

# UPDATE - 1

# DESCRIPTIVE AND DIAGNOSTIC ANALYSIS



# INTRODUCTION AND PROBLEM STATEMENT



## GOAL:

Help Jasmin maximize returns from peer-to-peer lending while managing risk through data-driven investment strategies.

## KEY DECISIONS:

- **Strategy Type:** Random, default-based, return-based, or combined
- **Return Metric:** Optimistic, intermediate, or pessimistic
- **Diversification:** Across loan grades and terms
- **Loan Evaluation:** Use borrower and loan characteristics to predict performance
- **Validation:** Assess if LendingClub grades align with actual risk/return

## SUCCESS METRICS (KPIs):

- |                                |                 |                |
|--------------------------------|-----------------|----------------|
| 1 Portfolio ROI                | 2 Default Rate  | 3 Sharpe Ratio |
| 4 Grade & Term Diversification | 5 Recovery Rate |                |

# DATA ACQUISITION AND PREPARATION



## SELECTION CRITERIA:

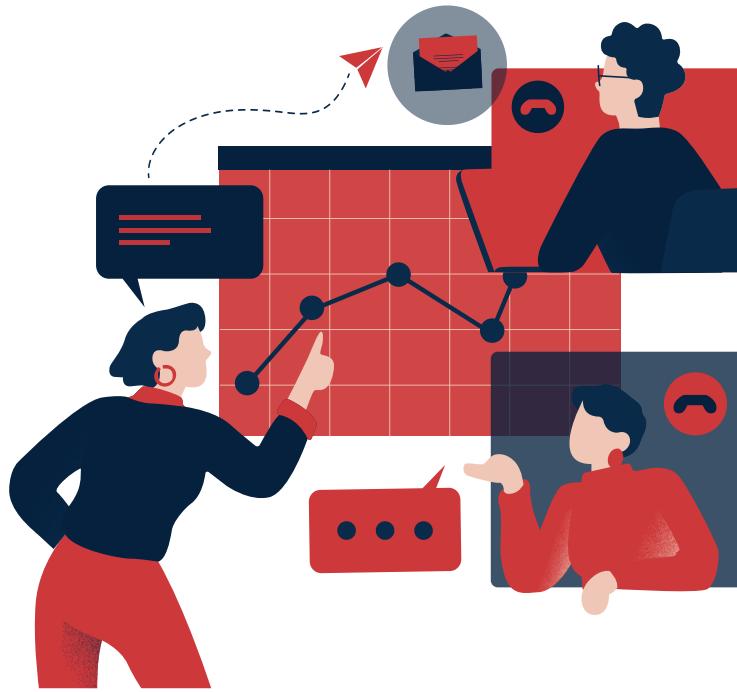
- Available at loan application time
- Relevant for default risk prediction
- Necessary for return calculation
- Useful for borrower segmentation/portfolio design

## SELECTION CRITERIA:

VARIABLE	JUSTIFICATION
inq_last_6mths	Indicates recent credit activity; flags potentially overleveraged borrowers
collections_12_mths_ex_med	Captures recent non-medical collections; strong signal for credit risk
acc_now_delinq	Current delinquencies; direct input for default risk modeling
total_rec_prncp	Required for actual return (principal repaid); part of ROI computation
total_rec_int	Captures total interest repaid; essential for return estimation

VARIABLE	JUSTIFICATION
total_rec_late_fee	Completes return picture for delayed payments; contributes to net ROI
application_type	Distinguishes between individual and joint applicants; useful for segmentation
total_acc	Total number of credit lines; proxy for borrower credit experience
last_credit_pull_d	Reflects recency of credit review; proxy for updated credit behavior
installment	Reflects monthly loan repayment amount; important for assessing borrower burden



