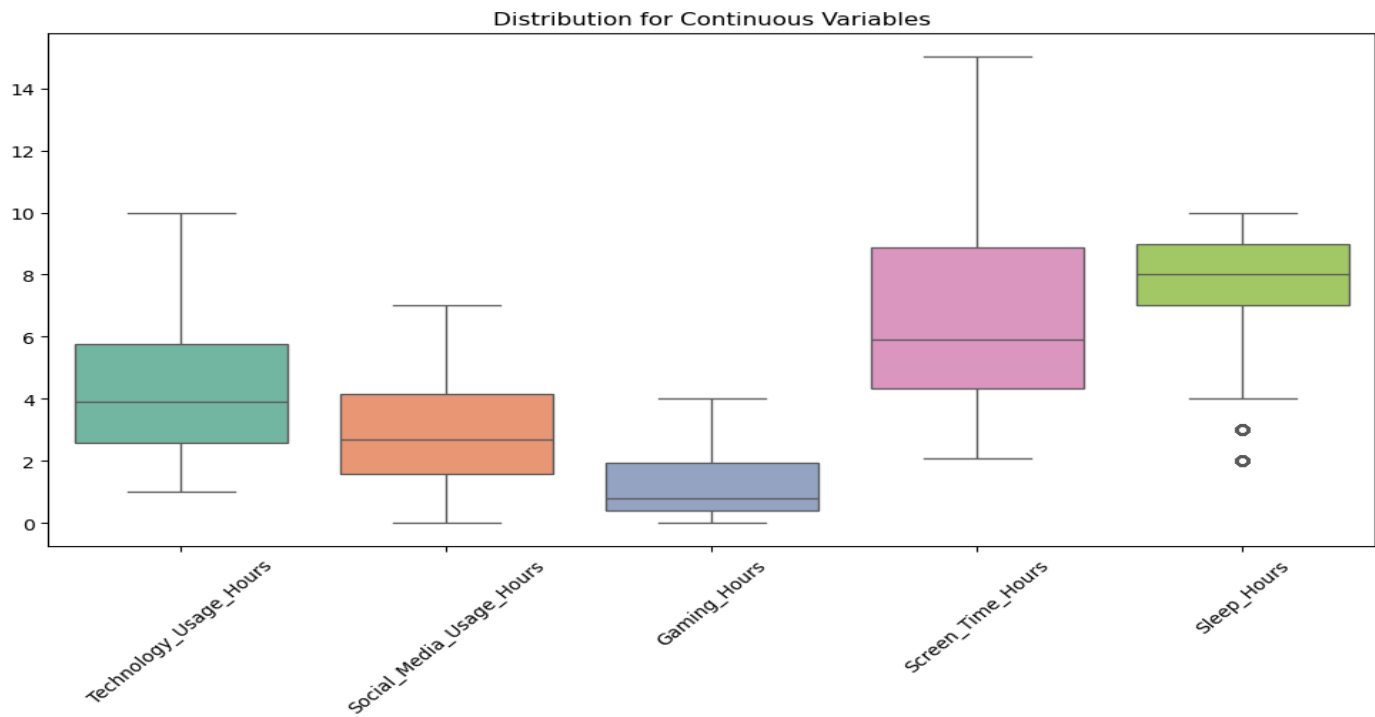
Figure 10

Box Plot: Sleep Hours and Mental Health.

- A box plot displayed the distribution of **Sleep_Hours** across different **Mental_Health_Status** categories. Individuals with fewer than 6 hours of sleep were more likely to report poorer mental health.
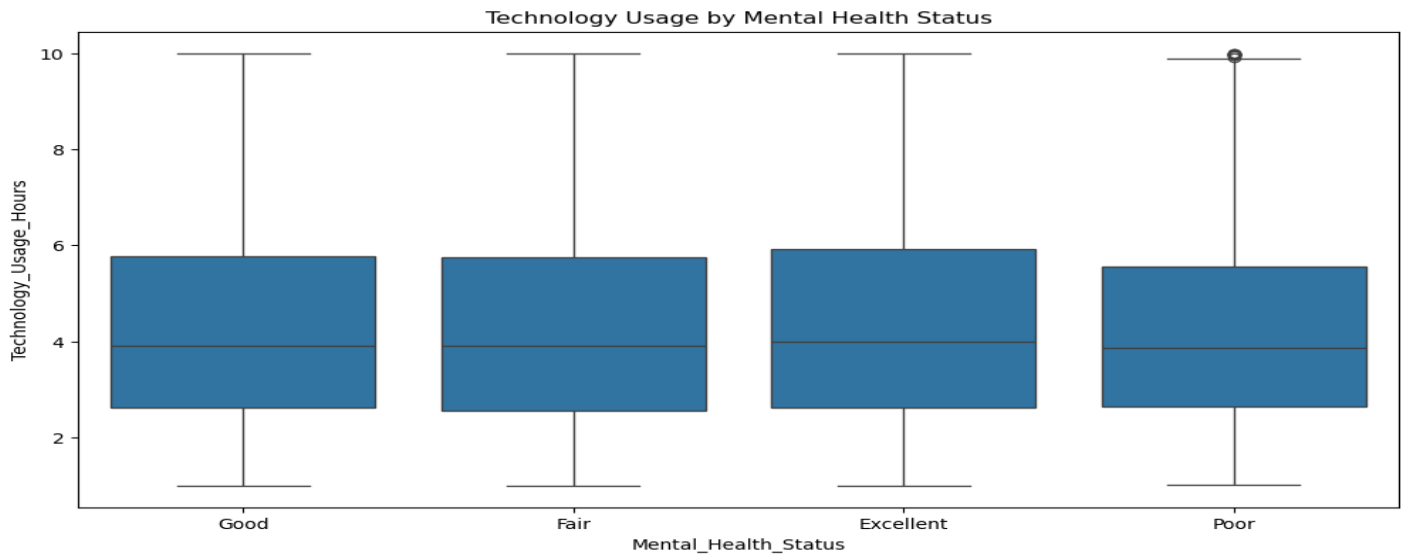

Technology Usage by Mental Health Status

Figure 11

Bar Chart: Stress Level Distribution by Gender.

- The bar chart illustrates the distribution of stress levels across genders: Female, Male, and Other.

