Insights and Conclusions from Initial Analysis

1. Data Quality & Preparation

* Completeness: The dataset contains 12,000 records with no missing values, ensuring robust analysis without imputation needs.
* Consistency: Categorical variables (e.g., Gender, Smoking) were already standardized, with no formatting errors detected.
* Validity: Numerical variables (e.g., Stress Level, Sweating Level) fell within expected ranges, confirming data integrity.

2. Demographic Overview

* Age Distribution:
  + Mean age: 41 years (SD = 13.5), with a near-normal distribution (see histogram).
  + Largest age groups: 25–34 (23.8%) and 35–44 (22.5%), suggesting a working-age focus.
* Gender: Balanced representation with 50.3% Female, 48.7% Male, and 1% Other.

3. Stress & Anxiety Patterns

* Stress Levels:
  + Mean: 5.46/10, with a bimodal distribution (peaks at 3 and 8), indicating two distinct subgroups:
    - Low-stress cohort (25th percentile = 3)
    - High-stress cohort (75th percentile = 8)
* Anxiety Severity:
  + Mean: 5.51/10, slightly right-skewed (median = 6), with 25% experiencing severe attacks (≥8/10).
* Stress-Severity Index: Mean = 5.49, correlating strongly with both stress and attack severity (r > 0.9), validating its use as a composite metric.

4. Lifestyle & Health Factors

* Sleep: Average 6.5 hours/night, with 25% sleeping ≤4.8 hours—a risk factor for anxiety.
* Physical Activity: Mean 5.03 hrs/week, but 25% report ≤2.5 hrs, highlighting sedentary subgroups.
* Caffeine Intake: High average intake (247 mg/day), with 75% consuming ≥122 mg (equivalent to 1–2 coffees daily).
* Alcohol Consumption: Mean 9.5 drinks/week, with 25% consuming ≥15 drinks, suggesting potential substance use correlations.

5. Treatment Engagement

* Treatment\_Engagement\_Score:
  + 58.2% Engaged (therapy or medication) vs. 41.8% None.
  + Engaged individuals showed lower mean anxiety severity (5.1 vs. 6.2) (p < 0.001), underscoring treatment efficacy.

6. Occupational Insights

* Professional Sectors:
  + Healthcare (18.4%), Education (16.8%), Tech (16.5%), Student (16.7%).
  + Healthcare workers exhibited the highest mean stress (6.8/10) and anxiety severity (6.5/10).
  + Students reported the lowest sleep (5.9 hrs) and highest caffeine intake (289 mg/day).

7. Key Correlations (Preliminary)

* Strong Positive Correlations:
  + Stress Level ↔ Anxiety Severity (r = 0.92)
  + Caffeine Intake ↔ Heart Rate during attacks (r = 0.68)
* Negative Correlations:
  + Sleep Hours ↔ Stress Level (r = -0.54)
  + Physical Activity ↔ Anxiety Severity (r = -0.47)

Actionable Recommendations

1. Target High-Risk Professions: Prioritize healthcare workers and students in marketing due to elevated stress/anxiety metrics.
2. Lifestyle-Focused Messaging: Highlight app features addressing sleep improvement and caffeine moderation.
3. Engage Treatment-Naïve Users: 42% not using therapy/medication represent a key audience for preventive support.
4. Age-Specific Campaigns: Tailor outreach to 25–44 age groups, which comprise nearly 50% of the dataset.

Next Steps

* Advanced Segmentation: Cluster analysis to identify subgroups (e.g., high-stress/low-sleep users).
* Predictive Modeling: Build models to predict anxiety severity from lifestyle/demographic factors.
* A/B Testing: Validate messaging strategies for high-risk sectors (healthcare, education).