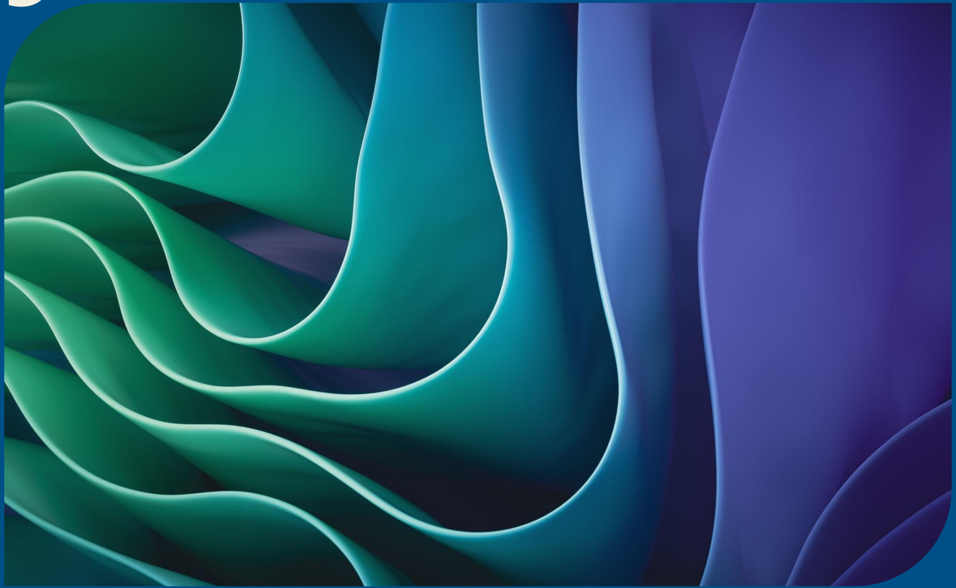
Lending Club Case Study

Amit Kumar
Tuyet Tu



Agenda

1. **Problem Statement**
2. **Data Approach**
3. **Key Findings**
4. **Recommendations**

PROBLEM STATEMENT

Leading online lending marketplace seeks to reduce credit losses by identifying key predictors of loan defaults.

Current Challenge: "Charged-off" loans represent significant financial risk and lost revenue.

Goal: Develop data-driven insights to strengthen loan screening and default risk reduction.

Data Processing Strategy

- Handle Missing Values
  - Drop columns with >20% missing data
  - Remove rows with <1% missing values in critical fields
  - Impute where logical:
    - Employment length → mode
    - Revolving utilization → mean
    - Bankruptcies → 0 (for 2% missing)
- Clean & Simplify
  - Remove administrative/index fields
  - Consolidate categories:
    - Home ownership: combine "none/other"
    - Employment length: create ranges
- Feature Engineering
  - Create default classification target
  - Develop income brackets
  - Calculate income-to-loan ratios

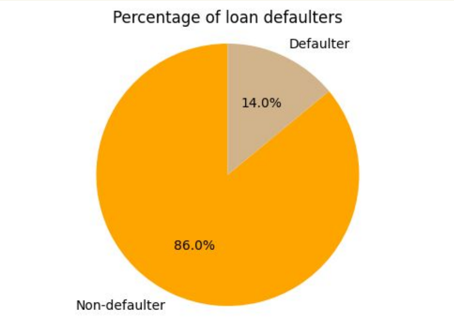
Goal: Prepare clean, structured dataset for meaningful default analysis

