



LENDING CLUB DATA ANALYSIS

Predicting a NonPerforming Loan



WHY REVISIT?

- Short timeframe, familiar with the dataset, plus lots of unexplored relationships
- Data is nice mix of continuous and categorical variables
- Historical data goes from 2007 to the present
- Will a loan turn out bad? Will look at sub grade variable, description variable and predict with logistic regression and random forest

THE DATASET

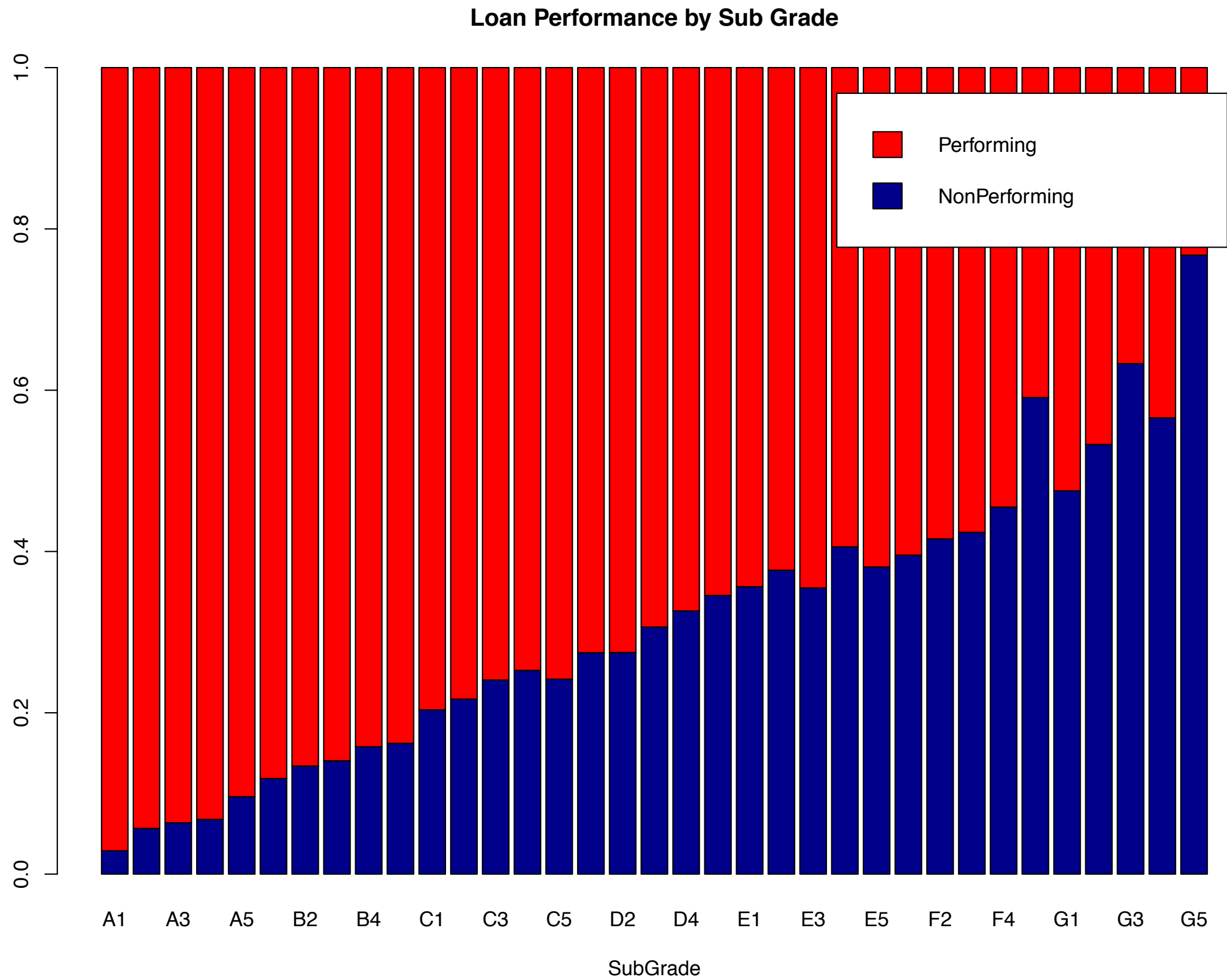
- Using dataset from 2007-2011. Data from 2012 to present can added later on.
- 42,485 observations. 110 variables
- Remove columns 85% NA ->down to 56 variables.
- Loans are either Good or Bad (performing or nonperforming).
Binary Classification. Mapped 7 categories of loans to 2 categories

