Motivation

The motivation behind choosing this dataset is straightforward, social media is a major part of daily life, especially for students. All of us use social platforms, and we're familiar with both the positive and negative effects they can have. This made the dataset relatable, relevant, and worth analysing. We also wanted to explore these effects more deeply and possibly contribute to and support healthier well-being in the digital age.

Project Goal

The main aim of this project is to explore the relationship between social media usage and mental health indicators. Specifically, we want to understand how digital habits (screen time, platform choice, time offline) and lifestyle factors (sleep, exercise, stress, happiness) influence overall well-being. The goal is not only to identify trends, but also to highlight the factors that may help individuals improve or maintain healthier mental wellness.

Social Media Wellbeing Analysis

We are working with the *Social Media and Mental Health Balance* dataset from Kaggle. This dataset examines how digital habits link to mental well-being. It includes information on daily screen time, social media platform preferences, sleep quality, happiness levels, stress scores, exercise frequency, and days spent without social media. These variables give us a broad view of how different lifestyle choices and digital behaviours may influence overall mental wellness. Through out this analysis we hope to answer the following questions:

Resources

For this project we will be using the following resources:

- GitHub
- Jupyter Notebook
- VS Code
- Dataset from Kaggle

Libraries we will explore:

- Pandas
- Matplotlib
- Numpy
- Seaborne
- Scikit - Learn

Target Audience

The findings from this project can benefit multiple groups:

1. **Social media companies** – to understand how their platforms may affect their users mental health and to consider features or adjustments that encourage healthier usage.
2. **Students and general social media users** – to increase awareness of how certain habits, screen time patterns, and platform choices may influence their well-being.
3. **Well-being educators** – to support guidance, awareness campaigns, or interventions aimed at promoting healthier online behaviours.

Overall, the project can help individuals make more informed choices about how they use social media and how their lifestyle habits may affect their mental state.

Q1. To what extent is the number of Days_Without_Social_Media related to an individual's overall Wellbeing_Score?

Hypothesis:

- Individuals who spend more days without social media will show a higher happiness index.

Relevant Variables:

- Days_Without_Social_Media (numerical)
- **Wellbeing_Score (numerical):**
  - Happiness_Index
  - Sleep_Quality
  - Stress_Level
  - Exercise_Frequency

Proposed Analysis Approach:

We will start by looking at the factors that would make up the wellbeing score. This will consist of factors such as happiness score, sleep quality, exercise frequency and stress level.

We will be combining the total of Sleep Quality, Exercise Frequency, and Happiness Index and then deducting the inverse of the Stress Level (e.g., $11 - \text{Stress Level}$) to get the final overall Wellbeing Score.

We will then analyze and see if there is a direct correlation and a linear relationship between the calculated Wellbeing Score and Days Without Social Media.

Q2. What is the typical Age and Gender profile for users of different Social_Media_Platform, and how does their average Daily_Screen_Time compare?

Hypothesis:

- Age and gender will significantly influence which social media platforms people use.

Relevant Variables:

- Age (numerical)
- Gender (categorical)
- Social_Media_Platform (categorical)
- Daily_Screen_Time(hrs) (numerical)

Proposed Analysis Approach:

We will begin by selecting the key variables (Age, Gender, Social Media Platform, and Daily Screen Time) and grouping users by platform. For each platform, we will look at the age and gender distributions to understand the typical profile of its users.

We will then compare the average daily screen time across different platforms to see whether certain groups (e.g., younger users or specific genders) tend to spend more time online.

To visualise these patterns clearly, we will use bar charts and box plots to show platform differences in age, gender, and screen-time behaviour. We will also include a multi-variable scatter plot that maps Age against Gender, with each point coloured by Social Media Platform (and possibly sized by screen time). This will help us visually identify clustering patterns and see whether age and platform choice relate to how long users spend online.

Q3. What are the most significant predictors of a user's Stress_Level, and can a model accurately predict whether a user falls into the high or low stress categories based on lifestyle factors?

Hypothesis:

- Higher daily screen time will be associated with higher stress levels.

Relevant Variables:

- Stress_Level (numerical → converted into a binary category)
- 0 = low stress (1-5)
- 1= high stress (6-10)

Proposed Analysis Approach:

We will start by converting the continuous Stress Level column into two distinct binary categories: High Stress(6-10) and Low Stress (1-5).

To identify the most significant predictors, we can use a correlation heatmap to visually see the linear relationship between Stress and other numerical factors (like Screen Time, Sleep Quality, Exercise Frequency).

We will then build a classification model that uses the identified lifestyle factors as inputs (predictors).

The model will be trained to predict which of the two categories ("High" or "Low" Stress) a user belongs to.