**Contents**

[**About The Dataset 1**](#_cudy2bhu4yvl)

[Dataset: Remote Work & Mental Health 1](#_htjr17gvq1a)

[Limitations: 2](#_jcyvzxxya5r1)

[**Data Documentation 2**](#_h8g6hsjzmj9x)

[Excel 2](#_eiveenvwv4d7)

[Tableau 2](#_dyqm7osabt8n)

## About The Dataset

### **Dataset:** [**Remote Work & Mental Health**](https://www.kaggle.com/datasets/waqi786/remote-work-and-mental-health)

* **Source:** Kaggle
* **Created:** September 23 2024.
* **Type:** AI-generated data
* **Size:** 5,000 records.
* **Key Variables Used:**
  + Work\_Location
  + Stress\_Level
  + Social\_Isolation\_Rating
  + Satisfaction\_with\_Remote\_Work
  + Region
  + Productivity Change

**Columns in Raw data**

* **Employee\_ID:**  
  A unique, anonymised identifier assigned to each employee.
* **Daily\_Working\_Hours:**The total number of hours an employee dedicates to work in a day.
* **Screen\_Time:**The total duration (in hours) an employee spends in front of digital screens (such as computers, tablets, or smartphones) during the workday.
* **Meetings\_Attended:**The count of meetings the employee participates in during a day.
* **Emails\_Sent:**The number of emails sent by the employee in a day, reflecting communication frequency and workload intensity.
* **Productivity\_Score:**A rating (typically on a scale from 1 to 10) that indicates the employee's perceived or measured productivity for the day.
* **Physical\_Activity\_Steps:**The total number of steps recorded by wearable devices, reflecting the employee’s physical activity level throughout the day.
* **Sleep\_Duration:**The number of hours an employee slept during the previous night, as recorded by wearable devices or self-report.
* **Work\_Life\_Balance\_Satisfaction:**A self-reported satisfaction score (on a scale from 1 to 10) that measures how well an employee feels they are balancing work demands with personal life.
* **Age** : The age of the employee.
* **Gender**: The gender identity.
* **Job\_Role**: The designated position or job title.
* **Industry**: The industry sector the employee belongs to.
* **Work\_Location**: The type of work arrangement: Remote, Hybrid, or Onsite.
* **Stress\_Level**: The self-assessed stress level (Low, Medium, High).
* **Mental\_Health\_Condition**: Reported mental health issues (e.g., Anxiety, Depression).
* **Social\_Isolation\_Rating**: A simulated score (1–5) reflecting feelings of workplace isolation.
* **Satisfaction\_with\_Remote\_Work**: Level of satisfaction with remote working arrangements (Satisfied, Neutral, Unsatisfied).

**Overview**

As flexible work arrangements become more common, it’s increasingly important to examine how different work settings affect employee well-being. This dataset, which contains synthetic data, was created to imitate trends related to remote, hybrid, and onsite work environments, with a focus on factors like stress, mental health, and job satisfaction.

Consisting of 5,000 AI-generated entries, the dataset offers a safe and controlled way for data analysts, HR professionals, and researchers to explore potential patterns and relationships without compromising any real individuals’ privacy.

### **Limitations:**

* The data is AI generated - so the trends may not match reality.
* Self-reported bias - stress, satisfaction and isolation ratings are subjective.
* A dataset of 5,000 records is not sufficiently representative - particularly for global analysis. A significantly larger sample size would be needed to ensure the data accurately reflects the working population worldwide

## 

## **Data Documentation**

### **Excel**

**Stress Level:**

* + Changed to numerical value
  + 1= low, 2= medium, 3 = high
  + Used formula: =IF(F2="High", 3, IF(F2="Medium", 2, IF(F2="Low", 1, "")))

**Satisfaction with Remote Work:**

* + Changed to numerical value
  + Encoded as: Satisfied (1), Neutral (0), Unsatisfied (-1).
  + Used Formula: =IF(K2="Unsatisfied", -1, IF(K2="Neutral", 0, IF(K2="Satisfied", 1, "")))

**Productivity**

* + Changed to numerical value
  + -1 = decrease, 0 = No Change, 1= increase
  + Used formula =IF(H2="Decrease", -1, IF(H2="No Change", 0, IF(H2="Increase", 1, "")))

**Age**

* + Min age was 22 , max age - 60
  + Grouped the ages like so: 20–29,30–39,40–49,50–60
  + Used the formula: =IF(A2<30, "20-29", IF(A2<40, "30-39", IF(A2<50, "40-49", "50-60")))

**Deleted Columns**

* + The following columns were removed from the spreadsheet as they were no longer relevant : gender, employee\_id, job\_role, industry, years\_of\_experience, number\_of\_virtual\_meetings , work\_life\_balance, mental\_health\_condition, access\_to\_ mental\_health\_resources, company\_support\_for\_remote\_work, physical\_activity, Age, emails\_sent, sleep\_duration, physical\_activity\_steps, emails\_sent, meetings\_attended, screen\_time.

### 

### **Tableau**

**Value formatting in Tableau**

* + Reviewed the field types of the columns in Tableau and noticed that the "Region" field was classified as a string. Updated its field type to "Region" accordingly.

**New Calculated Field - Average Stress Level**

IF AVG([Stress Level (Num)]) >= 2.5 THEN "High"

ELSEIF AVG([Stress Level (Num)]) >= 1.5 THEN "Medium"

ELSE "Low"

END

**New Calculated Field - Average Satisfaction with remote work**

* + Created this by by using this formula in Tableau:

IF AVG([Satisfaction with Remote Work]) > 0 THEN "Satisfied"

ELSEIF AVG([Satisfaction with Remote Work]) = 0 THEN "Neutral"

ELSE "Unsatisfied"

END

**New Calculated Field - Average Productivity Change**

Created this by using the formula in Tableau:

IF AVG([Productivity Change Score]) > 0 THEN "Increased"

ELSEIF AVG([Productivity Change Score]) = 0 THEN "No Change"

ELSE "Decreased"

END

**New Calculated Field - Average Social Isolation Rating**

IF AVG([Social Isolation Rating]) >= 4 THEN "High"

ELSEIF AVG([Social Isolation Rating]) >= 2.5 THEN "Medium"

ELSE "Low"

END

**New Calculated Field - Stress Level**

IF [Stress Level (Num)] = 3 THEN "High"

ELSEIF [Stress Level (Num)] = 2 THEN "Medium"

ELSEIF [Stress Level (Num)] = 1 THEN "Low"

END