Analysis on Anxiety Attack Factors, Symptoms, and Severity Dataset

Overview:

This analysis highlights the complexity of anxiety disorders and emphasizes the importance of a nuanced, personalized approach in both research and clinical practice. By understanding anxiety as a continuous spectrum influenced by multiple factors, healthcare providers will be able to develop more effective strategies to address this pervasive mental health issue.

source: https://www.kaggle.com/datasets/ashaychoudhary/anxiety-attack-factors-symptoms-and-severity/data
repo: https://github.com/Rongxuan-Zhou/EECE5642-Data-Visualization.git

1. Introduction

This analysis explores a dataset of over 12,000 records detailing factors related to anxiety attacks, including key features such as demographics, lifestyle habits, stress levels, and physiological responses. The study employs a variety of visualization techniques to uncover patterns and correlations in anxiety disorders, offering personal insights for mental health research and clinical practice.

Objective:
To explore patterns, triggers, and correlations in anxiety disorders. Potential use cases maybe mental health research, machine learning models, and healthcare analysis, etc.

2. Data Discription

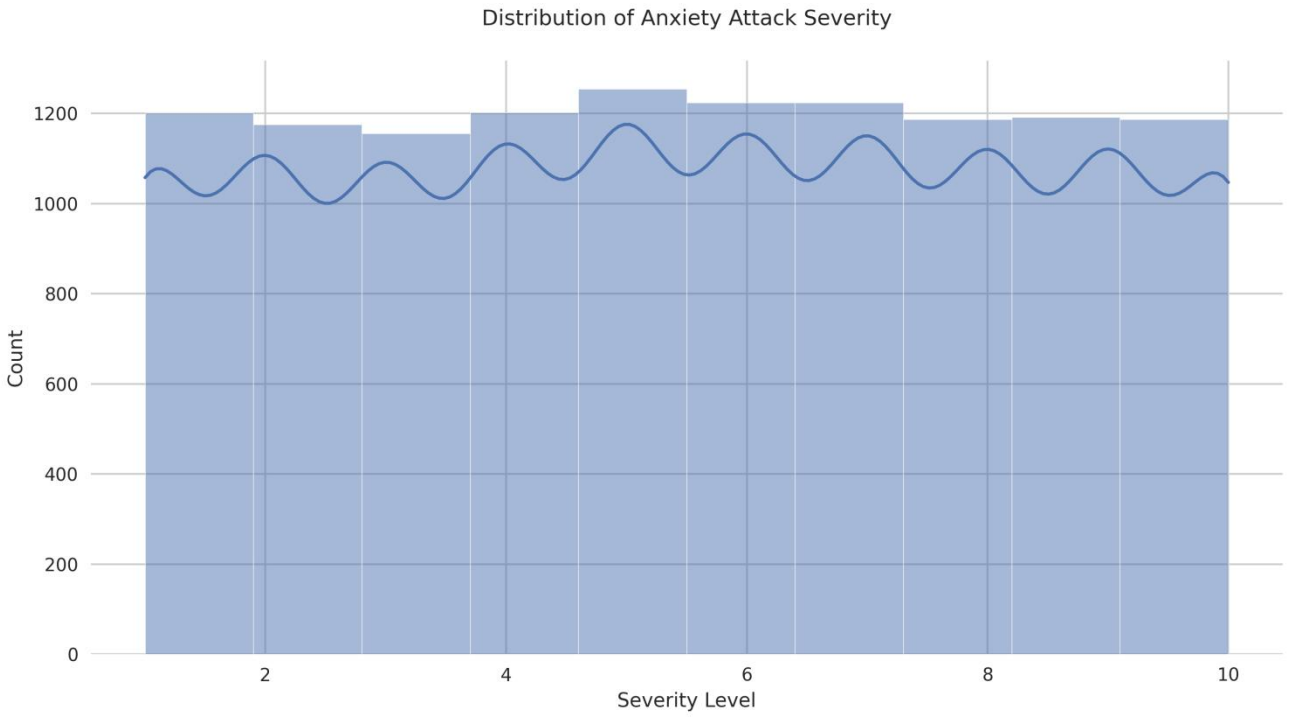
Data Dimensions: Over 12,000 records with multiple variables.
Data Types:

- Numerical (Continuous): Age, Sleep Hours, Physical Activity, Caffeine Intake, Alcohol Consumption, Heart Rate, Breathing Rate.
- Numerical (Discrete): Diet Quality, Sweating Level, Stress Level, Therapy Sessions, Anxiety Severity.
- Categorical (Nominal): Gender, Occupation.
- Categorical (Binary): Dizziness, Family History of Anxiety, Medication Use.

