Reducing Loan Defaults

A Strategic Data Analysis for a Fictional Lending Company

AUTHOR:
AREESHA SHAIKH

Data Analyst & Strategic Advisor

Linkedin

Github

EXECUTIVE SUMMARY

The fictional loan-lending company in this project was facing a sharp rise in customer defaults, threatening its profitability and cash flow. These defaults were often tied to approving high-risk borrowers without proper evaluation of key factors such as income, employment history, interest rates, and loan grades. This analysis was conducted to uncover the patterns behind these defaults and provide data-driven recommendations to improve future lending decisions.

Using borrower-level data, we explored both categorical features (such as loan grade, employment length, and loan purpose) and numerical metrics (like income, loan amount, and interest rate) to understand which factors were most closely associated with loan default.

Key findings included:

- Borrowers with **mid-tier loan grades (B, C, D)** showed significantly higher default volumes.
- **Loan amounts above PKR 10,000** were heavily concentrated with defaulters, with default rates rising as loan amounts increased.
- **Interest rates over 13%** showed a consistent and sharp rise in default rates, with the highest-risk group peaking above 80% defaults.
- While **lower income** brackets held the largest share of defaults, **income alone wasn't a reliable predictor** of risk.
- Common assumptions like **employment length** or **loan purpose** didn't show clear, standalone trends in predicting defaults.

Based on these insights, the project proposes strategic actions including stricter filters for high-risk segments, reevaluation of credit grading systems, and the inclusion of interaction-based scoring models to better assess borrower risk.

The outcome of this analysis gives the loan approval and risk teams a stronger analytical foundation to make smarter, safer lending decisions and reduce the financial impact of defaults moving forward.

TABLE OF CONTENTS

PROJECT OVERVIEW:

This project focuses on analyzing a fictional financial loan dataset to address the issue of increasing loan defaults. High default rates pose a major risk to loan-lending businesses, affecting both profitability and long-term sustainability.

The purpose of this analysis is to explore patterns in borrower data such as employment length, income levels, and loan grades to identify key factors linked to default behavior. By examining these trends, the project aims to support data-driven decisions in loan approvals and risk management.

This case study is designed for stakeholders in financial services particularly risk analysts, credit officers, and decision-makers who want to improve how loans are assessed and reduce future losses.

BUSINESS PROBLEM:

The fictional loan-lending company is facing a rise in customer defaults, which is directly affecting profitability and cash flow. These defaults often result from approving high-risk borrowers without a thorough evaluation of key factors such as income, employment history, and loan grades. This gap in assessment increases financial losses and weakens the overall lending strategy. To address this, the risk management and loan approval teams need data-driven insights that can improve how borrower risk is identified and loans are approved. This project aims to uncover patterns in past loan data to support more informed and strategic lending decisions.

OBJECTIVE & KEY QUESTIONS:

The objective of this project is to analyze borrower data and identify patterns associated with loan defaults. By uncovering these trends, the aim is to support more accurate risk assessment and informed loan approval decisions.

The analysis is guided by one core question:

What borrower characteristics are most common among defaulted loans?

To answer this, the project investigates:

- Whether borrower income levels and employment length are linked to default behavior
- How loan grades and sub-grades correlate with repayment outcomes
- The role of key financial variables such as interest rates and loan amounts in influencing default risk

These questions form the foundation for extracting insights that can help minimize loan defaults and strengthen the overall lending strategy.

DATA SUMMARY & CLEANING:

The dataset contains historical records of issued loans, including information on borrower income, employment length, loan amounts, interest rates, loan terms, credit grades, and repayment status. It also includes additional metadata such as dates of issue and repayment activity.

To ensure the quality of insights, the data was cleaned for consistency, accuracy, and relevance. Non-essential or overly granular fields that did not contribute meaningfully to the analysis were removed. Rows with missing values in critical fields were excluded to maintain the integrity of the patterns being explored. Key variables were selected based on their potential to influence loan repayment behavior and risk evaluation.