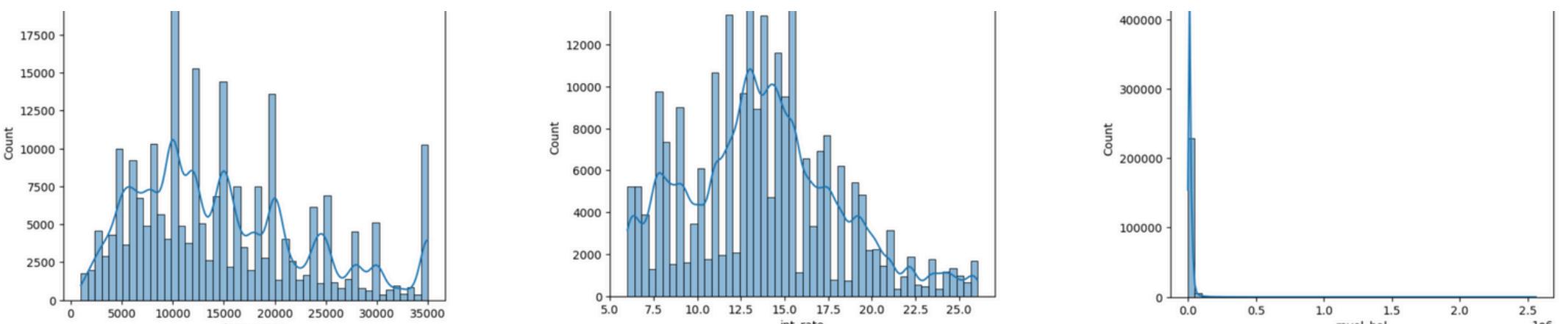


# DESCRIPTIVE STATISTICS AND FEATURE EXPLORATION

## UNIVARIATE DISTRIBUTIONS :

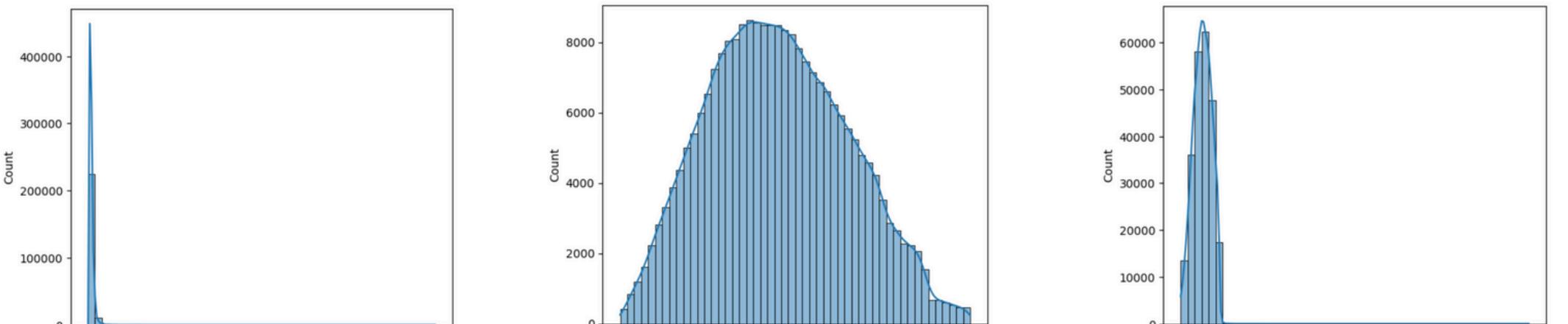
### LOAN\_AMNT AND INT\_RATE:

They show expected variation — most loans range from \$5K-\$15K, and interest rates are concentrated around 10-15%.



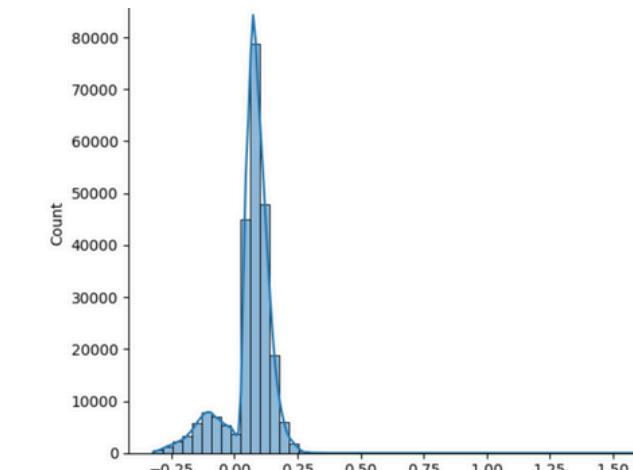
### ANNUAL\_INC AND REVOL\_BAL:

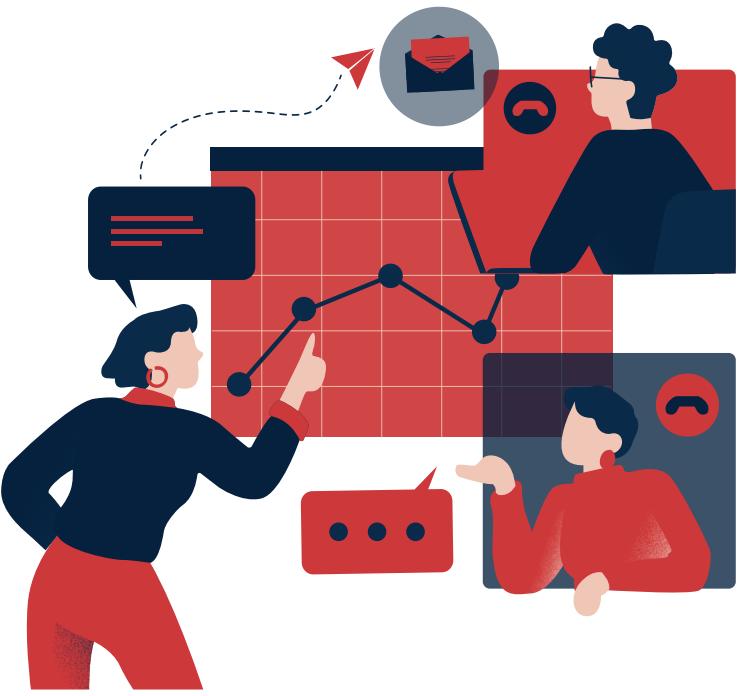
They are right-skewed with extreme outliers — visible again in the boxplots (mentioned in the code).



### DTI:

They appears to follow a near-normal distribution.





