# Synthetic Mental Health Dataset

The synthetic mental health dataset is designed to simulate patient-related information across various dimensions.

**Features included in the dataset:**

1. Patient\_ID: Unique identifier for each patient.
2. Age: Patient's age in years.
3. Gender: Patient's gender (Male/Female).
4. Ethnicity: Patient's ethnicity.
5. Socioeconomic\_Status: Socioeconomic status of the patient (Low/Middle/High).
6. Assessment\_Score: Scores from a standardized mental health assessment.
7. Diagnosis: Primary mental health diagnosis.
8. Diagnosis\_Severity: Severity of the mental health diagnosis.
9. Treatment\_History: History of received treatment (Medication/Therapy/Both/None).
10. Sleep\_Hours: Average hours of sleep per day.
11. Exercise\_Hours: Average hours of exercise per day.
12. Substance\_Use: Whether the patient engages in substance use (Yes/No).
13. Life\_Events: Significant life events affecting the patient.
14. Social\_Support: Level of perceived social support (Strong/Moderate/Weak).
15. Relationship\_Status: Patient's relationship status (Single/In a Relationship/Married).
16. Employment\_Status: Patient's employment status (Employed/Unemployed/Student).
17. Education\_Level: Patient's highest education level.
18. Visits\_to\_Professionals: Frequency of visits to mental health professionals.
19. Genetic\_Markers: Presence or absence of synthetic genetic markers.
20. Outcome\_Variable: Placeholder for a hypothetical outcome variable (e.g., mental health improvement).

**Possible Descriptive and Predictive Research Questions**

These questions cover a range of descriptive and predictive aspects, providing avenues for exploration and analysis within the synthetic mental health dataset.

**Descriptive Analysis**

1. What is the demographic distribution of patients in terms of age, gender, and ethnicity?
2. How is socioeconomic status distributed among the patient population?
3. What are the prevalent mental health diagnoses in the synthetic dataset?
4. How common is the use of medication, therapy, or both in the treatment history of patients?
5. What are the average sleep and exercise hours for patients in the dataset?

**Predictive Analysis**

1. Can we predict mental health assessment scores based on demographic and lifestyle factors?
2. How accurately can we predict the severity of mental health diagnoses using available features?
3. Is it possible to predict the effectiveness of a particular treatment based on patient characteristics?
4. What is the relationship between sleep, exercise, and mental health outcomes?
5. Can we predict the likelihood of substance use based on patient characteristics?
6. How do significant life events impact mental health outcomes over time?
7. Is there a temporal trend in perceived social support among patients?
8. How does the frequency of visits to mental health professionals change over time?
9. Can we predict mental health improvement based on initial patient characteristics?
10. Does the patient's education level influence treatment adherence?
11. How well does a comprehensive predictive model perform in forecasting mental health outcomes?

**Correlation and Relationships**

1. Is there a correlation between the average sleep hours and mental health assessment scores?
2. Are there specific genetic markers associated with certain mental health diagnoses?
3. Is there a gender difference in treatment response?
4. How does the prevalence of mental health diagnoses vary across different socioeconomic statuses?