Investment Strategies for Peer-to-Peer Lending

Update 1: Diagnostic & Descriptive Analysis

Team 18: Bhavesh Jain, Nikita Suryawanshi, Zarmeen

Agenda

1 Problem Overview
2 Goals & KPIs
3 Attributes' Summary
4 Hypotheses
5 Loan Grades' Analysis
6 Cluster Analysis
7 Principal Component Analysis

Problem Overview

- Jasmin Gonzales wants to diversify the investment portfolio and is looking at Peer-to-Peer (P2P) lending platforms like LendingClub.
- Challenges:
    1. There is high risk in this investment due to borrower defaults.
    2. There is a lack of clear action plan for maximizing returns while minimizing risks.
- Objective: Leverage machine learning and past data to identify the best investment strategies, balancing risk and return.

Goals & KPIs

Decisions for Jasmin to make:

- Whether she should invest in Peer-to-Peer (P2P) lending?
- Which loan should she invest in and by how much?
- Whether she should invest in a single loan at a time or multiple loans?

KPIs to evaluate success:

1. **Return:** Ensure the investment generates significantly profitable returns, and maximize recovery in case of a loss.
2. **Risk:** Minimize the risks associated with loan defaults and late repayments while maximizing returns.

Attributes' Summary

Categories

Borrower Profile & Credit

Critical for assessing credit risk. A stronger credit profile indicates higher likelihood of repayment.
Attributes: annual income, tax liens, home ownership, debt-to-income ratio

Loan Details

Helps to determine the cost of the loan and its risk-return profile and expected performance.
Attributes: loan amount, interest rate, funded amount, term, purpose, verification status

Payment Details

Helps measure the actual performance of the loan and track if payments are made on time & interest earned.
Attributes: last payment made, recoveries, total payment made

Other Credit Metrics

Metrics that add nuance to borrower's credit profile & provide additional details into credit history & refine risk assessment.
Attributes: credit card utilization, number of mortgage accounts

For more details on the categorization and their importance, see the file:
https://docs.google.com/spreadsheets/d/1x7uvzmYDW3W6E4R4N0G_KlBAJLePP6sXGX8MzoaXL-8/edit?usp=sharing

Hypotheses

Hypothesis 1

Return vs. Risk

Higher risk loans will bear higher returns, reflecting a positive correlation between the two.

Hypothesis 2.a

Borrower Profile & Default Rates

Borrower’s profile factors - homeownership status, geographic location, duration of employment, annual income - influences default rates.

Hypothesis 2.b

Borrower Credit & Default Rates

Borrower’s credit factors -delinquency/bankruptcy reports, debt-to-income ratio, time of opening credit line - influences default rates.

Hypothesis 3

Loan characteristics & Risk/Return

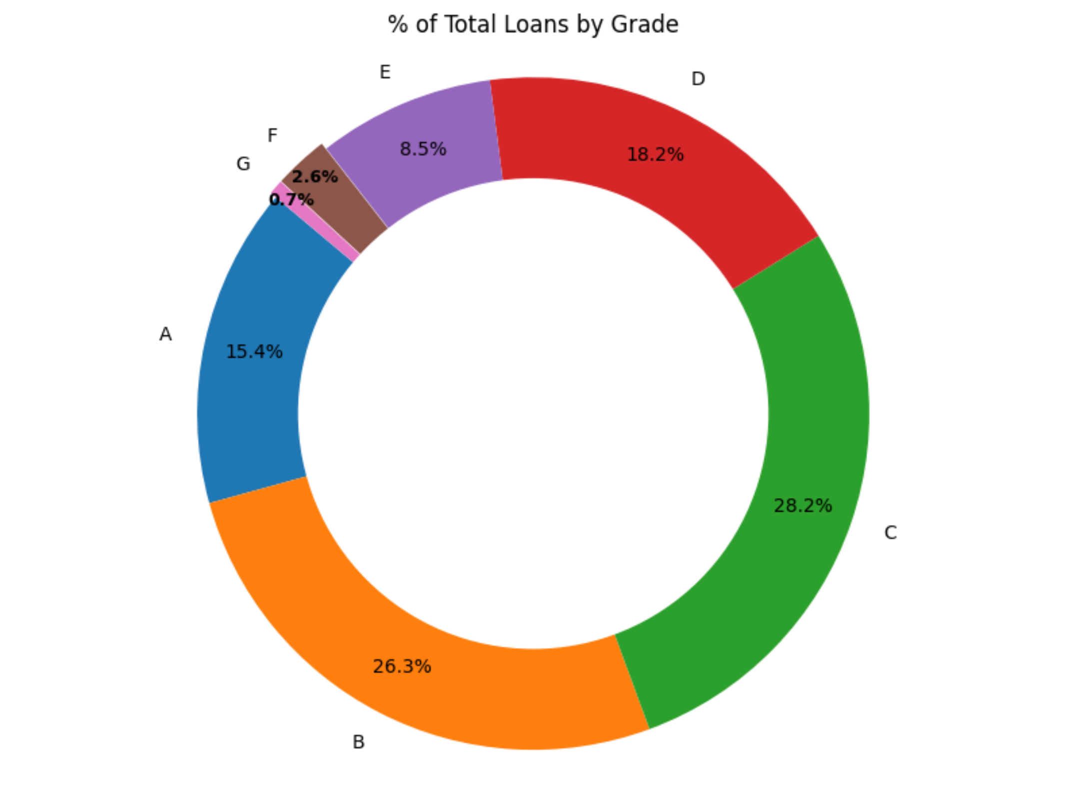
Loan related parameters, such as amount, term and purpose, positively influence the associated return and risk factors.

