

UPDATE 1



LendingClub Case

Yang Yue, Ted Wang, Goyo Lozano Palacio



Making a good financial decision by creating a **data-driven investment strategy**



More Return
Less Risk



Why Peer2Peer Lending?

- High returns = 6-23%
- Low defaults = 1-10%
- Vast, public data for ML

Jasmin Gonzales

Creating a **step-by-step detailed framework** to guide Jasmin's decisions

Decisions	Good vs Bad Investment	Proposed KPIs
<ul style="list-style-type: none">• Loan type• Capital allocation per loan• What borrower features matter most• Portfolio Construction	<ul style="list-style-type: none">• Higher Net Returns with lower risk• Better than baseline strategy performance• Track realized loan outcomes on test data	<ul style="list-style-type: none">• IRR• Sharpe ratio (excess return per risk unit)• Default rate• Herfindahl-Hirschman Index (portfolio concentration risk)

Choosing initial personal and **loan features** to help understand **drivers of risk and returns**

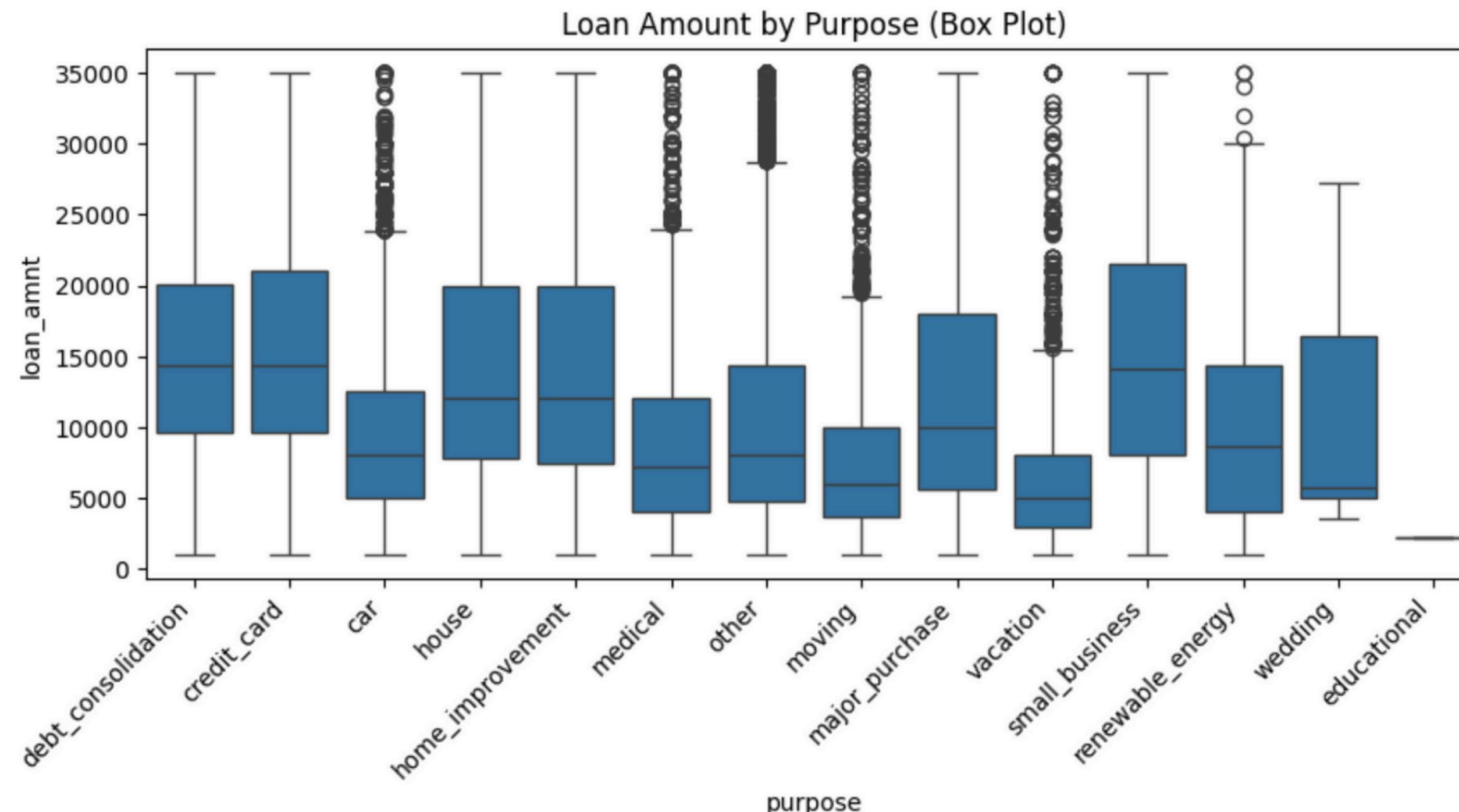
Current Features: Loan Details, Borrower Characteristics, Credit Profile & Risk Factors, Loan Performance

