**Scaled Data**

Data has been standardized using the StandardScaler.  
This ensures all numeric features have comparable ranges, preventing variables with larger scales from dominating machine learning models.

**Examples:**

* Features like **Age** and **Adherence to Treatment (%)** have been transformed to reflect standard deviations from the mean.
* The scaled values are now ready for use in ML algorithms.

**Average Metrics by Diagnosis**

A summary of average values across diagnoses suggests the following:

* **Bipolar Disorder** shows higher averages for **Symptom Severity**, **Sleep Quality**, and **Mood Score**.
* **Major Depressive Disorder** has the lowest averages in several indicators, such as **Symptom Severity** and **Sleep Quality**, indicating a pronounced negative impact on quality of life.
* **Panic Disorder** and **Generalized Anxiety** have varied outcomes but generally fare better than depression in these metrics.

**Gender and Diagnosis**

The distribution of cases by gender and diagnosis shows:

* **Men** are most frequently diagnosed with **Generalized Anxiety**, followed by **Major Depressive Disorder** and **Panic Disorder**.
* **Women** also predominantly experience **Generalized Anxiety**, but gender differences across other diagnoses are less pronounced.
* These trends highlight potential gender-based differences in the prevalence and diagnosis of disorders.

**Diagnosis and Medication**

Most commonly prescribed medications per diagnosis:

* **SSRIs** are widely used for **Panic Disorder** and **Generalized Anxiety**.
* **Mood Stabilizers** dominate for **Generalized Anxiety** and **Bipolar Disorder**.
* For **Major Depressive Disorder**, **Antidepressants** are the primary treatment, though other drugs like **Anxiolytics** are also frequently prescribed.

**Therapy Type and Diagnosis**

Therapy preferences by diagnosis reveal:

* **Cognitive Behavioral Therapy (CBT)** is the most commonly applied method for **depression** and **panic disorders**.
* **Mindfulness-Based Therapy** and **Dialectical Behavioral Therapy (DBT)** are popular for **bipolar** and **anxiety disorders**.
* **Interpersonal Therapy** is a common choice for treating **anxiety** and **depression**.

**Random Forest Model – Results**

**Classification of treatment outcomes:**

* The model’s accuracy is relatively low (27%), suggesting a need for further optimization.
* Performance for the categories:
  + **F1-scores** for *Deteriorated*, *Improved*, and *No Change* are comparable but far from ideal.
  + This may stem from data complexity or the lack of stronger predictive features.

**Feature Importance:**

* **Key features:**
  + **Adherence Index** (0.17): A critical indicator of treatment success.
  + **Treatment Effectiveness** (0.17): Strongly correlates with predicted outcomes.
  + **Mood Score**, **Stress Level**, and **Symptom Severity** also contribute significantly.

**Correlation Matrix**

Insights from the correlation analysis:

* **Strong correlation:**
  + **Gender** and **Gender Therapy Interaction (0.69)** indicate a significant influence of gender on therapy choice.
* **Moderate correlation:**
  + **Mood Score** and **Mood Improvement Ratio (0.37)** suggest mood improvement is closely tied to the initial mood score.
* **No significant correlation:**
  + **Age** shows little connection to most variables, implying it is not a key predictor.

**Recommendations**

**Model Optimization:**

* Experiment with alternative algorithms like **XGBoost** to improve performance.
* Refine feature selection and tune hyperparameters.

**Data Exploration:**

* Conduct in-depth analysis of therapy effectiveness based on gender differences.
* Include time-series analysis to track changes throughout treatment.

**Extend Variables:**

* Incorporate additional factors, such as social support, disease severity, or resource accessibility.