Derived Loan Return Features

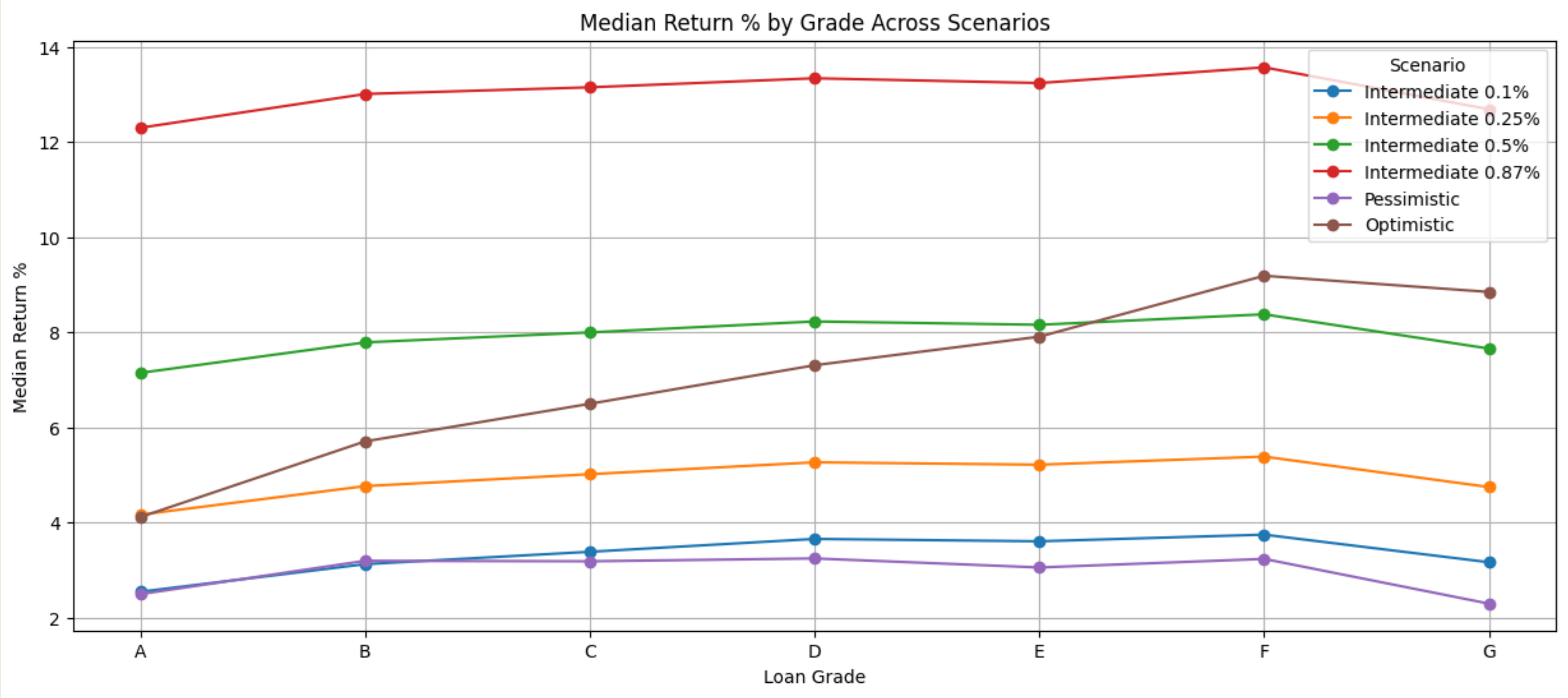
- **Method 1 (Pessimistic):** Assumes that once a loan is paid back, investor will not reinvest until the end of the term.
- **Method 2 (Optimistic):** Assumes that once loan is paid back, investor can reinvest in a loan with the exact same return.
- **Method 3 (Re-investment):** Assumes that once a loan is paid back, investor can reinvest at an annual interest rate of i%, compounded monthly.

