Healthcare Management: Patient Readmission Prediction

Business Introduction: Welcome to ReadmitGuard, a leading provider of analytics solutions for healthcare facilities. At ReadmitGuard, we specialize in analyzing data to forecast the likelihood of patient readmission within 30 days after discharge. Our goal is to empower healthcare providers to proactively intervene and prevent readmissions, thereby improving patient outcomes and reducing healthcare costs.

Mission Statement: Our mission at ReadmitGuard is to provide healthcare facilities with accurate and reliable insights that identify high-risk patients for readmission. By leveraging data-driven analysis, we aim to support healthcare providers in delivering proactive and personalized care to patients, ultimately enhancing the overall quality of care and reducing readmission rates.

Problem Statement: Analyze the likelihood of a patient being readmitted to a healthcare facility within 30 days after discharge. By identifying high-risk patients through data analysis, healthcare providers can proactively intervene to prevent readmissions, improve patient outcomes, and reduce healthcare costs.

Dataset Description: The dataset contains information about patients' demographics, medical history, admission details, procedures performed, medications prescribed, and discharge outcomes. Each record represents a unique patient encounter with attributes related to their healthcare journey.

Excel Tasks:

1. **Data Exploration:**
    - Make a statistical summary of pertinent features like the number of lab procedures, procedures, drugs, and furthermore.
    - Create a report showing the number of readmissions by gender and age group
    - Show race distribution in a chart
    - Plot count of patients for each category of readmitted

SQL Tasks:

2. **Data Loading:**
    - Create a schema named **healthcare**, set it as the default schema, and create tables using **diabetic_data.csv**
3. **Tasks:**
    - Calculate the total number of patient encounters in the healthcare dataset
    - Identify the top 10 most frequent diagnoses in the dataset
    - Calculate the average length of hospital stay for each admission type

- Determine the number of readmitted patients and the percentage of total encounters that they represent
- Identify the age distribution of patients
- Identify the most common procedures performed during patient encounters
- Calculate the average number of medications prescribed for patients in each age group
- Identify the distribution of readmission rates across different payer codes

Python Task:

3. **EDA Task:**

- Perform descriptive statistical analysis for numerical features
- Visualize the distribution of categorical features - race and gender
- Explore the relationship between readmission status and age
- Investigate correlations between numerical features
- Analyze the distribution of medication changes and total medications taken
- Examine the distribution of diagnoses categories
- Explore the distribution of patients across admission types, sources, and discharge dispositions
- Identify and visualize any outliers in the dataset, especially in numerical features
- Write an analysis report on performing exploratory data analysis (EDA) using Python in the context of building a fraud detection system for the healthcare management system

PowerBI Tasks:

4. **Interactive Dashboard Design in PowerBI:**

- Show the distribution of readmitted counts of diagnoses and emergency cases through bar graphs
  - Open Power BI Desktop and load the **diabetic_data.csv** file using the **Get Data > Text/CSV** option.
  - Navigate to the **Data view** and verify data types of required columns.
  - Switch to the Report view and insert **a Stacked Bar Chart** from the **Visualizations pane.**
  - Drag **Number Diagnoses** and **Number Emergency** to the **X-axis** field.
  - Drag **Readmitted** to the **Y-Axis** field.
  - Optionally, enable **Data Labels** under the Format pane for better readability.

- Create a visual capturing age bucket count by readmitted
  - Create a **new report page**.

- In the Data view, create a new column named **Age Numeric** using the DAX formula: **Age Numeric =**
  VALUE(LEFT(SUBSTITUTE(SUBSTITUTE('diabetic_data'[Age], "[", ""), ")", ""), 2))
- Create another new column named Age Bucket using the formula**: Age Bucket =**
  SWITCH(TRUE(),
  [Age Numeric] <= 10, "0-10",
  [Age Numeric] <= 20, "11-20",
  [Age Numeric] <= 30, "21-30",
  [Age Numeric] <= 40, "31-40",
  [Age Numeric] <= 50, "41-50",
  "51+"
)
  and click on **Apply** and **OK**
- Drag **Age Bucket** to **X-axis** and **Number Inpatient** to Y-axis
- Drop down on the **Number Inpatient and** change it from **Sum** to **Count**
- Drag **Readmitted** to **legend.**
- Optionally, enable **Data Labels** under the Format pane for better readability.

- Visualize the number of medications, procedures, diagnoses, emergencies, time in the hospital by readmitted
  - Click on new **report page**
  - Drag **Readmitted** to **X-axis** and **Num Medications**, **Num Procedures**, **Number Diagnoses**, **Number Emergency** and **Time In Hospital** to **Y-axis**
  - Power BI will automatically stack them for visual comparison.
  - Optionally, use **Legend** to distinguish between the metrics.

- Create a bubble chart showing the number of medications, Lab procedures and Diagnoses and use Readmitted as color code
  - Click on new **report page.**
  - Insert a **Scatter Chart** from the **Visualizations** pane.
  - Drag **Num Medications** to the **X-axis**.
  - Drag **Num Lab Procedures** to the **Y-axis**.
  - Drag **Number Diagnoses** to the **Size** field.
  - Drag **Readmitted** to the **Legend (Color)** field.

- Create a dashboard with all the visualizations
  - Click on new **report page.**
  - Copy the visuals form each report (**Readmitted**, **Age Bucket**, **Multiple Readmitted Data**, and **Bubble Chart)** and paste them and readjust on dashboard area.