Lending Club Case Study

Loan Default Analysis & Risk Assessment

Exploratory Data Analysis (EDA)
Group 7

Rahul Yadav, Kotresh T M, Prasanth Padmanaban, KHAN MOHAMMED FASEEHULLA, Pradhan Prangya

January 2025

Executive Summary

- Analyzed 39,717 loan applications (2007-2011) with 14.4% default rate
- Identified key risk factors: Credit grades B, C, D account for majority of defaults
- 60-month terms show significantly higher default rates than 36-month terms
- Geographic risk concentration in CA, FL, NY states
- Low income (<$40k) and high DTI ratios (>16) strongly predict defaults
- Debt consolidation loans represent highest risk category
- Developed risk-based pricing framework with 4-tier classification system

Business Problem & Objectives

- **Challenge:**
- • Lending Club faces critical loan approval decisions
- • Rejecting good applicants = Lost business opportunity
- • Approving risky applicants = Credit losses

- Objectives:
- • Identify applicants at risk of defaulting
- • Understand driver variables behind loan default
- • Enable data-driven loan approval decisions
- • Reduce credit losses through better risk assessment

Dataset Overview

- 📊 Total Records: 39,717 loan applications
- 📅 Time Period: 2007-2011
- 💳 Focus: Approved loans with definitive outcomes
- ✅ Fully Paid: 32,950 loans (84.5%)
- ❌ Charged Off: 5,627 loans (14.4%)
- 🔍 Excluded: 'Current' status loans (ongoing)
- 🏦 Source: Lending Club historical data

Methodology

- **1. Data Understanding & Cleaning**
  - Removed 54 columns with only NA values
  - Handled outliers using IQR method
  - Imputed missing employment length values
  - Standardized categorical variables
- **2. Exploratory Data Analysis**
  - Univariate analysis of key variables
  - Bivariate analysis vs loan status
  - Correlation analysis
  - Multivariate risk assessment
- **3. Tools & Technologies**
  - Python, pandas, numpy
  - matplotlib, seaborn for visualization
  - Jupyter Notebook environment

Loan Status Distribution

📈 **Overall Default Rate: 14.4%**

✅ **Fully Paid: 32,950 loans (84.5%)**

- Successful loan completions
- Generated interest income

❌ **Charged Off: 5,627 loans (14.4%)**

- Failed to complete payment
- Resulted in credit losses

🎯 **Target: Reduce default rate through better risk assessment**

Risk Factor #1: Credit Grades

- 🎯 **Key Finding: Grades B, C, D account for majority of defaults**

- **Grade A (Lowest Risk):**
  - Minimal default rates (<5%)
  - Highest creditworthiness

- **Grades B, C, D (High Risk):**

  Grade B: 1,352 charged-off applications

  Sub-grades B3, B4, B5: Highest default likelihood

  Significant contributor to credit losses

- 📊 **Business Impact:**

  Grade is the strongest predictor of default

  Clear risk segmentation opportunity

  Basis for risk-adjusted pricing

Risk Factor #2: Loan Terms & Interest Rates

- ⏰ **Loan Term Impact:**

  60-month loans: Higher default rates

  36-month loans: Lower default rates

  Extended terms = Increased risk exposure

- 💰 **Interest Rate Correlation:**

  13%-17% rate range: Highest defaults (2,025 cases)

  Higher rates often reflect higher risk borrowers

  Strong predictor of loan performance

- 🎯 **Strategic Implications:**

  Limit 60-month terms for lower-grade borrowers

  Risk-adjusted pricing based on term length

  Enhanced criteria for longer-term loans

Risk Factor #3: Geographic Distribution

- 🗺 **High-Risk States:**

  California (CA): 1,055 charged-off applicants

  Florida (FL): High concentration of defaults

  New York (NY): Significant risk exposure

- 📊 **Geographic Patterns:**

  Risk clustering in specific regions

  Economic factors influence default rates

  State-specific risk assessment needed

- 💡 **Strategic Response:**

  State-specific risk adjustments

  Monitor regional economic indicators

  Adjust pricing by geographic risk

Risk Factor #4: Income & Debt-to-Income Ratio

- 💰 **Income Analysis:**

  Low income (<$40k): 1,561 charged-off cases

  Higher income correlates with lower default rates

  Income verification impacts risk assessment

- 📊 **Debt-to-Income Ratio:**

  DTI > 16: Very high risk category (1,178 defaults)