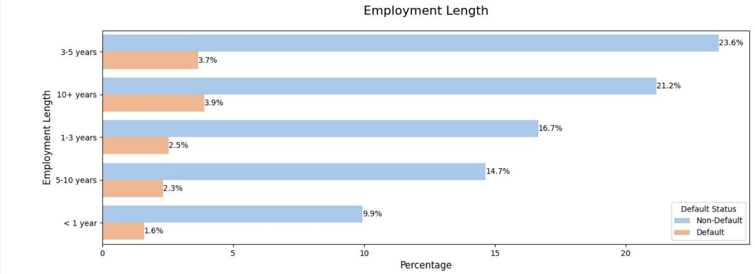
Key Findings

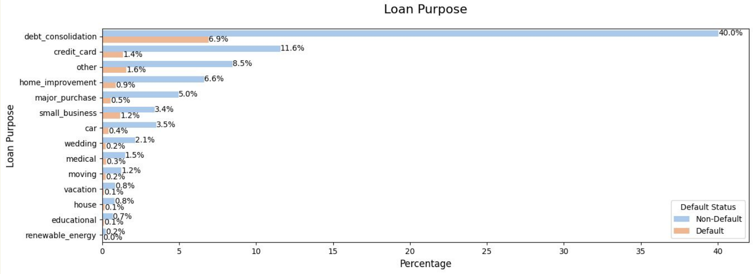
Predictors of loan defaults.

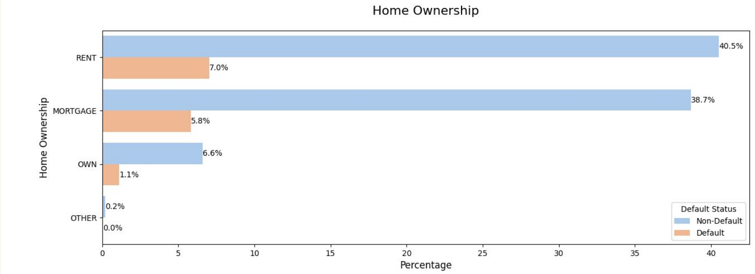
Borrower Profile

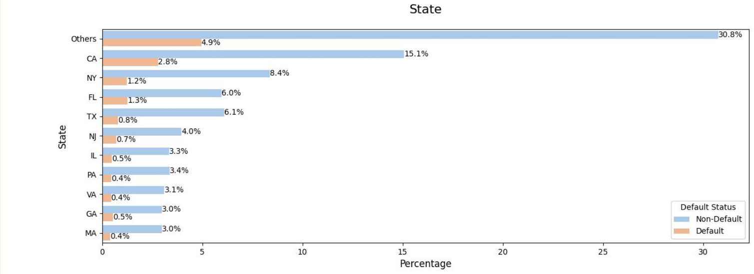
- 14% of loans are defaulters
- Employment length: borrowers with 10+ years of experience form the largest group, suggesting **employment stability is a key factor.**
- Housing status data: renters and mortgage holders are the predominant borrowers, with similar default patterns across housing categories.
- The state-wise distribution shows concentrated lending in major states like CA, NY, TX, and FL, with varying default rates across regions, suggesting the need for geography-specific risk assessment strategies.





