

FINAL PROJECT REPORT

CREDIT RISK ANALYSIS AND LOAN DEFAULT PATTERN IDENTIFICATION

Sector: Financial Analytics / Credit Risk Modeling

Team: Shourya, Animesh, Punit, Rachit, Prakhar, Dev

Institute: Newton School Of Technology

Faculty:

Date: 18/02/2026

1. Executive Summary

- Financial institutions face increasing pressure to expand unsecured lending while maintaining portfolio quality. Rising default rates directly affect profitability, capital adequacy, and long-term financial stability.

- This project analyzes 32,584 consumer loan records to identify key drivers of loan default. Using structured exploratory analysis, segmentation techniques, and dashboard-based risk visualization, the study evaluates how borrower demographics, income stress, credit history, housing status, and loan characteristics influence repayment behavior.
 - Key findings show that loan-to-income ratio, prior default history, short credit history length, rental housing status, and certain loan purposes significantly increase default probability. Risk factors compound, meaning borrowers exhibiting multiple stress indicators demonstrate disproportionately higher default likelihood.
 - The interactive dashboard enables dynamic risk segmentation, supporting data-driven underwriting decisions, pricing refinement, and exposure control strategies. The analysis provides a practical framework for improving credit risk management while maintaining sustainable lending growth.
-

2. Sector & Business Context

2.1 Financial Lending Sector Overview

The consumer lending sector plays a crucial role in economic growth by providing individuals access to credit for education, medical needs, and

personal expenses. However, lending institutions operate in a high-risk environment where inaccurate borrower assessment can lead to increased non-performing assets.

Key operational considerations in this sector include:

- Credit risk evaluation
- Interest rate pricing
- Portfolio diversification
- Regulatory compliance

2.2 Current Challenges

Despite established credit evaluation systems, lenders continue to face several challenges:

- Limited visibility into borrower affordability
- Rising unsecured personal loan exposure
- High default rates among specific borrower segments
- Difficulty in balancing growth and risk control

2.3 Problem Motivation

Loan default prediction is fundamental to:

- Reducing financial losses
- Improving capital efficiency
- Enhancing credit accessibility
- Supporting responsible lending practices

This project was chosen to analyze default behavior patterns and provide insights that can strengthen credit risk frameworks.

3. Problem Statement & Objectives

3.1 Problem Statement

Despite established grading and screening mechanisms, loan defaults persist due to compounded borrower-level risk factors that are not evaluated holistically.

Core Question:

How can borrower demographic, financial, and credit attributes be systematically analyzed to identify high-risk profiles and reduce default exposure?

3.2 Project Scope

Included:

- Cross-sectional borrower risk analysis
- Default vs non-default comparison
- Risk segmentation across loan grades and purposes
- Income stress evaluation

Excluded:

- Time-series repayment modeling
 - Real-time credit scoring systems
 - Behavioral transaction-level analysis
-

3.3 Objectives

1. Analyze borrower financial and demographic characteristics
 2. Identify key default drivers
 3. Compare risk across loan grades and intents
 4. Evaluate impact of affordability stress
 5. Develop actionable risk insights
-

3.4 Success Criteria

- Clear differentiation between defaulted and non-defaulted loans
 - Quantifiable insights supported by data
 - Alignment with practical credit risk decision-making
-

4. Data Description

Dataset Source: [credit_dataset](#)

Dataset obtained from:- **Kaggle**

4.1 Dataset Overview

The dataset used in this project consists of **32,584 preprocessed loan records**. Each record represents an individual borrower and loan application outcome.

