Hospital Patient Analysis Assignment

Introduction

In this assignment, you will perform a comprehensive analysis of hospital patient data using three CSV files: drug_code.csv, diagnostic_code.csv, and patient_data.csv. The goal of this analysis is to gain insights into patient demographics, drug prescriptions, diagnoses, and mortality rates. You will answer a series of questions and enhance the assignment with more visualizations.

Dataset Description

drug_code.csv: Contains information about different drug codes, including their descriptions and severity levels.
diagnostic_code.csv: Contains diagnostic codes (ICD-9 codes) along with their descriptions and severity levels.
patient_data.csv: Contains patient information, including subject IDs, genders, ages, drug prescriptions, and diagnostic codes.

Questions and Tasks

- Question 1: Identifying High-Risk Patients

- Identify all the patients who were prescribed drugs with a high severity level and had a high mortality rate. Return a list of subject IDs and their genders for these high-risk patients.

- Question 2: Gender Distribution among Expired Patients

- Calculate the distribution of genders (male, female) among patients who have expired. Provide the percentage of male and female patients among the expired cases. Visualize this distribution using a pie chart.

- Question 3: Common Diagnoses for High-Risk Patients

- Among the high-risk patients identified in Question 1, determine the most common ICD-9 diagnosis codes. Return the top 3 most common diagnosis codes along with their counts.

- Question 4: Mortality Rate by Diagnosis Type

- Calculate the mortality rate for each diagnosis type (ICD-9 code). Return a list of diagnosis types with their corresponding mortality rates (number of deaths divided by the number of patients with that diagnosis type). Visualize this information using a bar chart.

- Question 5: Age of Patients with Common Diagnoses

- Using the diagnostic codes from Question 3, calculate the average age of patients for each of the top 3 common diagnosis codes. Visualize this information using a bar chart.

- Question 6: Drug Prescription Trends

- Identify the top 5 most prescribed drug codes across all patients. Provide the drug codes along with their descriptions and the total number of times they were prescribed. Visualize this information using a horizontal bar chart.

- Question 7: Gender Disparity in Critical Diagnoses

- Analyze the gender distribution among patients with critical diagnoses (diagnosis codes associated with high severity). Calculate the percentage of male and female patients among these critical cases. Visualize this information using a stacked bar chart.

Submission Requirements

- Your submission should include the following components:
- Clearly written code in a programming language of your choice (e.g., Python) that performs the necessary analysis and answers the questions.
- Inline comments explaining your code's logic and major steps.
- Visualizations for Questions 2, 4, 5, 6, and 7.
- A report or document discussing your findings and insights gained from the analysis. This should include interpretations of the results, trends, and any significant observations.
- Use Following data set.

    1. Data Set Patient- https://drive.google.com/file/d/1vA3LE_H3qiEll2bYGEKqWhrsjgO0bkiK/view?usp=sharing

    2. Data Set Diagnostic code- https://drive.google.com/file/d/1o-eFKu_tE3mWduOjh6-fW_K_a8k_ugTt/view?usp=sharing

3. Data Set Drug code- https://drive.google.com/file/d/1ScyTEIIQrn-pI1o4yGJ0CgolCqbKHE2N/view?usp=sharing