In addition to the three interest rates for method 3 (ret_INTa, ret_INTb, ret_INTc), we have added an additional interest rate, ret_INTd, assumed to be ***0.873%*** *monthly (equivalent to a 10.99% annual median interest rate).*

Loan Grades' Analysis

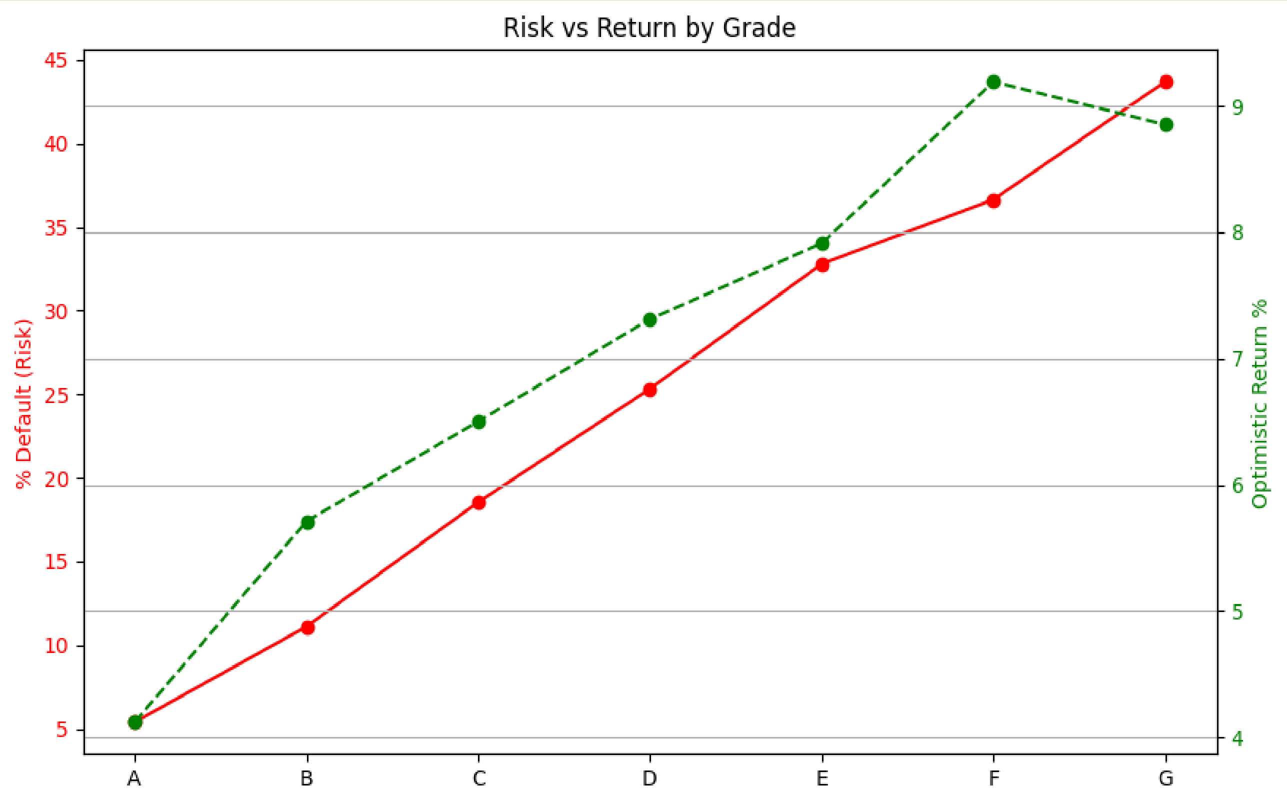


- The highest number of loans are in Grade C, closely followed by Grade B, accounting for more than half of the loans.
- The lowest number of loans are in Grade G, followed by Grade F.
- The top three slices (Grade C, B and D) account for roughly 72.7% of the dataset, highlighting a concentrated distribution in a few categories

