🎯 EduFin SQL Training - Complete Problem Statement

# 📋 Training Overview

Company: EduFin Credit Solutions  
Business: Education loan provider (₹500 crore portfolio)  
Crisis: Default rates spiked from 8% to 12.4%  
Your Role: Senior Data Analyst  
Timeline: 7 days to master corporate SQL  
Tools: Databricks Community Edition + 1M+ record dataset

# 🏢 Business Background

## EduFin Credit Solutions

Founded: 2019  
Portfolio: ₹500+ crores across 150,000 loans  
Customers: 100,000+ students from 5,000+ institutions  
Geographic Reach: 500+ cities across India  
Current Crisis: Default rate jumped from 8% to 12.4% in Q1 2024

## Database Structure (1M+ Records)

-- Main Tables  
customers (100K records) -- Customer demographics and financials  
loans (150K records) -- Loan details and status  
payments (800K records) -- Payment history and behavior  
institutions (5K records) -- Educational institutions  
defaults\_collections (20K records) -- Default and collection data  
economic\_indicators (240 records) -- Economic data by quarter  
geographic\_demographics (50K records) -- City-wise economic data

# 📅 7-Day Training Structure

## Pre-Training Requirements

Students must complete before Day 1:  
- Set up Databricks Community Edition  
- Import EduFin dataset (provided)  
- Review SQL basics cheatsheet (provided)  
- Read business context document

## Training Schedule

- Day 1 (Monday): Version 1 - Recorded session (Portfolio Risk Analysis)  
- Day 2 (Tuesday): Independent practice and support  
- Day 3 (Wednesday): Version 2 - Live session (Collection Strategy)  
- Day 4 (Thursday): Independent research and development  
- Day 5-7 (Fri-Sun): Version 3 - Strategic project (Business Intelligence)

# 🔥 VERSION 1: Portfolio Risk Crisis Analysis

## Day 1 - Recorded Session (90 minutes)

### The Crisis Call (Monday 9:47 AM)

CEO Rajesh Khanna:  
"Emergency board meeting at 1 PM. Default rates jumped 34% last quarter. Three investors threatening to pull out.  
I need comprehensive portfolio analysis showing exactly where we're bleeding money. You have 3 hours. Company survival depends on your analysis."

# 📊 Analysis Requirements

## 📌 Question 1: Portfolio Health Dashboard

* - Loan distribution by status (Active, Closed, Defaulted)
* - Total exposure and average loan size by category
* - Month-over-month trend analysis
* - Geographic concentration of risks

## 📌 Question 2: Risk Segmentation Analysis

* - Default rates by customer demographics (age, income, city tier)
* - Institution-wise performance analysis
* - Employment type vs default correlation
* - CIBIL score effectiveness in predicting defaults

## 📌 Question 3: Geographic Risk Mapping

* - State and city-wise default rates
* - Urban vs rural performance comparison
* - Economic indicator correlation with defaults
* - Regional expansion risk assessment

## 📌 Question 4: Temporal Pattern Analysis

* - Seasonal default patterns
* - COVID-19 impact analysis (2020-2021)
* - Economic cycle correlation
* - Early warning indicator identification

## 📌 Question 5: Financial Impact Assessment

* - Revenue loss calculation by segment
* - Collection efficiency analysis
* - Provision requirement estimation
* - Profitability impact by product type

# ⚙️ Technical Requirements

- Handle 800K+ payment records efficiently  
- Use advanced SQL: CTEs, window functions, complex joins  
- Optimize queries for performance  
- Handle data quality issues (missing values, duplicates)

# 📦 Deliverables

- Executive summary (2 pages, board-ready)  
- Detailed analysis with SQL code  
- Top 3 risk factors with evidence  
- Immediate action recommendations  
- 30-day tactical plan