LendingClub

UPDATE - 3

JASMIN'S INVESTMENT PLAYBOOK: OPTIMIZING RISK TO MAXIMIZE RETURNS

Group 21:
Angela Pan
Poorvi Phulwani
Sai Priya Sure

OUR INVESTMENT ADVICE FOR YOU

Our Journey to Find the Best Strategy

Update 1:

We analyzed thousands of loans to understand risk patterns

- Identified key borrower characteristics that predict loan success
- Created a method to measure realistic returns on loans

Update 2:

We built prediction models and tested strategies

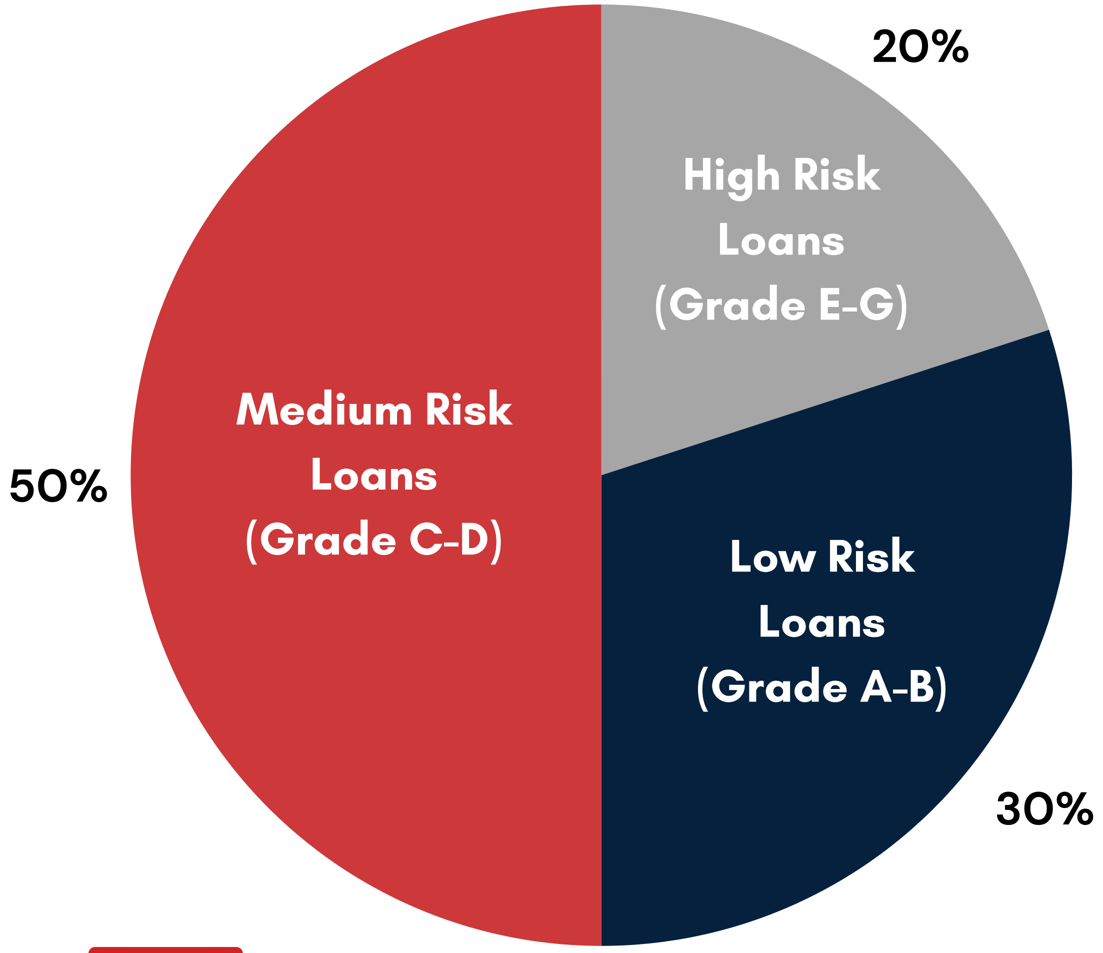
- Discovered interest rate is the strongest predictor of default
- Found temporal patterns provide 13.55% better returns
- Learned portfolios need at least 100 loans for stability

Update 3:

We created your optimal investment mix

- Used 7 borrower clusters to better understand risk levels
- Found the perfect balance across risk categories and created a portfolio that outperforms all previous strategies

Next Steps?
Start investing with our balanced portfolio!



The Winning Strategy
$1,000,000 across 100 loans
Expected Return: 4.42%

CLUSTERING SIMILAR LOANS HELPS ESTIMATE RISK BY GROUPING BORROWERS WITH SHARED FINANCIAL PROFILES

To support smarter investment decisions, we grouped borrowers into categories using K-Means Clustering based on shared financial characteristics.

What we did?

We used K-Means clustering on features available at the time of loan application, such as:

- Loan Amount
- Interest Rate
- Annual Income
- Debt-to-Income Ratio (DTI)
- Employment Length
- Home Ownership Status

They indicate how likely a borrower is to repay — essential for risk profiling.

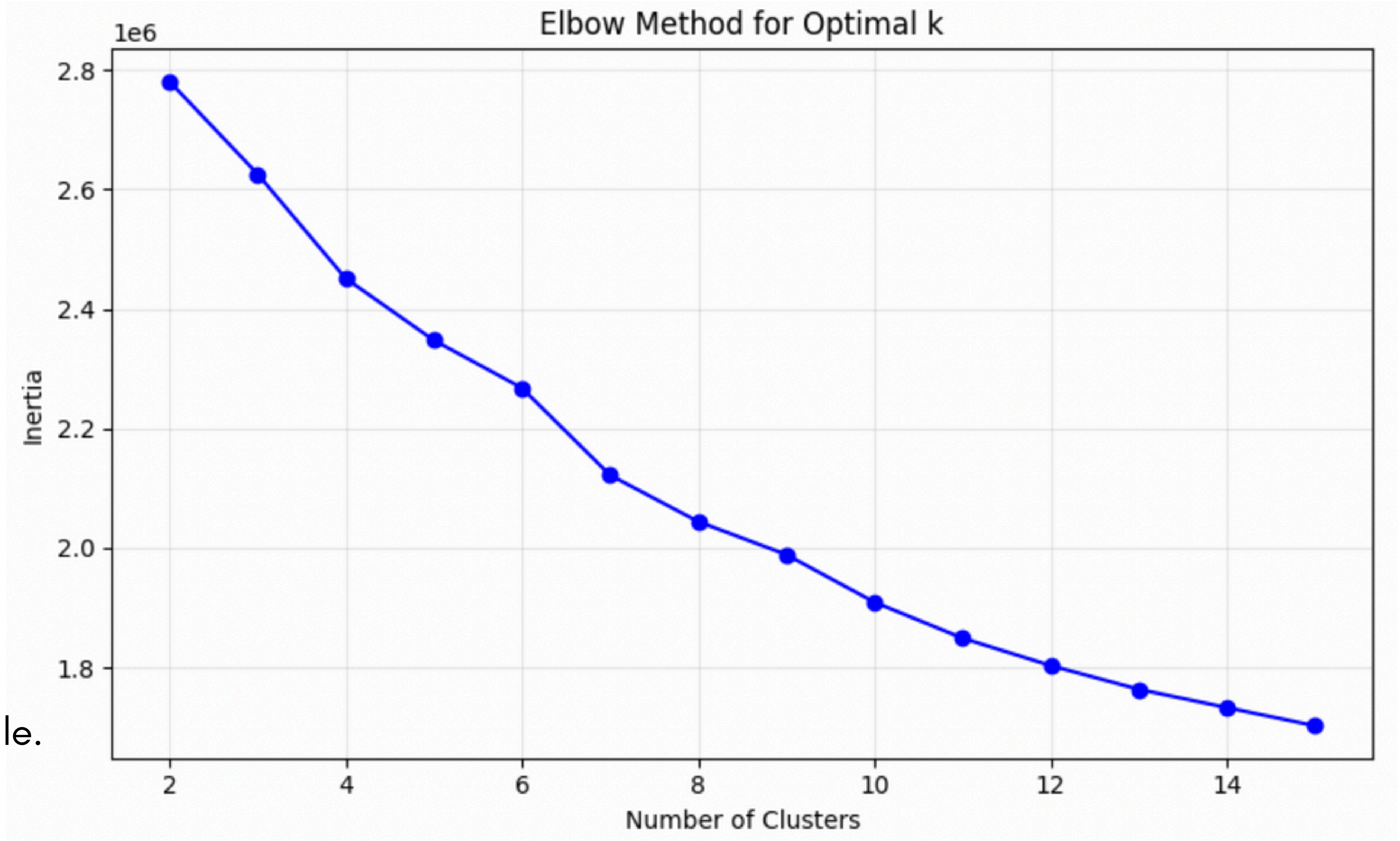
Why Clustering?

