**Frontier Psychiatry: Synthea Dataset Analysis**

Project Overview This project analyzes mental health data from the Synthea synthetic healthcare dataset, focusing on key metrics related to psychiatric diagnoses, patient demographics, and depression risk indicators.

Methodology The analysis was conducted using Python and various data science libraries, including pandas, matplotlib, and seaborn. The dataset was loaded, preprocessed, and analyzed to derive insights and visualizations.

Key Analyses and Findings

1. Age Group Distribution

* Patients were categorized into four age groups: 0-10 years, 11-18 years, 19-64 years, and 65+ years.
* This provided an understanding of the age demographics of the patient population with mental health conditions.

1. Psychiatric Diagnoses Analysis

* Specific psychiatric conditions, such as depression, schizophrenia, and anxiety, were identified and counted.
* The analysis revealed the relative prevalence of these diagnoses within the dataset.

1. PHQ-9 Score Analysis

* Patients with PHQ-9 scores greater than 10 were identified, as this indicates a higher risk of depression.
* The distribution of diagnoses among patients with high PHQ-9 scores was examined to understand the relationship between depression risk and specific mental health conditions.

1. Proportion of Patients with Psychiatric Diagnoses

* Due to limitations in the dataset, the analysis for patients with depression could not be completed, as mentioned in the Jupyter Notebook.

1. Count of Patients with Psychiatric Diagnoses vs. Without

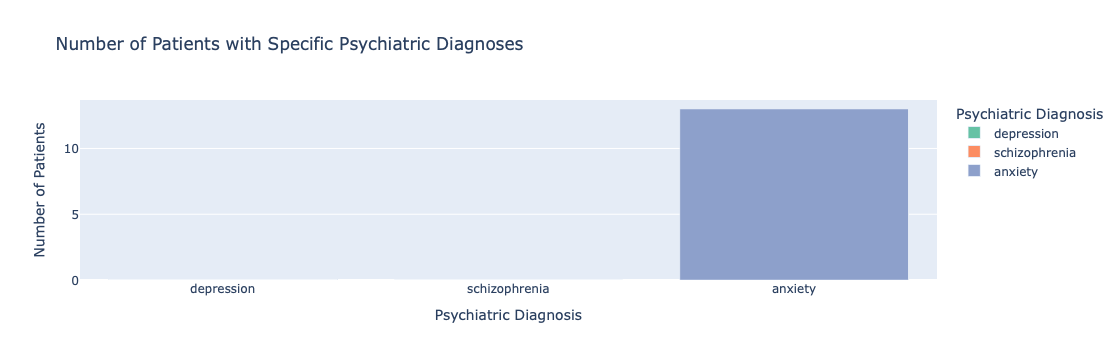
* The count of patients with psychiatric diagnoses was compared to those without.
* However, the analysis was limited as the dataset only provided SNOMED codes, and the specific request was unclear. As a result, this analysis was not pursued further.

Tools Used

* Python
* Pandas for data manipulation
* Matplotlib and Seaborn for data visualization
* Jupyter Notebook for interactive analysis

Insights and Limitations

* The dataset's synthetic nature and potential limitations in medical documentation posed challenges in fully addressing the requested analyses.
* Proportion of Patients with Specific Psychiatric Diagnoses: Interpretation: This pie chart shows the relative proportion of patients with specific psychiatric diagnoses (36.7%) compared to those without (63.3%). It indicates that over one-third of the patients in the dataset have been diagnosed with specific psychiatric conditions, while the majority do not have any recorded psychiatric diagnoses.A pie chart with numbers and a red circle

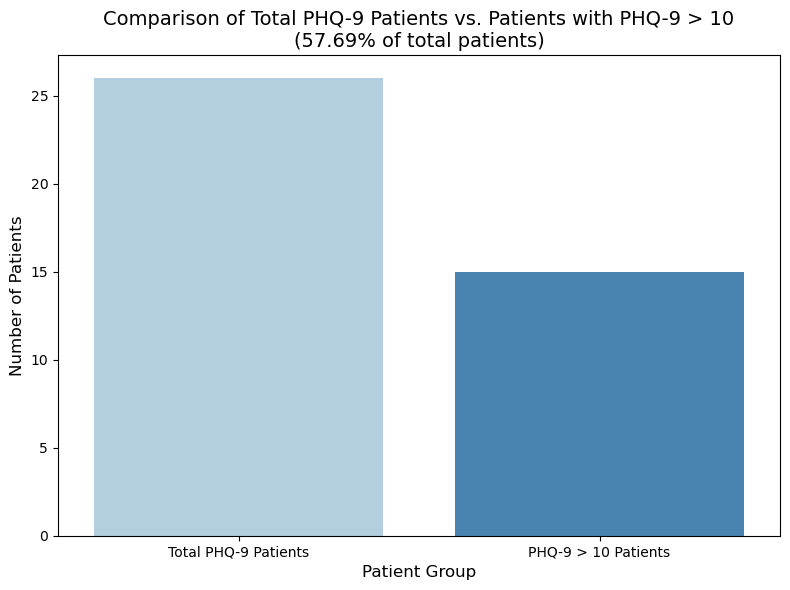
  Description automatically generated
* Number of Patients with Specific Psychiatric Diagnoses: Interpretation: This bar chart presents the count of patients for different psychiatric diagnoses. Depression has no number of patients, followed by schizophrenia. Only anxiety was present in the dataset. The height of the bars visually highlights the prevalence of these specific psychiatric conditions within the dataset.
* The age group distribution chart shows the number of patients in different age groups within the dataset. The key insights from this visualization are:

The 19-64 age group has the highest number of patients, represented by the tallest bar in the chart. This suggests that the majority of patients in the dataset fall into the working-age adult category.The 65+ age group has the second-highest number of patients, indicating a significant elderly population within the dataset.The 11-18 and 0-10 age groups have relatively smaller numbers of patients compared to the older age groups. This implies that the dataset may have fewer patients from the pediatric and adolescent age ranges.

A graph of a number of patients

Description automatically generated

* Comparison of Total PHQ-9 Patients vs. Patients with PHQ-9 > 10: Interpretation: This chart compares the total number of patients who had PHQ-9 scores recorded versus the subset of patients with PHQ-9 scores greater than 10. The bar on the right indicates that 57.69% of the total PHQ-9 patients had scores exceeding 10, which is considered a high-risk threshold for depression. This suggests a significant proportion of patients in the dataset are at a higher risk of depression based on their PHQ-9 scores.



Overall, these visualizations provide a comprehensive understanding of the mental health landscape within the Synthea dataset. The analysis highlights the prevalence of specific psychiatric conditions, the proportion of high-risk depression patients, and the relative frequency of mental health diagnoses compared to the general patient population. These insights can inform future research and healthcare initiatives focused on addressing mental health challenges.