Additional Features:

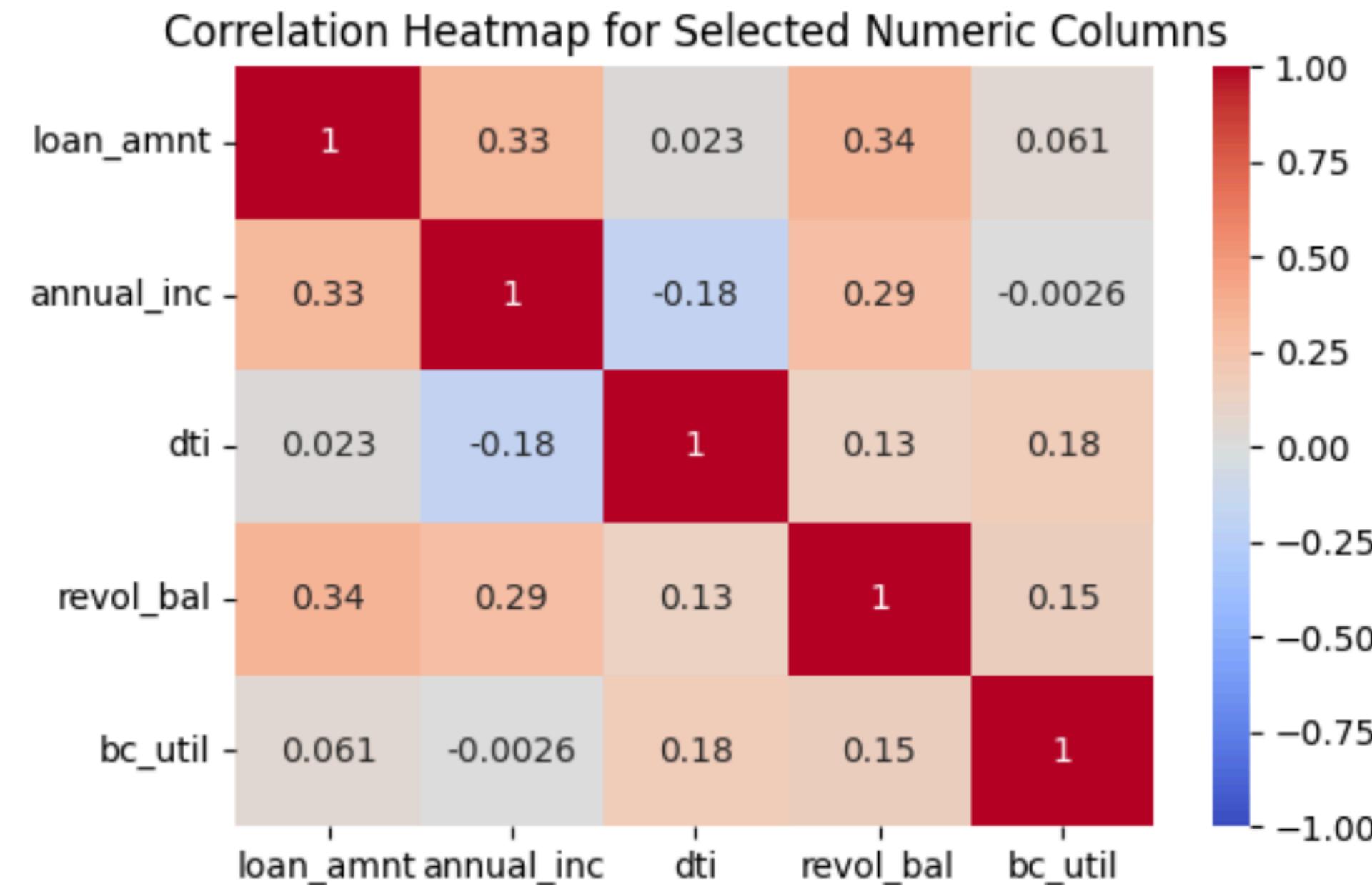
Recency of delinquency	higher credit risk
Current Delinquencies	Immediate risk
Credit Card Utilization	Shows financial strain
Mortage Accounts	Finacial responsabilty
Past bankruptcies	red flags



Initial Insight: People **borrow more for small businesses and credit card** debt, and less for cars, weddings, or education.

