



Bank Loan Default & Credit Risk Analysis

(SQL | MySQL)

Project Overview

This project analyzes bank loan, customer, repayment, and credit history data to assess **loan default risk** and identify **early warning signals**. Using structured SQL queries, the analysis evaluates portfolio-level risk, borrower segments, loan characteristics, and repayment behavior to support **data-driven credit risk management**.

The project is designed to simulate **real-world banking risk analysis**, focusing on prevention, monitoring, and smarter decision-making rather than post-default recovery.

Business Objectives

- ★ Measure the **baseline loan default rate**
 - ★ Identify **high-risk customer and loan segments**
 - ★ Analyze the impact of **credit score, income, loan size, and tenure**
 - ★ Detect **early warning signals** from repayment behavior
 - ★ Segment loans into **Low / Medium / High Risk** buckets
 - ★ Provide **actionable business recommendations** for risk mitigation
-

Dataset Description

The analysis uses five structured tables:

Table Name	Description
customers	Customer demographic and financial profile
credit_history	Prior loans, defaults, and credit utilization
loans	Loan details including amount, type, term, and approval date

repayments	Repayment behavior (due date, payment date, paid vs due)
defaults	Loan default flags and reasons

Data Volume

- ❖ Customers: **500**
 - ❖ Credit History Records: **500**
 - ❖ Loans: **800**
 - ❖ Repayment Records: **29,172**
 - ❖ Defaults: **44**
-

Analysis Workflow

1 Step 1–2: Database Setup

- ❖ Created normalized schema with primary & foreign keys
- ❖ Loaded CSV data into MySQL
- ❖ Ensured referential integrity

2 Step 3: Data Understanding & Sanity Checks

Key validations performed:

- ✓ No duplicate records
- ✓ No orphan records
- ✓ No missing or corrupt data
- ✓ Defaults occur only **after loan approval**
- ✓ Behavioral risk signals confirmed (late payments > partial payments)

Baseline Default Rate: 5.50%

3 Step 4: Core Business Risk Analysis

Key Findings:

- ❖ **44 out of 800 loans defaulted → 5.50% default rate**
- ❖ **Low credit score customers** show **~19.75% default rate**
- ❖ **Low-income customers** default at **~9.30%**
- ❖ **Auto & Personal loans** are most prone to default (~6–6.5%)
- ❖ **Large loans** default **1.5× more than medium** and **4.5× more than small loans**
- ❖ **Loan tenure > 24 months** significantly increases default risk
- ❖ **NORTH region** shows the highest default rate (**~7.50%**)

4 Step 5: Advanced Risk Signals & Early Warning Analysis

Behavioral Risk Insights:

- ❖ **Late payments** are a strong early-warning indicator
- ❖ **Partial payments** often appear months before default
- ❖ **Prior defaults strongly predict future defaults**
- ❖ Combined risk logic simulates a **real bank risk dashboard**
- ❖ **High-risk loan count $\approx 2\times$ Medium-risk loans**

Loans were classified into:

- ❖ **Low Risk**
 - ❖ **Medium Risk**
 - ❖ **High Risk**
-



Key Business Insights

- ★ Default risk is **highly concentrated**, not random
 - ★ Credit score and income remain strong **foundational risk drivers**
 - ★ Repayment behavior provides **actionable early signals**
 - ★ Large, long-term loans increase exposure significantly
 - ★ Regional concentration requires **localized risk strategies**
-



Business Recommendations

- ★ Strengthen approval criteria for **low credit score & prior default borrowers**
 - ★ Introduce **early warning monitoring** using late/partial payment flags
 - ★ Apply **risk-based pricing** for high-risk segments
 - ★ Reassess **long-term and high-value loan products**
 - ★ Shift focus from recovery to **early prevention & monitoring**
-



Overall Business Value Delivered

- ★ Established a **clear baseline credit risk profile**
 - ★ Enabled **early detection of high-risk loans**
 - ★ Reduced potential losses through proactive monitoring
 - ★ Improved decision-making for loan approval & pricing
 - ★ Demonstrated how SQL can power **real-world banking analytics**
-

SQL Concepts Used

- ★ Complex JOINS
 - ★ CASE-based segmentation
 - ★ Common Table Expressions (CTEs)
 - ★ Aggregations & window logic
 - ★ Risk scoring models
 - ★ Behavioral analytics
-

Tools Used

- ★ MySQL
 - ★ MySQL Workbench
 - ★ SQL
-

Author

Anirudha Das

Aspiring Data Analyst | SQL | Business Analytics | Risk Analysis

Focused on solving business and operational problems using data analytical tools

 Bolpur | WB | India