Data Format and Style: Structured data in a tabular format (CSV).

3. Visualization Results and Analysis

Visualization 1: Distribution of Anxiety Severity
This combination of a histogram and a density curve (KDE) effectively visualizes the distribution of anxiety severity levels.

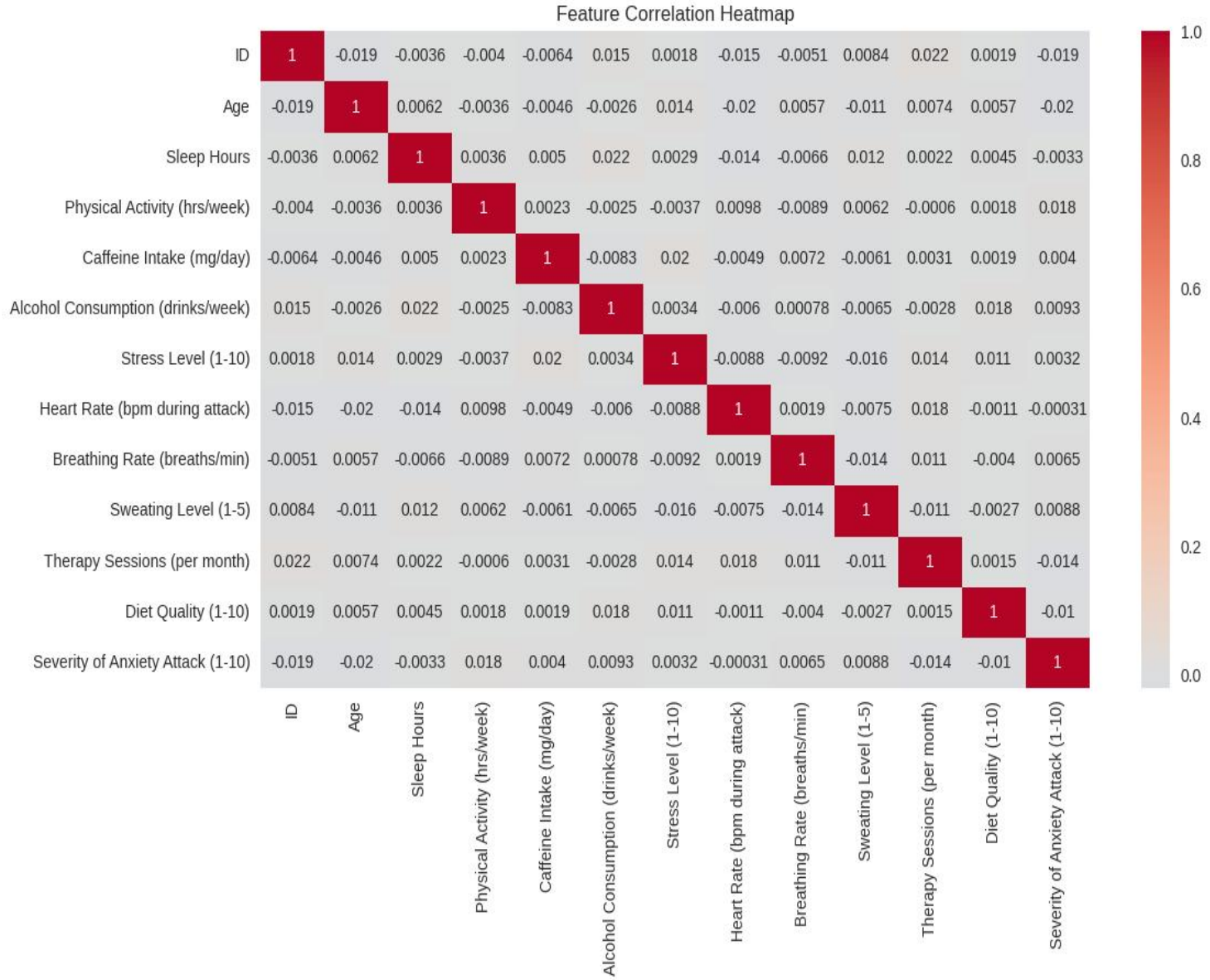


Observations:

- The ample sample size (1000+ samples per severity level) ensures robust statistical analysis.
- The even distribution across severity levels indicates good representativeness of the collected samples.
- The lack of floor or ceiling effects suggests that the dataset captures the full spectrum of anxiety severity.
- Slightly higher counts in the moderate range (5-6) compared to the extremes (1-2 or 9-10) might reflect a natural distribution of anxiety severity in the population.