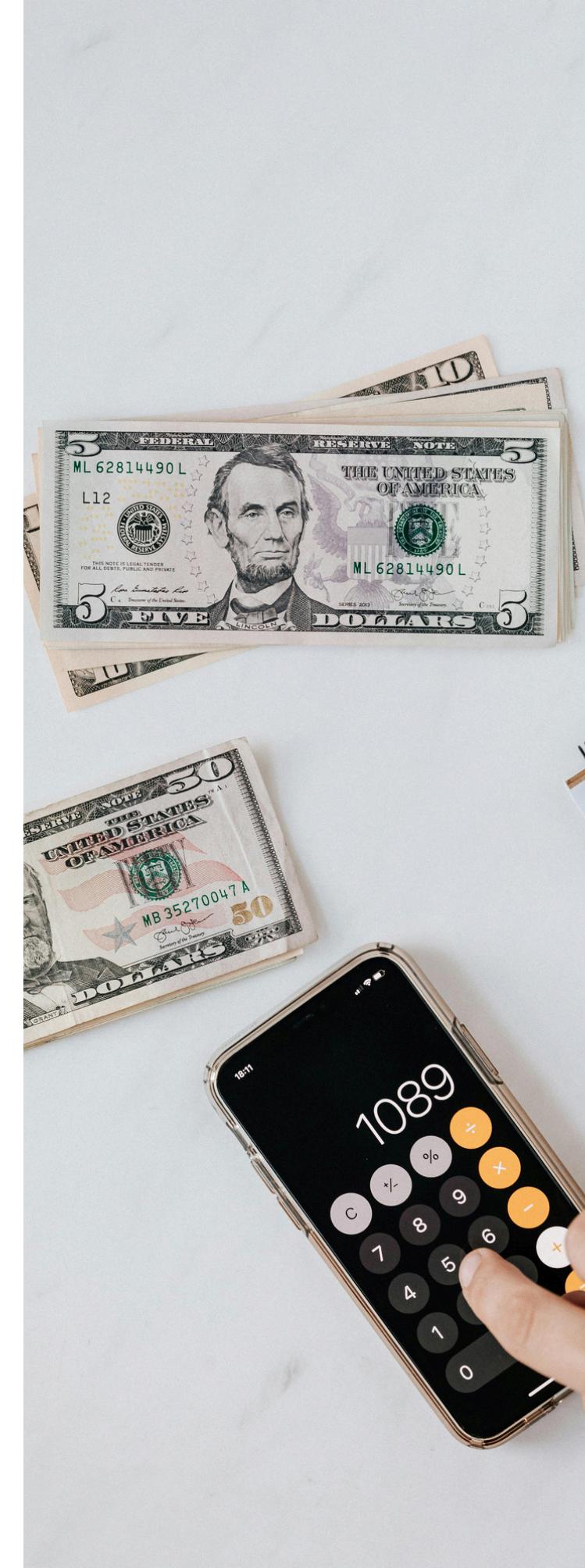


Initial diagnostic: people who **borrow more** also tend to **have higher incomes** and **more credit card debt**