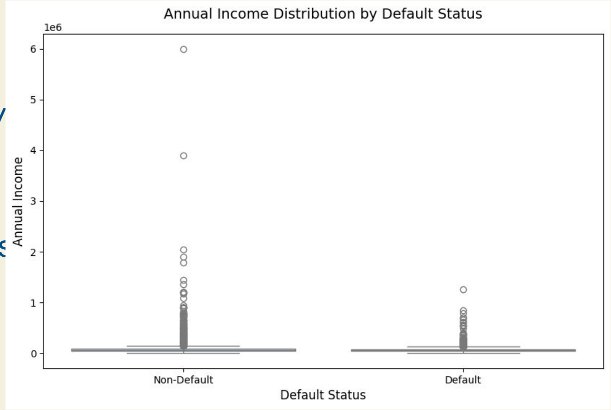






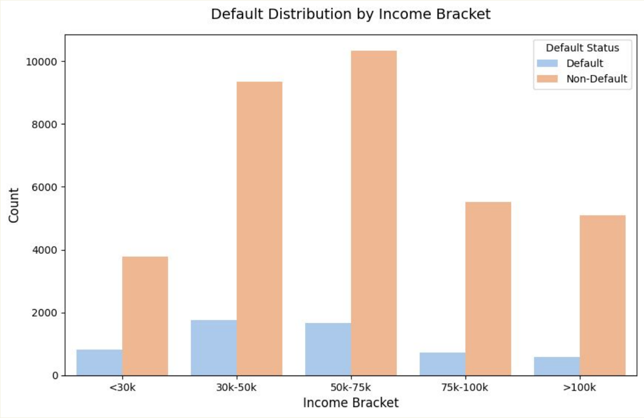
Annual Income Range

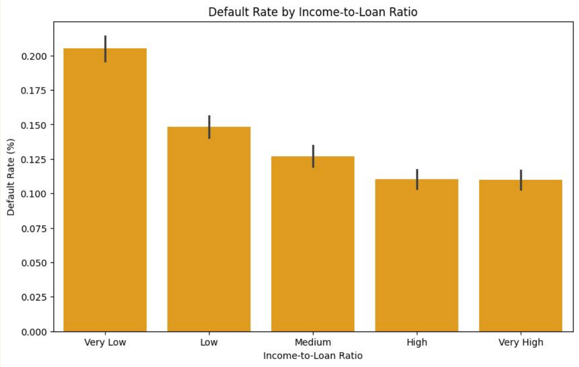
- Even though the non-defaulting group includes a few outliers with extremely high incomes, this doesn't significantly impact the overall default patterns.
- Most borrowers are clustered in lower income ranges with middle-income brackets ($30k-75k) representing the highest concentration of both defaults and non-defaults.
- The default ratio remains relatively consistent across income brackets, suggesting that income alone is not a strong predictor of default risk.



Income-to-Loan Ratio:

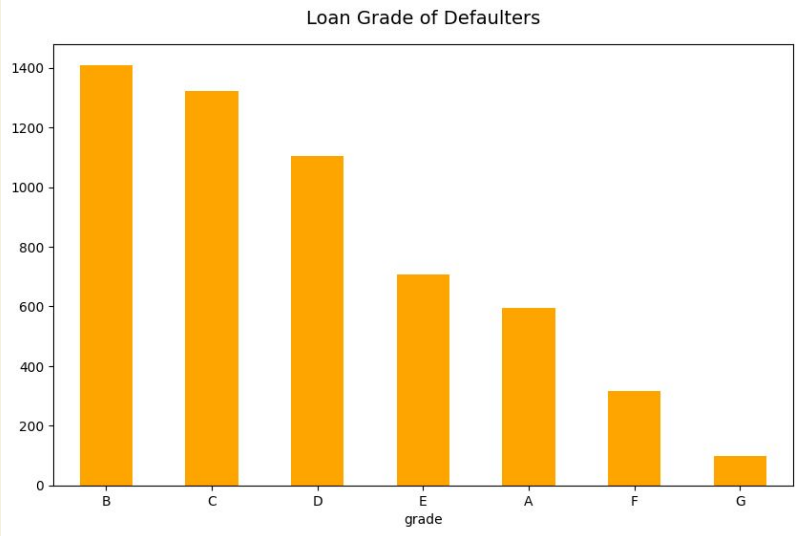
- Very strong negative correlation with default rates as Default risk steadily decreases as income-to-loan ratio increases
- Borrowers with very low income-to-loan ratios have highest default rate (~20%)