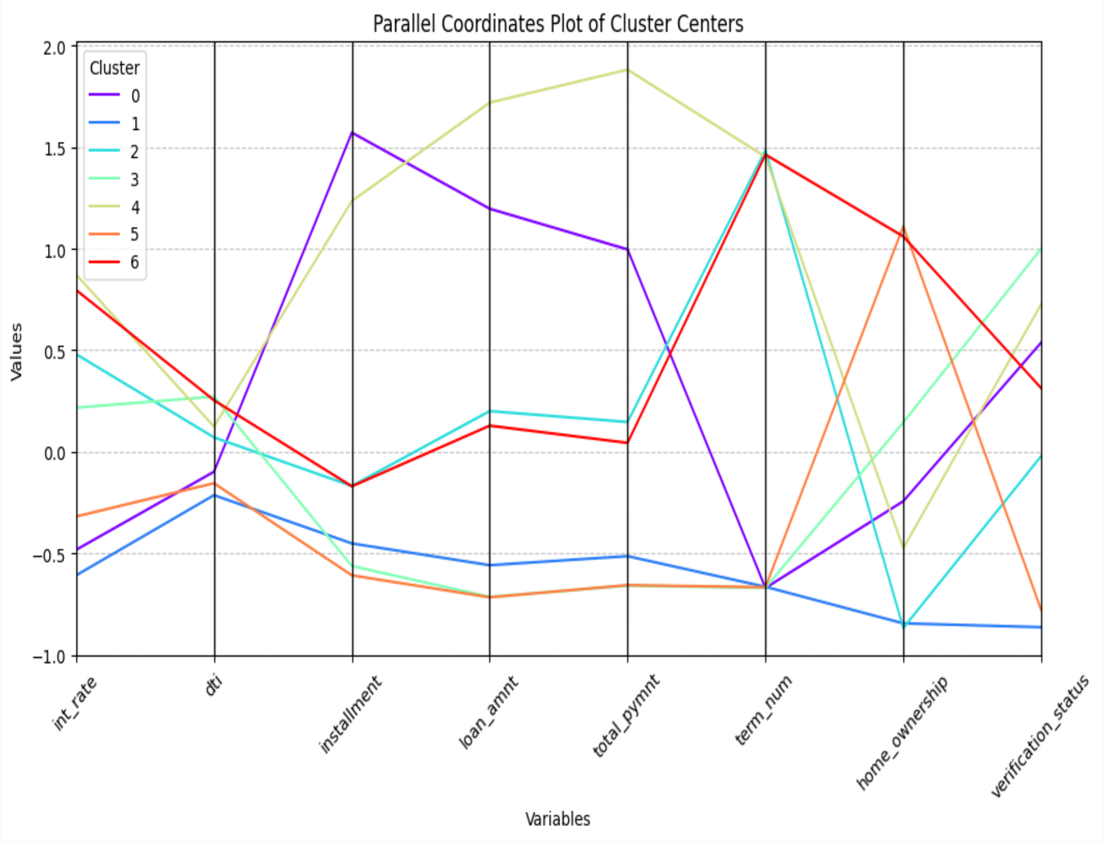


- Returns tend to rise from Grade A to about D/E, indicating potentially greater payoff for higher-risk loans.
- The median interest rate of 10.99% shows to yield promising returns for all grades.
- The optimistic method has a relatively low return for all grades but F and G, while the pessimistic method has low returns for all.
- The gap between Optimistic and Pessimistic grows for riskier grades (E, F), suggesting increased volatility or uncertainty in outcomes.

- There is a direct correlation observed between risk and return for all grades.
- A higher risk in investment promises a higher return.
- Risk rises consistently from A to G, showing that lower-grade loans carry significantly higher risk of default.
- The chart clearly highlights that Grade A/B offer lower yields with minimal default, whereas Grades F/G can yield greater returns but dramatically raise default risk.



Clustering Analysis



See speaker notes for additional details

Cluster 0 : Grade A/B - Short -Term, Larger Loans at Lower Rates: Low interest rate but considerable amount and installments for smaller terms.