# DESCRIPTIVE STATISTICS AND FEATURE EXPLORATION

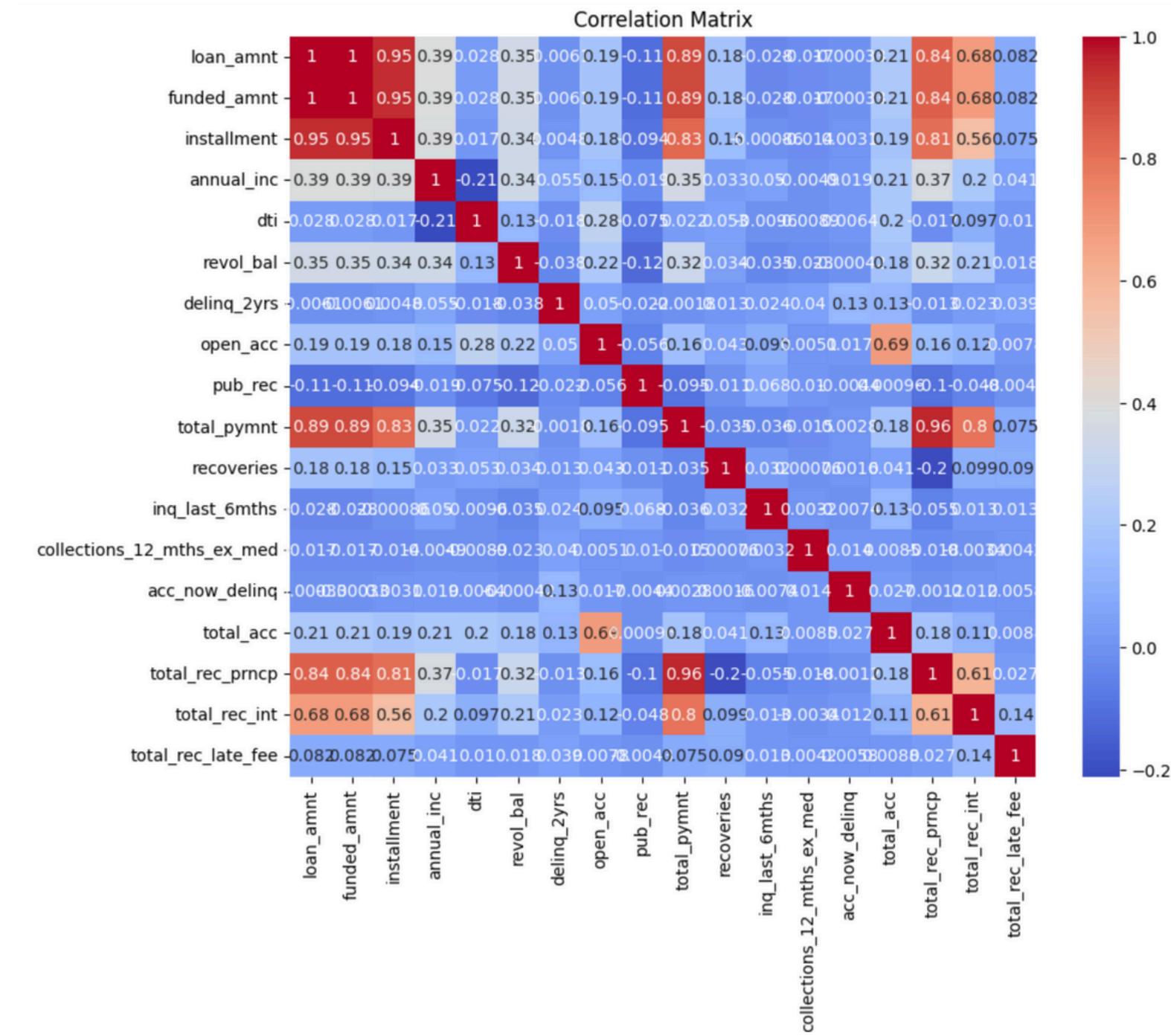
## CORRELATION MATRIX HIGHLIGHTS:

### HIGH CORRELATIONS:

- loan\_amnt, funded\_amnt, installment (expected: structurally linked)
- total\_rec\_prncp with total\_pymnt and total\_rec\_int (return components)

### LOW CORRELATIONS:

Default-related variables (inq\_last\_6mths, acc\_now\_delinq, collections\_12\_mths\_ex\_med) and return metrics — supports their inclusion as independent risk signals.