Figure 12

Bar Chart: Average Technology Usage Hours by Sleep Status and Age Group.

- The bar chart illustrates how average technology usage hours vary across different sleep statuses (Bad, Fair, Well) and age groups (Young Adult, Adult, Old Adult).
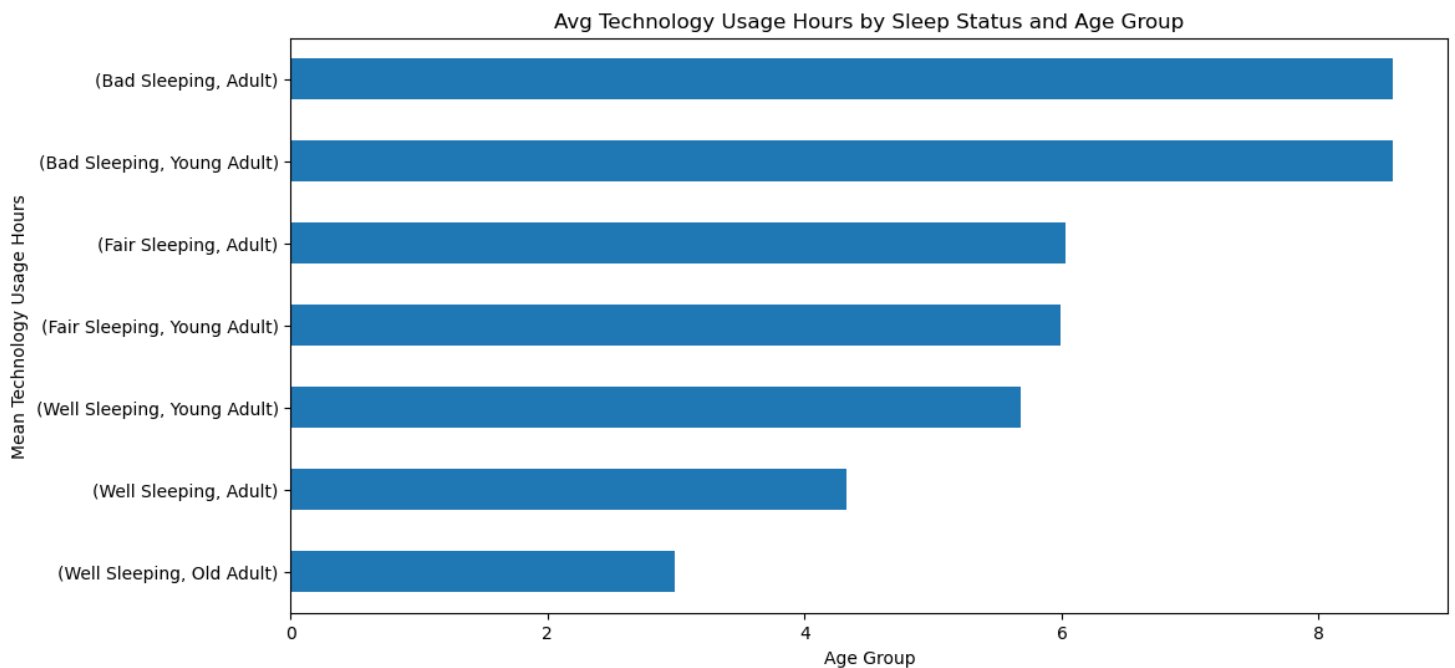
Figure 13

line Chart: Average sleep hours per night for different age groups and genders.

- The x-axis represents the age groups: Adult, Old Adult, and Young Adult. The y-axis represents the average sleep hours. The different lines represent the genders: Female, Male, and Other.
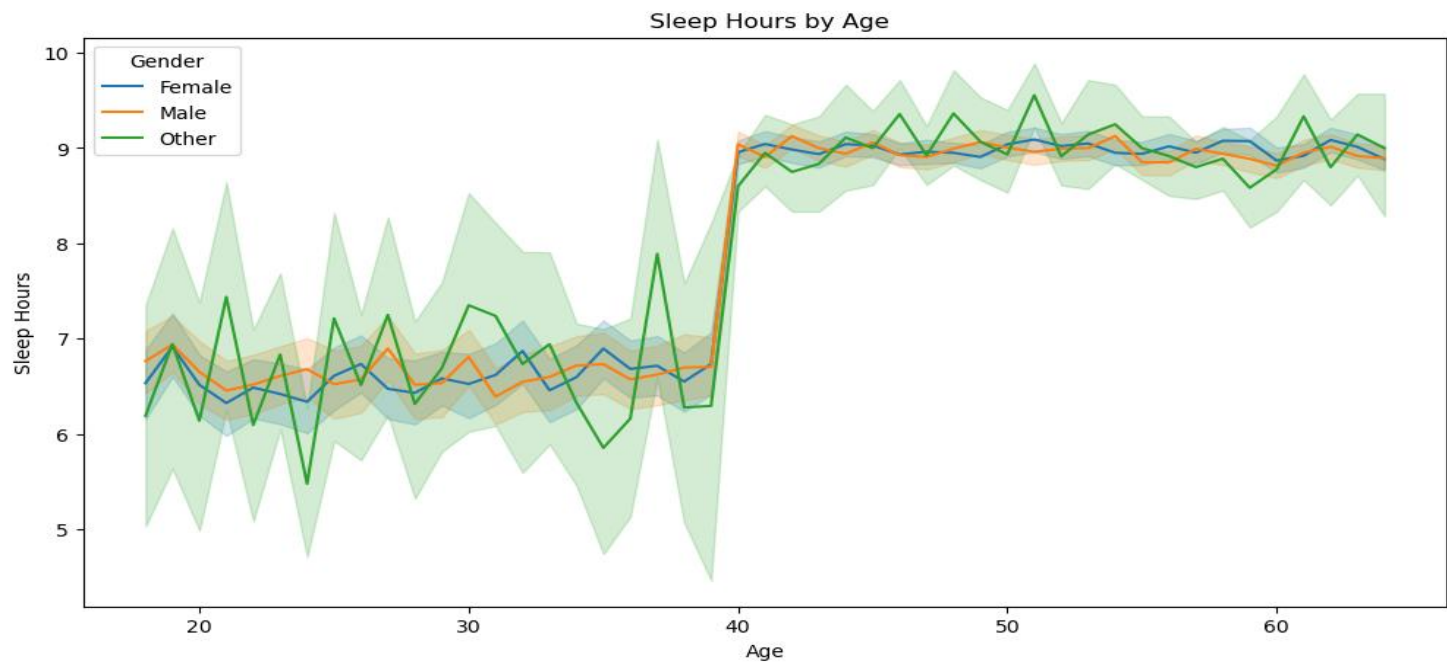
Figure 14

Horizontal Bar Chart: Average technology usage hours per day for different age groups and sleep statuses.