- Treats similar loans similarly — helps spot consistent trends in borrower behavior.
- Reduces noise by focusing on types of borrowers, not just isolated loans.
- The Elbow Method showed that 7 clusters gave the best fit — conveniently aligning with LendingClub's Grade A–G system, which you're already familiar with.

What Changed After Clustering

- Each loan now belongs to one of 7 borrower types with a shared risk profile.
- We can now assign risk levels (standard deviation of predicted return) to loans based on their cluster.
- It helps compare risk and return across loan options more clearly.

Identifying 7 Borrower Clusters Using the Elbow Method



Clustering gives you a smart, scalable way to assess borrower quality — like picking neighborhoods, not individual houses. It sets the foundation for a data-driven, risk-aware investment strategy.

ASSIGNING RISK USING RETURN VARIABILITY WITHIN EACH BORROWER GROUP REVEALS HIGH-RETURN, LOW-RISK OPPORTUNITIES

What we did?

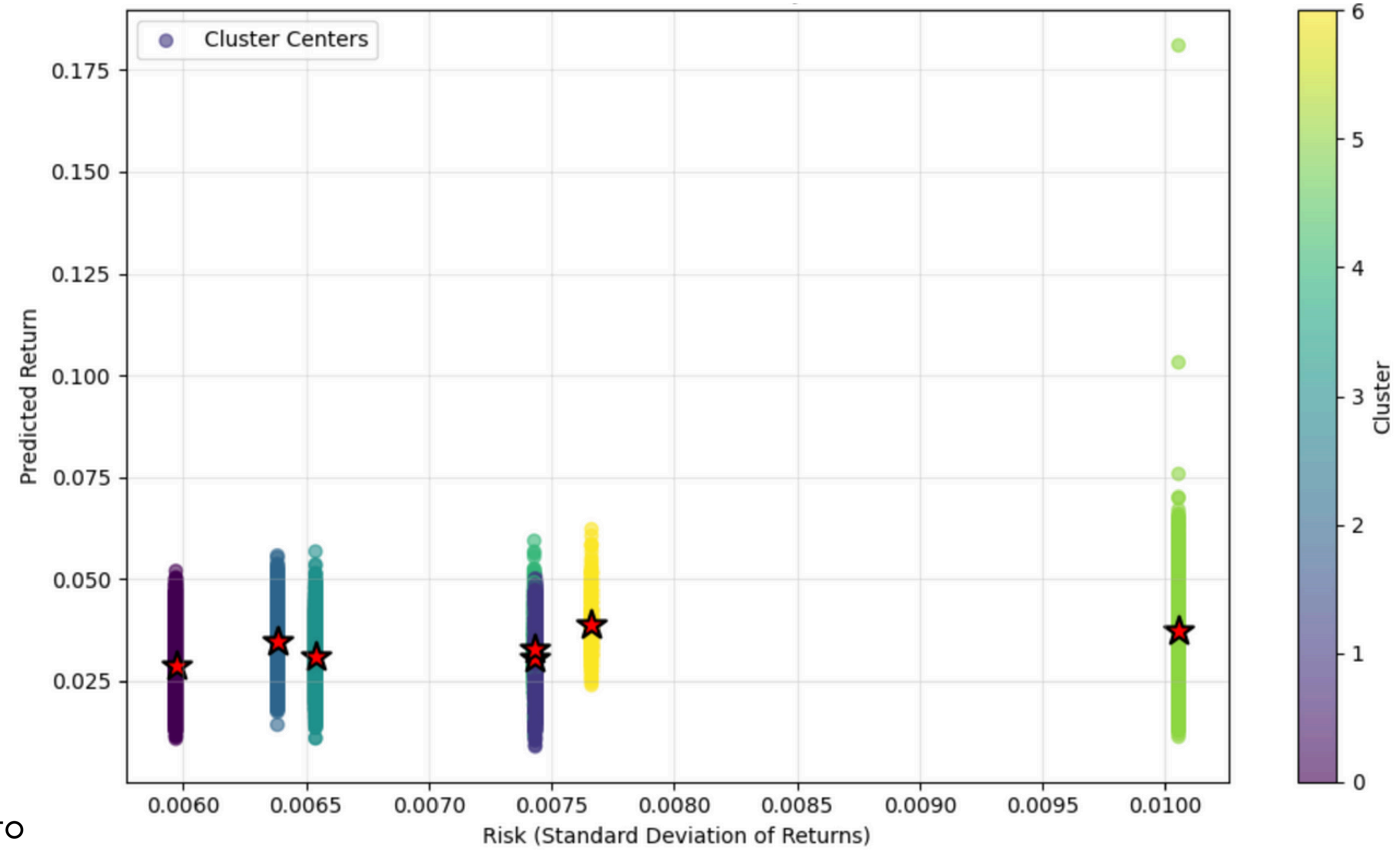
- We grouped borrowers into 7 clusters based on financial similarity.
- For each cluster, we calculated the standard deviation of predicted returns (using our Ridge regression model) to estimate risk.
- Each loan in the test set was assigned to the closest cluster based on borrower features.
- That cluster's return variability became the risk level for the loan.
- This created a structured framework to evaluate loans based on both predicted return and associated risk, instead of relying on one metric alone.

What we observed:

- **Cluster 2** had the highest Sharpe Ratio (return-to-risk), making it the most attractive for risk-adjusted returns.
- **Cluster 6** showed the highest predicted returns, ideal for risk-tolerant investors like you, if you're looking to maximize upside.
- **Cluster 0** had the lowest risk, best suited for conservative strategies.