The target variable is:

- **Loan Status:**
 - 1 – Default
 - 0 – Non-default

4.2 Key Variables

The dataset includes the following categories of variables:

Demographic Variables

- Person Age
- Employment Length

Financial Variables

- Annual Income
- Loan Amount
- Loan Percentage of Income

Loan Characteristics

- Loan Intent
- Loan Grade
- Interest Rate
- Loan Status

Credit History Indicators

- Credit History Length
- Previous Default Indicator

Housing Information

- Home Ownership Status

4.3 Data Limitations

- Absence of macroeconomic indicators
 - No post-loan repayment timeline data
 - No transactional or behavioral data
 - Binary default outcome only
-

5. Data Cleaning and Preparation

Step No.	Category	Action Performed	Variables Affected	Reason / Justification
1	Missing Values	Removed records with missing interest rate values	loan_int_rate	Interest rate is a critical risk indicator and cannot be reliably imputed
2	Missing Values	Retained records with missing employment length when other attributes were valid	person_emp_length	Prevented unnecessary data loss while preserving analytical integrity
3	Error Handling	Removed spreadsheet calculation errors such as #DIV/0!	Invalid columns	Such values have no business meaning and disrupt numerical analysis
4	Column Cleaning	Dropped unnamed and empty columns	Unnamed columns	Columns contained no analytical relevance
5	Outlier Treatment	Identified and removed unrealistic values	person_age, person_income, person_emp_length	Prevented distortion of statistical measures and risk patterns
6	Ratio Validation	Removed invalid or infinite	loan_percent_income	Ensured mathematical and financial validity

		loan-to-income ratios	t_income	
7	Data Type Fixing	Converted variables to numeric format	Age, income, loan amount, interest rate, ratios	Required for statistical analysis and visualization
8	Categorical Standardization	Standardized category labels and removed malformed entries	Home ownership, loan intent, grade, default history	Ensured consistent grouping and comparison
9	Target Validation	Standardized loan outcome into binary format (0/1)	loan_status	Required for default classification analysis
10	Consistency Check	Ensured non-negative and valid credit history values	cb_person_credit_history_length	Credit history length cannot be negative
11	Duplicate Removal	Removed duplicate borrower records	All fields	Prevented bias in default rate calculations

6. KPI & Metric Framework

The following key performance indicators were used to assess credit risk:

- **Overall Default Rate:** Approximately **21.9%**
- **Average Loan Amount:** ₹9,876
- **Average Interest Rate:** 11.19%
- **Average Loan-to-Income Ratio:** 18%

- **Average Age:** 28 years

These metrics establish the baseline for further comparative analysis.

7. Exploratory Data Analysis (EDA)

7.1 Demographic Analysis

The median borrower age is **26 years**, with most borrowers falling between **23 and 30 years**. Default rates are observed to be higher among younger borrowers, indicating limited credit experience and financial stability.

Suggested Visualization: Histogram of borrower age

Insight: Higher default density in lower age groups

7.2 Income & Affordability Analysis

The income distribution is right-skewed, with a median income of **₹55,000**. Borrowers with higher loan-to-income ratios show a significantly increased likelihood of default, highlighting income stress as a critical risk factor.

Suggested Visualization: Box plot of loan-to-income ratio vs loan status

Insight: Defaulted loans cluster at higher income burden levels

7.3 Loan Amount & Interest Rate Analysis

Loan amounts range from ₹500 to ₹35,000. Defaulted loans generally have higher loan amounts and are associated with higher interest rates, reaching up to **23.22%**.

This trend validates risk-based pricing practices used by lenders.

7.4 Loan Grade Analysis

Loan grades show a strong inverse relationship with repayment performance. Lower grades consistently exhibit higher default rates, confirming the effectiveness of grading mechanisms in risk differentiation.

Suggested Visualization: Bar chart of default rate by loan grade

7.5 Loan Intent Analysis

Medical and personal loans display higher default incidence compared to education loans. This suggests that necessity-driven borrowing may involve higher financial stress.

7.6 Credit History & Past Defaults

Borrowers with prior defaults demonstrate substantially higher current default rates. Shorter credit history length also correlates with increased risk.

Suggested Visualization: Stacked bar chart of past default vs loan status

8. Advanced Analytical Insights

8.1 Risk Segmentation

Borrowers categorized as:

Low Risk:

- Low LTI
- No prior default
- Higher grade
- Longer credit history

High Risk:

- High LTI
- Prior default
- Lower grade
- Short credit history

High-risk segments show compounded default probability.

8.2 Exposure Concentration

High-risk borrowers receive higher average loan amounts (~₹10.9K), increasing loss severity potential.