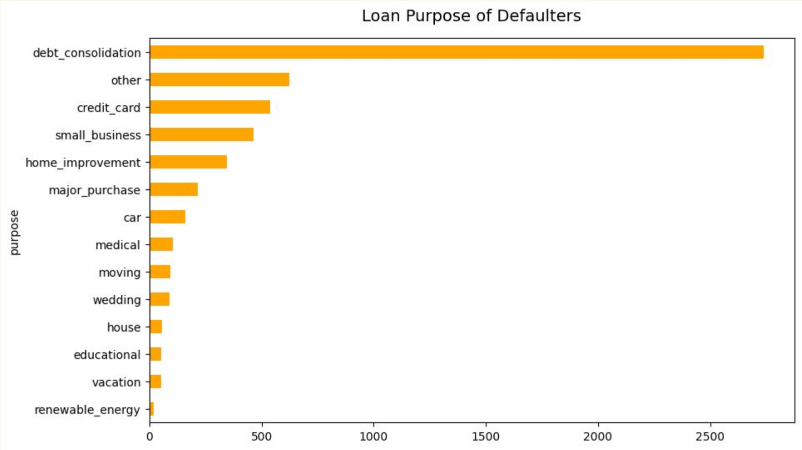




Loan Distribution

- Overally, the pattern from lower-risk grades (B>C>D) to higher-risk grades (F>G) shows the strategy to approve more for low-risk grades loan. However, Grade B and C loans - higher quality loan grades - account for the highest number of defaults, suggesting potential flaws in the risk assessment model.
- Debt consolidation significantly dominates the default cases with approximately 16,000 defaults, followed distantly by credit card refinancing.
- The data suggests a need for more stringent evaluation of debt consolidation loan applications, even for applicants qualifying for better loan grades.





Verification status

The bank requires verification for a large portion of loans (around 57%)

verification_status	count	is_default mean	int_rate mean	loan_amnt mean
Not Verified	16882	0.125	11.201	8426.736
Source Verified	9973	0.142	12.101	10102.868
Verified	12789	0.159	13.031	15794.100

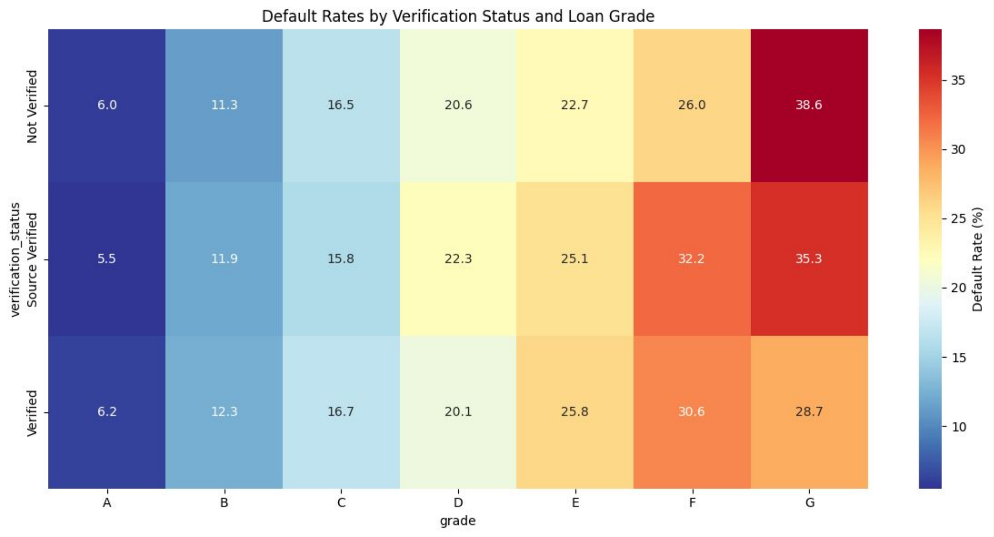
Verification status x Loan amount

Higher-risk applications with high loan amount likely trigger verification requirements, resulting in higher interest rate & default rates, in particular:

- Higher loan amount: 2x difference between verified & non-verified
- Interest rates increases by 0.9% with each verification level
- Default rates increases by 1.7% with each verification level

Verification status x Grade loan

Even in the same high-quality grades, verified loans tend to have a higher default rates.