Status	Mapping	Description
Fully Paid	Performing	Loan has been fully repaid
Current	Performing	Loan is up to date on all payments
In Grace Period	NonPerforming	Loan is between 0 and 15 days past due
Late (16 - 30 days)	NonPerforming	Loan is between 16 and 30 days past due
Late (31 - 120 days)	NonPerforming	Loan is between 31 and 120 days past due
Default	NonPerforming	Loan is over 121 days past due
Charged Off	NonPerforming	Loan for which there is no reasonable expectation of additional payments

THE DATASET

- Variable Subgrade : Lending Club maps borrowers to a series of grades [A-F] and subgrades [A-F][1-5] based on their **risk profile**. Loans in each subgrade are then given appropriate interest rates. The specific rates will change over time according to market conditions, but generally they will fall within a tight range for each subgrade.

PERCENTAGE BAD LOANS PER SUB GRADE

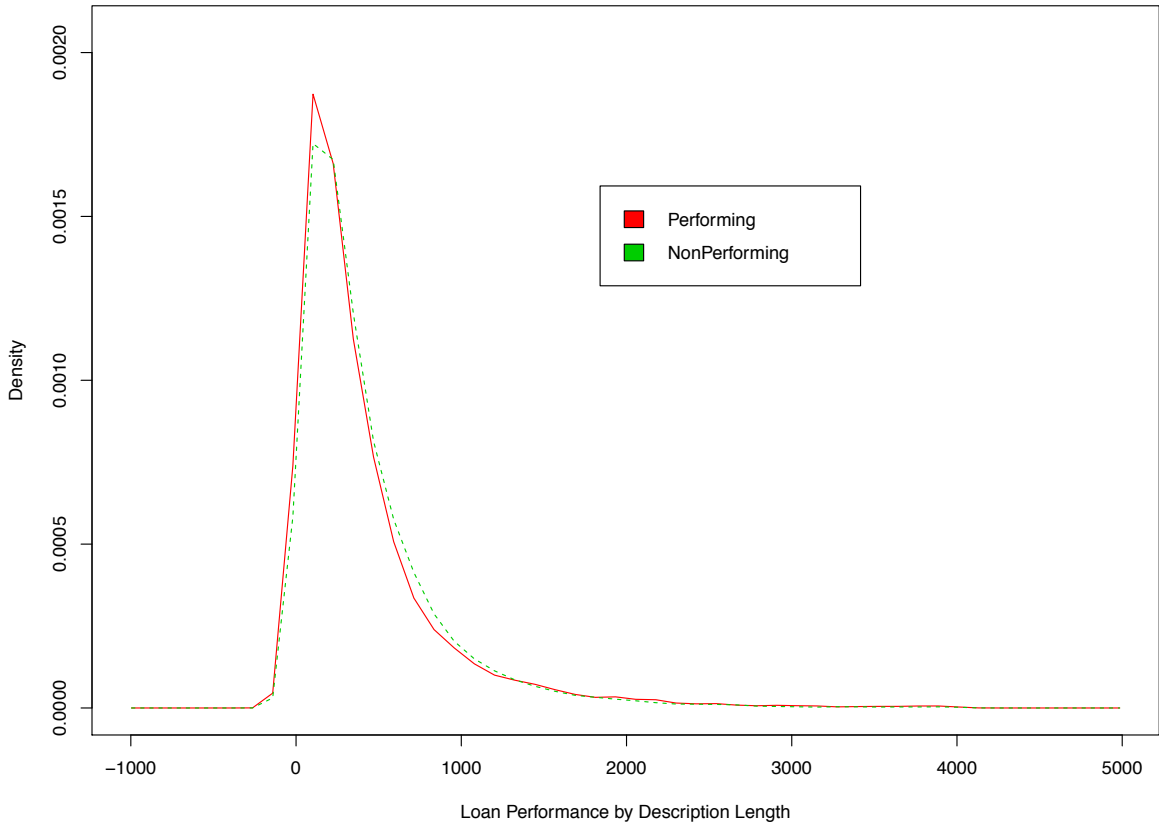


WHAT IS DESCRIPTION FIELD?

I plan on combining three large interest bills together and freeing up some extra each month to pay toward other bills. I've always been a good payor but have found myself needing to make adjustments to my budget due to a medical scare. My job is very stable, I love it.

DENSITY OF LOANS VS. DESCRIPTION LENGTH

Loan Performance by Description Length



Loan Performance by Description Length > 1000

