



Casestudy: Lending Club Dataset

- Loan data from 2007 up until 2015 including rejected applications and accepted applications.
- Of the 500k accepted applicants about 160k loans have either been completely paid off or defaulted.
- There are about 4 million applicants in the rejected loans dataset.
- **Use Case 1:** Predict the likelihood of a user defaulting based on the information supplied when applying for a loan.
- **Use Case 2:** Determine the interest rate Lending Club would have offered the user based on the information supplied when applying for a loan.
- Full Data: <https://www.kaggle.com/wendykan/lending-club-loan-data>
- H2O Subset: <https://s3.amazonaws.com/h2o-public-test-data/bigdata/laptop/lending-club/loan.csv>

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

Case Study: Lending Club Dataset

- Loan data from 2007 up until 2015 including rejected applications and accepted applications.
- Of the 500k accepted applicants about 160k loans have either been completely paid off or defaulted.
- There are about 4 million applicants in the rejected loans dataset.
- **Use Case 1:** Predict the likelihood of a user defaulting based on the information supplied when applying for a loan.
- **Use Case 2:** Determine the interest rate Lending Club would have offered the user based on the information supplied when applying for a loan.
- Full Data: <https://www.kaggle.com/wendykan/lending-club-loan-data>
- H2O Subset: <https://s3.amazonaws.com/h2o-public-test-data/bigdata/laptop/lending-club/loan.csv>