Risk is therefore amplified both by probability and exposure size.

8.3 Compound Risk Effect

Default likelihood increases non-linearly when multiple stress factors overlap:

- High LTI + Short credit history
 - Prior default + High interest rate
 - Rental status + Low employment duration
-

9. Dashboard Design

The dashboard was implemented in Google Sheets using pivot tables, slicers, and interactive filters.

Objective

Enable dynamic risk segmentation for decision-makers.

Components

- Portfolio Snapshot Panel
- Default vs Non-default Distribution
- Interest Rate Comparison
- Risk Segment vs Loan Amount
- Home Ownership Risk View
- Experience-Based Default Analysis
- Loan Intent Risk Mapping
- Key Risk Signals Summary

Filters include:

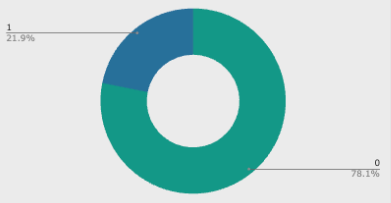
- Loan Grade
- Home Ownership
- Experience Type
- Loan Intent

The dashboard supports interactive credit risk evaluation.

Customers Total	Loan Amount Average	Person income Average	Percent Income Average	Loan Interest Rate Average	Age Average
32411	9,876.69	67,531.05	0.18	11.19	28

Out of 32,411 customers, ~21.9% have defaulted, indicating a moderate but significant credit risk in the portfolio.
The majority (~78.1%) are non-defaulters, showing overall portfolio stability.

Default vs Non-Default share



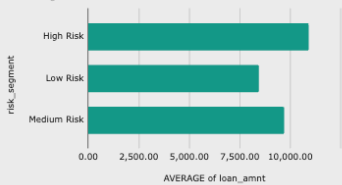
*Defaulters are charged higher interest rates (~12.9%) compared to non-defaulters (~10.5%).
*This suggests risk-based pricing, but also indicates that higher interest burden may contribute to defaults.

Business takeaway :
High interest rates may increase short-term revenue but can elevate default probability.

Average Interest Rate: Default vs Non-Default



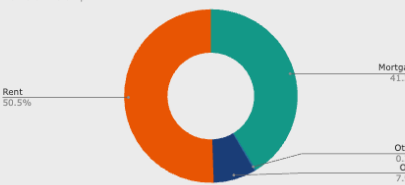
Average Loan Amount according to risk evolved



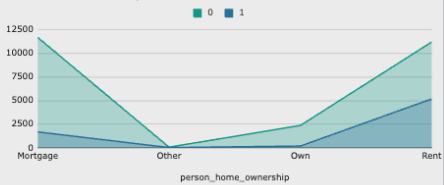
*High-risk customers receive the highest average loan amount (~₹10.9K).
*Low-risk customers receive smaller loans (~₹8.5K).

Risk Observation :
Lending larger amounts to high-risk customers increases potential loss exposure.

Customer Distribution home ownership



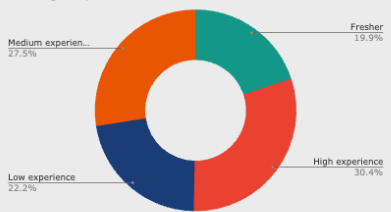
Comparion of loan status Based on Home Ownership



*Renters show the highest number of defaults, followed by mortgage holders.
*Customers who own homes have comparatively lower default counts.

Business takeaway
Home ownership is a strong indicator of financial stability and lower default risk.

Loan Distribution according to experience



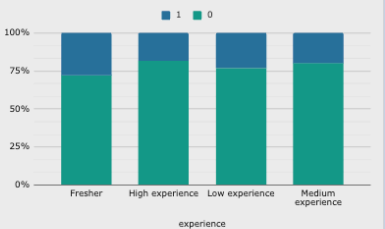
*Freshers show a relatively higher default proportion compared to experienced customers.
*Customers with high and medium experience exhibit better repayment behavior.

Risk Interpretation :
Work experience correlates positively with loan repayment reliability.

