

# **Project Proposal**

## **Loan Grade and Default Loan Prediction**

# Loan data from Lending Club

## Description

This data set represents thousands of loans made through the Lending Club platform, which is a platform that allows individuals to lend to other individuals. Of course, not all loans are created equal. Someone who is a essentially a sure bet to pay back a loan will have an easier time getting a loan with a low interest rate than someone who appears to be riskier. And for people who are very risky? They may not even get a loan offer, or they may not have accepted the loan offer due to a high interest rate. It is important to keep that last part in mind, since this data set only represents loans actually made, i.e. do not mistake this data for loan applications!

## Usage

loans\_full\_schema

## Features:

A data frame with 10,000 observations on the following 55 variables.

- emp\_title

Job title.

- emp\_length

Number of years in the job, rounded down. If longer than 10 years, then this is represented by the value 10.

- state

Two-letter state code.

- home\_ownership

The ownership status of the applicant's residence.

- annual\_income

Annual income.

- verified\_income

Type of verification of the applicant's income.

- debt\_to\_income

Debt-to-income ratio.

- annual\_income\_joint

If this is a joint application, then the annual income of the two parties applying.

- verification\_income\_joint

Type of verification of the joint income.

- debt\_to\_income\_joint

Debt-to-income ratio for the two parties.

- delinq\_2y

Delinquencies on lines of credit in the last 2 years.

- months\_since\_last\_delinq

Months since the last delinquency.

- earliest\_credit\_line

Year of the applicant's earliest line of credit

- inquiries\_last\_12m

Inquiries into the applicant's credit during the last 12 months.

- total\_credit\_lines

Total number of credit lines in this applicant's credit history.

- open\_credit\_lines

Number of currently open lines of credit.

- total\_credit\_limit

Total available credit, e.g. if only credit cards, then the total of all the credit limits. This excludes a mortgage.

- total\_credit\_utilized

Total credit balance, excluding a mortgage.

- num\_collections\_last\_12m

Number of collections in the last 12 months. This excludes medical collections.

- num\_historical\_failed\_to\_pay

The number of derogatory public records, which roughly means the number of times the applicant failed to pay.

- months\_since\_90d\_late

Months since the last time the applicant was 90 days late on a payment.

- current\_accounts\_delinq

Number of accounts where the applicant is currently delinquent.

- total\_collection\_amount\_ever

The total amount that the applicant has had against them in collections.

- current\_installment\_accounts

Number of installment accounts, which are (roughly) accounts with a fixed payment amount and period. A typical example might be a 36-month car loan.

- accounts\_opened\_24m

Number of new lines of credit opened in the last 24 months.

- months\_since\_last\_credit\_inquiry

Number of months since the last credit inquiry on this applicant.

- num\_satisfactory\_accounts

Number of satisfactory accounts.

- num\_accounts\_120d\_past\_due

Number of current accounts that are 120 days past due.

- num\_accounts\_30d\_past\_due

Number of current accounts that are 30 days past due.

- num\_active\_debit\_accounts

Number of currently active bank cards.

- total\_debit\_limit

Total of all bank card limits.

- num\_total\_cc\_accounts

Total number of credit card accounts in the applicant's history.

- num\_open\_cc\_accounts

Total number of currently open credit card accounts.

- num\_cc\_carrying\_balance

Number of credit cards that are carrying a balance.

- num\_mort\_accounts

Number of mortgage accounts.

- account\_never\_delinq\_percent

Percent of all lines of credit where the applicant was never delinquent.

- tax\_liens

a numeric vector

- public\_record\_bankrupt

Number of bankruptcies listed in the public record for this applicant.

- loan\_purpose

The category for the purpose of the loan.

- application\_type

The type of application: either individual or joint.

- loan\_amount

The amount of the loan the applicant received.

- term

The number of months of the loan the applicant received.

- interest\_rate

Interest rate of the loan the applicant received.

- installment

Monthly payment for the loan the applicant received.

- grade

Grade associated with the loan.

- sub\_grade

Detailed grade associated with the loan.

- issue\_month

Month the loan was issued.

- loan\_status

Status of the loan.

- initial\_listing\_status

Initial listing status of the loan. (I think this has to do with whether the lender provided the entire loan or if the loan is across multiple lenders.)

- disbursement\_method

Disbursement method of the loan.

- balance

Current balance on the loan.

- paid\_total

Total that has been paid on the loan by the applicant.

- paid\_principal

The difference between the original loan amount and the current balance on the loan.

- paid\_interest

The amount of interest paid so far by the applicant.

- paid\_late\_fees

Late fees paid by the applicant.

## Business Problem:

Lending Club managers need to make good lending decisions based on the right loan grade. So they need to know:

- What are the most important features in grade classification?
- What are the most important features in defaulting loans?
- Particularly which features help greatly hedge the risk of the lender?