# HYPOTHESIS FORMATION AND FEATURE IMPORTANCE



## HIGHER ROI

Higher loan amounts and interest rates

## HIGHER DEFAULT RATE

Higher interest rates + lower income

## HIGHER SHARPE RATIO

Lower default rates + more consistent returns

## HIGHER GRADES

Moderate interest rates + low default rates

## SHORT-TERM LOANS

Lower default rate but lower return

## HIGHER RECOVERY RATES

Lower net losses + higher ROI

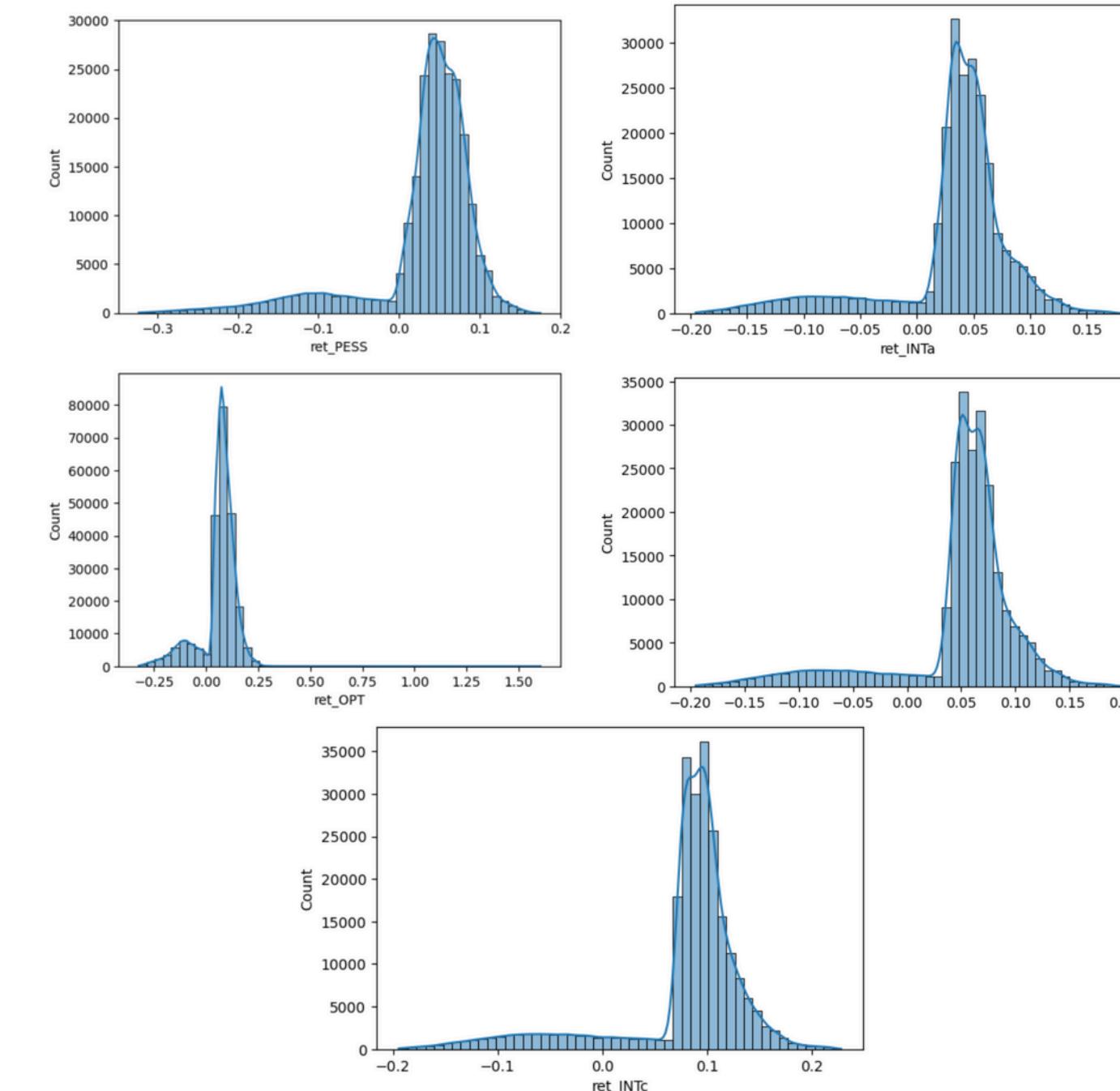


## PURPOSE:

Estimate investor return under different assumptions to support strategy development.