Verification status doesn't significantly reduce default rates. Loan grade appears to be a stronger predictor of default than verification status

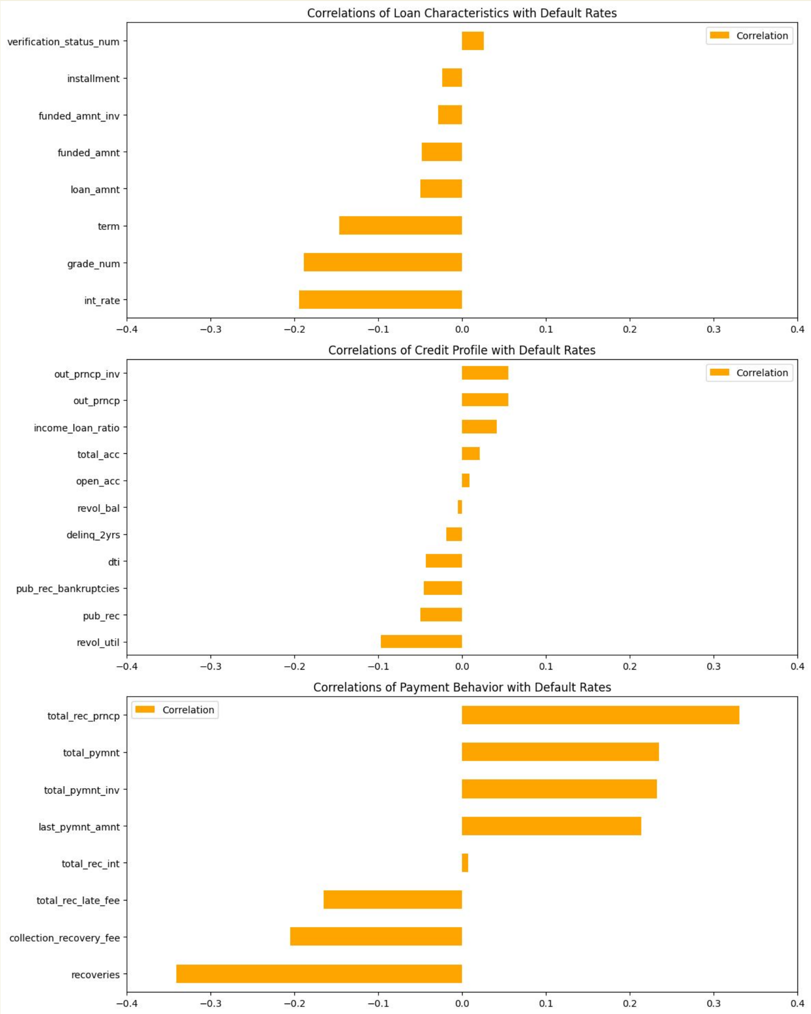
Correlations

<u>**Loan Characteristics**</u>

- Most features have negative correlations with default rates, of which **interest rate, grade loan and term** shows strongest impact.
- Verification status has a slight positive correlation
- All others has minimal predictive power to the defaulting probability

Loan characteristics are not as strong predictor of default compared as to metrics of payment history, while those of credit profiles have very slight impact.

Suggesting that we should consider loan characteristics only for new borrowers and focus more on past payment patterns to predict the defaulting probability