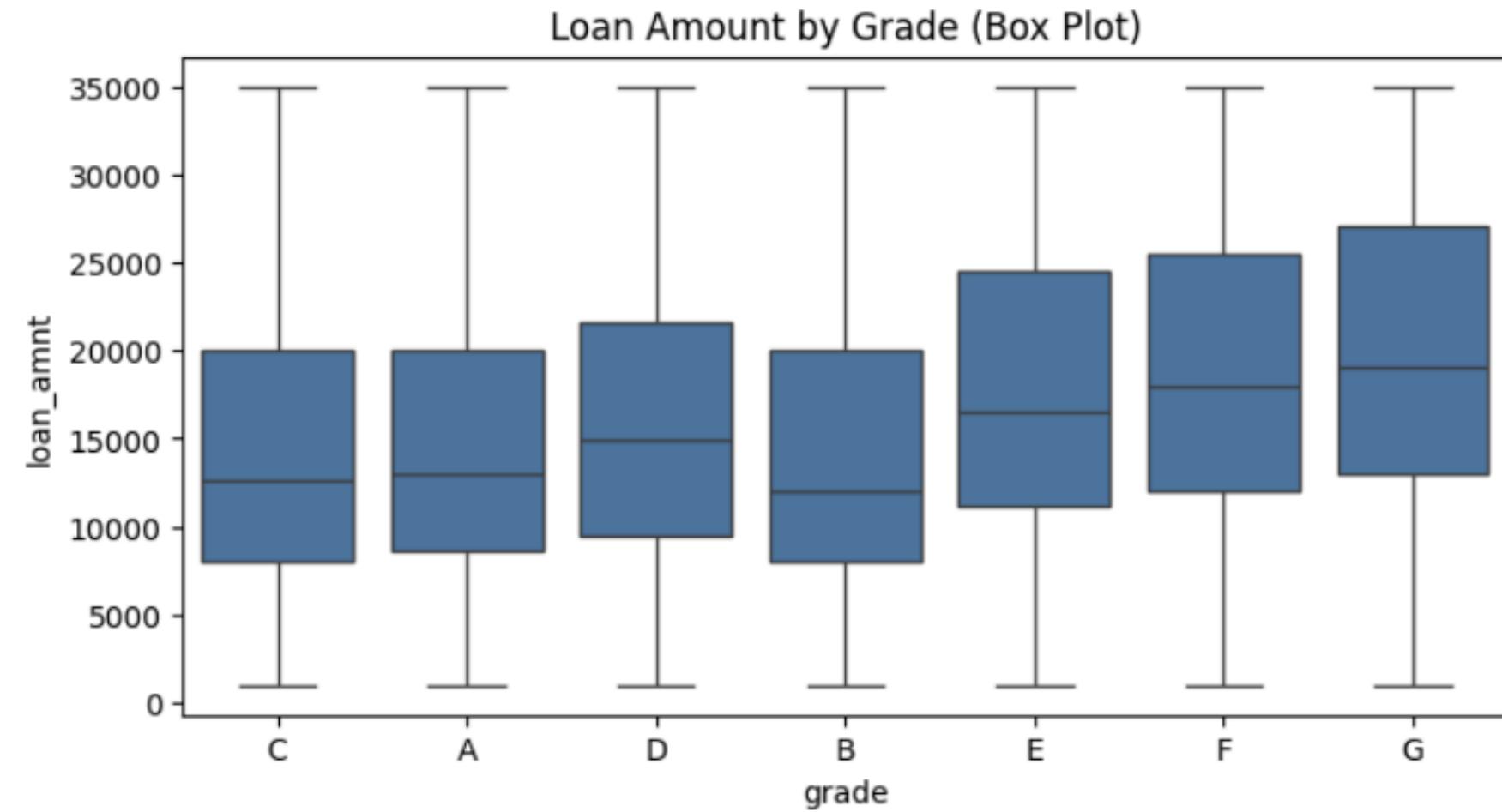
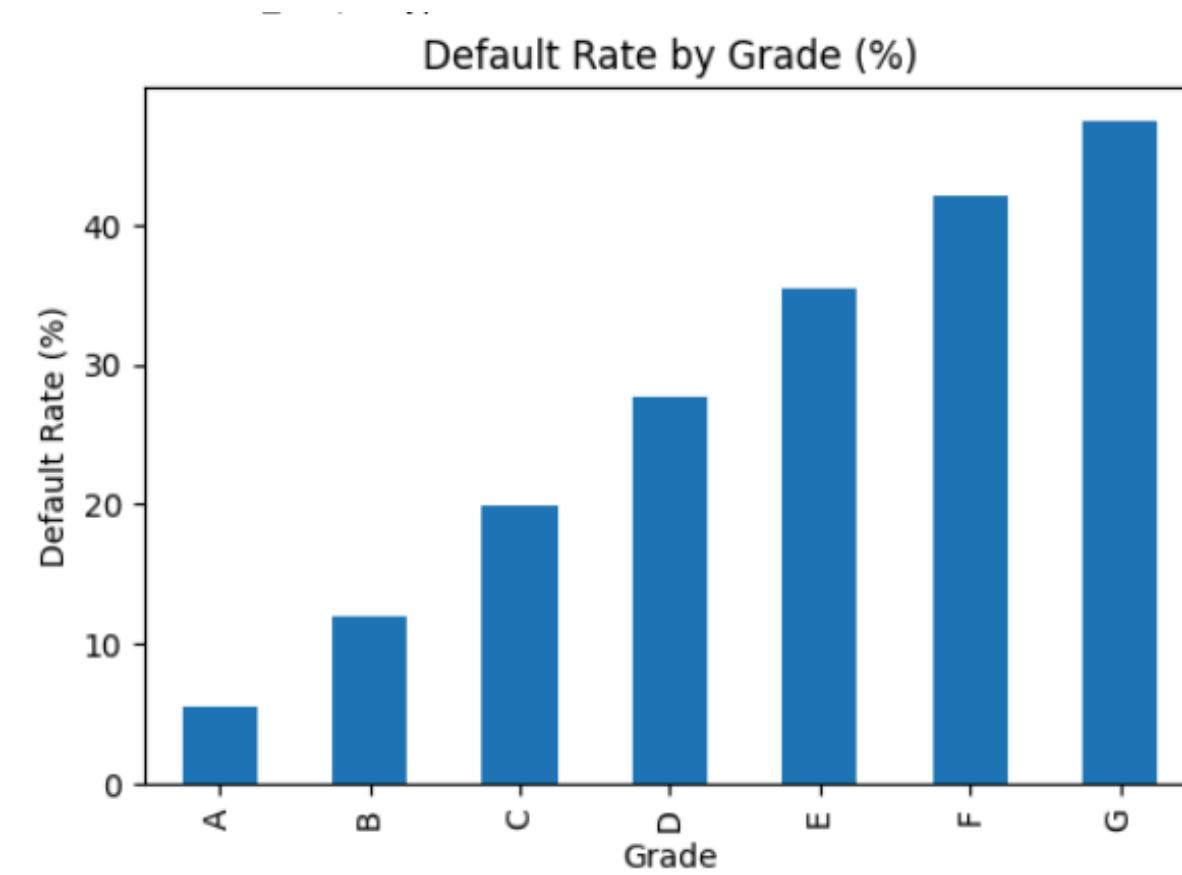