Metric	Type	Assumption	Observation
ret_OPT	Optimistic	Full repayment with immediate reinvestment	Right-skewed; most loans ~5-10%, some >100%
ret_PESS	Pessimistic	Full term held, no reinvestment	More stable; captures losses from defaults
ret_INTa/b/c	Intermediate	Reinvestment at fixed monthly rates (0.1%, 0.25%, 0.5%)	Higher rate → higher and tighter

# DERIVED LOAN RETURN FEATURES





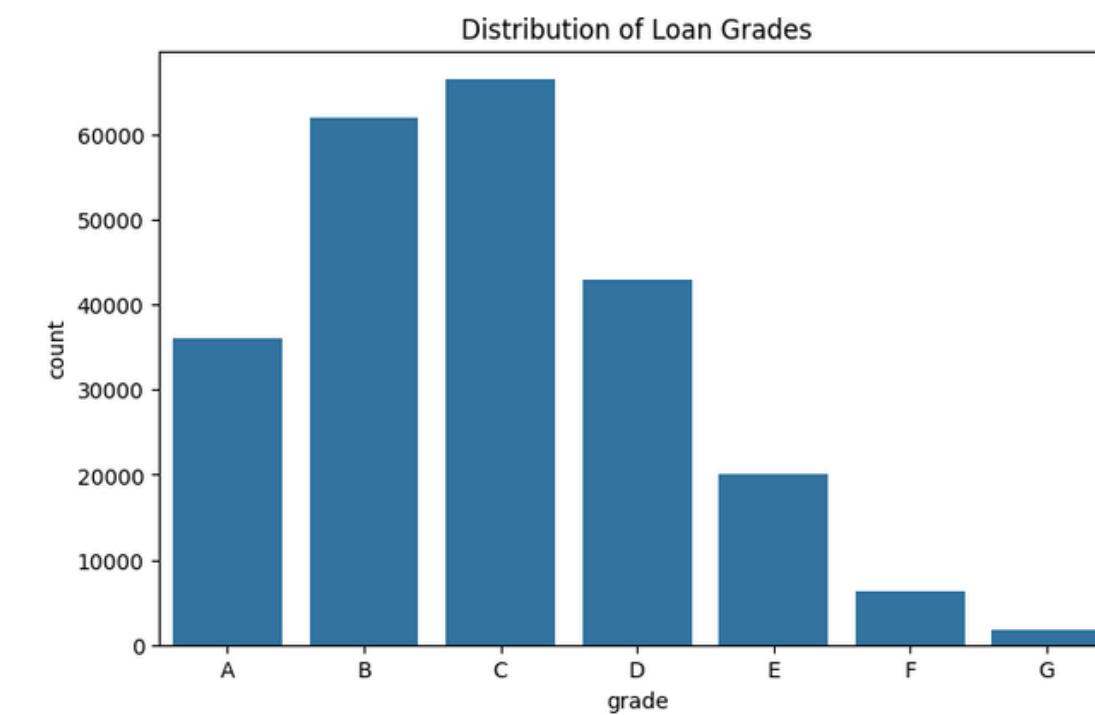
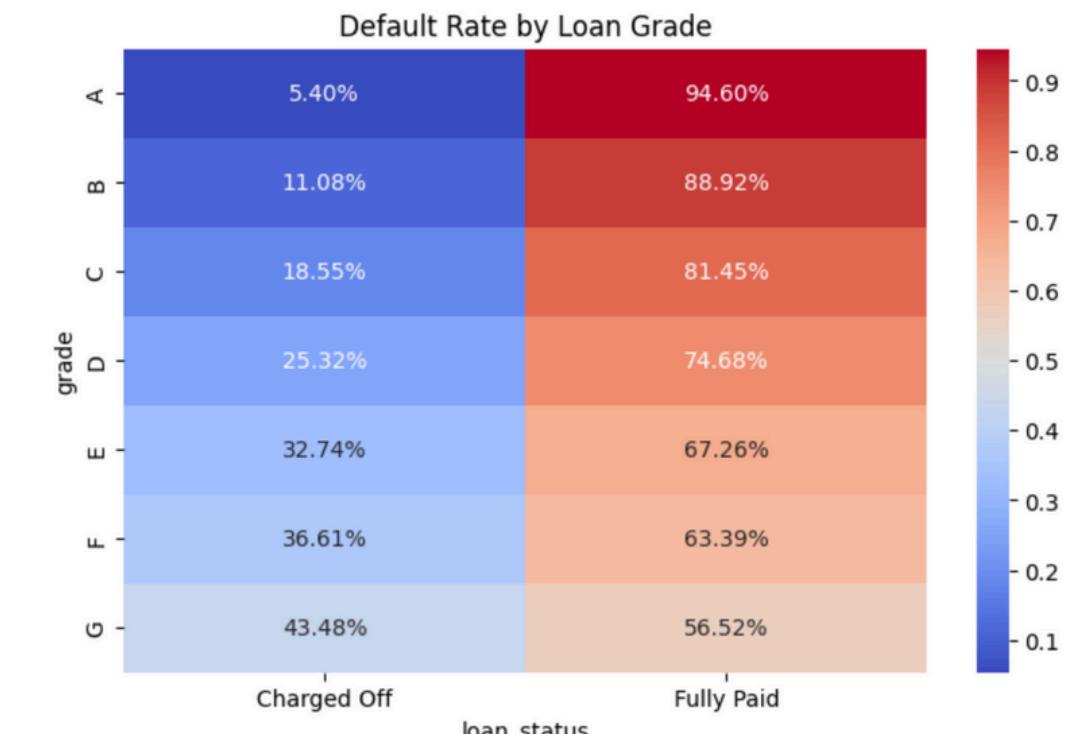
# INTERPRETING LENDINGCLUB LOAN GRADES

