Business Value through Integrative Analytics
Team 43

Smart Investing in Peer-to-Peer Loans: From Data to Action

Kojo Dokyi
Srushti Kadam
Nivedha Gautham Raj

Cluster-Based Return Prediction Simplifying Portfolio Strategy

- Trained a return prediction model using 2014 Lending Club loans
- Grouped loans into **20 clusters** based on shared borrower traits
- Each cluster has its own **average return** and **risk level (standard deviation)**
- 2015 loans are matched to their **nearest 2014 cluster**
- New loans inherit the cluster's **risk/return profile**
- Enables **quick, interpretable risk assignment** without retraining
- Helps investors **balance return vs. volatility** when building portfolios

Jasmin's Investment Strategy – Data-Backed & Risk-Aware

- Jasmin wants **high returns**, **low defaults**, and **clear decision-making**
- She's seeking a **data-driven strategy** to avoid unnecessary risk while capturing upside
- We **clustered loans by risk profiles** to group similar borrower behaviors
- Used **XGBoost to predict returns** based on borrower features
- Assigned each loan a **risk score** using volatility (standard deviation)
- This approach helps Jasmin build a portfolio that **balances risk and reward**
- The strategy is **clean, interpretable**, and grounded in real data

- Selected **25 key features** from Lending Club data (2014–2015)
- Engineered 3 return types - used **intermediate return** for realism
- Found **Grades E–G** have high default risk
- **High income & job stability** reduce default likelihood
- Clustered loans into **3 borrower segments** by risk profile
- Recommended focus on **Grade B–C loans**
- Emphasized **portfolio diversification** to balance risk and return

Update 2 – Key Insights

- Predicted defaults using only **borrower-supplied data** to avoid signal leakage
  Chose **Logistic Regression** for interpretability, **XGBoost** for performance
- Key predictors: **loan term, DTI, income**
- Used **ret_INTc** return metric for realistic risk–reward balance
  Custom Strategy 4: 21.87% return with **0% defaults** – best performer
- Other strategies: Low-default (safe), High-return (risky), Random (baseline)
- **Hybrid** and **Interest Rate Bin** strategies scaled well to larger portfolios
  Performance drop in 2015 showed **need for time-aware retraining**

From Rules to Optimization: Smarter Portfolio Design

We move beyond rules by balancing risk, return, and constraints

Why Optimize?

- Rules are rigid and ignore budgets or diversification.
- Investors like Jasmin need:
  - High return
  - Acceptable risk
  - Portfolio flexibility (size, grade mix, etc.)

Our Optimization Approach

- **Objective**: Maximize total expected return
- **Constraints**:
  - Budget limit (e.g., $300K)
  - Max loan count (e.g., 100)
  - Optional diversification (e.g., spread across grades)
- **Risk**: Defined as standard deviation of predicted return within clusters
  - Clustering captures real-world loan similarity

What the Optimized Portfolio Delivers

Balanced. Scalable. High-performing.

Optimized Portfolio Results

- **100 loans selected** – maximize return while respecting risk constraints
- **9.17% average return** with controlled risk.
- Risk exposure is aggregated from **cluster-level volatility**, providing a smarter measure than grade or interest rate alone.

Metric	Value
Loans Selected	100
Total Investment	$166,500
Total Expected Return	$15,266.78
Average Return	9.17%
Total Risk Exposure	5.04

The optimized portfolio proves that algorithmic strategies can outperform simple rules – especially at scale.

Robust Across Portfolios and Budgets

Optimization adapts to constraints while preserving strong returns

Loans	Budget	Investment	Total Return	Avg Return	Total Risk	Feasibility
50	$150K	$65,525	$5,789	8.83%	2.62	✅
75	$250K	$104,075	$8,912	8.56%	3.91	✅
100	$300K	$165,450	$15,325	9.26%	5.14	✅
125	$150K	—	—	—	—	❌ Infeasible

Scalable Returns

Larger portfolios yield higher expected returns — up to **9.26%** for 100 loans.

Risk-managed

Total risk exposure grows gradually with size, but remains within acceptable bounds.

Feasibility-aware

Optimization flags infeasible scenarios under strict budgets (e.g., 125 loans at $150K).

What Makes These Loans Optimal?

Our picks follow clear, interpretable patterns

How we ensure explainability

- Predicted return comes from an XGBoost model trained on application-time features
- Risk is derived from clusters of similar loans, not arbitrary rules
- Each cluster has a nameable profile (e.g., “low DTI, moderate income”)

Common Patterns in Selected Loans

- Mostly Grades B and C
- Annual income > $60K
- DTI < 20
- Term = 36 months more frequent than 60
- Lower revolving utilization (<60%)

What Makes These Loans Optimal?

Our picks follow clear, interpretable patterns

How we ensure explainability

- Predicted return comes from an XGBoost model trained on application-time features
- Risk is derived from clusters of similar loans, not arbitrary rules
- Each cluster has a nameable profile (e.g., “low DTI, moderate income”)

Common Patterns in Selected Loans

- Mostly Grades B and C
- Annual income > $60K
- DTI < 20
- Term = 36 months more frequent than 60
- Lower revolving utilization (<60%)

Our Final Recommendation: A Risk-Aware, Diversified Investment Strategy

Balanced Strategy with Strong Returns and Controlled Risk

Key Decision Rules Summary

- **Select 100 Loans**: A manageable and diversified portfolio size
- **Budget Cap of $200,000**: Ensures total loan investment stays within a practical funding limit.
- **Grade Diversification Constraint**: Maximum 30% of loans from any single grade (A through G) to prevent overconcentration
- **Data-Driven Filtering**: We only looked at loans that were small (under $5,000), had decent borrower income, and weren't overloaded with debt.

Recommendation Box

Number of Loans	100
Money Invested	$165,450
Total Profit Expected	$15,325
Average Return	9.17%
Risk Level	Low to Medium
Loan Grades Chosen	A, B, C, D, E

- **Diversified**: The money isn't concentrated in one kind of loan. It's spread across safer and riskier ones.
- **Stable Borrowers**: Most loans went to people with manageable debt and steady finances.

What Could Go Wrong? And What's Next

Key Risks and Assumptions

What We're Assuming	Why It Matters
We assume past loan behavior predicts future outcomes	But the economy or borrower behavior could change suddenly.
We use predicted returns from our best model	But even the best model can't be 100% right — especially in real life.
Risk is based on clusters of similar loans	But new loans might not match old patterns exactly
We chose 100 loans and $200K budget as the base	But Jasmin's actual preferences might be different.

Next Steps

What We'd Like to Explore	Why It Helps
Try different levels of risk aversion (β)	Helps tailor the portfolio for cautious vs. bold investors
Test other numbers of clusters (K)	Improves how we define and understand risk
Incorporate credit trends over time	Adds insight from macroeconomic shifts or borrower patterns
Build a dynamic dashboard for Jasmin	Gives her an easy way to track performance and adjust investments