Hypotheses about feature importance in determining loan grades

- 1 Loans with higher grades (A, B) will have lower default rates and better returns compared to lower grades (E, F).
- 2 Higher interest rates are associated with higher default rates.
- 3 A higher debt-to-income ratio increases the likelihood of default.
- 4 Higher annual income will be inversely related to default rates.
- 5 A higher number of delinquencies will correlate with a higher default rate



Lower grades have a higher charged/paid off ratio



- The lower-grade loans (such as F and G) tend to have higher default rates compared to lower grade loans

Higher-grade loans tend to be smaller in size on average, reflecting that the lower-risk borrowers are typically asking for smaller loans. Conversely, higher-risk borrowers (Grades E, F, and G) are requesting larger loans on average, which could correlate with a higher likelihood of default.

Lower grades also have higher delinquency and interest rates

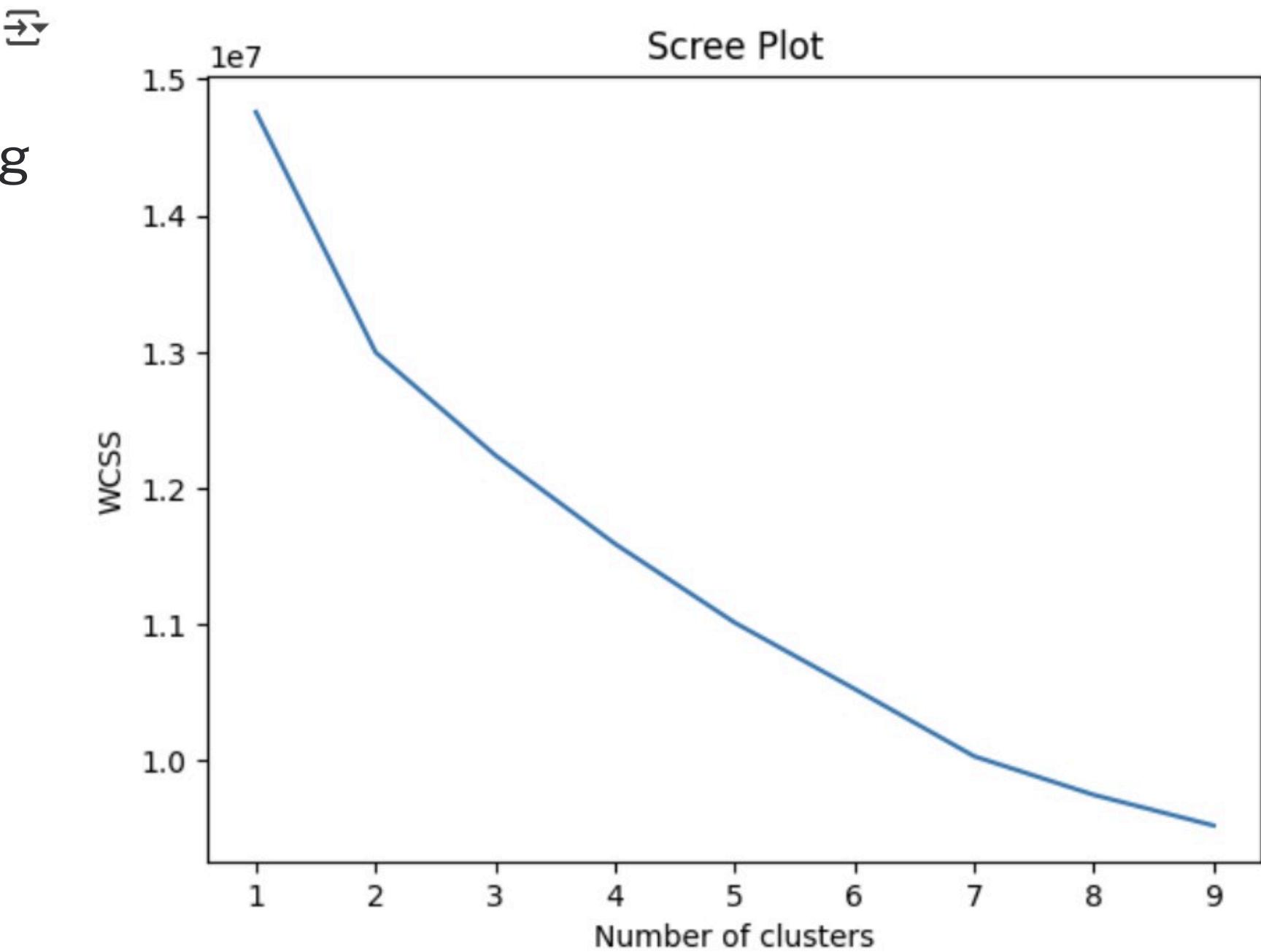
Grouped Descriptive Statistics for delinq_2yrs by grade:								
	count	mean	std	min	25%	50%	75%	max
grade								
A	109284.0	0.218083	0.679751	0.0	0.0	0.0	0.0	18.0
B	179160.0	0.352925	0.918744	0.0	0.0	0.0	0.0	22.0
C	186621.0	0.379775	0.970528	0.0	0.0	0.0	0.0	30.0
D	105247.0	0.384182	0.980697	0.0	0.0	0.0	0.0	39.0
E	54817.0	0.379627	0.967661	0.0	0.0	0.0	0.0	27.0
F	15935.0	0.375965	0.936968	0.0	0.0	0.0	0.0	17.0
G	3811.0	0.412490	1.084315	0.0	0.0	0.0	0.0	19.0

