Frontier Psychiatry Data Analysis Project

Overview

This project analyzes the Synthea synthetic health dataset to provide insights into patient demographics, psychiatric diagnoses, PHQ-9 scores, depression treatment gaps, and CPT code utilization related to psychiatric conditions. The analysis includes visualizations to aid understanding and is intended to showcase data analysis and visualization skills.

Table of Contents

Data Handling

- **Dataset**: Synthea synthetic health data.
- **Files Used**: Multiple CSV files including patients, conditions, observations, medications, encounters, etc.
- **Preprocessing**:
    - Data cleaning and normalization.
    - Handling missing values.
    - Date parsing and calculations.

Requirements

- **Python 3.x**

- **Libraries**:
    - pandas
    - matplotlib
    - plotly
- **Optional**:
    - Jupyter Notebook (for interactive exploration)

Project Structure

kotlin

Copy code

project/

├── data/

│ ├── allergies.csv

│ ├── careplans.csv

│ ├── claims.csv

│ ├── claims_transactions.csv

│ ├── conditions.csv

│ ├── devices.csv

│ ├── encounters.csv

│ ├── imaging_studies.csv

│ ├── immunizations.csv

│ ├── medications.csv

│ ├── observations.csv

│ ├── organizations.csv

│ ├── patients.csv

│ ├── payer_transitions.csv

│ ├── payers.csv

│ ├── procedures.csv

│ ├── providers.csv

│ └── supplies.csv

├── frontier_analysis.py

├── README.md

└── visualizations/

    ├── age_group_distribution.png

    ├── psychiatric_diagnoses.png

    ├── phq9_scores.png

    └── cpt_code_utilization.png

How to Run

1. **Clone the Repository**:

bash

Copy code

git clone https://github.com/yourusername/frontier-psychiatry-project.git

2. **Navigate to the Project Directory**:

bash

Copy code

cd frontier-psychiatry-project

3. **Install Required Libraries**:

bash

Copy code

pip install pandas matplotlib plotly

4. **Ensure Data Files are in Place**:
    - Place all the CSV files in the data/ directory.
5. **Run the Analysis Script**:

bash

Copy code

python frontier_analysis.py

6. **View Outputs**:
    - Visualizations will be saved in the visualizations/ directory.
    - Insights will be printed to the console.

Analysis Approach

1. Age Group Distribution

- **Objective**: Visualize the number of patients in different age groups: 0–10, 11–18, 19–64, 65+ years.
- **Method**:
    - Calculate age by subtracting birth date from the current date.
    - Categorize patients into defined age groups.
    - Count the number of patients in each group.

2. Psychiatric Diagnoses

- **Objective**: Show the count of patients with psychiatric diagnoses such as depression, anxiety, schizophrenia.
- **Method**:
    - Filter conditions for relevant psychiatric diagnoses.
    - Count occurrences of each diagnosis.

3. PHQ-9 Scores

- **Objective**: Display patients with a PHQ-9 score greater than 10 by diagnosis.
- **Method**:
    - Filter observations for PHQ-9 assessments.
    - Select records where the score is greater than 10.
    - Merge with conditions to associate scores with diagnoses.

4. Depression Treatment Analysis

- **Objective**: Identify patients diagnosed with depression who did not receive antidepressant treatment within the first 90 days.
- **Method**:
    - Filter conditions for depression diagnoses.
    - Merge with encounters and medications.
    - Calculate the days between diagnosis and treatment.
    - Identify patients with no treatment within 90 days.

5. CPT Code Utilization (Bonus)

- **Objective**: Analyze the frequency of CPT codes used in patients with psychiatric-related ICD-10 codes.
- **Method**:
    - Filter encounters for psychiatric diagnoses.
    - Count the frequency of CPT codes associated with these encounters.

Visualizations

The following visualizations are generated and saved in the visualizations/ directory:

1. **Age Group Distribution** (age_group_distribution.png)
    - Bar chart showing the number of patients in each age group.
2. **Psychiatric Diagnoses** (psychiatric_diagnoses.png)
    - Bar chart showing counts of psychiatric diagnoses.
3. **PHQ-9 Scores by Diagnosis** (phq9_scores.png)
    - Bar chart displaying the number of patients with PHQ-9 scores >10 by diagnosis.
4. **CPT Code Utilization** (cpt_code_utilization.png)
    - Bar chart showing the frequency of CPT codes used in psychiatric encounters.

Insights

- **Age Group Distribution**:
    - Majority of patients fall within the 19–64 age group.
- **Psychiatric Diagnoses**:
    - The most common psychiatric diagnosis is anxiety.
- **PHQ-9 Scores**:
    - A significant number of patients with high PHQ-9 scores also have stress-related diagnoses.
- **Depression Treatment Analysis**:
    - Identified patients diagnosed with depression who did not receive timely antidepressant treatment.
- **CPT Code Utilization**:
    - Certain CPT codes are more frequently used in psychiatric-related encounters.

Assumptions and Limitations

- **Data Quality**:
    - The analysis assumes the synthetic data accurately represents realistic patient data.
- **Missing Data**:
    - Any missing values were handled by omission or imputation where appropriate.
- **Date Calculations**:
    - Current date is assumed to be the date when the script is run.
- **Diagnosis Matching**:
    - Diagnoses were matched based on keyword searches in the description fields.

- **Treatment Identification**:
    - Antidepressant treatments were identified based on medication descriptions; some cases might be missed due to naming variations.

Contact Information

For any questions or further information, please contact:

- **Name**: Azmeer Mohammed
- **Email**: azmeermohammed1998@gmail.com
- **GitHub**: Azmeer22