FINANCIAL ANALYSIS:

To investigate the root causes behind increasing loan defaults, a targeted financial analysis was conducted using selected borrower and loan variables. The focus was placed on identifying patterns and risk indicators that consistently align with loan repayment outcomes. This section examines key dimensions such as income level, employment history, loan grading, and sub-grade performance — all of which are critical for evaluating borrower credibility and informing more strategic loan approval decisions.

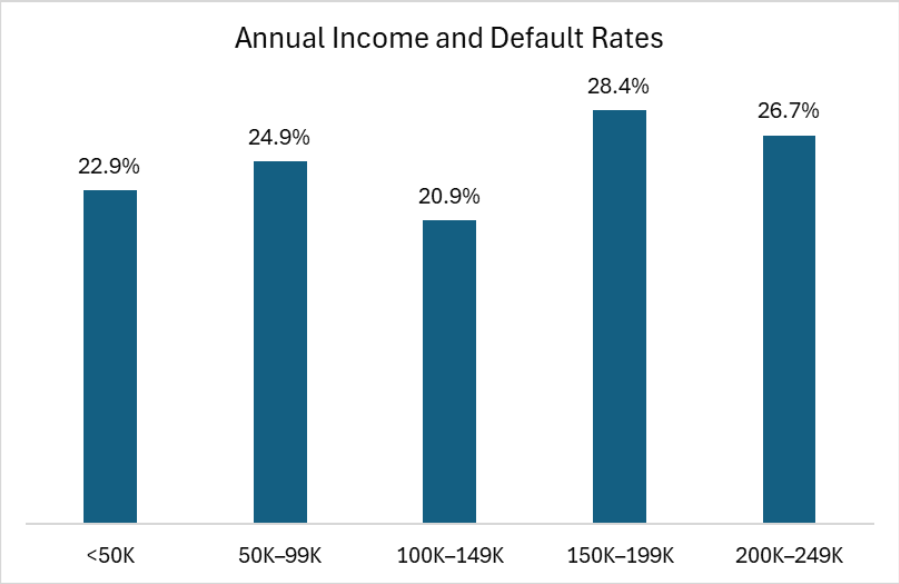
1. **What borrower characteristics are most common among defaulted loans?**

Categorical Pattern Analysis:

Attributes	Most Defaulted Group(s)
Grade	B, C, D
Sub Grade	B3, B5, C1, C2
Employment Length	<1 Year, 2 Years, 10+ Years
Home Ownership	Rent, Mortgage
Loan Purpose	Debt Consolidation, Credit Card, Small Business, Other Purposes
Term	36 Months
Verification Status	Not Verified

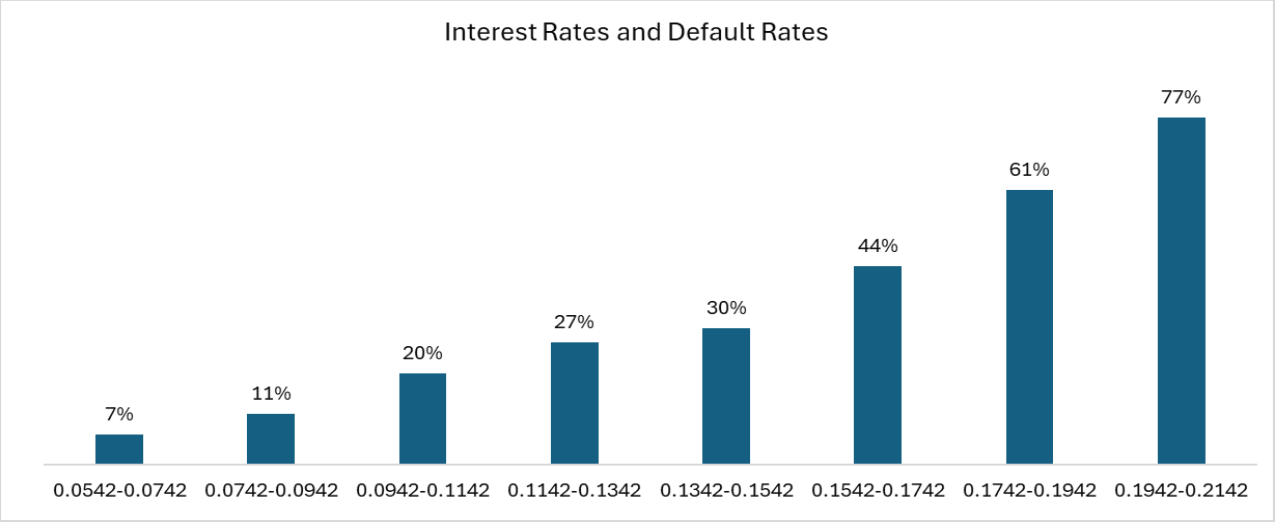
Numerical Pattern Analysis:

1. Annual Income vs Default Rate:



The highest volume of defaults came from borrowers earning under PKR 100K annually. While default rates ranged from **20.9% to 28.4%**, no strong correlation between income and default was observed. However, borrowers in the **50K–199K bracket** formed the **bulk of defaulters**, reflecting broader loan distribution patterns.

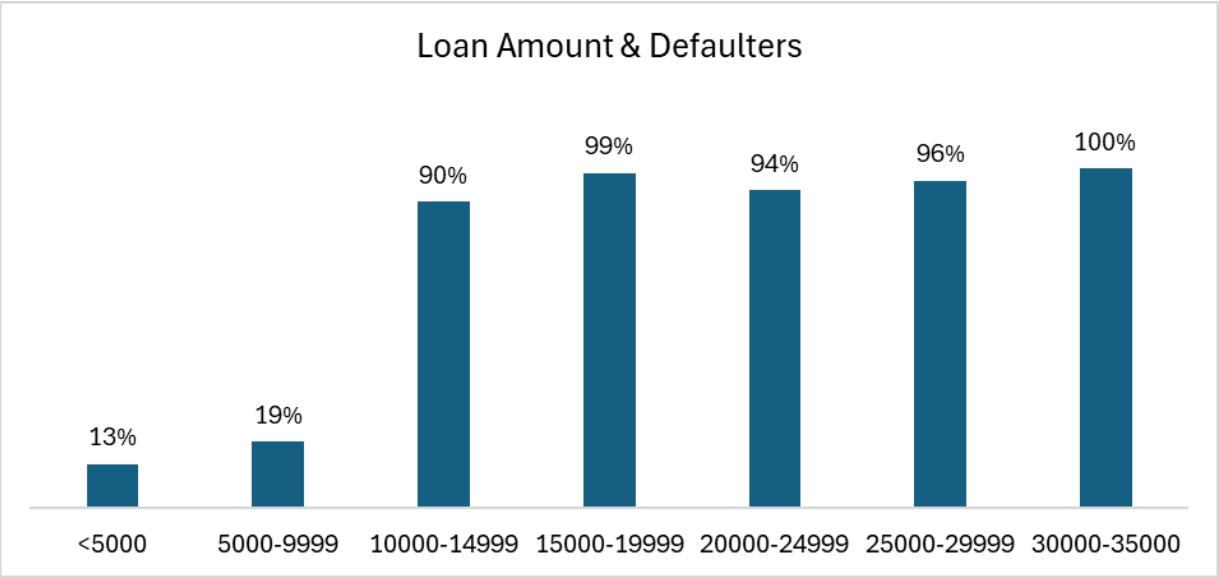
2. Interest Rate vs Default Rate:



Higher interest rates = higher risk of default.

Default rates jump from just **7% at low interest rates (5–7%)** to over **60%** once rates cross **17%**. The sharpest spike begins **after 13%**, showing that interest rate is **a strong indicator of credit risk**. Lenders likely charge higher interest to riskier borrowers — and it shows in the data.