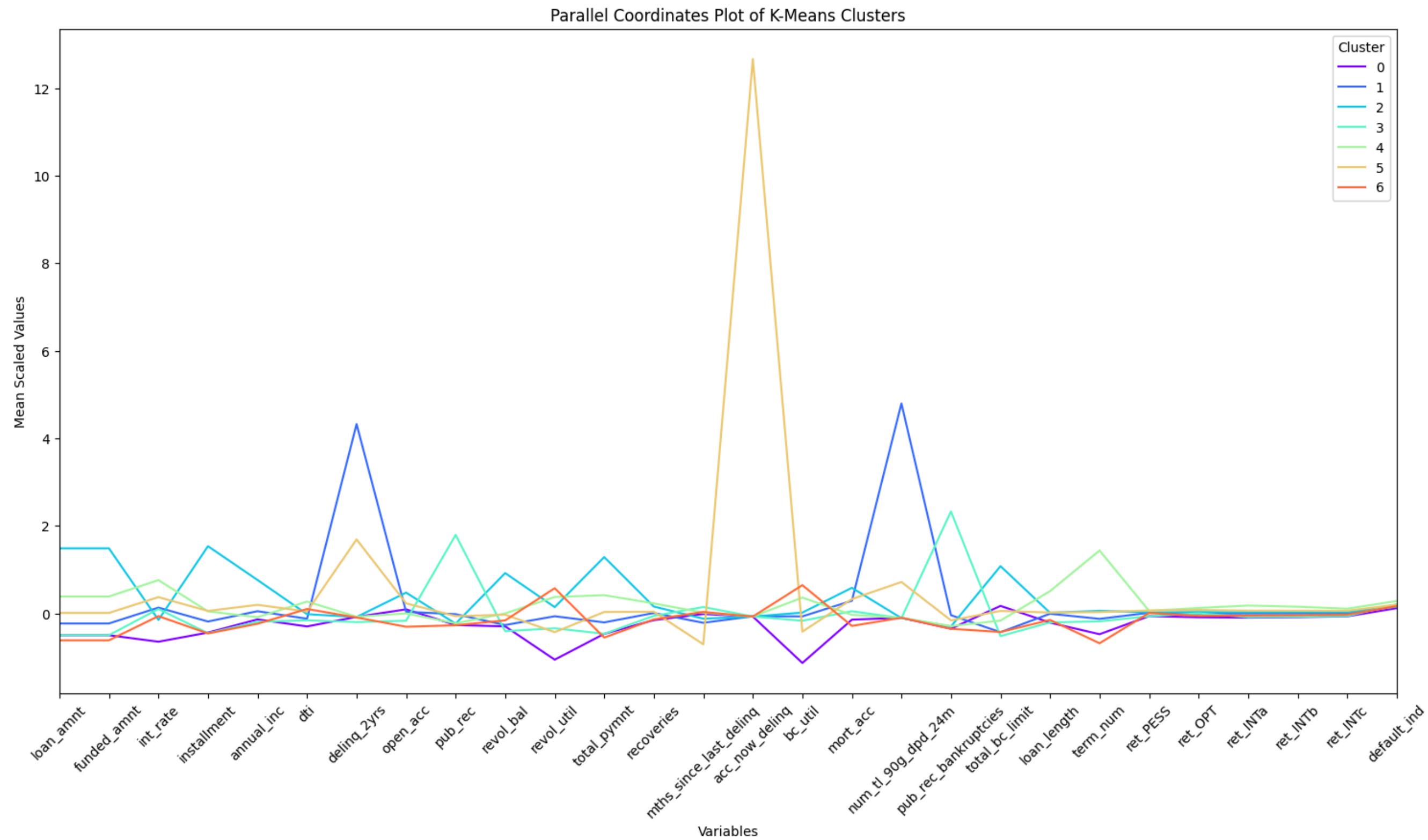
Grouped Descriptive Statistics for int_rate by grade:								
	count	mean	std	min	25%	50%	75%	max
grade								
A	109284.0	7.116395	0.922050	5.32	6.39	7.26	7.89	8.90
B	179160.0	10.439345	1.329716	6.00	9.17	10.64	11.53	13.53
C	186621.0	13.601161	0.961131	6.00	12.69	13.65	14.33	16.24
D	105247.0	16.882958	1.013670	6.00	15.99	16.99	17.57	19.97
E	54817.0	19.680268	1.249084	6.00	18.55	19.52	20.49	23.40
F	15935.0	23.856631	1.211182	6.00	22.99	23.99	24.99	26.06
G	3811.0	26.400559	0.986805	6.00	25.83	25.89	26.77	28.99