- The x-axis represents the mean technology usage hours, and the y-axis represents the age groups and their sleep statuses.
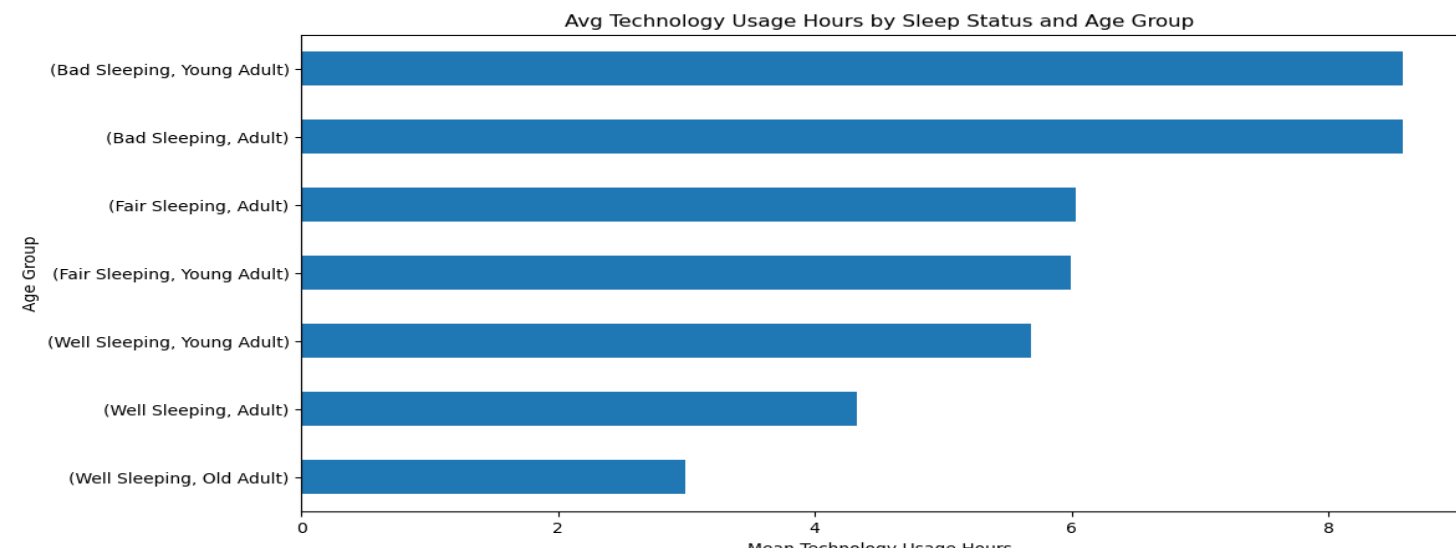
Figure 15

Bar Chart: Users' Stress levels based on their high social media usage.

- The x-axis represents the stress levels: High, Medium, and Low. The y-axis represents the number of users.
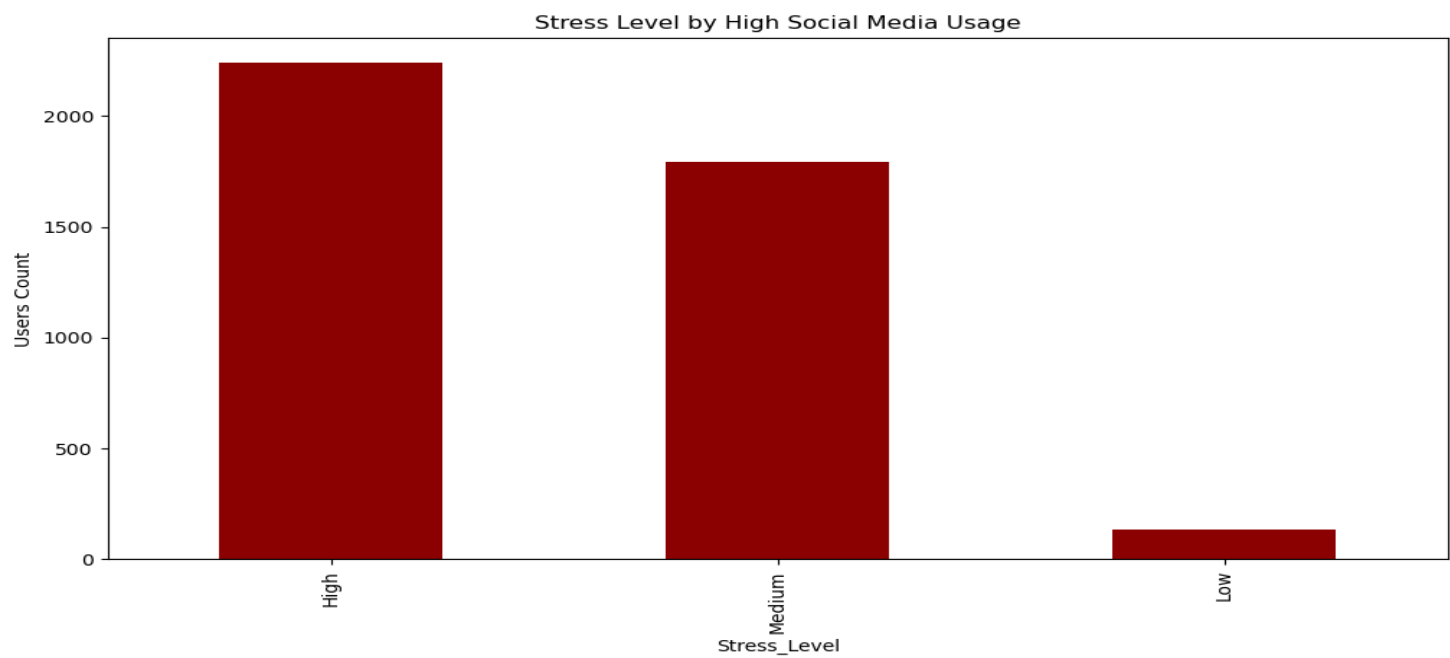
Figure 16

Horizontal Bar Chart: Mental health status of users with support.

