**EHR Analysis Project: Predicting Outcomes from Lake County Hospitalization Data**

**About the Repository**

This repository showcases an Electronic Health Records (EHR) analysis project focused on hospitalization discharge rates in Lake County, Illinois. The project demonstrates a comprehensive data analytics pipeline, from data acquisition and cleaning to exploratory analysis, setting the stage for potential predictive modeling of patient outcomes.

A diagram of a health record

AI-generated content may be incorrect.

**Dataset**

The project utilizes hospitalization discharge rate data from Lake County, Illinois, obtained from the U.S. Government's open data portal [Link](https://catalog.data.gov/dataset/hospitalization-discharge-rates-49dd7). This rich dataset provides a foundation for analyzing critical care statistics and patient outcomes.

**Data Preprocessing**

The data preprocessing phase involved several crucial steps:

* Downloading the raw dataset from the government portal
* Initial data cleaning and wrangling using Excel
* Importing the cleaned data into Python for further analysis
* Encoding categorical variables
* Feature optimization to create new variables that may be useful for analysis.

**Data Analysis**The analysis presents a detailed exploratory data analysis, including:

* Overview of patient demographics (age, gender, ethnicity)
* Analysis of admission sources and diagnoses
* Distribution of hospital and ward IDs
* Examination of key metrics such as admission height and weight
* Discharge statistics and outcomes

**Key findings:**

1. Patient Demographics and Diversity

The majority of patients are elderly, with an average age of 65 years. The age distribution shows a right skew, indicating more elderly patients than younger ones. A significant number of patients are male, suggesting a possible gender disparity in hospitalization rates.

1. Critical Care Metrics

The average hospital stay duration is approximately 6 days, but a significant number stay longer. The discharge locations reveal that most patients return home, but a noticeable proportion are transferred to skilled nursing facilities.

1. Admission Weight vs. Discharge Weight

Patients tend to gain or maintain weight during hospitalization, with average admission weight at 83 kg and discharge weight at 84.5 kg. However, some patients show significant weight fluctuations which could be due to fluid retention, critical illness or nutritional needs.

1. Admission Diagnoses & Critical Conditions

The most common admission diagnoses include hypertension, acute myocardial infarction (heart attack), and cerebrovascular diseases (stroke-related conditions). A significant portion of patients undergo surgical interventions such as coronary artery bypass grafting (CABG) and carotid endarterectomy, reinforcing the dataset’s focus on cardiac and vascular health.

1. Hospital Source & Unit Admission Trends

The Emergency Department (ED) is the primary source of hospital admissions, indicating that most patients arrive in critical conditions requiring immediate attention. A notable proportion of patients come from operating rooms, meaning many are undergoing planned surgeries or urgent procedures.

**Modeling and Evaluation**

After a thorough exploratory analysis, the modeling phase was initiated, and the primary goal was to predict hospital discharge status. The data was first preprocessed by addressing missing values and encoding target variable features 'hospitaldischargestatus' using labelEncoding methods. The refined dataset was then divided into training and testing subsets (8:2 split). A RandomForestClassifier was used to train the data, achieving a test accuracy of approximately 98%. Detailed classification reports were generated to further validate the model, such as confusion matrices and feature importance analysis to understand the drivers behind the predictions. The finalized model, along with its preprocessing components, was saved for seamless integration into the EHR system, paving the way for real-time clinical decision support.