Mental Health and Social Media Balance Dataset

2025-10-25

Summary of the "Mental Health and Social Media Balance" Dataset This dataset provides a valuable snapshot of the complex relationship between digital habits and personal well-being. It is designed to help students and researchers explore how modern technology use correlates with mental health outcomes. The data appears to be cross-sectional, meaning it captures information from different users at a single point in time. The dataset includes key demographic variables such as Age and Gender. The core of the data focuses on user behavior, specifically Daily_Screen_Time (measured in hours) and the primary Social_Media_Platform each user prefers (like TikTok, Facebook, or Instagram). To measure well-being, the dataset uses three simple, self-reported metrics on a 1-to-10 scale: Sleep_Quality, Stress_Level, and a general Happiness_Index. This quantitative data makes it ideal for statistical analysis. Furthermore, the dataset includes other important lifestyle factors that might influence these outcomes. These are Exercise_Frequency (how many times per week) and Days_Without_Social_Media, which can be used to analyze the potential benefits of a "digital detox." This dataset is quite suitable, especially for academic projects or entry-level data analyses. It allows for the investigation of direct relationships between lifestyle choices, such as screen time and exercise, and self-reported mental health metrics. I can easily apply foundational statistical methods, like correlation matrices or simple t-tests, to compare group outcomes. By analyzing these variables, patterns regarding digital balance can be uncovered and hypotheses can be tested. For instance, it provides a solid foundation for discussing the extent to which time spent offline or the use of different platforms impacts overall happiness and stress. The findings can spark critical conversations about digital literacy and the formation of healthier personal technology habits, making it an extremely valuable resource for understanding contemporary well-fbeing challenges in a connected world.

```r
install.packages("readr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```r
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.5'
## (as 'lib' is unspecified)
```

```r
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
data <- read_csv("Mental_Health_and_Social_Media_Balance_Dataset.csv")
```

```
## Rows: 500 Columns: 10
## -- Column specification -------------------------------------------------
```

```
## Delimiter: ","
## chr (3): User_ID, Gender, Social_Media_Platform
## dbl (7): Age, Daily_Screen_Time(hrs), Sleep_Quality(1-10), Stress_Level(1-10...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
#Control
glimpse(data)
```

```
## Rows: 500
## Columns: 10
## $ User_ID                 <chr> "U001", "U002", "U003", "U004", "U005", "U0~
## $ Age                     <dbl> 44, 30, 23, 36, 34, 38, 26, 26, 39, 39, 18,~
## $ Gender                  <chr> "Male", "Other", "Other", "Female", "Female~
## $ `Daily_Screen_Time(hrs)` <dbl> 3.1, 5.1, 7.4, 5.7, 7.0, 6.6, 7.8, 7.4, 4.7~
## $ `Sleep_Quality(1-10)`   <dbl> 7, 7, 6, 7, 4, 5, 4, 5, 7, 6, 7, 5, 7, 5, 7~
## $ `Stress_Level(1-10)`    <dbl> 6, 8, 7, 8, 7, 7, 8, 6, 7, 8, 6, 7, 10, 7, ~
## $ Days_Without_Social_Media <dbl> 2, 5, 1, 1, 5, 4, 2, 1, 6, 0, 2, 3, 7, 4, 4~
## $ `Exercise_Frequency(week)` <dbl> 5, 3, 3, 1, 1, 3, 0, 4, 1, 2, 0, 2, 1, 0, 3~
## $ Social_Media_Platform   <chr> "Facebook", "LinkedIn", "YouTube", "TikTok"~
## $ `Happiness_Index(1-10)`  <dbl> 10, 10, 6, 8, 8, 8, 7, 7, 9, 7, 7, 9, 8, 8,~
```

```r
head(data)
```

```
## # A tibble: 6 x 10
##   User_ID   Age Gender `Daily_Screen_Time(hrs)` `Sleep_Quality(1-10)`
##   <chr>   <dbl> <chr>                    <dbl>                 <dbl>
## 1 U001       44 Male                       3.1                     7
## 2 U002       30 Other                      5.1                     7
## 3 U003       23 Other                      7.4                     6
## 4 U004       36 Female                     5.7                     7
## 5 U005       34 Female                     7                       4
## 6 U006       38 Male                       6.6                     5
## # i 5 more variables: `Stress_Level(1-10)` <dbl>,
## #   Days_Without_Social_Media <dbl>, `Exercise_Frequency(week)` <dbl>,
## #   Social_Media_Platform <chr>, `Happiness_Index(1-10)` <dbl>
```

```r
tail(data)
```

```
## # A tibble: 6 x 10
##   User_ID   Age Gender `Daily_Screen_Time(hrs)` `Sleep_Quality(1-10)`
##   <chr>   <dbl> <chr>                    <dbl>                 <dbl>
## 1 U495       43 Female                     5.6                     8
## 2 U496       23 Male                       6.9                     5
## 3 U497       43 Female                     5.6                     7
## 4 U498       41 Male                       7.7                     5
## 5 U499       23 Male                       4.2                     9
## 6 U500       43 Female                     5.9                     5
## # i 5 more variables: `Stress_Level(1-10)` <dbl>,
## #   Days_Without_Social_Media <dbl>, `Exercise_Frequency(week)` <dbl>,
## #   Social_Media_Platform <chr>, `Happiness_Index(1-10)` <dbl>
```