- The x-axis represents the number of users, and the y-axis represents the mental health status: Poor, Excellent, Good, and Fair.
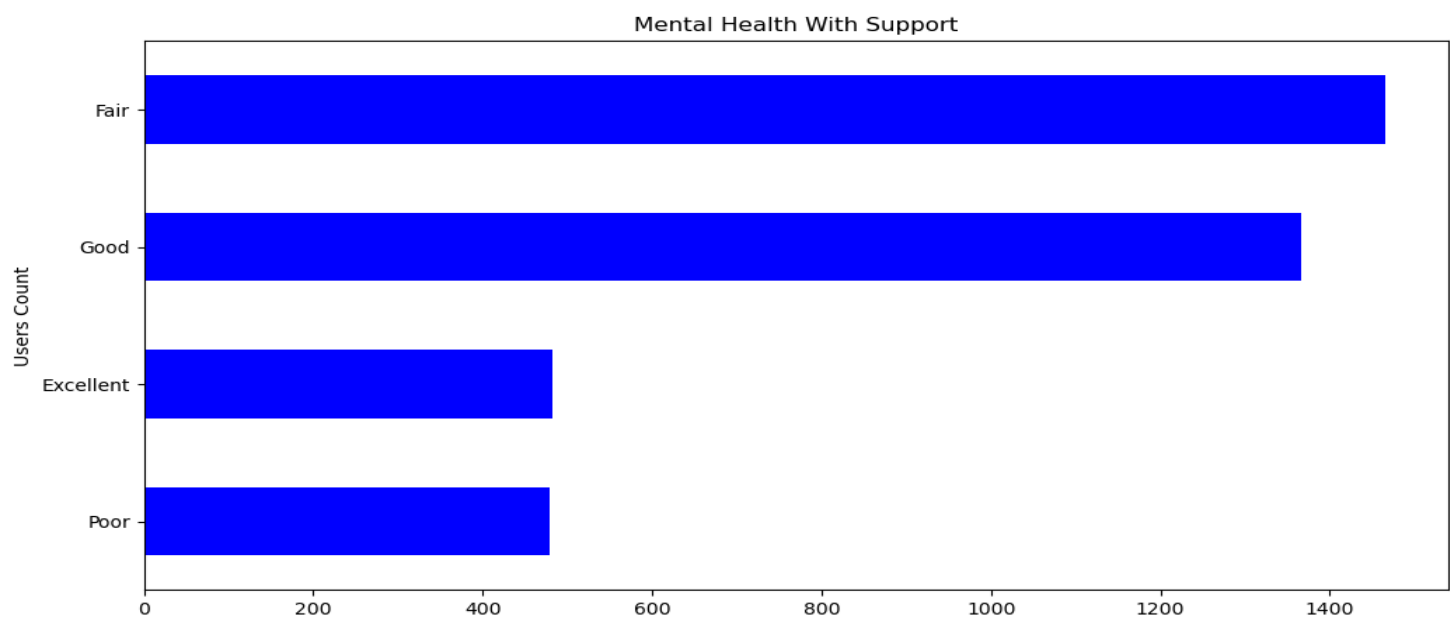
Figure 17

Horizontal Bar Chart: Impact of high stress or high tech usage on the work environment.

- The x-axis represents the number of users, and the y-axis represents the work environment impact: Positive, Neutral, and Negative.
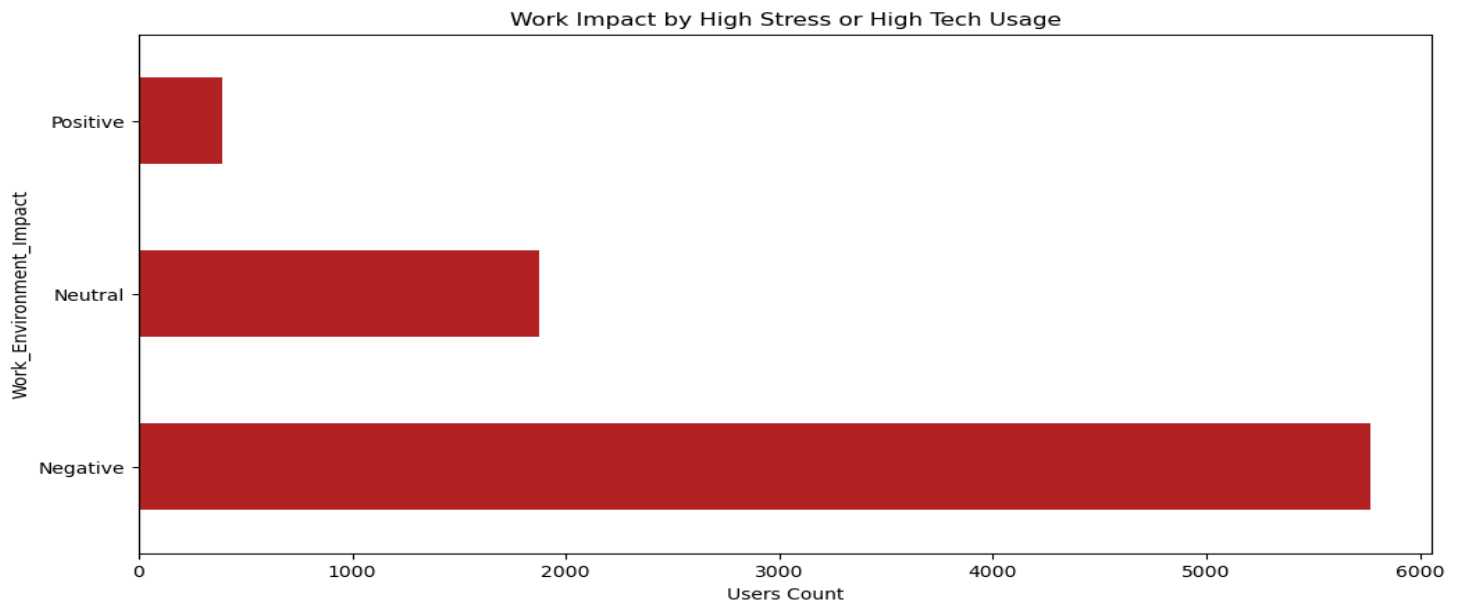
Figure 18