## Grade Distribution & Default Trends

1. Most loans fall under Grades B and C
2. Default Rate ranges from:
  - a. ~5.4% (Grade A) → ~43.5% (Grade G)

## Return Patterns by Grade

- Grades A-C: Lower, stable returns with fewer outliers
- Grades D-G: Higher average returns, but with greater volatility and losses





# K-MEANS CLUSTERING TO RECREATE LOAN GRADES

## GOAL:

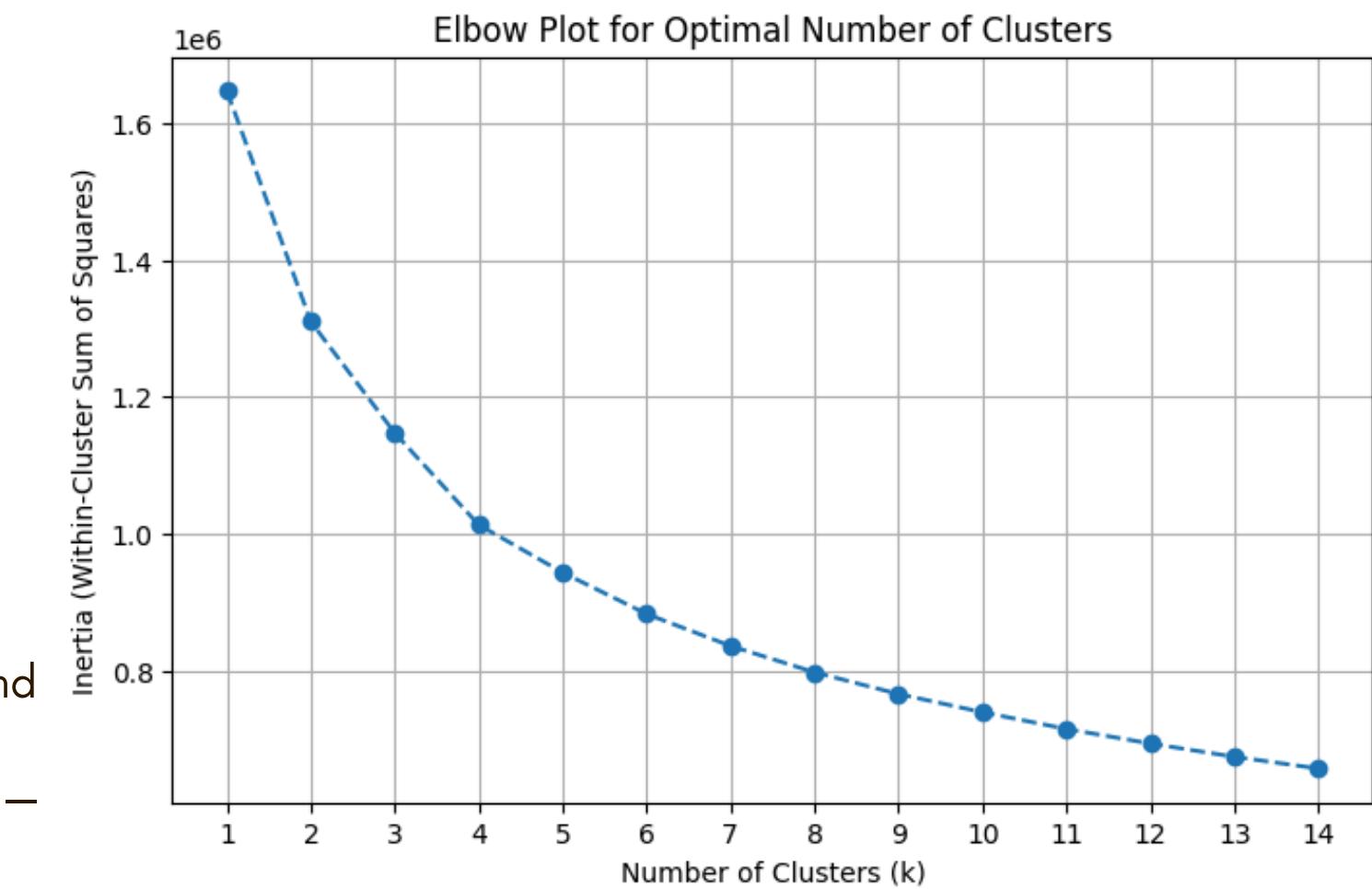
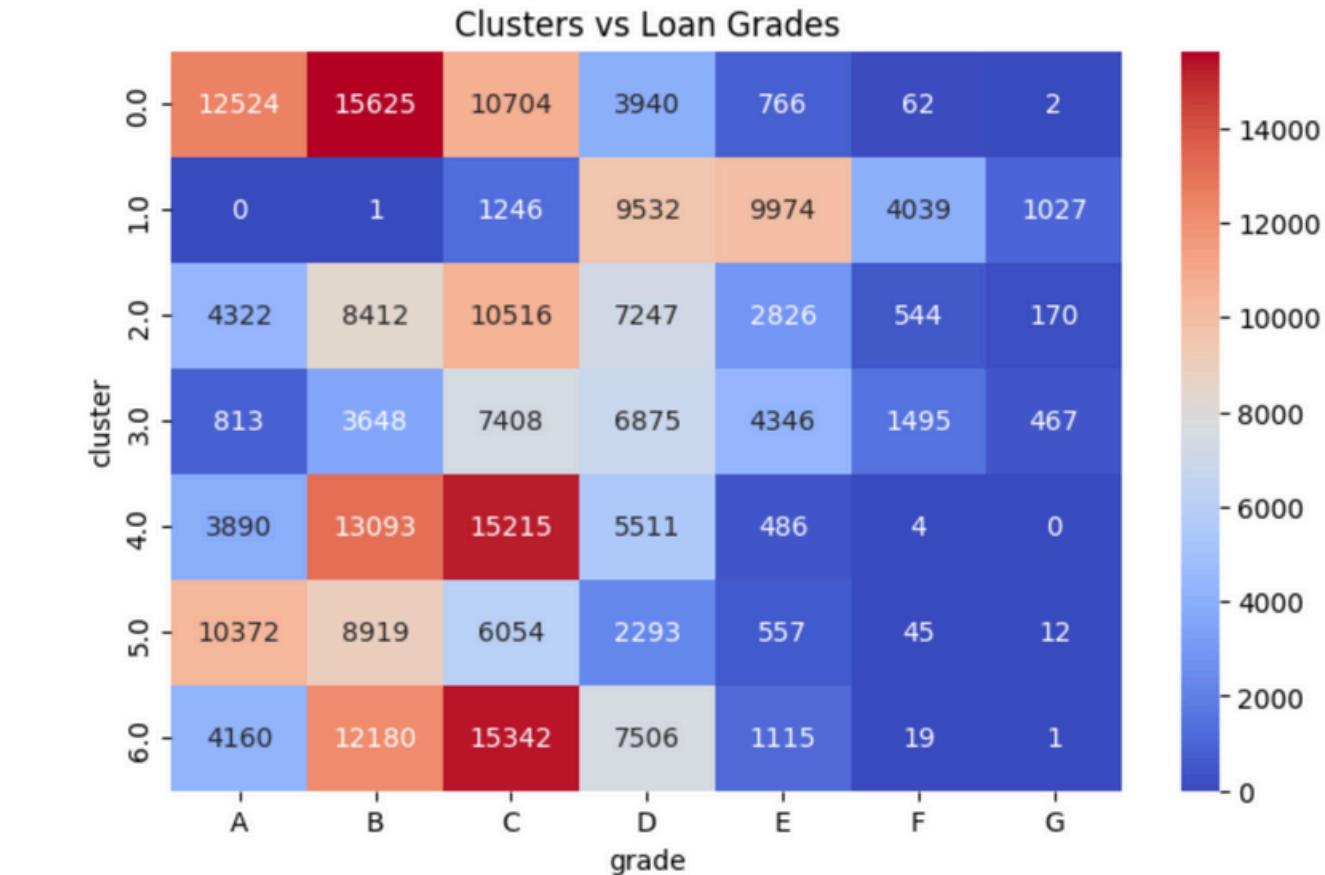
To replicate LendingClub's loan grade segmentation (Grades A-G) using unsupervised clustering.

