



Casestudy: Lending Club Dataset

	Column Name	Description	Unit	Role
1	loan_amnt	Requested loan amount	US dollars	Predictor
2	term	Longest term length	Months	Predictor
3	int_term	Recommended interest rate	Rate	Response
4	emp_length	Employment length	Years	Predictor
5	home_ownership	Housing status	Categorical	Predictor
6	annual_inc	Annual income	US dollars	Predictor
7	purpose	Purpose for the loan	Categorical	Predictor
8	addr_state	State of residence	Categorical	Predictor
9	dti	Debt to income ratio	Percent	Predictor
10	delinq_2yrs	Number of delinquencies in the past 2 years	Count	Predictor
11	revol_util	Revolving credit line utilized	Percent	Predictor
12	total_acc	Number of active accounts	Count	Predictor
13	bad_loan	Bad loan indicator	Boolean	Response
14	longest_credit_length	Age of oldest active account	Years	Predictor
15	verification_status	Income verification status	Boolean	Predictor

Case Study: Lending Club Dataset

Predictor Column Name	Transformation
loan_amnt	Manage high and low amounts
term	Recode as 0/1 binary
emp_length	Create missing value indicator
home_ownership	Combine categories
annual_inc	Manage high and low annual incomes
purpose	Cross-Validated Mean Target Encoding
addr_state	Cross-Validated Mean Target Encoding
dti	Manage high and low debt-to-income ratios
delinq_2yrs	Manage high and low delinquency counts
revol_util	Manage high and low utilization
total_acc	Manage high and low number of accounts
longest_credit_length	Manage high and low credit lengths
verification_status	Recode as 0/1 binary

Case Study: Lending Club Dataset

	Column Name	Description	Unit	Role
1	loan_amnt	Requested loan amount	US dollars	Predictor
2	term	Longest term length	Months	Predictor
3	int_term	Recommended interest rate	Rate	Response
4	emp_length	Employment length	Years	Predictor
5	home_ownership	Housing status	Categorical	Predictor
6	annual_inc	Annual income	US dollars	Predictor
7	purpose	Purpose for the loan	Categorical	Predictor
8	addr_state	State of residence	Categorical	Predictor
9	dti	Debt to income ratio	Percent	Predictor
10	delinq_2yrs	Number of delinquencies in the past 2 years	Count	Predictor
11	revol_util	Revolving credit line utilized	Percent	Predictor
12	total_acc	Number of active accounts	Count	Predictor
13	bad_loan	Bad loan indicator	Boolean	Response
14	longest_credit_length	Age of oldest active account	Years	Predictor