📊 Healthcare Analytics

🎯 Objective

This project focuses on analyzing a **healthcare dataset** using **SQL for data processing** and **Power BI/Tableau for visualization**. The goal is to extract meaningful insights, identify trends, and present findings effectively for data-driven decision-making in the healthcare sector.

By the end of this project, learners will:
✅ Conduct **exploratory data analysis (EDA)** to uncover trends and patterns.
✅ Perform **data cleaning and transformation** using SQL.
✅ Create **interactive Power BI/Tableau dashboards** for visualization.
✅ Provide **data-driven insights and recommendations** for healthcare improvements.

📌 Project Overview

The dataset contains **detailed healthcare information**, including **patient demographics, diagnoses, treatments, and costs**. The analysis aims to identify patterns, trends, and inefficiencies in healthcare services.

Learners will:

- Import and preprocess the dataset using **SQL**.
- Clean and standardize data to ensure consistency.
- Perform **exploratory data analysis (EDA)** to extract insights.
- Create **interactive dashboards in Power BI/Tableau** for better visualization.
- Provide **actionable recommendations** for optimizing healthcare operations.

📂 Dataset Details

📁 **Dataset Name:** Healthcare Data Analysis
📁 **Dataset Link:** CSV / Excel
📊 **Total Records:** 55,501+

📌 Columns Overview:

Patient Information:

- **Patient_ID**: Unique identifier for each patient.
- **Age**: Age of the patient.
- **Gender**: Male / Female.
- **Region**: Geographic location of the patient.

Medical Details:

- **Diagnosis**: Medical condition diagnosed.
- **Treatment_Type**: Type of treatment received.
- **HospitalStayDays**: Duration of hospital stay.
- **Doctor_ID**: ID of the treating doctor.

Financial Metrics:

- **Medical_Cost**: Cost of treatment.
- **Insurance_Covered**: Amount covered by insurance.
- **Out_of_Pocket**: Patient's direct expense.

🛠️ Tasks to be Performed

📌 Task 01: Data Preparation

Import & Load Data

🔹 Upload the dataset into an SQL database using `LOAD DATA` or import tools.

Clean and Preprocess

✅ Remove duplicates and missing values.
✅ Standardize **gender, diagnosis, and region** fields.
✅ Convert non-numeric fields (e.g., **HospitalStayDays**) into numeric formats.
✅ Normalize **financial metrics** for consistent analysis.

📌 Task 02: Exploratory Data Analysis (EDA) with SQL

SQL Queries:

✅ Analyze **patient distribution** by region and age group.
✅ Identify the **most common diagnoses and treatments**.

✅ Examine **average hospital stay duration** based on diagnosis.
✅ Analyze **cost trends and insurance coverage** across treatments.

📌 Task 03: Interactive Dashboards in Power BI/Tableau

📄 Page 1: Healthcare Overview Dashboard

📊 **KPIs:**
✔️ Total Patients
✔️ Average Hospital Stay
✔️ Total Medical Cost

📉 **Visualizations:**
📌 **Bar Chart:** Most common diagnoses
📌 **Pie Chart:** Patient distribution by gender
📌 **Line Chart:** Monthly treatment trends

📄 Page 2: Financial Analysis Dashboard

📊 **Visualizations:**
📌 **Scatter Plot:** Insurance Coverage vs. Medical Cost
📌 **Heatmap:** Hospital stay trends by region
📌 **Matrix:** Treatment type vs. average cost

🔍 **Filters & Slicers:**
✅ Region-based analysis
✅ Age group segmentation
✅ Diagnosis & treatment filters

📌 Task 04: Data Analysis & Insights

🔹 Use SQL to answer key business questions:
✅ What are the **top 5 diagnoses** contributing to **high medical costs**?
✅ How does **hospital stay duration** vary by **treatment type**?
✅ What is the **average out-of-pocket expense** for different patient groups?

📊 **Insights Example:**
💡 **Patients with chronic conditions** tend to have **higher hospital stay durations**.
💡 **Regions with higher insurance coverage** have **lower out-of-pocket expenses**.
💡 **Elderly patients** tend to have **higher treatment costs** and longer stays.

📌 Task 05: Recommendations

✅ **Optimize Resource Allocation**: Focus on regions with **higher hospitalization rates**.
✅ **Improve Insurance Coverage**: Reduce **out-of-pocket expenses** for low-income groups.
✅ **Enhance Early Diagnosis Programs**: Reduce hospital stays by **preventive care measures**.
✅ **Cost Optimization**: Identify treatments with **high expenses** and suggest cost control strategies.

📜 Final Report & Submission Guidelines

📂 Deliverables:

📊 **SQL Outputs** (Saved as `.csv` or `.xlsx` files).
📊 **Interactive Power BI/Tableau Dashboards** for stakeholders.
📑 **A detailed report** documenting analysis, insights, and recommendations.

✅ Submission Checklist:

✔️ **SQL Scripts:** Queries for data processing & analysis.
✔️ **Power BI Dashboard:** Fully functional with visualizations.
✔️ **Report Structure:**
📄 **Introduction & Objective**
📊 **EDA & SQL Queries**
📉 **Key Findings & Trends**
📌 **Visual Analysis Screenshots**
📜 **Conclusion & Business Insights**

📌 Evaluation Criteria:

💯 **Data Cleaning & SQL Queries (25%)**
📊 **EDA & SQL Insights (25%)**
📉 **Power BI/Tableau Dashboards (25%)**
📑 **Report Quality & Interpretation (25%)**

📩 Submission Details:

📄 **Report Format:** PDF
📊 **Dashboard File:** `.pbix` (Power BI) / `.twbx` (Tableau)
📝 **SQL Code:** Documented `.sql` file
📌 **GitHub Repository:** Provide a **link with organized project files**
🛑 **Submit all files and links via email to**: `projects@emergingindiagroup.com`