## METHODODOLOGY:

- Applied K-Means clustering with K = 7 to align with the 7-grade system
- Selected key numerical variables: Interest Rate, Loan Amount, Debt-to-Income Ratio, Revolving Utilization, etc.
- Used the Elbow Method to validate K=7 as the optimal balance between model fit and interpretability
- Conducted cross-tabulation of clusters vs. grades to assess alignment

## INSIGHTS:

- Elbow plot confirms K=7 as a strong balance between cluster compactness and interpretability
- The Heatmap reveals that several clusters strongly align with LendingClub grades – especially for high-risk and low-risk extremes
- Clustering reveals distinct borrower segments by financial risk profile



# INTERPRETING LOAN CLUSTERS

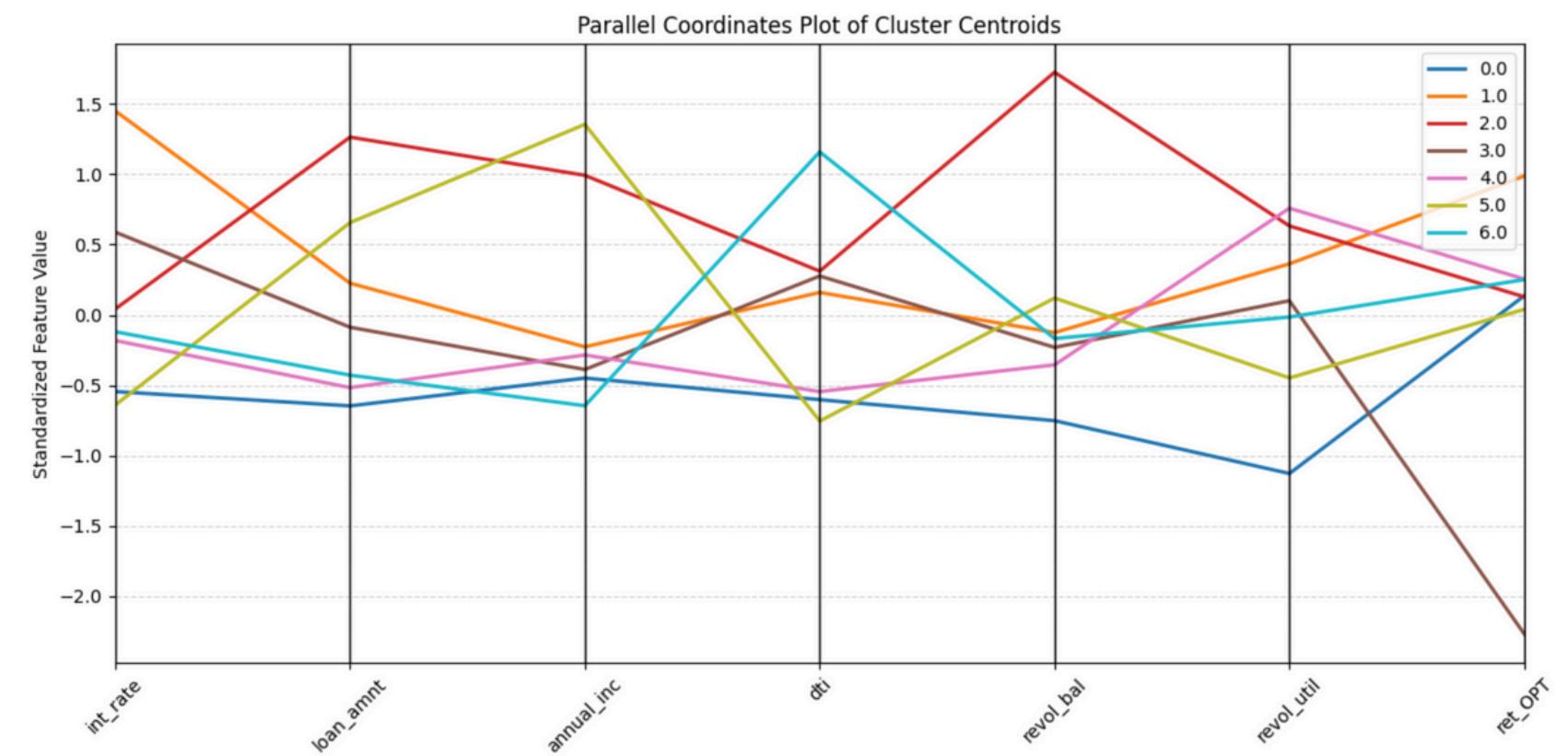
**GOAL:** To describe and name clusters based on key characteristics.

## METHODOLOGY:

- Analyzed cluster means for financial patterns.
- Created parallel coordinates plot to visualize cluster behavior.
- Mapped clusters to LendingClub grades based on patterns.

## INSIGHTS:

VARIABLE	JUSTIFICATION
Cluster 0	High-income, low risk borrowers (Grade A)
Cluster 1	High-risk, high return borrowers (Grade D)
Cluster 2	Financially stressed borrowers (Grade E)
Cluster 3	Premium borrowers (Grade B)
Cluster 4	Balanced borrowers with stable repayment (Grade C)
Cluster 5	High income outliers with mixed repayment (Grade F)
Cluster 6	Volatile, High Variance borrowers (Grade G)