Each cluster effectively represents a portfolio personality, helping you decide not just which loans to invest in, but what kind of investor you want to be. Helps balance risk and return by diversifying across borrower profiles.

Risk vs. Predicted Return by Cluster



Looking at return alone doesn't tell the full story. By understanding risk through predicted return variability, you can pick loans that align with your personal comfort level — whether you want stability, growth, or a balance of both.

20 CLUSTERS UNCOVERED DETAILED LOAN PATTERNS, BUT MANY WERE UNSTABLE OR TOO SMALL FOR PRACTICAL DECISIONS

What we did?

We tested a higher-resolution segmentation by grouping loans into 20 clusters using K-Means. For each cluster, we calculated:

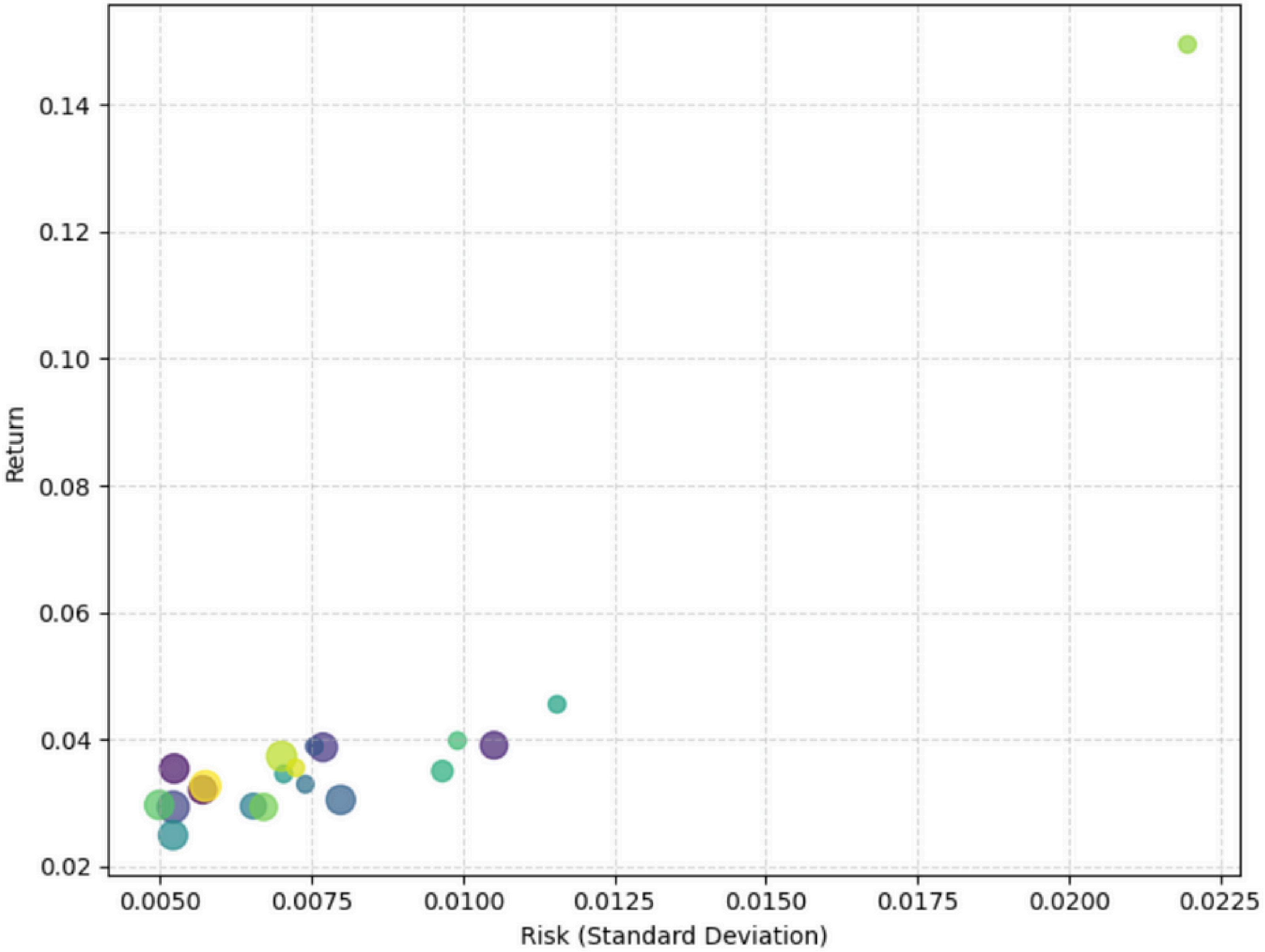
- Mean predicted return
- Risk as standard deviation of predicted return
- Sharpe ratio (return divided by risk)

What we observed from the model:

- Some clusters, like Cluster 1 and Cluster 16, showed excellent Sharpe ratios (6.7+), hinting at promising return-risk profiles.
- However, tiny cluster sizes (e.g., Cluster 16 = 3 loans) made those insights unreliable.
- Risk distribution was wide and scattered, adding complexity without better returns.
- LendingClub grades were scattered across clusters, reducing interpretability.

Pros of Using 20 Clusters	Cons of Using 20 Clusters
Captures subtle differences between borrowers	Many clusters too small to be investable
Reveals high Sharpe ratio clusters	Overfitting risk due to noise and fragmentation
May expose underpriced risk pockets	Harder to interpret, especially for decision-making

Risk vs. Return for 20 Clusters



20 clusters gave us detail — but too much of it. Some groups were promising but too small or unstable to trust. This complexity risks overwhelming investors rather than guiding them.

7 CLUSTERS SIMPLIFY INVESTMENT INSIGHTS WHILE PRESERVING KEY RETURN-RISK DISTINCTIONS

What we did?

We compared our 20-cluster model with a refined 7-cluster model, chosen using the Elbow Method and aligned with LendingClub's existing A-G grade system. For each, we analyzed:

