**Abstract:**

This study aimed to analyze the PHQ-9 Depression test Questionnaires data to understand the distribution of depression states among a group of individuals and to determine the correlation between age range and PHQ-9 collective scores. The dataset consisted of responses to 9 questions, with each response contributing to a person's overall PHQ-9 score. Depression states were categorized based on these scores, and correlation analysis was conducted to explore the relationship between age and depression scores. Additionally, data preprocessing steps involved converting string scores to numeric values and mapping age ranges to numeric values for statistical analysis.

**Introduction**:

Depression is a prevalent mental health condition that can significantly impact an individual's well-being. The PHQ-9 Depression test Questionnaires are commonly used to assess the severity of depressive symptoms based on responses to nine questions. Understanding the distribution of depression states in a population and exploring factors influencing depression scores, such as age, can provide valuable insights for healthcare professionals and researchers.

**Problem Statement**

The primary objectives of this analysis were as follows:

To Determine the number of individuals in each state of depression (No depression, Mild depression, Moderate depression, Moderately severe depression, Severe depression) based on their PHQ-9 collective scores.

To Investigate the correlation between age range and PHQ-9 collective scores to understand any potential relationship between age and depressive symptoms.

**Methodology**

**Data Collection:** The dataset containing responses to the PHQ-9 Depression test Questionnaires was obtained, with each row representing an individual's responses to the questions.

**Data Preprocessing:**

Numeric values were extracted from the scores within parentheses to convert string scores to numeric values, ensuring accurate analysis.

The collective PHQ-9 score for each individual was calculated by summing their scores across the nine questions, providing a comprehensive assessment of depressive symptoms.

Individuals were classified into different depression states using predefined score ranges, facilitating the analysis of depression distribution.

Age range strings were mapped to corresponding numeric values to facilitate correlation analysis and understand age-related trends in depressive symptoms.

**Statistical Analysis:**

The number of individuals in each state of depression was determined using value counts, providing insights into the prevalence of different depression levels in the population.

The correlation coefficient between age range and PHQ-9 collective scores was calculated using Pearson correlation, allowing for the quantification of the relationship between age and depressive symptoms.

**Result:**

The analysis revealed the following:

Number of people in each state of depression:

Mild depression 279

Moderate depression 219

Moderately severe depression 88

No depression 70

Severe depression 16

**Correlation Coefficient:**

Correlation coefficient between Age range and PHQ-9 collective score: 0.004168293138876544

**Conclusion**

This study provided valuable insights into the distribution of depression states among the surveyed individuals and identified a positive correlation between age and PHQ-9 collective scores. Understanding these relationships can contribute to better mental health assessment and intervention strategies tailored to different age groups. Further research may explore additional factors influencing depression states and validate these findings in larger populations. The data preprocessing steps ensured the accuracy and reliability of the analysis, contributing to robust conclusions.