3. Loan Amount vs Default Rate:



Default rates are very low for loans below PKR 10,000 of around 13–19%. But once loan amounts cross PKR 10,000, default rates skyrocket, peaking at 99% in the PKR 15K– 20K bracket. Even with smaller sample sizes, the spike is sharp and consistent, suggesting higher loan amounts come with significantly higher risk.

BUSINESS INSIGHTS:

Note: These insights are based on groups with enough data to show reliable patterns. Any exceptions (with very few cases) are mentioned separately.

1. High-Risk Borrower Profiles Are Linked to Mid-Tier Loan Grades

Loans issued to borrowers with **grades B, C, and D** represent a major portion of defaults, with sub-grades like **B5, C1, and C2** showing the highest default volume. These mid-tier credit grades may carry more risk than currently accounted for.

2. Default Risk Rises Sharply Beyond PKR 10,000 Loan Amount

Borrowers with loans **above PKR 10,000** show a sharp increase in default rates, with **90%+ defaults observed** in the PKR 10,000–25,000 range. While the highest bracket (PKR 30,000–35,000) shows 100% default, this is based on a small number of loans and should be interpreted with caution.

3. High Interest Rates Are Strongly Tied to Defaults

Default rates increase consistently with interest rates above **13%**, peaking at **84%** in the 21%–23% bracket. While 100% defaults were observed at the 23%–25% tier, this was based on very few cases.

4. Lower-Income Groups Dominate Default Volume, But Not Always the Rate

Borrowers earning **below PKR 100,000 annually** accounted for a large share of defaults, though higher-income brackets showed scattered spikes in default percentage, often due to small sample sizes.

5. Employment Tenure and Loan Purpose Alone Offer Weak Predictive Value

Borrowers with **10+ years of employment** still had high default volumes. Similarly, **debt consolidation** loans dominated defaults, but this may reflect popularity, not risk level.

STRATEGIC RECOMMENDATIONS:

1. **Reevaluate Mid-Tier Loan Grades (B, C, D)**
   The highest default volumes came from borrowers in these grades. Consider tightening approval criteria or introducing additional checks (e.g., income verification, credit utilization) for these groups.

2. **Set Stricter Approval Filters for Loans Above PKR 10,000**
   Since defaults rise sharply beyond this point, implement more rigorous borrower assessments (like debt-to-income thresholds or employment stability requirements) for mid-range loan amounts.

3. **Cap Interest Rates or Apply Affordability Checks**
   Defaults spike when interest rates exceed 13%. Consider placing a cap, or at least running affordability stress tests before issuing high-interest loans.

4. **Don’t Rely Solely on Income as a Risk Indicator**
   Income alone doesn’t consistently predict repayment behavior. Use it alongside other variables like purpose, term, and historical payment behavior.

5. **Enhance Scoring Models with Interaction Between Features**
   Employment length and loan purpose alone didn’t provide clear signals. Future credit risk models should consider **combinations** of variables rather than relying on individual fields.

CONCLUSION:

This project set out to help a fictional loan-lending company tackle rising loan defaults by uncovering patterns in borrower data. Through a focused analysis of both categorical and numerical features, we found that most defaults occurred among borrowers with mid-tier loan grades (B, C, and D), loan amounts above PKR 10,000, and interest rates exceeding 13%. While low income and long employment tenure were common among defaulters, these factors alone were not strong predictors of default risk.

Our insights challenge the current loan approval process, which often relies too heavily on individual factors like income or employment length. By identifying which combinations of borrower characteristics actually correlate with defaults, this analysis provides a stronger foundation for risk assessment. The recommendations outlined aim to tighten approval criteria, refine interest rate thresholds, and reduce default rates.

This analysis equips the lending team with actionable, data-driven insights to support more effective and financially sound loan approval strategies.