- Delinquency rates increase as the loan grade decreases.
- Grades A and B have the lowest delinquency rates, while Grades E, F, and G have noticeably higher delinquency rates, reinforcing that higher-risk borrowers are more prone to financial struggles.
- Lower-grade loans (C to G) have higher mean loan amounts, with Grade G having a mean loan amount of \$20,314.
- Higher-risk borrowers are taking on larger loans, possibly due to less financial stability or greater borrowing needs.

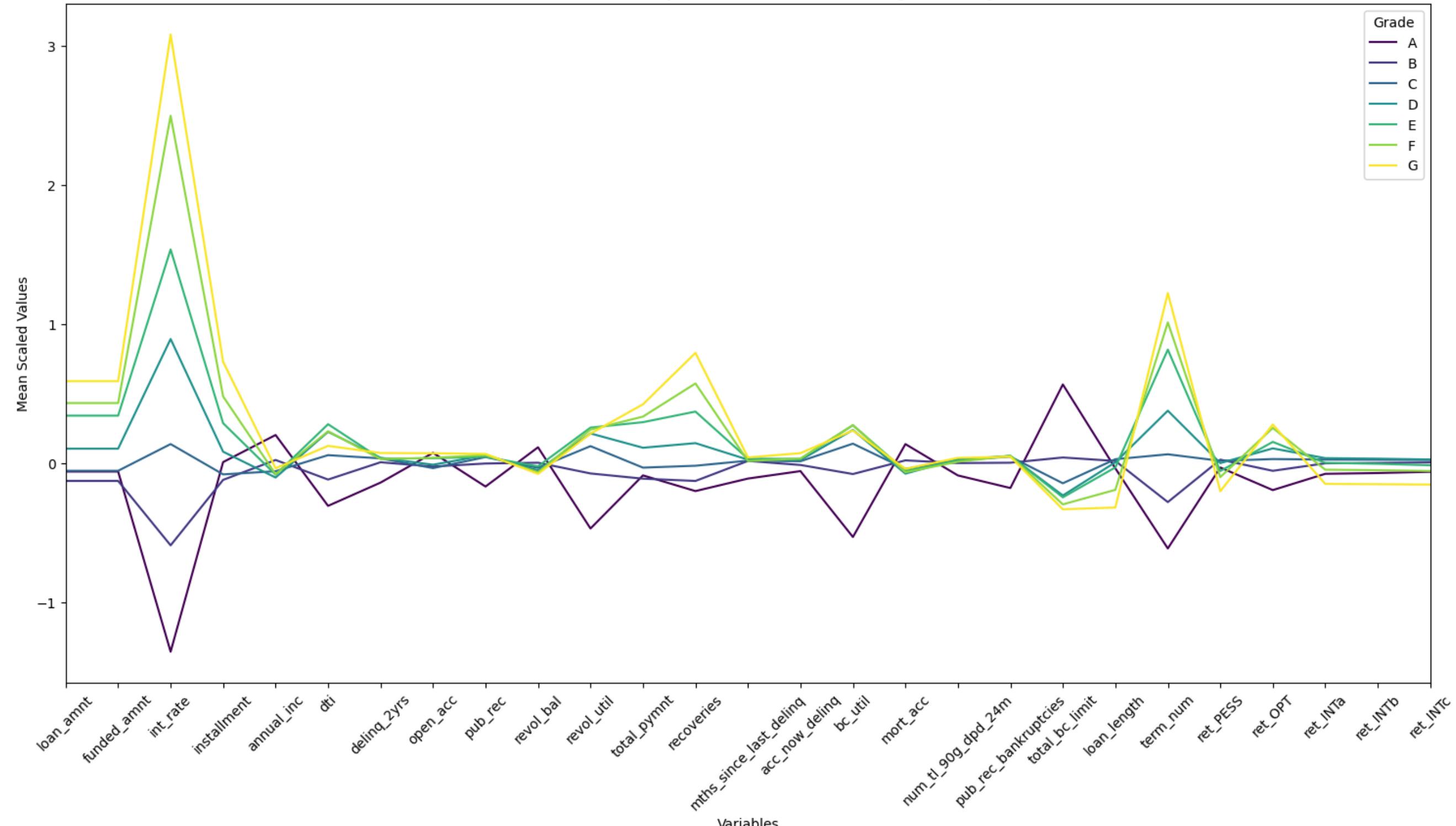
Clustering reveals 7 groups similar to LendingClub grades