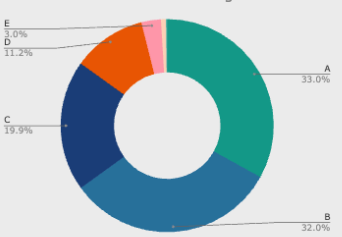
*Medical, personal, and venture loans show higher default counts.
*Education and home improvement loans perform relatively better.

Strategic Insight
Loans for non-income-generating or emergency purposes carry higher default risk.

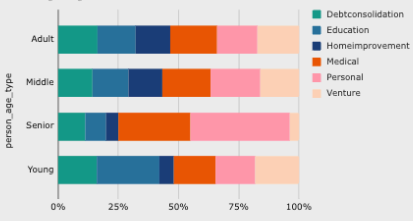
Default and Non Default users distribution



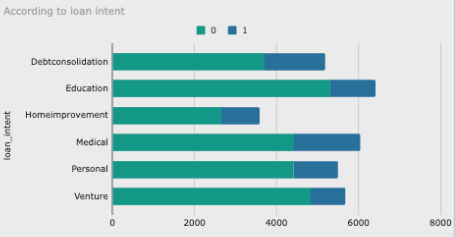
Customer distribution on loan grade



Loan purpose analysis According to age



Customer Distribution According to loan intent



Key Risk Signals Summary

- High interest rates linked to higher defaults
- Renters and freshers are the most vulnerable segments
- Medical & personal loans carry higher default risk
- High loan amounts issued to high-risk customers increase exposure

Conclusion:
The dashboard highlights that default risk is strongly influenced by interest rate, employment experience, home ownership, and loan purpose. Optimizing loan size and pricing for high-risk segments can significantly improve portfolio performance.

10. Key Insights Summary

1. Overall default rate is approximately 22%
2. Loan-to-income ratio is a strong predictor of default
3. Higher interest rates align with higher risk
4. Loan grades effectively differentiate borrower risk
5. Younger borrowers default more frequently
6. Past defaults strongly predict future defaults
7. Medical and personal loans carry higher risk
8. Renters show higher default incidence
9. Short employment history increases risk.
10. Risk factors compound across borrower profiles

11. Recommendations

Recommendation 1: Income Stress Thresholds

Introduce stricter caps on loan-to-income ratios to reduce default exposure.

Recommendation 2: Grade-Based Pricing Refinement

Enhance interest rate differentiation within mid-risk loan grades.

Recommendation 3: Enhanced Screening for High-Risk Loan Intents

Apply additional verification for medical and personal loans.

Recommendation 4: Exposure Control for High-Risk Segments

Cap loan size for borrowers with compounded risk indicators

Recommendation 5: Multi-Factor Risk Approval Matrix

Implement risk-weighted approval matrix combining multiple signals

12. Impact Estimation

A 10% reduction in high-risk loan approvals could lead to:

- 3–5% reduction in defaults
 - Improved portfolio stability
 - Better capital utilization
-

13. Limitations

- No time-series repayment behavior
 - No macroeconomic impact modeling
 - No behavioral transaction data
 - Cross-sectional analysis only
-

14. Future Scope

- Machine learning-based default prediction
 - Explainable AI for credit scoring
 - Integration of macroeconomic variables
 - Real-time risk monitoring dashboards
-

15. Conclusion

This project demonstrates how structured data analysis and dashboard-driven segmentation can uncover meaningful default risk patterns in consumer lending.

By identifying key risk drivers and exposure concentration, the study provides actionable strategies for improving underwriting quality, refining pricing models, and strengthening portfolio stability.

The framework balances growth with responsible risk management, offering practical value for financial institutions.

16. Contribution Matrix

Team Member	Dataset & Sourcing	Cleaning	KPI & Analysis	Dashboard	Report Writing	PPT	Overall Role
Shourya	✓	✓	✓	✓	✓		Project Lead
Animesh		✓	✓		✓		Analysis Lead
Punit		✓	✓	✓			Dashboard Lead
Rachit	✓	✓				✓	Data and Quality Lead
Prakhar	✓				✓	✓	Strategy and PPT Lead
Dev							