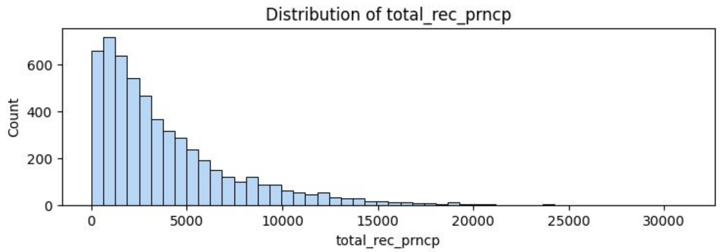


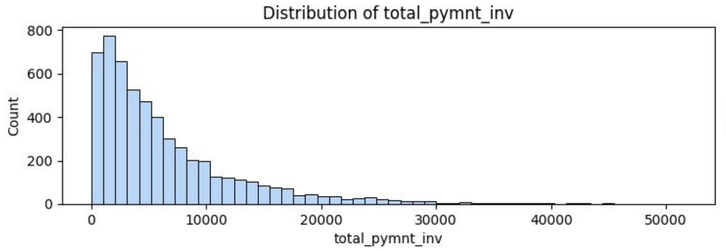
Payment Distribution

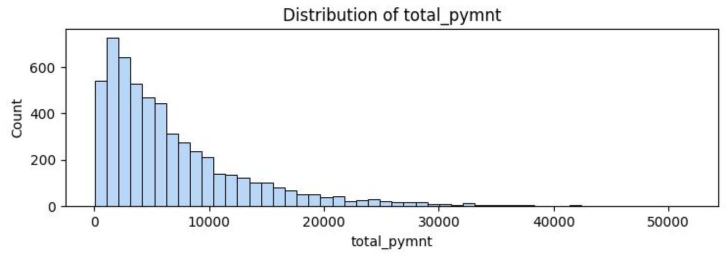
Default Risk Pattern (Based on Payment Distributions of high correlated variables)

- The heavily right-skewed distributions of payment data show most loans cluster at lower payment amounts
- High count of low-value payments suggests significant default risk in the portfolio
- The sharp drop-off in payment distributions indicates a clear threshold (at about 20,000$) where borrowers either pay consistently or stop paying entirely

This threshold is significant to cut off the risk for default reduction or loan segmenting



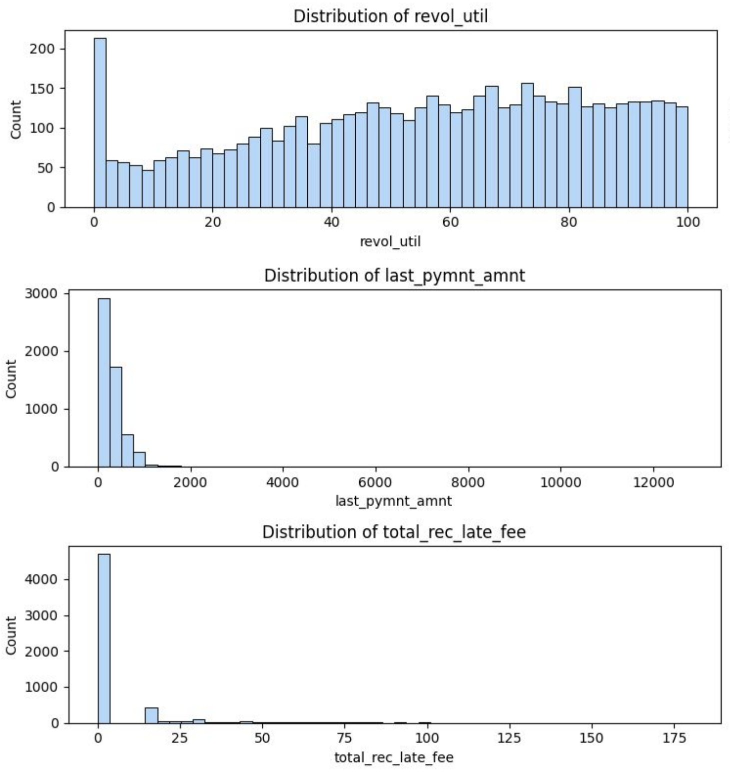




Payment Distribution

Default Risk Pattern (Based on Payment Distributions of high correlated variables)