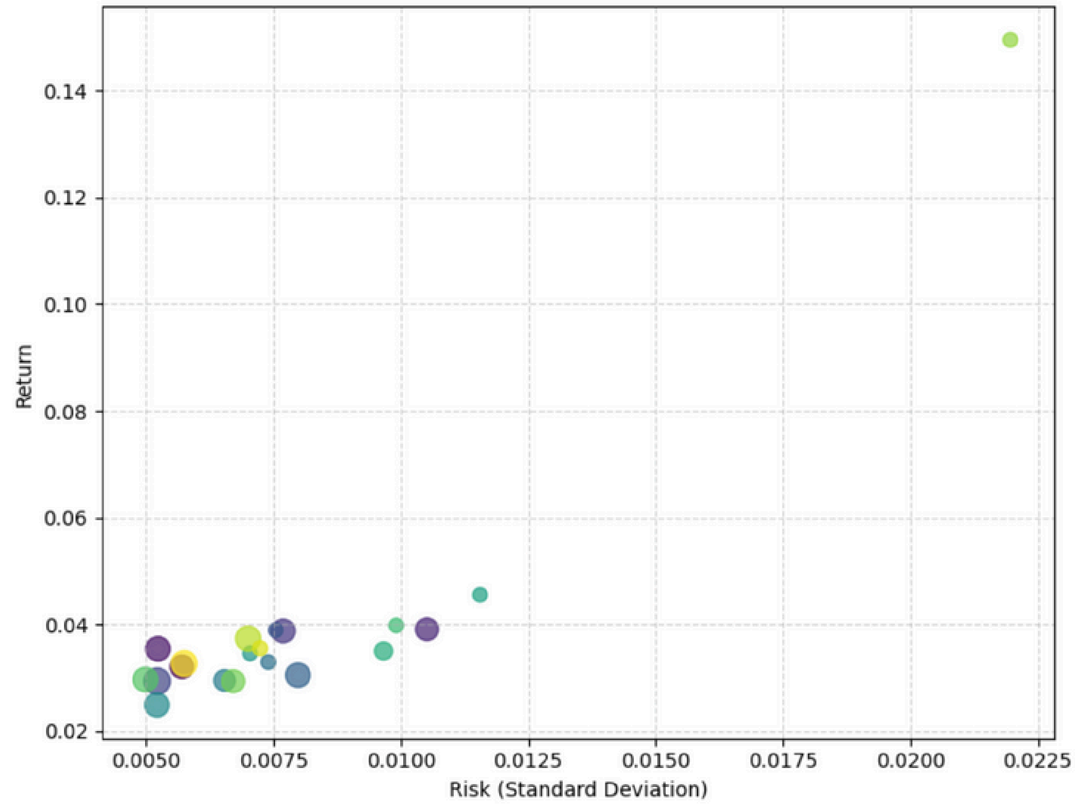
- Mean predicted return
- Risk (standard deviation of predicted return)
- Cluster size and separation

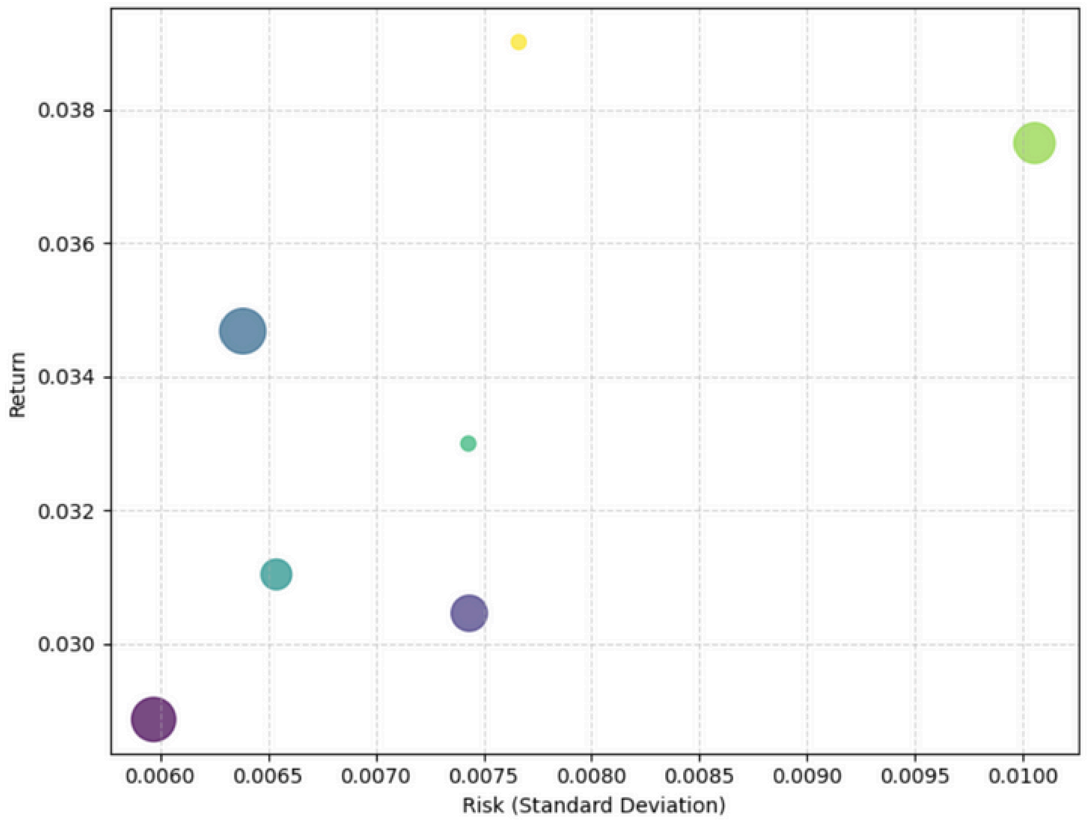
What we observed from the model:

- 7 clusters offered a clearer view of loan risk-return tradeoffs, with larger, more stable groups.
- The risk distribution in 7 clusters was tighter, improving consistency.
- The 7-cluster model still captured meaningful variation in Sharpe ratios — but with fewer outliers and simpler decision paths.
- Risk-return bubbles for 20 clusters were more noisy and included tiny, unreliable clusters.

Pros of Using 7 Clusters	Cons of Using 7 Clusters
Stable cluster sizes improve decision reliability	May overlook niche segments
Better alignment with LendingClub grades	Slightly lower peak Sharpe ratios in some clusters
Easier to interpret and apply	-

