**Objective:** Predictive Modelling and Feature Analysis on Electronic Health Records (EHR) Data

Dataset: <https://www.kaggle.com/datasets/prasad22/healthcare-dataset>

**Tasks**:

1. **Dataset Acquisition**: Download the synthetic EHR dataset from the provided link.
2. **Data Preprocessing**:
   * Load the dataset into a Pandas DataFrame and display the first 5 rows.
   * Convert Date of Admission to datetime format.
   * Handle missing values using forward fill for temporal data.
3. **Feature Engineering**:
   * Create a new feature: Length of Stay (Discharge Date - Date of Admission in days).
   * Create a binary feature: High Billing (1 if Billing Amount > median, 0 otherwise).
4. **Exploratory Data Analysis (EDA)**:
   * Visualize the distribution of Length of Stay by Medical Condition using a Box Plot.
   * Plot a bar chart showing the count of readmissions by Gender and Medical Condition.
5. **Predictive Modeling**:
   * Split the dataset into training (80%) and testing (20%) sets.
   * Train a logistic regression model to predict Readmission Risk using features: Age, Gender, Medical Condition, Billing Amount, and Length of Stay.
   * Evaluate the model using accuracy, precision, recall, and F1-score.
6. **Feature Importance Analysis**:
   * Extract and rank feature importance (e.g., coefficients from logistic regression).
   * Plot a bar chart of feature importance scores.