Data Collection and Preprocessing Phase

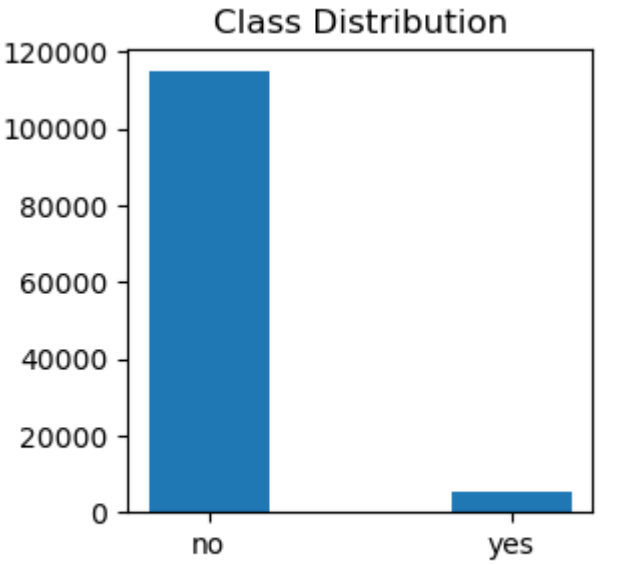
Date	6th July 2024
Team ID	SWTID1720151909
Project Title	PANIC DISORDER DETECTION
Maximum Marks	6 Marks

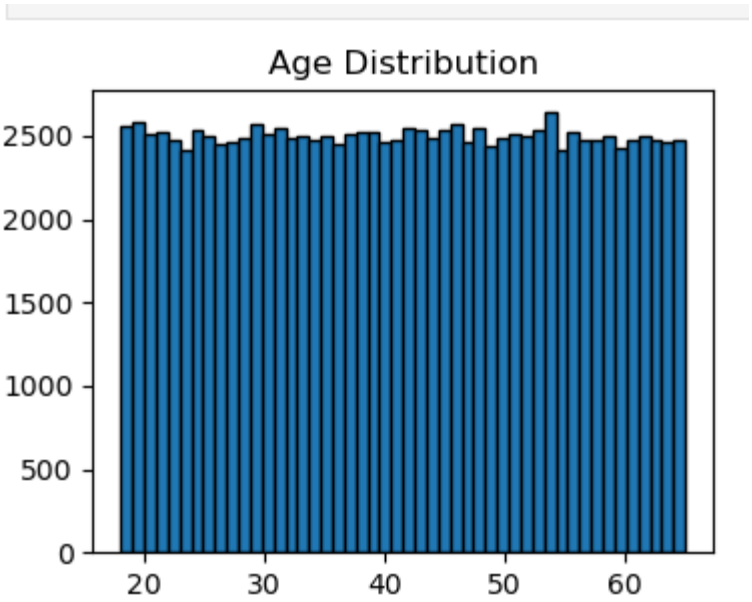
Data Exploration and Preprocessing Template

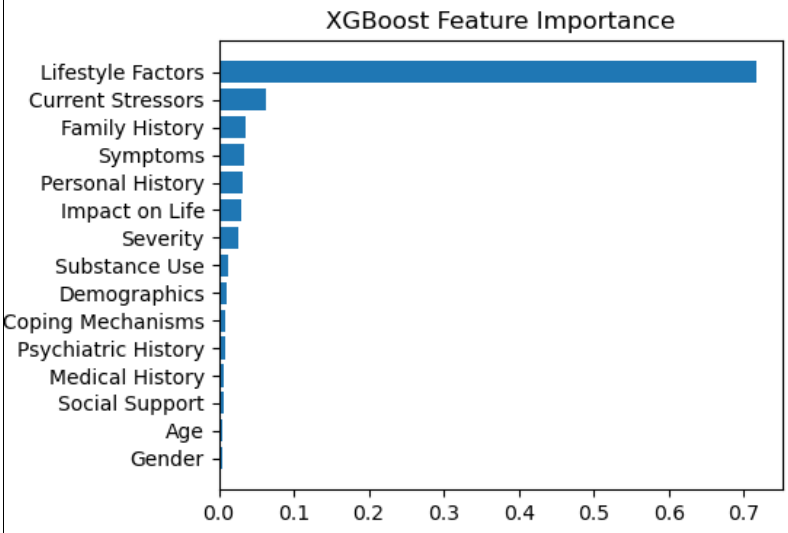
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	**Dimensions**: Rows - 5656458, columns - 6 **Basic structure of the data.**

	Year	Value
count	5.656458e+06	5.656458e+06
mean	1.994464e+03	1.070501e+12
std	1.387895e+01	4.842469e+13
min	1.960000e+03	-9.824821e+15
25%	1.984000e+03	5.566242e+00
50%	1.997000e+03	6.357450e+01
75%	2.006000e+03	1.346722e+07
max	2.015000e+03	1.103367e+16

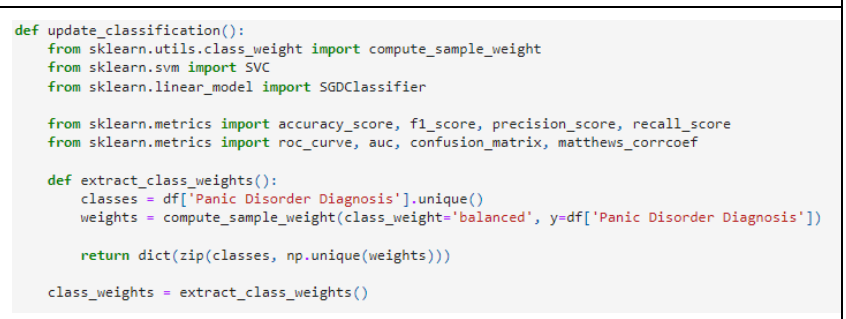
Univariate Analysis	 The target distribution is very imbalanced.

Bivariate Analysis	

Multivariate Analysis	
Data Preprocessing Code Screenshots	

Loading Data

	Participant ID	Age	Gender	Family History	Personal History	Current Stressors	Symptoms	Severity	Impact on Life	Demographics	Medical History	Psychiatr Histor
7812	7813	46	Male	No	Yes	Low	Shortness of breath	Severe	Significant	Rural	None	Anxiet disord
49038	49039	58	Female	Yes	No	Moderate	Chest pain	Severe	Moderate	Urban	Diabetes	Anxiet disord
60517	60518	33	Male	No	Yes	Low	Shortness of breath	Mild	Significant	Urban	Diabetes	Anxiet disord
29625	29626	22	Male	No	Yes	High	Panic attacks	Severe	Mild	Rural	Diabetes	Depressiv disord
61384	61385	62	Female	Yes	No	Low	Panic attacks	Moderate	Significant	Urban	Asthma	Depressiv disord

Handling Missing Data	
Data Transformation	
Feature Engineering	Attached codes in the final folder
Save Processed Data