Risk vs. Return Comparison: 20 Clusters vs. 7 Clusters

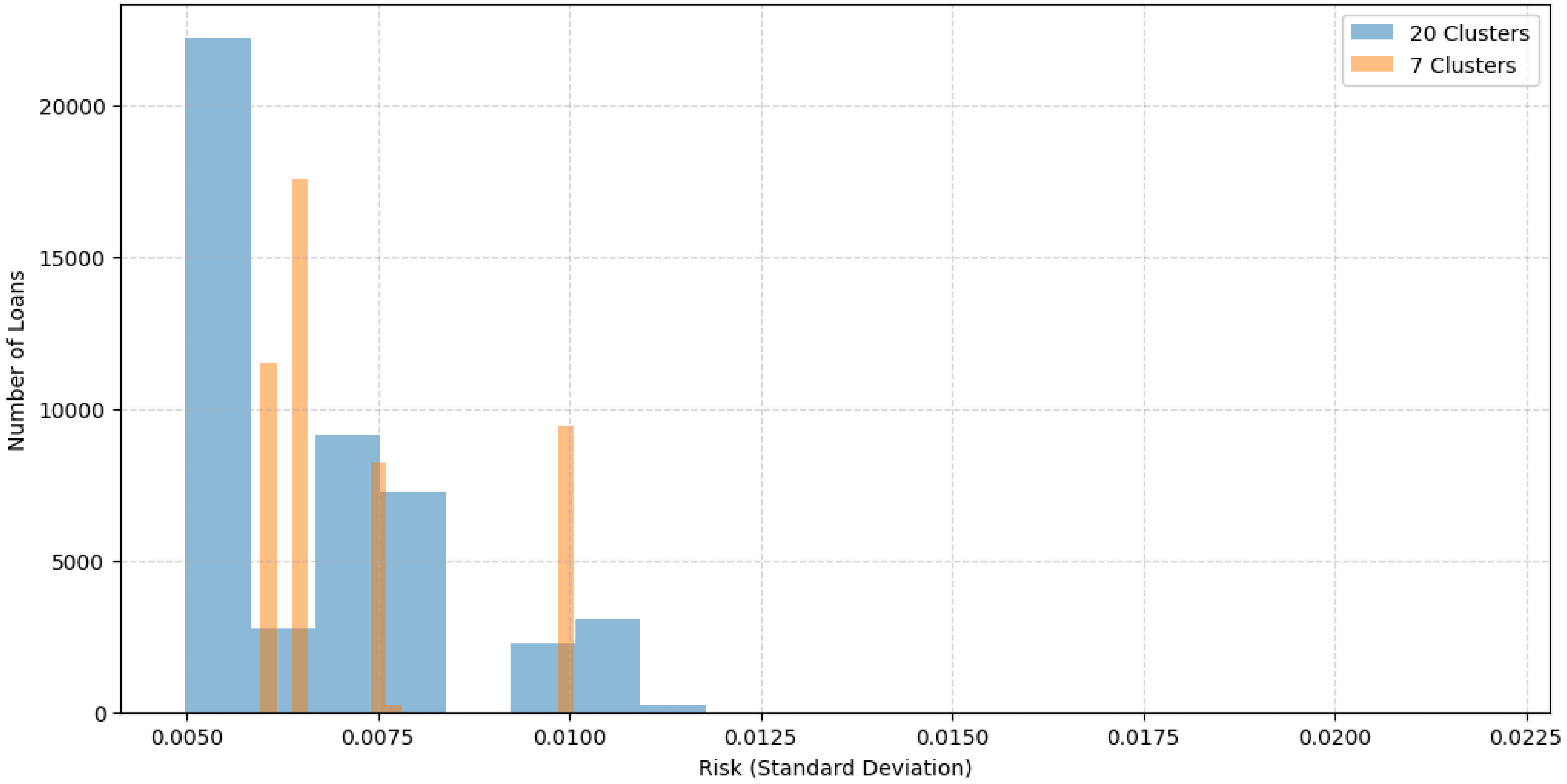




7 borrower types give you fewer but stronger signals — they're easier to track, understand, and trust. That means smarter, more consistent investment choices for you and your grandmother.

FEWER, CLEARER CLUSTERS IMPROVE RISK SEPARATION AND PREDICTABILITY

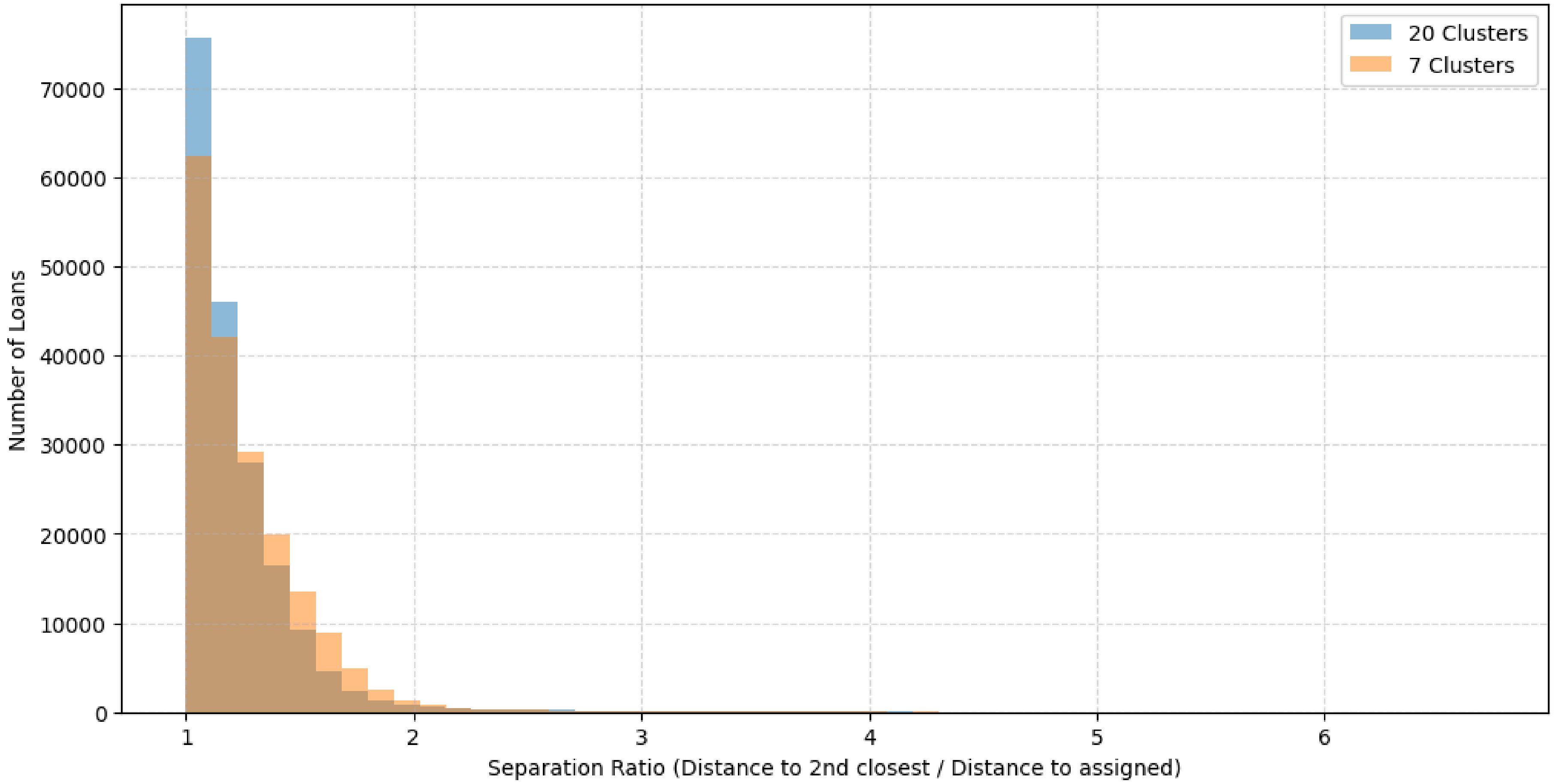
Distribution of Risk Measures



Risk Consistency:

- 7 clusters showed tighter, cleaner groupings of risk values, making risk scores more consistent and interpretable.

CLuster Seperation Quality



Stronger Cluster Boundaries:

- Loans in the 7-cluster model were more distinctly assigned — reducing overlap and ambiguity in risk tagging.

Clearer clusters = clearer investment decisions!