- Optimal number of clusters $k = 7$ according to scree plot
- 7 resulting clusters can be cross-tabulated with LendingClub grades to “validate” whether clusters correlate directly with grade groupings





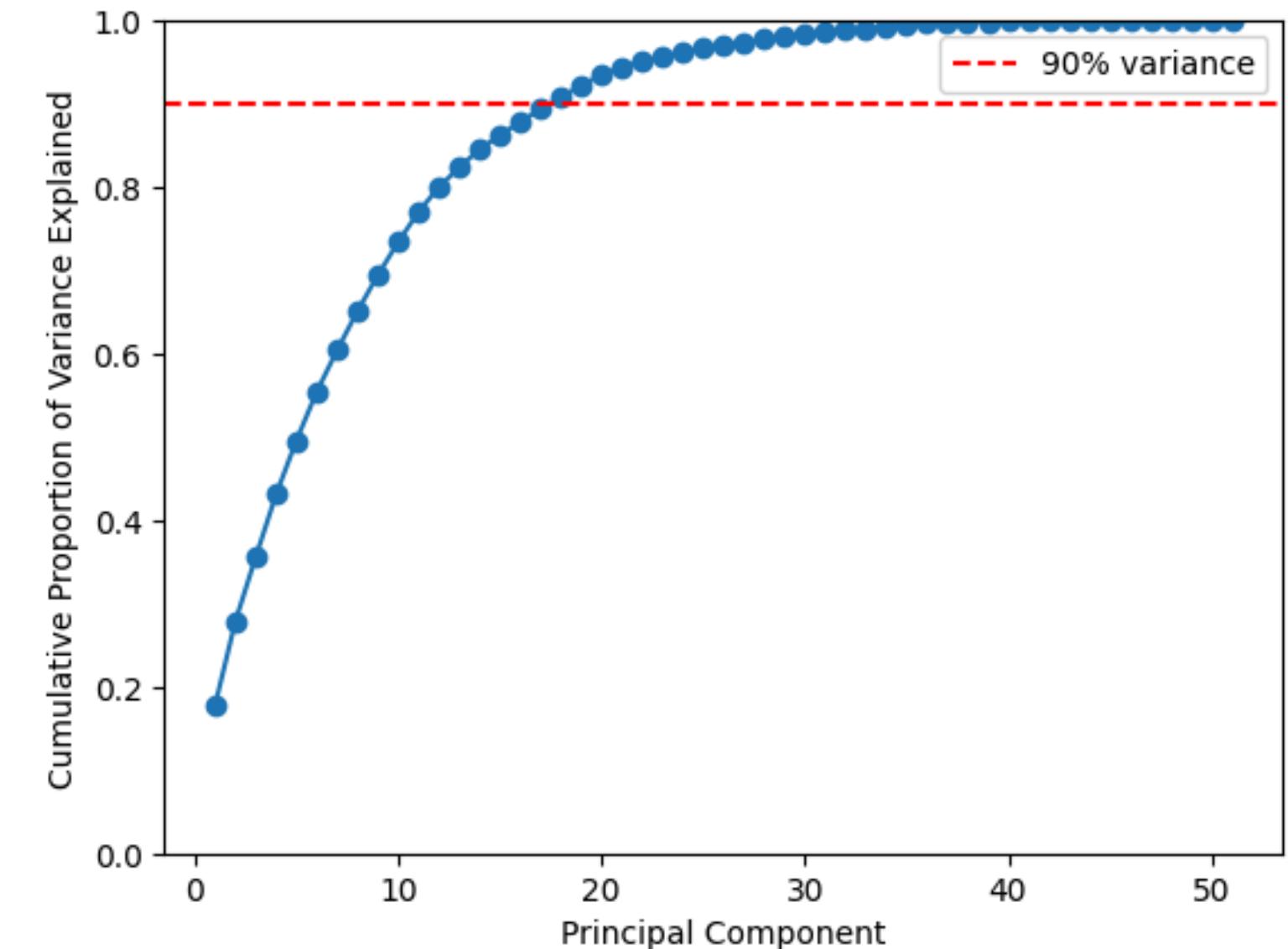
Parallel Coordinates Plot by Loan Grade (Actual LendingClub Categories)



90% of the variance in data can be explained by 18 PCs

From an initial feature set of 47 columns, we were able to reduce dimensionality down to 18 principal components. This will allow Jasmin to simplify the data and model without losing value and info.

This PCA plot shows that 18 components are enough to explain 90% of the variance in the data.



PCA Analysis