Visualization 2: Feature Correlation Heatmap

The symmetric matrix format effectively displays the correlation coefficients between pairs of variables, with color gradient from deep red (strong positive correlation) to light blue (strong negative correlation) visually representing the strength and direction of relationships.



Observations:

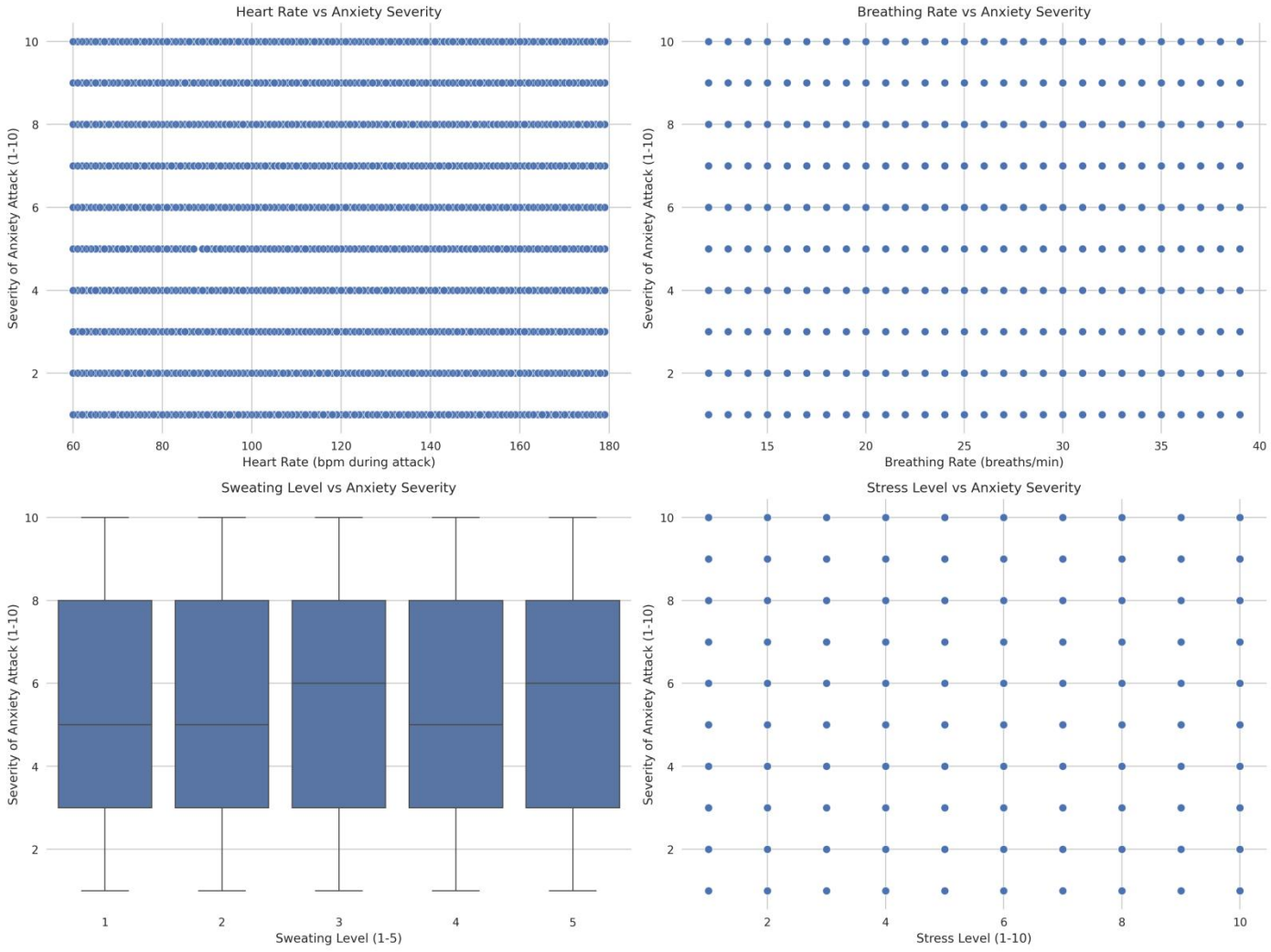
- The predominance of light-colored cells suggests weak correlations between most variables, indicating their relative independence.
- The absence of strong correlations (except on the diagonal) implies that no single variable overwhelmingly influences another.
- The weak positive correlation between Physical Activity and Anxiety Severity (0.018) and the weak negative correlation between Age and Anxiety Severity (-0.02) are notable but not strong enough to draw definitive conclusions.