7 CLUSTERS OFFER STABLE SHARPE RATIOS AND STRONGER ALIGNMENT WITH LENDINGCLUB'S CREDIT GRADES

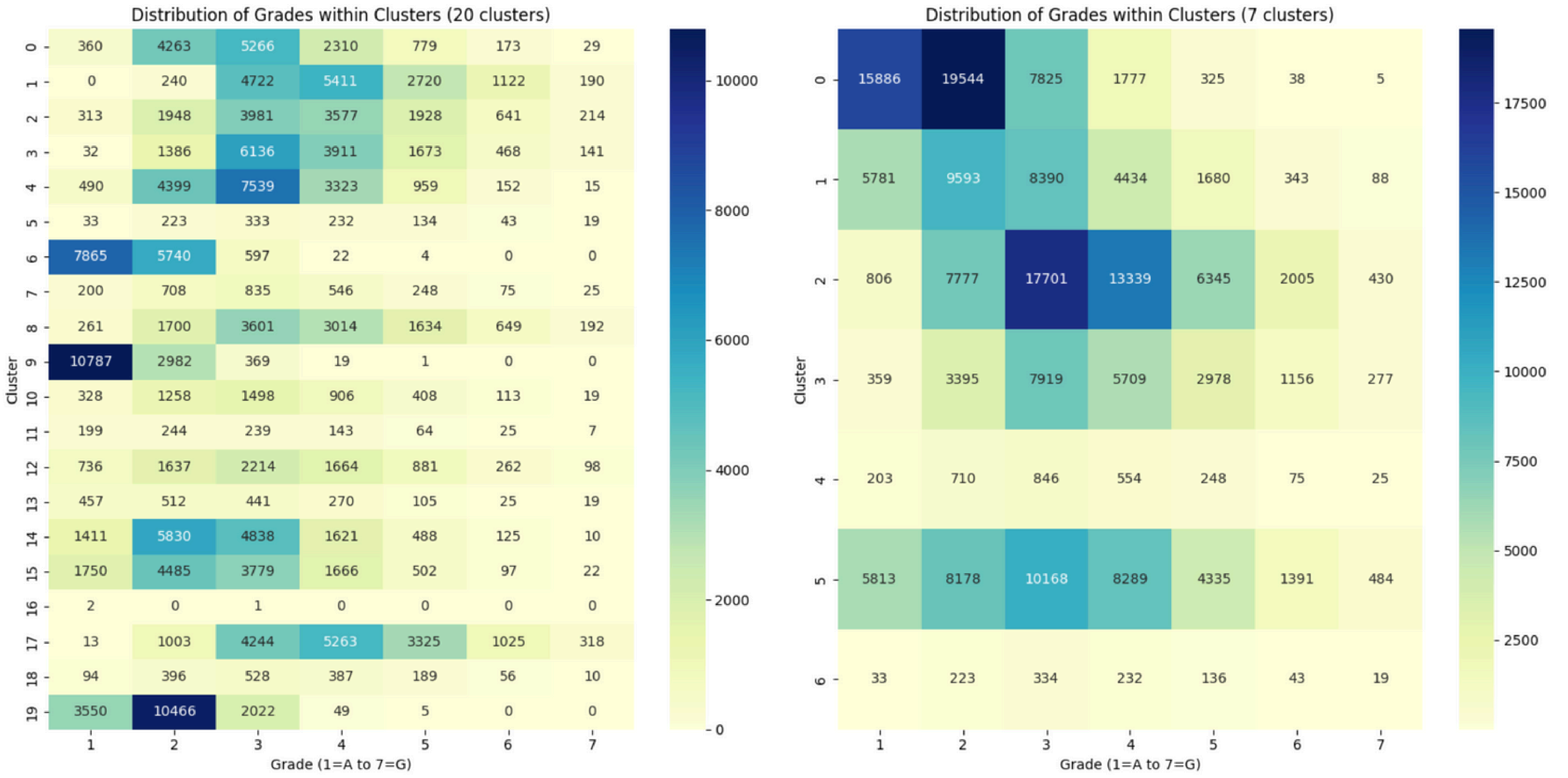
What we did?

- Compared the Sharpe ratios (return divided by risk) across clusters in both models.
- Analyzed the distribution of LendingClub credit grades (A-G) within clusters to evaluate interpretability.
- Assessed whether more clusters (20) led to better differentiation — or just more noise.

What we observed?

- 20 clusters had slightly higher Sharpe ratio peaks in some small clusters (e.g., Cluster 1 & 16), but their size and reliability were questionable.
- The 7-cluster model showed consistent Sharpe ratios across larger borrower groups — less volatile, more dependable.
- Grade heatmaps showed that 20 clusters fragmented credit grades, while 7 clusters aligned clearly with LendingClub's original loan classification system.
- Simpler segmentation also reduced model complexity — aiding future expansion or automation.

How Grades Align with Borrower Clusters



High Sharpe ratios are great — but only when you can act on them confidently. With 7 stable borrower types, you get dependable performance and a structure that matches how real-world lenders think.

SMARTLY BALANCING LOAN GRADES LEADS TO THE HIGHEST OVERALL RETURNS

Portfolio Overview

- Amount Invested: $1,000,000
- Total expected return: 4.42
- Average return per loan: 0.0442
- Balanced Grade Distribution
  - Grade A-B: 30%(low risk)
  - Grade C-D: 50% (medium risk)
  - Grade E-G: 20%(high risk)