# PRINCIPLE COMPONENT ANALYSIS

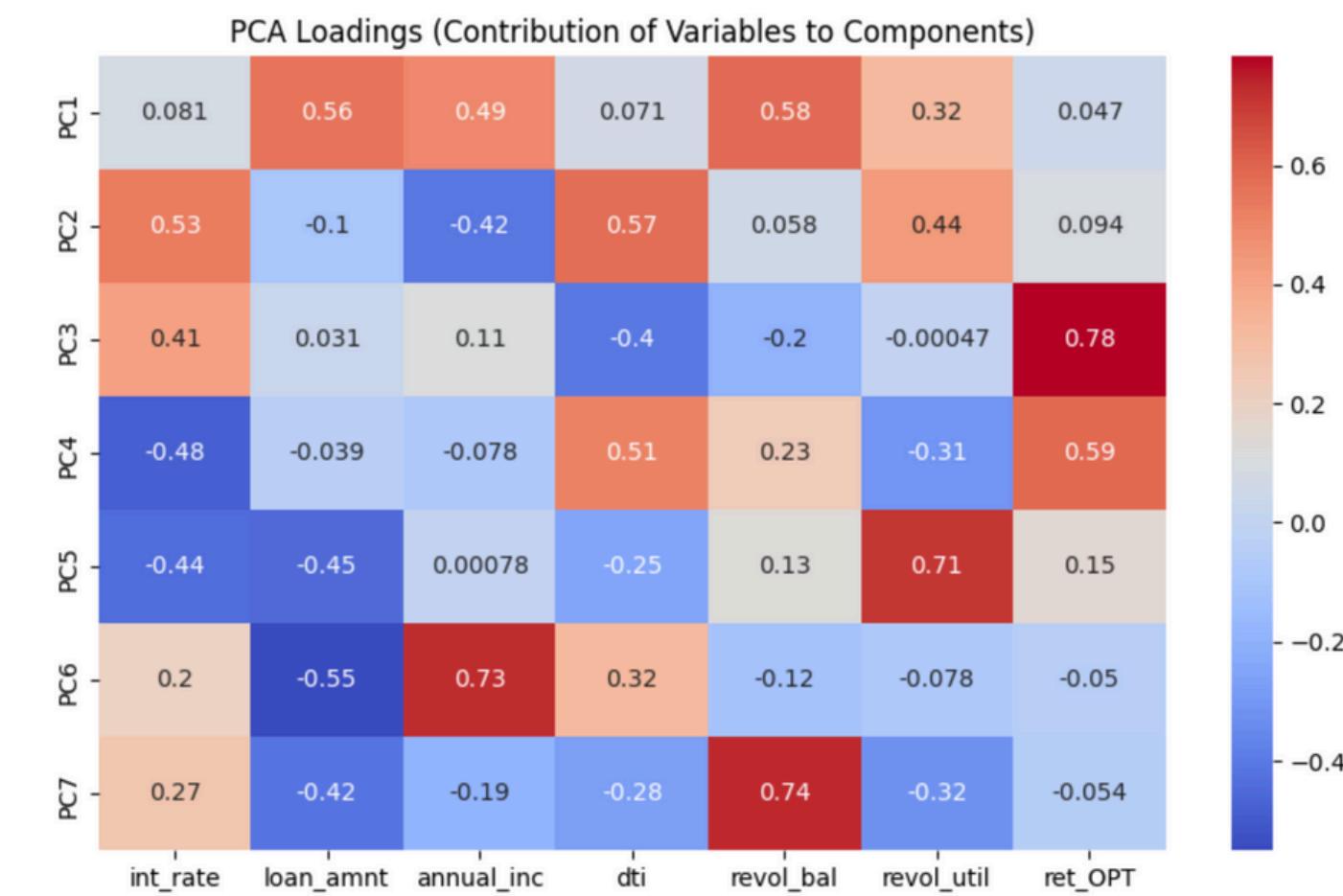
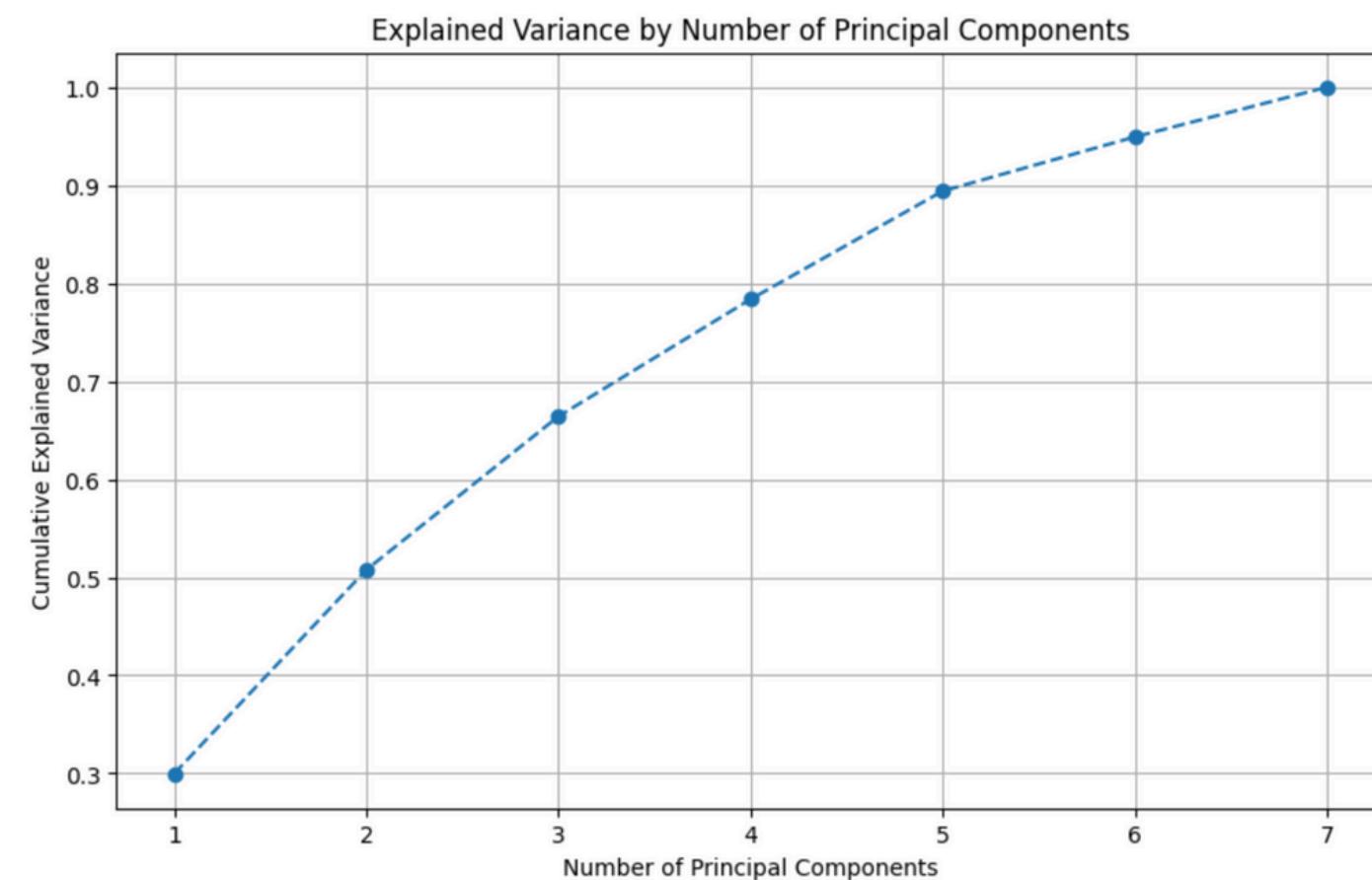
**GOAL:** To identify the key drivers of loan performance.

## METHODOLOGY:

- Performed PCA on same set of cluster variables.
- Explained variance plot to determine number of components.
- Analyzed PCA loadings to identify key financial drivers.

## INSIGHTS:

- **PC1** – Financial Scale: Loan amount, income, revol\_bal → Borrower capacity & loan size
- **PC2** – Risk Profile: Interest rate, DTI, revol\_util → Borrower leverage & credit risk
- **PC3** – Return Signal: ret\_OPT → Investor return potential, independent of risk
- **PC4** – Leverage & Efficiency: High DTI + returns, low rate/util → Efficient but leveraged borrowers
- **PC5 & PC7** – Credit Usage: revol\_bal, revol\_util → Credit usage intensity & behavior





# THANK YOU