Fact Beyond Expectation:

- The near-zero correlation between Sleep Hours and Stress Level (0.0029) indicates that these factors do not significantly influence each other, which is contradictory to common belief that personal mode and stress level is positively related to a lack of sleep hours.
- The slight negative correlation between Caffeine Intake and Alcohol Consumption (-0.0083) suggests a minimal inverse relationship.
- The very weak positive correlation between Heart Rate and Breathing Rate (0.0019) implies a negligible relationship.

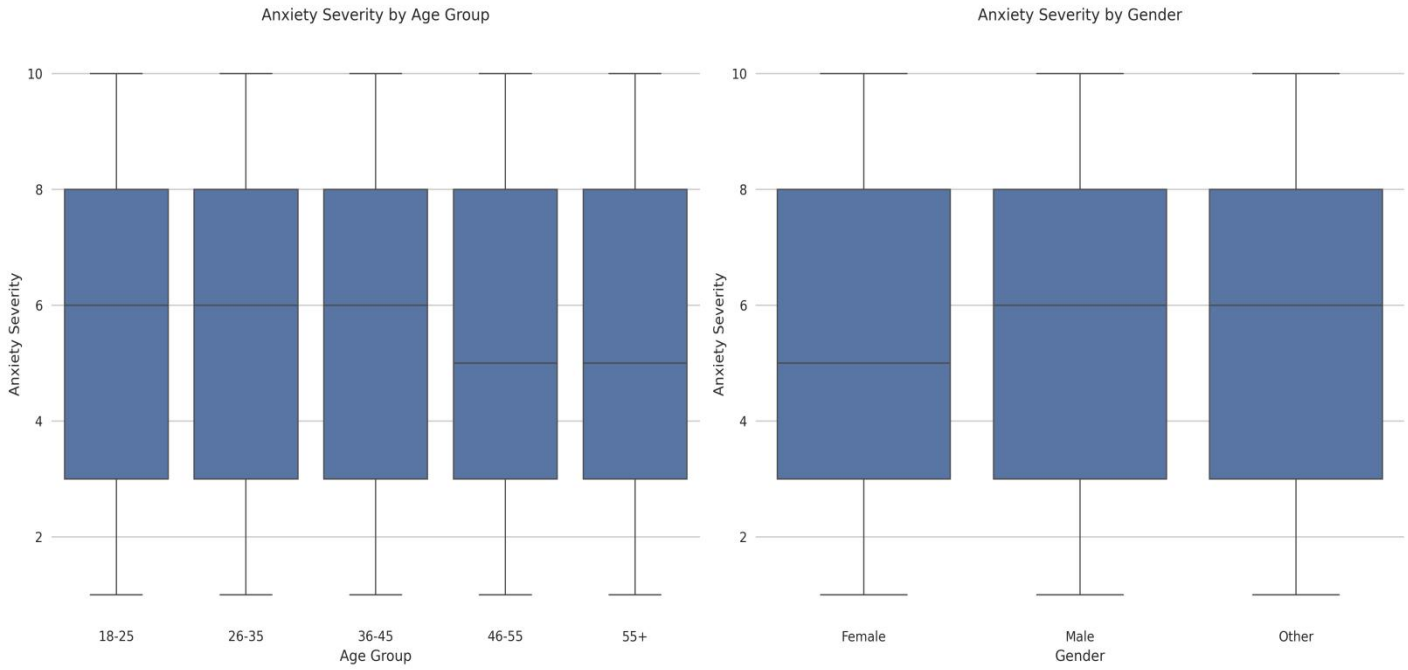
Practical Insights:
In the feature correlation heatmap, the general lack of strong correlations underscores the complexity of anxiety, which suggests that it is influenced by multiple independent factors. The complexity indicates that anxiety cannot be attributed to a single cause, necessitating a multifaceted approach in both research and treatment.