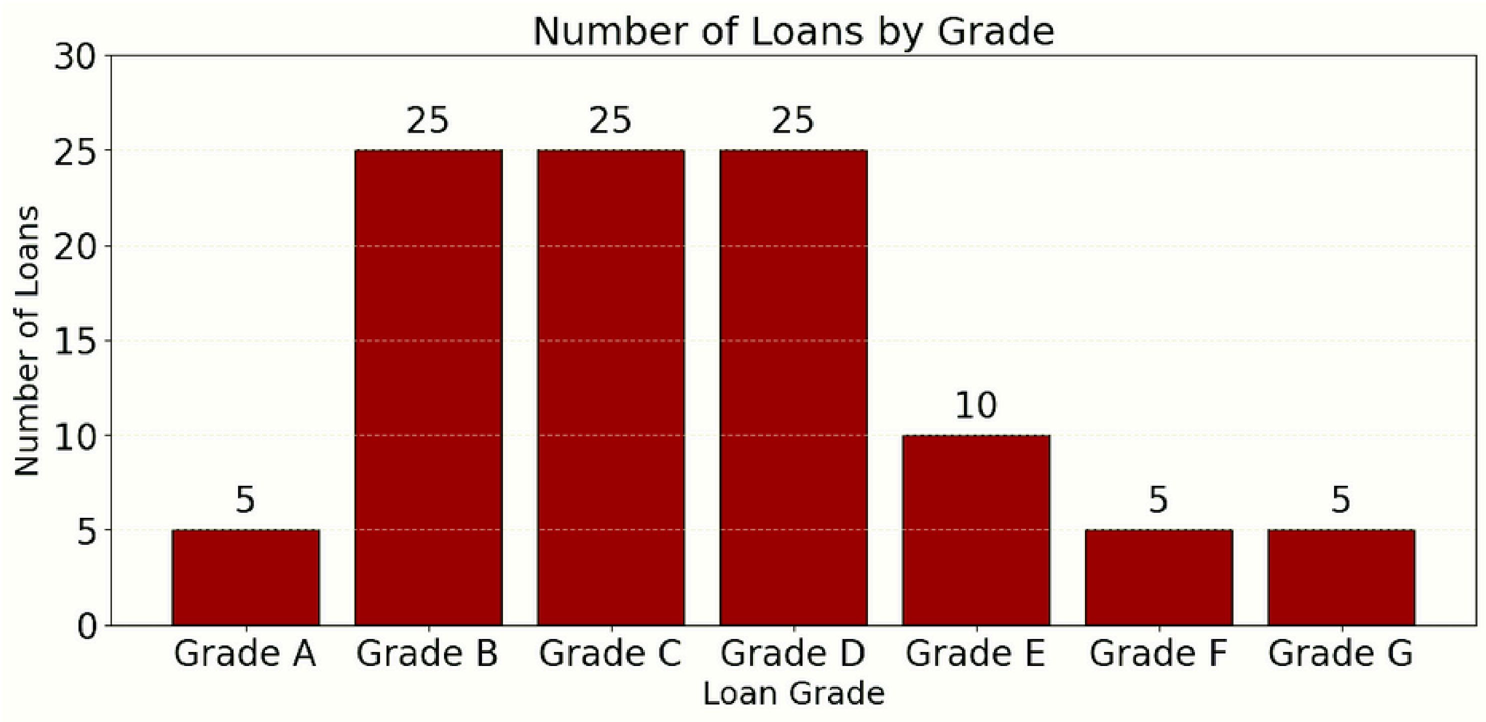
Return Comparison

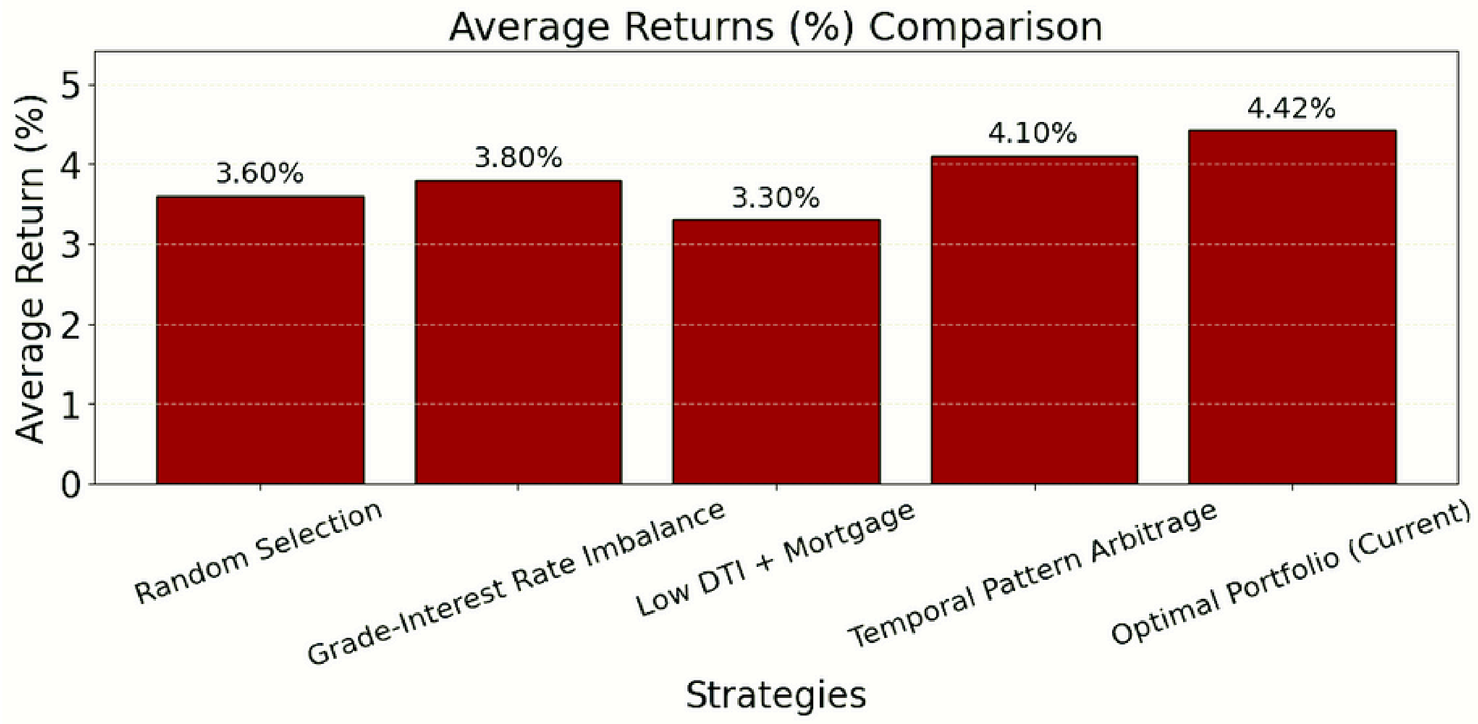
- The optimal portfolio's 4.42% return outperforms all the Week 4 strategies, including the best-performing Temporal Pattern Arbitrage strategy (4.1%).

Optimization Advantages

- Balanced Risk Diversification: The optimal portfolio includes loans across all grades (A through G) in a specific distribution, which suggests a more sophisticated risk diversification approach than the single-rule strategies.
- Multi-factor Optimization: Unlike the Week 4 strategies that focused on 1-2 variables, the optimization likely considers multiple factors simultaneously to find the ideal combination.
- This 100-loan portfolio size appears to be strategically chosen based on earlier findings