Cluster 1 : Grade A - Under -Verified, Low -Value Borrowers: Most loan parameters are low and status is not verified.

Cluster 2 : Grade D/E - Moderate -Rate, Long-Term, Verified Renters: Average loans, moderate interest rates, high terms with no home ownership and mostly verified.

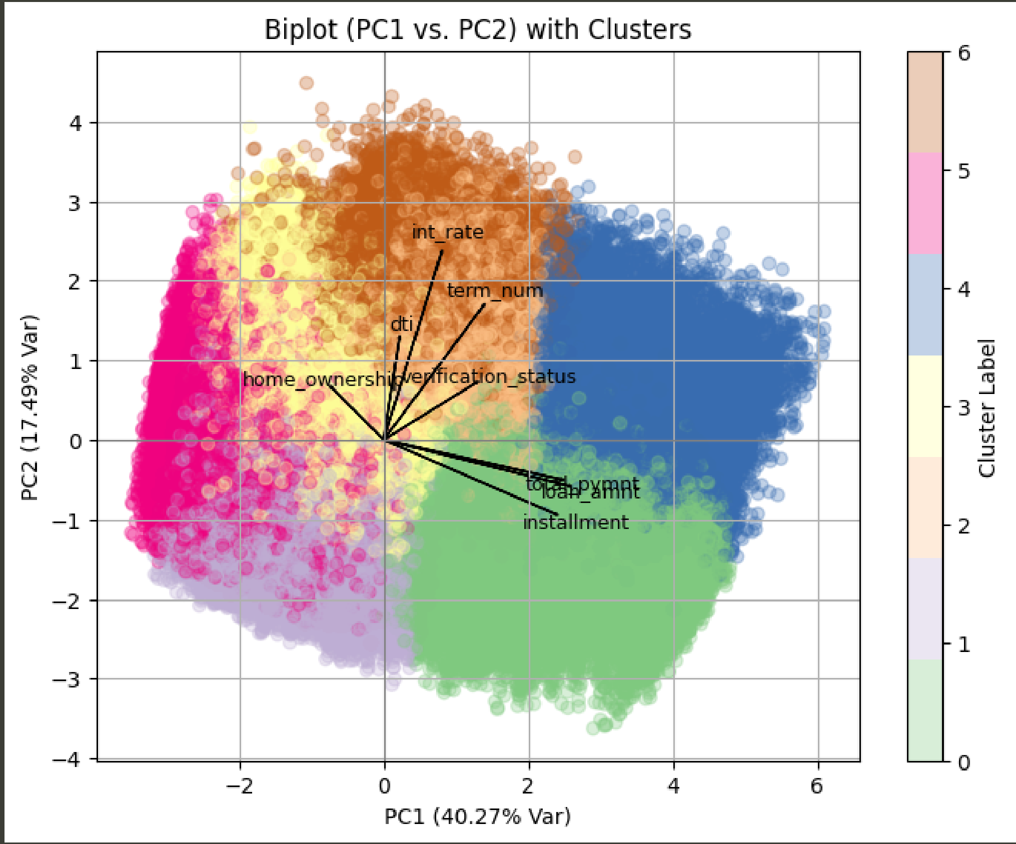
Cluster 3 : Grade C/D- High -DTI, Short Term, Well -Verified Borrowers: Highest dti, low loan parameters for short terms and most verified.

Cluster 4 : Grade G- Large-Principal, High -Rate, High -DTI Borrowers: Highest interest rates, loan amounts, very high dti, low home ownership & long terms.

Cluster 5 : Grade B/C- Modest -Parameter Homeowners : Most loan parameters near or below average, may/may not be verified and most own their homes.

Cluster 6 : Grade F/G- High -Interest, High -DTI, Risky Borrowers: Very high interest rates for average loan amounts, highest dti and very long

Principal Component Analysis



- PC1 reflects a loan/payment dimension (influenced by parameters like loan amount, total payment, installment, etc).
- PC2 reflects a borrower profile/risk dimension (influenced by parameters like debt-to-income ratio, term, verification status, etc).
- They primarily segregate loans by how large and how risky they are.
- On a visual basis:
  1. points on the right are larger loans with larger installments and total payments.
  2. points on the left are smaller loans with smaller installments and total payments.
  3. points on the top have higher dti, longer terms and unverified status.
  4. points on the bottom have lower dti, shorter terms and verified status.

THANK YOU!