  DTI 8-12: Optimal performance range

  Strong predictor of repayment ability

- 🎯 **Risk Management:**

  Set DTI limits based on credit grade

  Enhanced income verification for high-risk segments

  DTI-based pricing tiers

Correlation Analysis

- 📈 **Strong Correlations:**

  Installment ↔ Loan amount, Funded amount

  Term length ↔ Interest rate

  Annual income ↔ Loan amount

- 📉 **Weak Correlations:**

  DTI has weak correlation with most fields

  Employment length shows minimal correlation

- 🔻 **Negative Correlations:**

  Public record bankruptcies ↔ Most variables

  Annual income ↔ DTI (expected inverse relationship)

- 💡 **Business Insights:**

  Combined variables provide stronger predictive power

  Multi-factor risk scoring approach recommended

Key Risk Indicators Summary

- 🥇 **Top Risk Factors (in order of importance):**

1️⃣ Credit Grade: Strongest predictor (Grades B, C, D)

2️⃣ Loan Term: 60-month loans higher risk than 36-month

3️⃣ Interest Rate: 13%-17% range shows highest defaults

4️⃣ DTI Ratio: >16 significantly increases default probability

5️⃣ Annual Income: <$40k associated with higher defaults

6️⃣ Loan Purpose: Debt consolidation highest risk category

7️⃣ Home Ownership: Renters show higher default rates

- 📊 **Risk Range: Default rates vary from <5% to >25%**
- 🎯 **Predictive Power: Combined factors provide strong assessment**

Business Recommendations

- 🎯 **Immediate Actions:**
  - Enhanced grade-based pricing for Grades C & D
  - Limit 60-month terms for lower-grade borrowers
  - Implement DTI threshold limits (max 16 for high-risk)
  - Apply state-specific risk adjustments
- 📈 **Strategic Initiatives:**
  - Develop composite risk scoring model
  - Dynamic pricing based on multiple risk factors
  - Portfolio diversification across risk segments
  - Early warning systems for borrower monitoring
- ⚡ **Quick Wins:**
  - Stricter criteria for debt consolidation loans
  - Enhanced income verification for <$40k applicants
  - Premium pricing for high-risk combinations

Risk Mitigation Framework

- 🟢 **Tier 1 (Low Risk):**
  - Grade A, 36-month terms, DTI < 12, Income > $60k
  - Standard processing, competitive rates
- 🟡 **Tier 2 (Medium Risk):**
  - Grade B, Mixed terms, DTI 12-16, Standard processing
  - Moderate risk adjustments
- 🟠 **Tier 3 (High Risk):**
  - Grades C-D, Enhanced scrutiny, Limited 60-month terms
  - Risk-adjusted pricing, additional verification
- 🔴 **Tier 4 (Highest Risk):**
  - Multiple negative factors combined
  - Premium pricing or decline recommendation
  - Enhanced monitoring if approved

Conclusions & Business Value

- 🎯 **Key Achievements:**
  - Identified clear risk patterns in 39,717 loan dataset
  - Developed actionable risk assessment framework
  - Created data-driven pricing recommendations
- 💡 **Critical Insights:**
  - Credit grade is the strongest default predictor
  - Multi-factor approach provides superior risk assessment
  - Geographic and demographic patterns enable targeted strategies
- 📈 **Business Impact:**
  - Potential to reduce 14.4% default rate significantly
  - Enhanced profitability through risk-based pricing
  - Improved portfolio performance and risk management
- 🚀 **Next Steps: Implement risk tiers and monitor performance**