Loan Grade Distribution and Strategy Returns



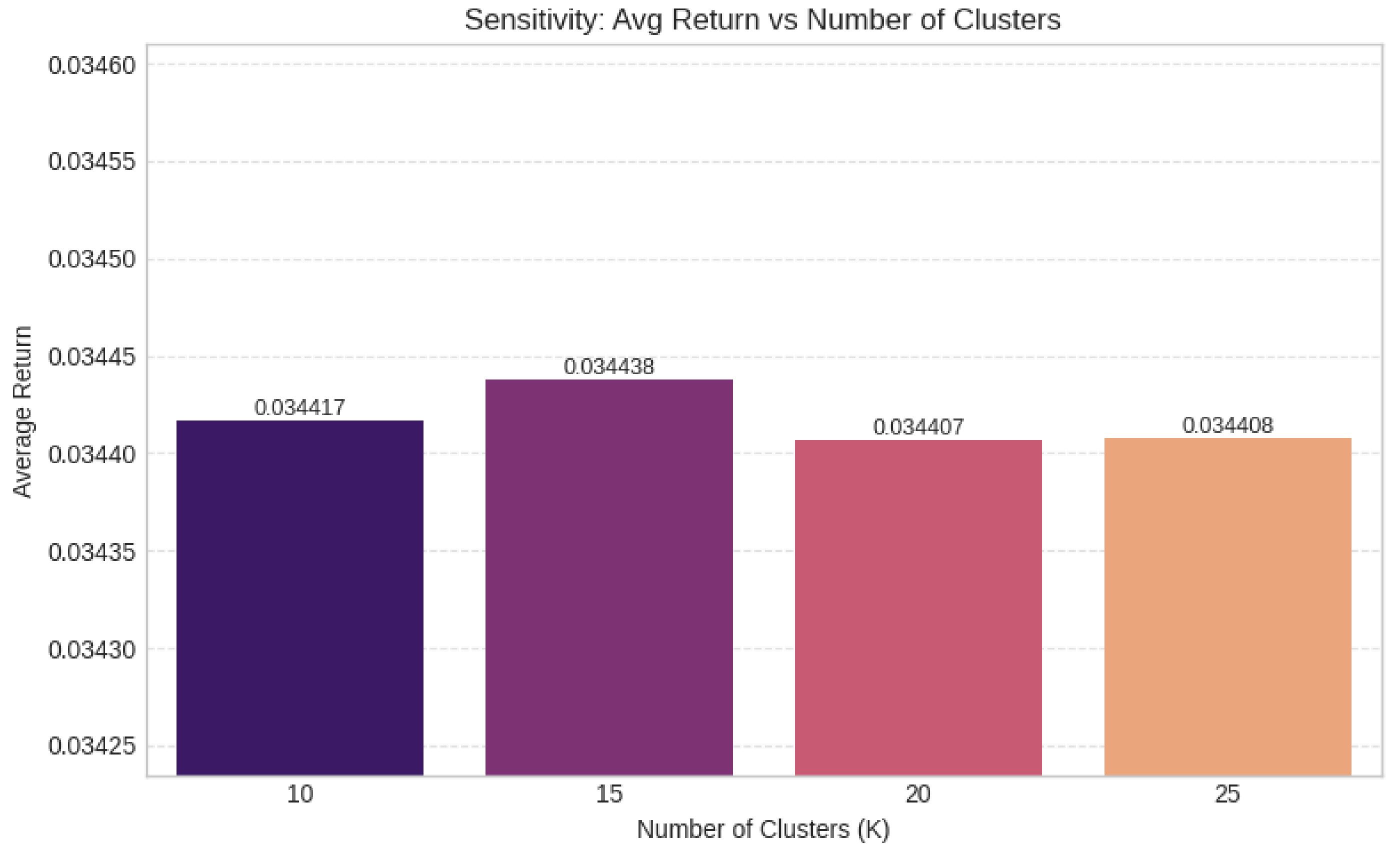


TESTING PORTFOLIO SIZE AND RISK MODELS CONFIRMS 100 LOANS AS THE OPTIMAL STRATEGY

Average Return vs Number of loans by budget

Our analysis grouped loans by risk:

- Group A: Low risk, lower return
- Group B: Medium risk, highest return
- Groups C & D: Higher risk



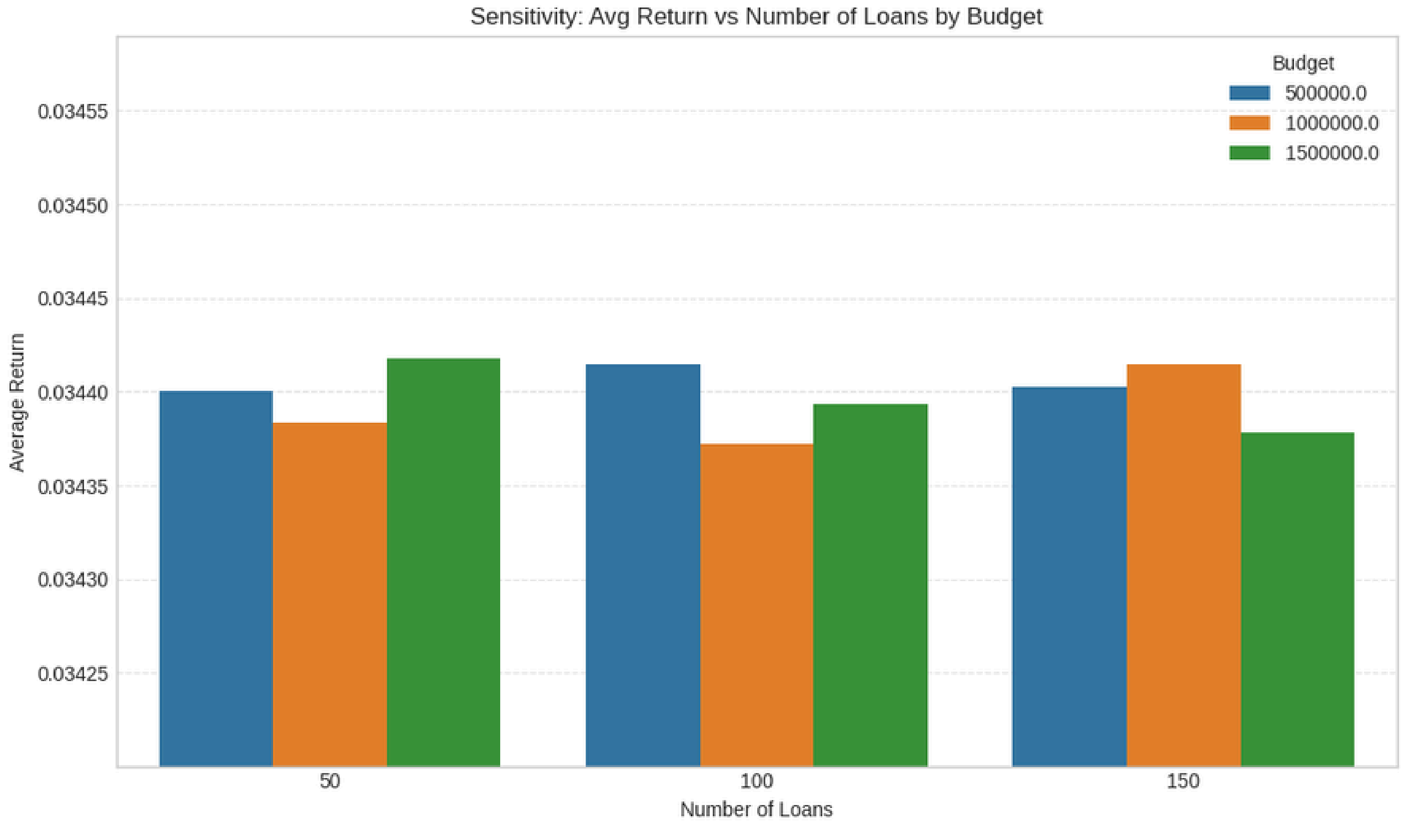
Average Return vs Number of Clusters

Small portfolios are risky:

- More variation in returns
- Higher impact if a loan defaults

100+ loans provides stability:

- More consistent returns
- Better protection against losses



What This Means For Your Investment:

- **Sweet spot: 100 loans provides optimal balance of return and diversification**
- **Medium-risk grouping (Group B) yields the highest returns at 3.44%**

WHEN THE MODEL BREAKS: LIMITATIONS OF OUR STRATEGY

Potential Challenges

- **Economic Downturns:** During recessions, default patterns may change dramatically across all clusters
- **Interest Rate Fluctuations:** Our models were built on historical data with specific interest rate environments
- **Regulatory Changes:** Lending rules can shift, affecting borrower behavior and loan availability
- **Small Sample Limitations:** Some clusters may have insufficient loan data for reliable predictions

When Jasmin Should Use Her Judgment

- **New Loan Types:** When LendingClub introduces loan products not in our training data
- **Unusual Borrower Profiles:** When loans don't fit cleanly into our 7 cluster framework
- **Market Disruptions:** During financial crises or unprecedented events
- **Concentration Risk:** If available loans become too concentrated in particular clusters

Future Enhancements

- Regular model retraining with new data
- Adaptive risk thresholds that respond to market conditions
- Additional borrower attributes for more precise clustering
- Dynamic portfolio rebalancing based on performance feedback

The 4.42% return is based on historical patterns that may not persist!

LendingClub

THANK YOU