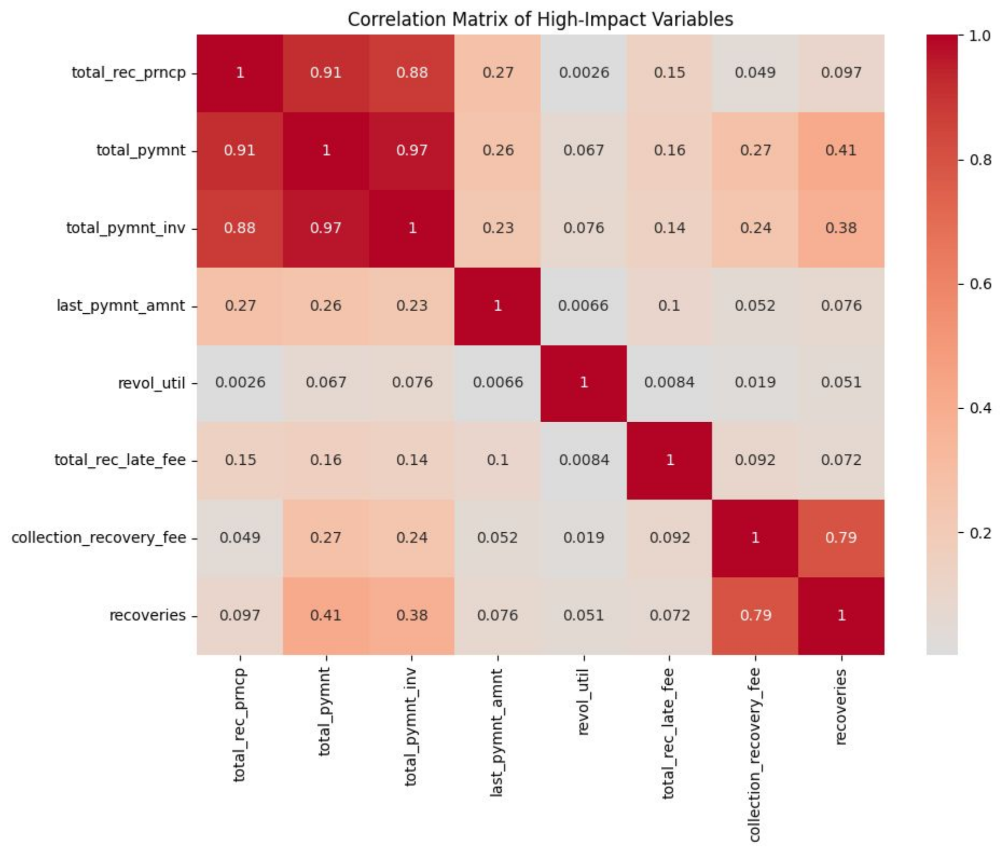
- Fairly uniform distribution across different utilization levels, Suggests that default risk isn't strongly tied to any specific utilization level
- Most defaulted loans have very low or no late fees, indicate that most defaults happen without going through extended late fee periods. Many borrowers default early in their loan term.



Payment Distribution

Default Risk Pattern (Based on Payment Distributions of high correlated variables)

- Strong positive correlations of 3 types of payment amounts because they represent different aspects. They are closely related and might be redundant in analysis
- Last payment amount shows moderate correlation, could be a moderate indicator of overall payment behavior
- Late fees or revolving utilization appear to independent factors
- Recoveries are most closely tied to collection recovery fees, which makes sense as they're both part of the debt recovery process.



Strategic Recommendations for Default Risk Reduction

High-Priority Actions:

- **Enhance Early Risk Detection**
  - Implement real-time payment pattern monitoring system or trigger for irregular payment patterns
  - Focus intervention in first 6 months of loan term
    (Early defaults show minimal late fees, indicating rapid progression to default)
- **Revise Lending Criteria**
  - Enforce strict income-to-loan ratio thresholds
  - Strengthen debt consolidation loan screening
  - Develop region-specific risk models for major markets
    (20% default rate in low income-to-loan ratios, highest defaults in debt consolidation)
- **Optimize Risk Assessment Model**
  - Prioritize payment history data for assessment
  - Reform grade B/C loan evaluation with stricter verification for high-quality grades
  - Set $20,000 payment threshold as risk indicator
    (Payment patterns are stronger default predictors than credit profiles, Higher-grade loans show unexpected default rates)

Implementation Priority: Focus on early payment pattern monitoring and income-to-loan ratio controls for immediate risk reduction.