Visualization Dashboard

Interface

- The image conveys the potential negative effects of technology usage on mental health. The overwhelmed person and the rain cloud visually emphasize the stress and anxiety associated with excessive technology use. The text boxes are part of a larger presentation or resource that provides more details and statistics about this topic.
  - **Overall Statistics**: This section summarizes key metrics such as average sleep hours, average age, total users, and all metrics that measure the distribution of variables.
  - **Mental Health**: This section offers a comprehensive and interactive way to explore the complex relationships between **mental health**, technology usage, and other factors. It allows users to drill down into specific areas of interest and gain insights into the data.
  - **Stress Level**: This section offers a comprehensive and interactive way to explore the complex relationships between **Stress Level**, technology usage, and other factors. It allows users to drill down into specific areas of interest and gain insights into the data.
  - **Work Impact**: This section offers a comprehensive and interactive way to explore the complex relationships between **Work Impact**, technology usage, and other factors. It allows users to drill down into specific areas of interest and gain insights into the data.

Figure 19

Overall Statistics:

- Gaming **Hours & Sleeping Hours for Mental Health:** This chart shows the relationship between gaming hours and sleep hours for different mental health statuses.
- **Age Distribution:** This chart illustrates the distribution of ages among the participants.
- **Stress Level:** This section provides information about the distribution of stress levels among different genders and age groups.
- **Mental Health Status:** This section shows the distribution of mental health statuses (e.g., excellent, fair, good).
- **Sleep Status:** This section displays the distribution of sleep statuses (e.g., well sleeping, fair sleeping, bad sleeping).
- **Screen Time & Sleep Hours for Work Impact:** This chart explores the relationship between screen time, sleep hours, and work impact.
- **Work Impact:** This section presents the distribution of work impact (e.g., positive, negative, neutral).
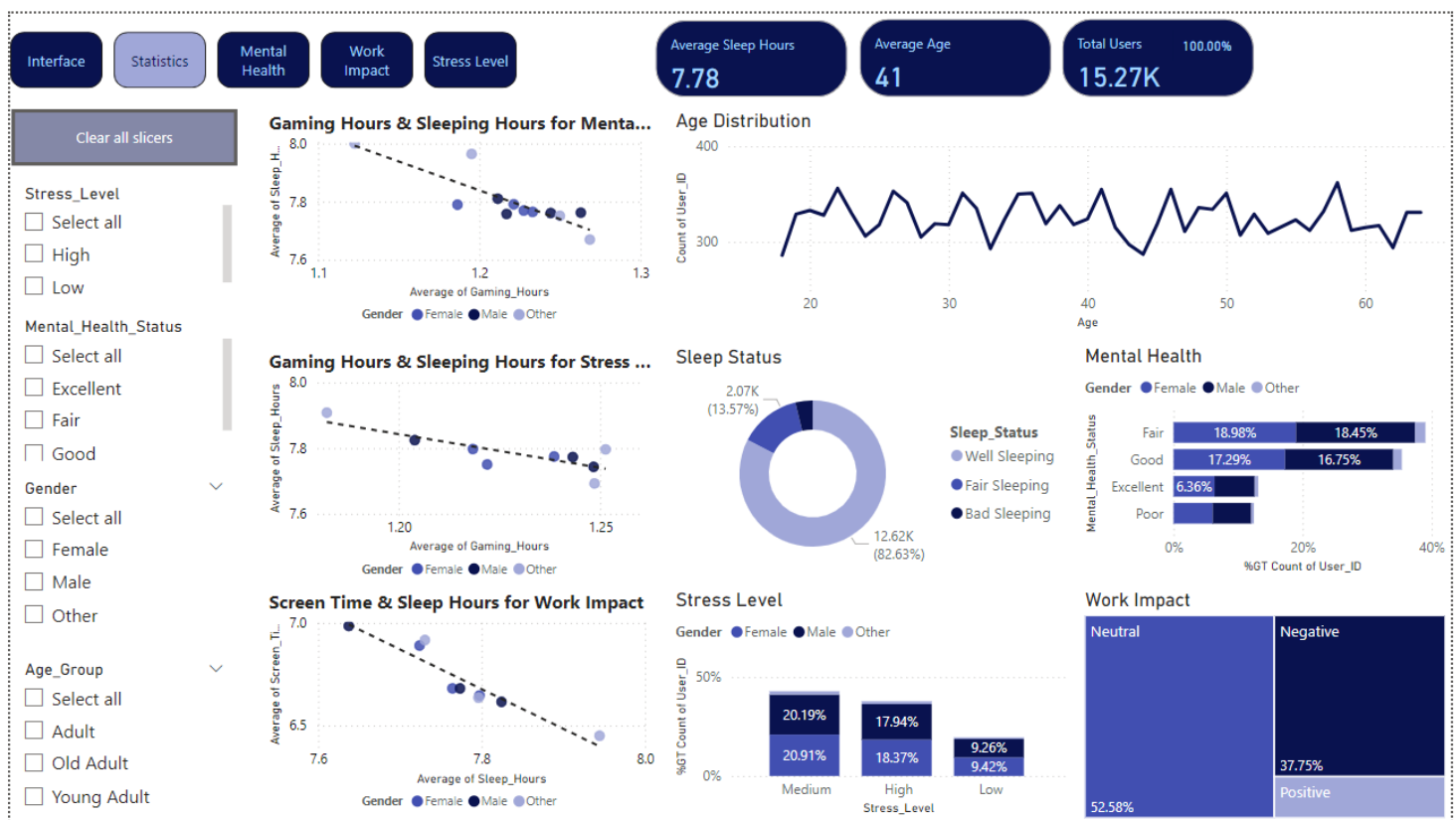

Figure 20

Mental Health:

- **Gaming & Mental Health:** This chart shows the relationship between gaming habits and mental health indicators.
- **Mental Health by Work Impact:** This chart explores the impact of work on mental health.
- **Physical Activity & Mental Health:** This chart analyzes the relationship between physical activity and mental health.
- **Social media & Mental Health:** This chart investigates the connection between social media usage and mental health.
- **Support System & Mental Health:** This chart examines the role of support systems in influencing mental health.
- **Sleeping & Mental Health:** This chart explores the relationship between sleep habits and mental health.
- **Screen Time & Mental Health:** This chart analyzes the impact of screen time on mental health.
- **Stress Level & Mental Health:** This chart investigates the connection between stress levels and mental health.
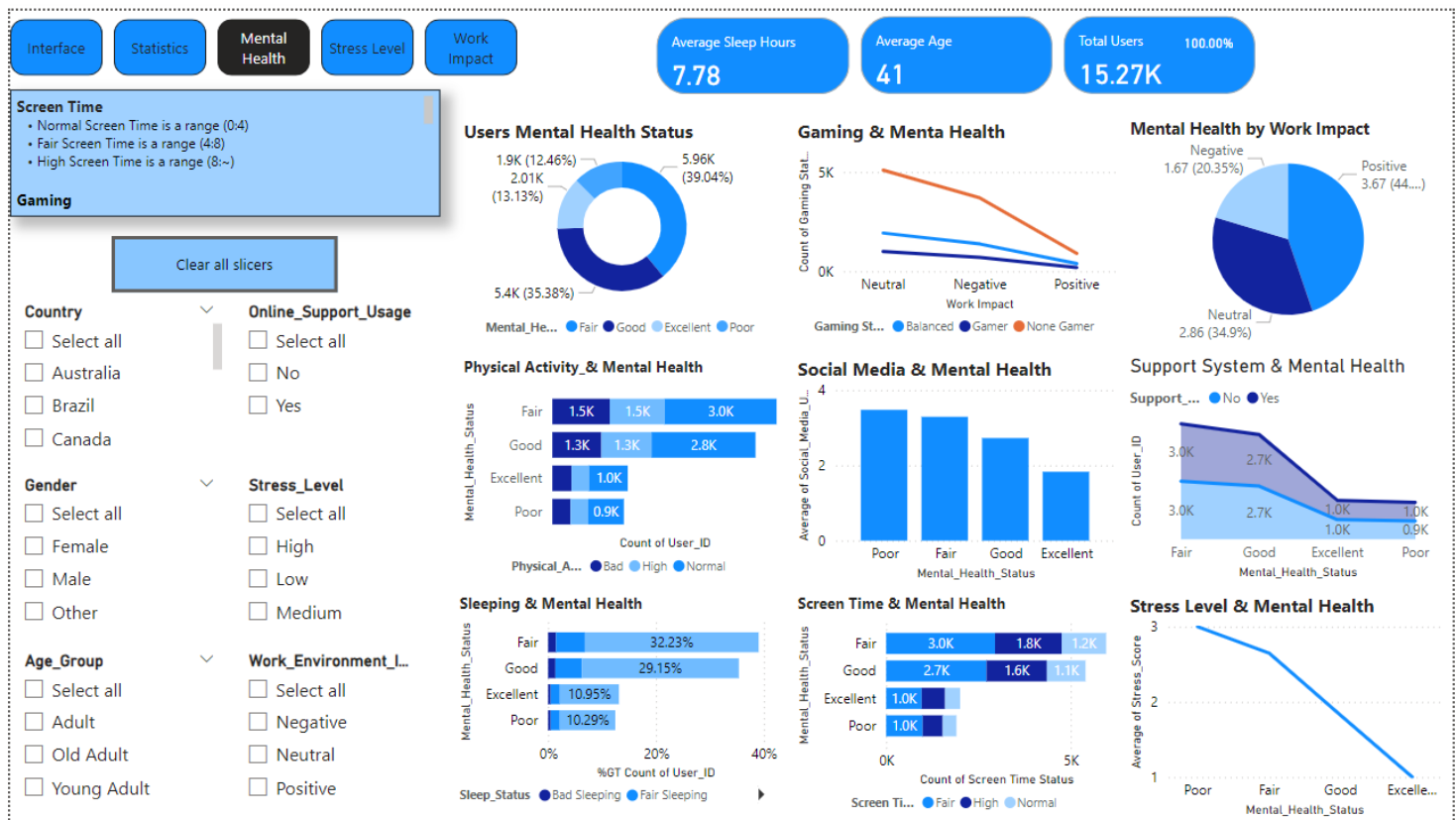

Figure 21

Stress Level:

- **Gaming & Work Impact:** This chart explores the relationship between gaming habits and work impact.
- **Stress Level by Mental Health:** This chart analyzes the distribution of stress levels across different mental health statuses.
- **Physical Activity & Stress Level:** This chart investigates the connection between physical activity and stress levels.
- **Social Media & Stress Level:** This chart examines the relationship between social media usage and stress levels.
- **Support System & Stress Level:** This chart explores the role of support systems in influencing stress levels.
- **Sleeping & Stress Level:** This chart analyzes the relationship between sleep habits and stress levels.
- **Screen Time & Stress Level:** This chart investigates the impact of screen time on stress levels.
- **Stress Level & Work Impact:** This chart explores the relationship between stress levels and work impact.
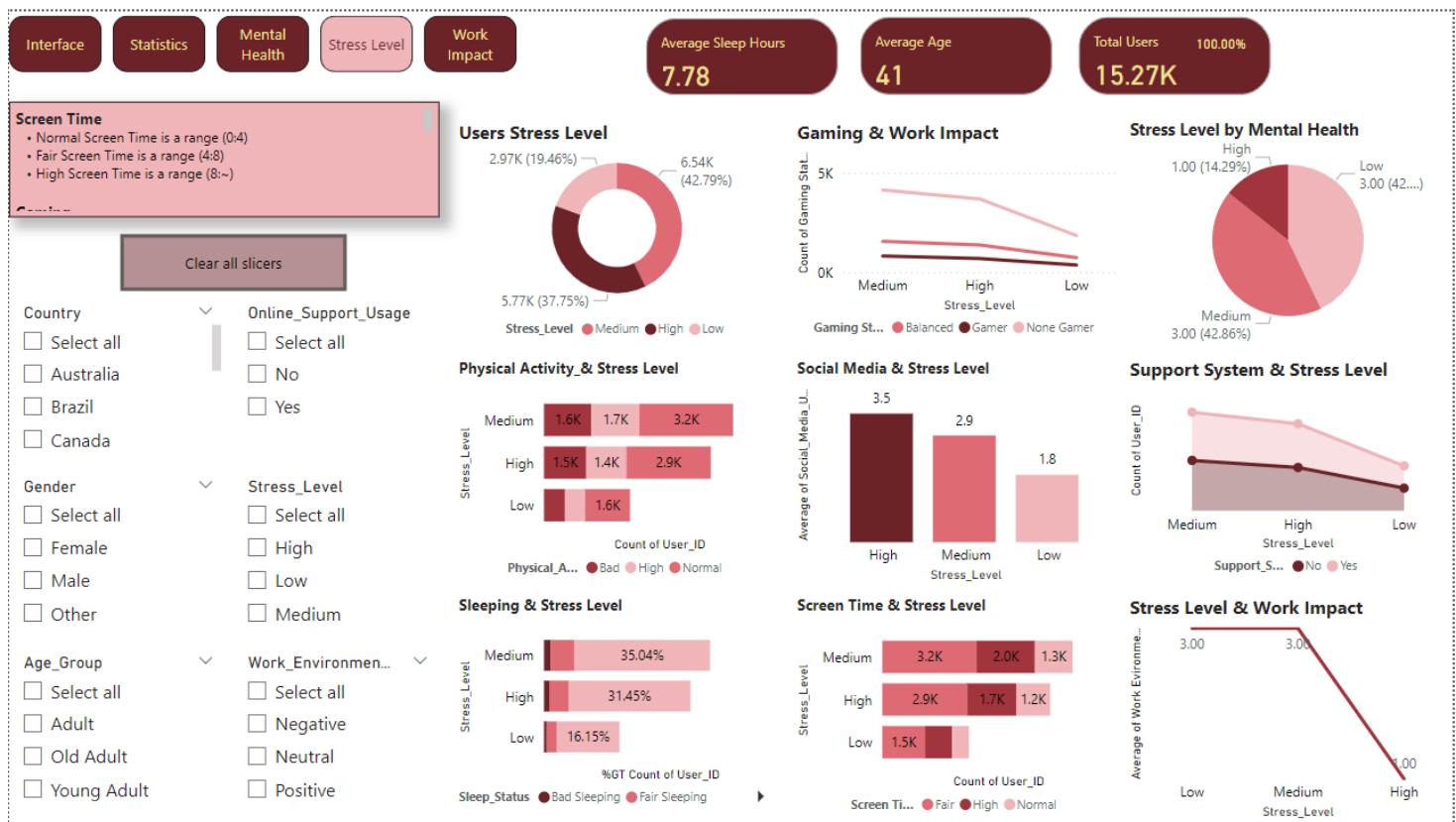

Figure 22

Work Impact:

- **Users Work Impact:** This chart shows the distribution of work impact among users.
- **Gaming & Work Impact:** This chart explores the relationship between gaming habits and work impact.
- **Work Impact on Mental Health:** This chart analyzes the impact of work on mental health based on different mental health statuses.
- **Physical Activity & Work Impact:** This chart investigates the connection between physical activity and work impact.
- **Social media & Work Impact:** This chart examines the relationship between social media usage and work impact.
- **Support System & Work Impact:** This chart explores the role of support systems in influencing work impact.
- **Sleeping & Work Impact:** This chart analyzes the relationship between sleep habits and work impact.
- **Screen Time & Work Impact:** This chart investigates the impact of screen time on work impact.
- **Stress Level & Work Impact:** This chart explores the relationship between stress levels and work impact.
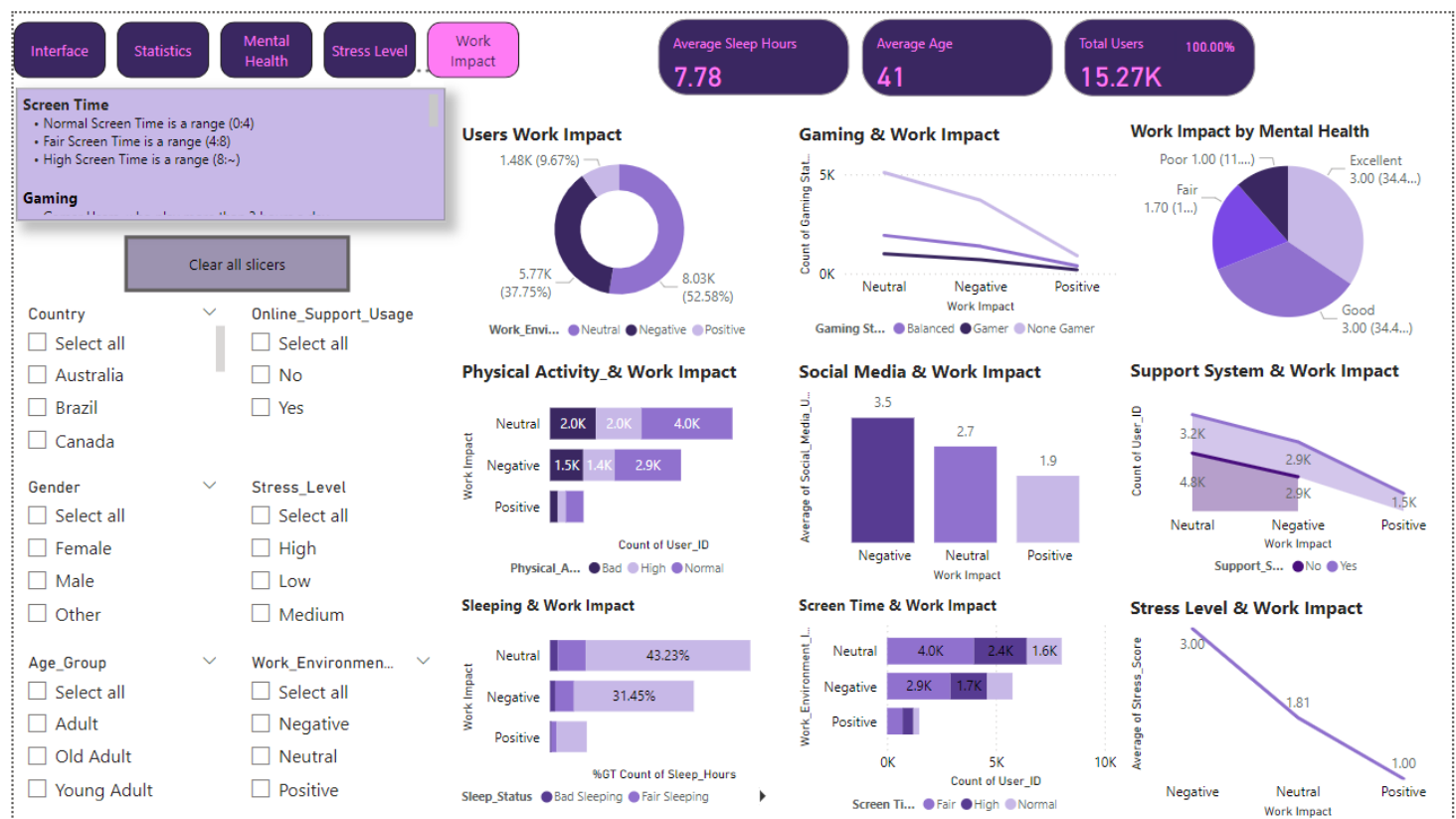

