Power BI Dataset Assessment Form

Name: Renin Rappai

Dataset's Link (Drive link):

https://www.kaggle.com/datasets/sonalshinde123/social-media-mental-health-indicators-dataset/data

Project Goal

The project aims to identify patterns and relationships between **screen time, sleep habits, physical activity, social media interactions, and mental health indicators** such as stress, anxiety, and mood, enabling data-driven insights for better digital well-being and decision-making.

Data Quality

Handling Missing Values

Missing values in the mental health dataset are handled using a logical and analysis-driven approach:

- Records are removed only when critical fields such as Person_ID, Mental_State, Stress_Level, or Platform are missing.

- For non-critical fields (e.g., sleep hours, physical activity, screen time), missing values are:

    o Imputed using average or category-based values, or

    o Labeled as "Unknown" where appropriate.

Data Source(s)

Dataset Description

- **Source:** Kaggle

- **Dataset Type:** CSV (Mental Health & Social Media Usage Dataset)

- **Nature:** Synthetic / survey-based behavioral and mental health data

- **Use Case:** Analysis of stress, anxiety, mood, sleep, physical activity, and social media usage patterns.

Data Content Validation

Data Consistency

- All columns maintain **consistent data types**.

- No mixed data types exist within individual fields.

Calculated / Derived Fields

Yes, the dataset includes **derived fields created in Power BI**, such as:

- Sleep Category

- Age Group

- Average Stress Level

- Average Anxiety Level

- Average Screen Time

- Positive and Negative Interaction Percentages

Categorical Variables

Yes, the dataset includes several categorical variables:

- Mental State

- Platform

- Sleep Category

- Age Group

These are used for **segmentation, comparison, and filtering** across dashboard pages.

Time-Based Data

- Date fields are formatted as **Date / Month / Year**

- Used for:

    o Monthly trend analysis

    o Time-based filtering

    o Behavioral pattern tracking

Outliers & Ranges

- Numerical fields such as screen time and social media time may contain high values.

- **Outliers are:**
  - Reviewed visually
  - Retained if behaviorally valid
  - Aggregated at platform or category level to avoid skewed insights

Dataset Size

- **Dataset Size:** 364 KB