Visualization 3: Physical Factors Analysis
This is a four separate scatter/box plots arranged in a 2x2 grid focusing on physical factors affecting anxiety level.



Heart Rate vs Anxiety Severity: No clear linear correlation between heart rate and anxiety severity, suggests heart rate alone is not a reliable predictor of anxiety severity.
Breathing Rate vs Anxiety Severity: Similar to heart rate, shows no strong correlation, indicates breathing rate is not independently predictive of anxiety levels
Sweating Level vs Anxiety Severity: Median anxiety severity remains relatively constant, suggests sweating intensity doesn't strongly indicate anxiety severity
Stress Level vs Anxiety Severity: Slight positive correlation visible, indicates stress level may be a better indicator of anxiety severity

Visualization 4: Demographic Analysis
This chart illustrates the distribution of anxiety severity across different age groups and genders



Age Group Analysis:

- 18-25 years: Highest anxiety severity, likely due to academic, career, and social pressures.
- 26-35 years: Slightly lower anxiety levels, possibly related to career development and family responsibilities.
- 36-45 years: Further decline in anxiety, potentially due to increased life and work experience.
- 46-55: Lower anxiety levels, possibly reflecting life stability.
- 55+ years: Lowest anxiety severity, potentially due to reduced life pressures post-retirement.

Gender Analysis:
For female, hiigher anxiety severity compared to males, possibly influenced by biological, social, and psychological factors. For male, lower anxiety levels but still significant. Anxiety levels intermediate between females and males, indicating gender identity may also impact anxiety.

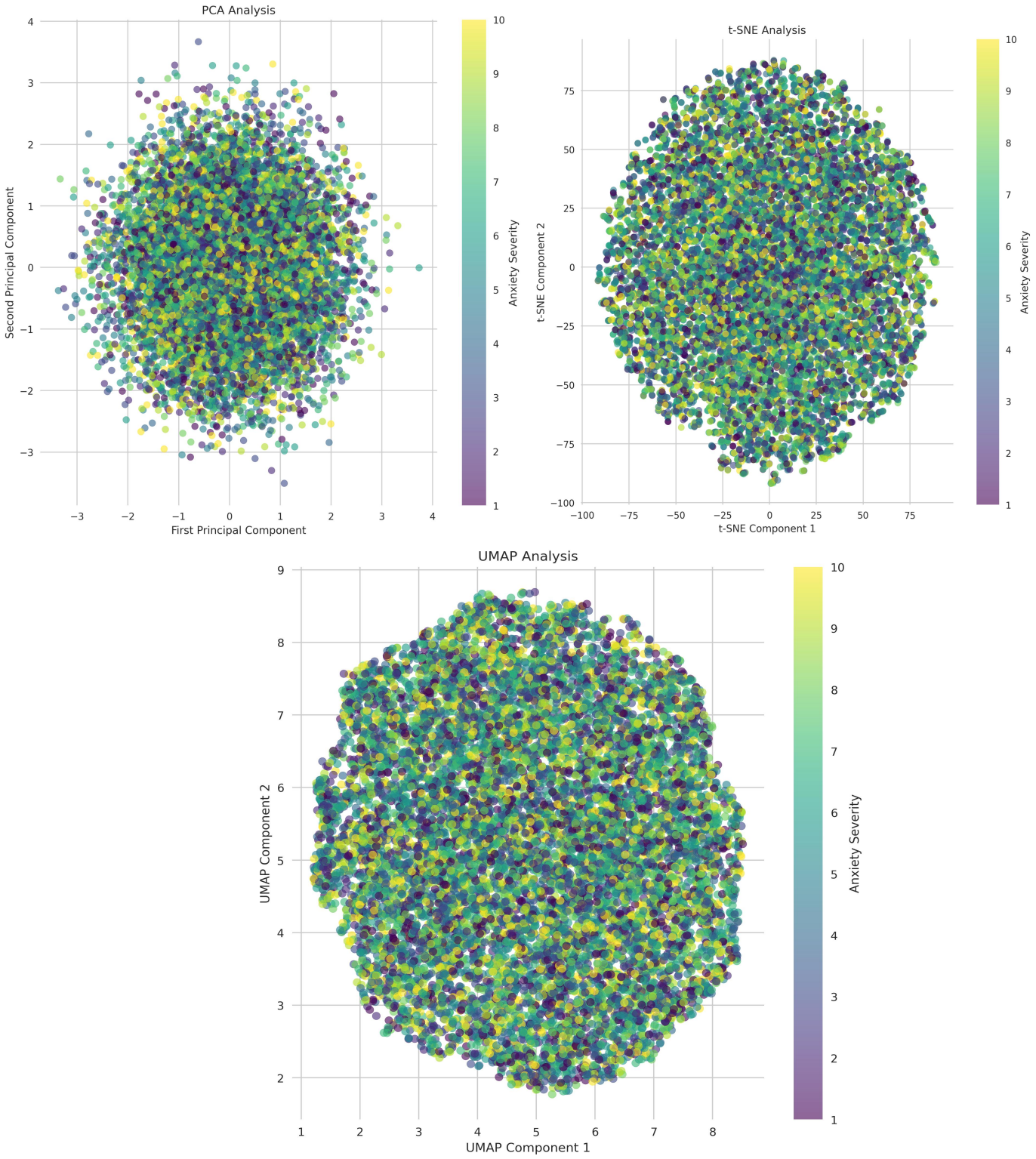
Trends:

- Anxiety severity decreases with age.
- Females exhibit higher anxiety levels across all age groups.

Implications:

- Tailored mental health interventions are needed for different age groups and genders.
- Increased support and resources should be directed towards younger individuals and females.

Visualization 5: Dimensionality Reduction Analysis
This visualization of anxiety disorder data utilizing three different dimensionality reduction techniques (PCA, t-SNE, and UMAP), revealing several vital insights into the structure and distribution of the data.



PCA (Principal Component Analysis): The data exhibits an elliptical distribution. No distinct clustering is observed, suggests that there may be linear relationships between anxiety severity and various features, but these relationships are not particularly strong.
t-SNE (t-Distributed Stochastic Neighbor Embedding): The data shows a more uniform, circular distribution, and local structures are well-preserved. However, different levels of anxiety symptoms (indicated by color) do not form distinct clusters, suggesting a continuous spectrum rather than discrete categories.
UMAP (Uniform Manifold Approximation and Projection): The distribution is similar to t-SNE but retains more global structure. Again, no clear clustering is evident. Samples with different anxiety levels are uniformly mixed, reinforcing the idea of a continuous spectrum.

Observations:
Anxiety severity (represented by colors from purple to yellow, 1-10) is distributed continuously in the space, without clear boundaries.
The relationships between different features are complex, potentially nonlinear.
The absence of distinct clusters suggests that anxiety symptoms form a continuous spectrum rather than discrete categories.
The similar patterns observed across PCA, t-SNE, and UMAP enhance the reliability of these findings.

Implications:
The assessment and treatment of anxiety disorders may require more personalized and continuous methods rather than rigid categorization, and clinicians should consider the combined effects of multiple factors rather than focusing on single indicators.

4. Summary and Perspective

Key Findings:

The dataset captures a full spectrum of anxiety severity, with a slightly higher prevalence of moderate severity levels (5-6), suggests that anxiety severity follows a natural distribution in the population, with no significant floor or ceiling effects.

Most variables exhibit weak correlations, however notably, physical activity and age show weak correlations with anxiety severity, while sleep hours and stress levels demonstrate minimal influence on each other, challenging common assumptions.

Physiological factors such as heart rate, breathing rate, and sweating level show no strong correlation with anxiety severity, suggesting they are not reliable standalone predictors. However, stress levels exhibit a slight positive correlation, indicating their potential as a better indicator of anxiety severity.

Anxiety severity is highest among younger age groups (18-25 years) and decreases with age, likely due to reduced life pressures and increased stability. Females consistently report higher anxiety levels across all age groups, highlighting the need for gender-specific interventions.

Techniques like PCA, t-SNE, and UMAP reveal that anxiety symptoms form a continuous spectrum rather than discrete categories. The absence of distinct clusters underscores the complexity of anxiety disorders and supports the need for personalized, continuous treatment approaches.

Perspective:

Anxiety disorders require individualized treatment plans that consider the combined effects of multiple factors, rather than rigid categorization.

Clinicians should adopt a multifaceted approach, addressing biological, psychological, and social factors to effectively manage anxiety.

Younger individuals and females may benefit from increased support and resources, given their higher reported anxiety levels.