Figure 23

Definitions:

These definitions provide a standardized way to categorize individuals based on their behaviors and habits for technology usage, mental health, or work performance analysis.
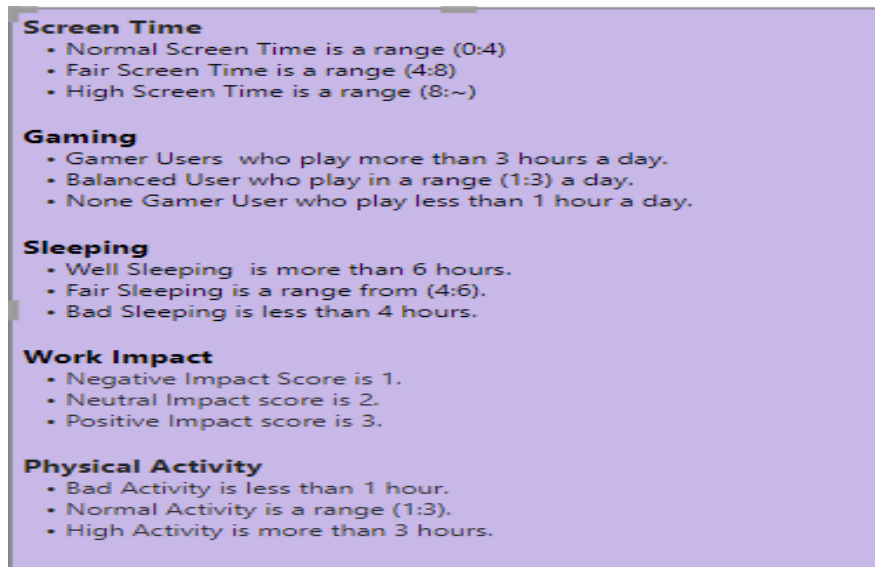
Screen Time
- Normal Screen Time is a range (0:4)
- Fair Screen Time is a range (4:8)
- High Screen Time is a range (8:~)

Gaming
- Gamer Users who play more than 3 hours a day.
- Balanced User who play in a range (1:3) a day.
- None Gamer User who play less than 1 hour a day.

Sleeping
- Well Sleeping is more than 6 hours.
- Fair Sleeping is a range from (4:6).
- Bad Sleeping is less than 4 hours.

Work Impact
- Negative Impact Score is 1.
- Neutral Impact score is 2.
- Positive Impact score is 3.

Physical Activity
- Bad Activity is less than 1 hour.
- Normal Activity is a range (1:3).
- High Activity is more than 3 hours.

Figure 24

Columns Created:

- Stress Level Score: this code assigns a numerical value to the Stress_Score based on the categorical value in the Stress_Level column.

```
1  Stress_Score =
2  SWITCH(
3      mental_health[Stress_Level],
4      "High", 3,
5      "Medium", 2,
6      "Low", 1,0
7  )
```

Figure 25

- Physical Activity: this code assigns a categorical label ("High", "Normal", or "Bad") to the Physical_Activity variable based on the number of hours spent on physical activity.

```
1  Physical_Activity = SWITCH(
2      TRUE(),
3      mental_health[Physical_Activity_Hours] >=3, "High",
4      mental_health[Physical_Activity_Hours] >=1, "Normal",
5      "Bad"
6      )
```

Figure 26

- Work Environment Score: this code assigns a numerical value to the Work Environment Score based on the categorical value in the Work_Environment_Impact column.

```
1  Work Evironment Score = SWITCH(
2      TRUE(),
3      mental_health[Work_Environment_Impact] = "Negative", 1,
4      mental_health[Work_Environment_Impact] = "Nuetral", 2,
5      3
6  )
```

Figure 27

- Screen Time Status: this code assigns a categorical label ("High", "Fair", or "Normal") to the Screen Time Status variable based on the number of hours spent on screen time.

```
1  Screen Time Status = SWITCH(
2      TRUE(),
3      mental_health[Screen_Time_Hours] >= 8,
4      "High",
5      mental_health[Screen_Time_Hours] >= 4,
6      "Fair",
7      "Normal"
8  )
```

Figure 28

- Gaming Status: this code assigns a categorical label ("Gamer", "Balanced", or "None Gamer") to the Gaming Status variable based on the number of hours spent gaming.

```
1  Gaming Status = SWITCH(
2      TRUE(),
3      mental_health[Gaming_Hours] >= 3, "Gamer",
4      mental_health[Gaming_Hours] >= 1, "Balanced",
5      "None Gamer"
6  )
```

Figure 29

Measures:

- Average Sleep Hours: This measure calculates the average of the values in the Sleep_Hours column of the mental_health DataFrame.

```
1  Average Sleep Hours = AVERAGE(
2  |    mental_health[Sleep_Hours])
```

Figure 30

- Average Age: This measure calculates the average of the values in the Age column of the mental_health DataFrame.

```
1  Average Age = AVERAGE(
2  |    mental_health[Age]
3  )
```

Figure 31

- Total Users: This measure calculates the total of users of the mental_health DataFrame.

```
1  Total Users = COUNT(mental_health[User_ID])
```

- User Percentage: This measure calculates the percentage of users of the total users of mental_health DataFrame.