LendingClub

LendingClub: would you invest?

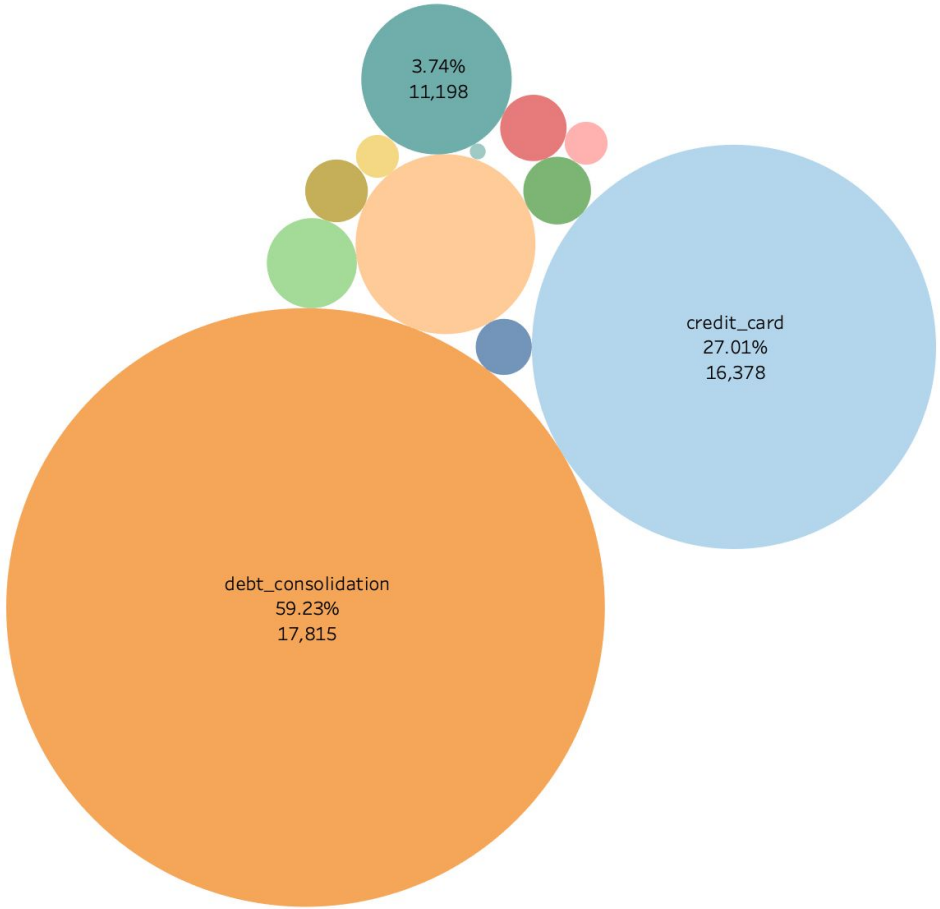
DSI Project 6
June 9, 2020
Alina

About LendingClub

- Peer-to-peer lending platform headquartered in San Francisco, CA
- Investors could be individuals or companies
- Investors: Min investment $1,000
- Borrowers: loan amount ranges from $1,000 to $40,000
  - Mean loan amount for the dataset $16,515
- Risk diversification: min $25 per note, investment can be spread out among as many notes as possible
- Interest rates for personal loans range from 10.68% to 35.89%
  - Mean interest for the dataset 13.21%
- Loan term in dataset: 36 months and 60 months
- Unconventional variety of reasons to borrower

LendingClub

Transactions by Purpose



Purpose
- car
- credit_card
- debt_consolidation
- home_improvement
- house
- major_purchase
- medical
- moving
- other
- renewable_energy
- small_business
- vacation

Problem Statement

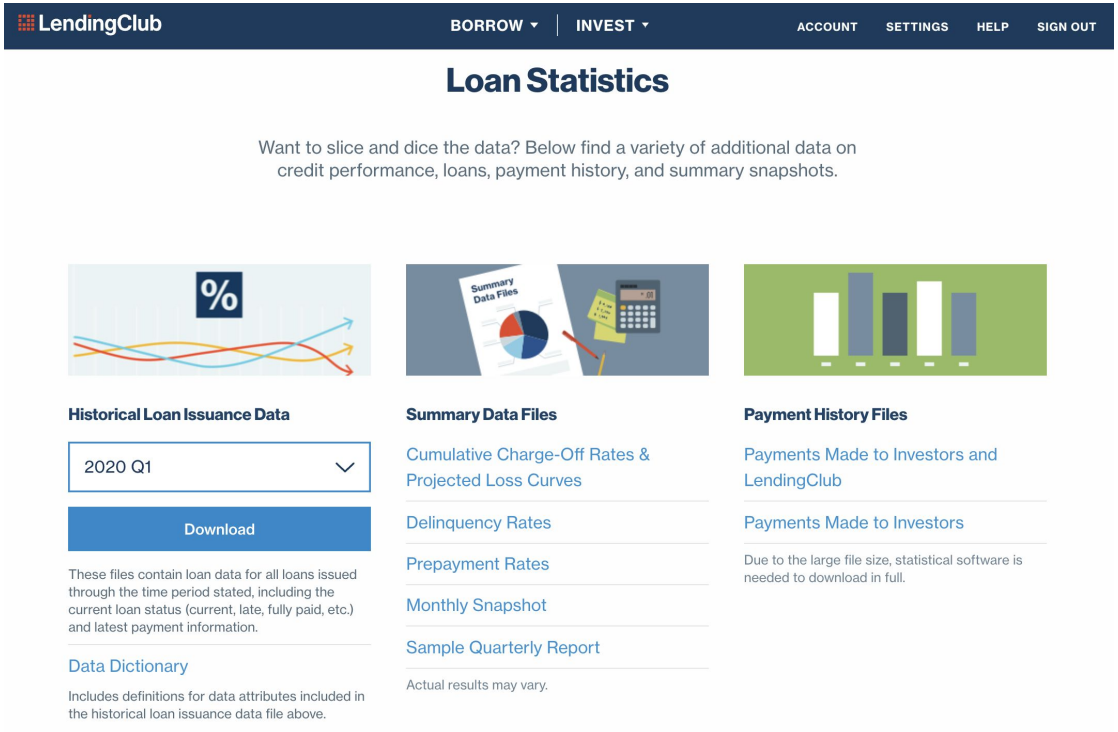
- LendingClub looks like an attractive way to invest while helping others, and bypassing the traditional banking system. The idea of people being able to help each other achieve some of their financial goals is cool.

The goal of the project is to deep dive the LendingClub's most recent dataset and answer the following questions:

- Is it a good idea to invest through LendingClub?
- At origination, what are the top factors that would affect the longevity of the loan?
- How well can we predict the possibility of a loan fallout?

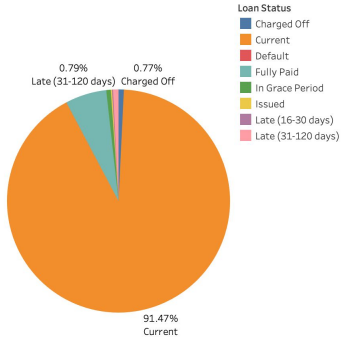
Gathering Data

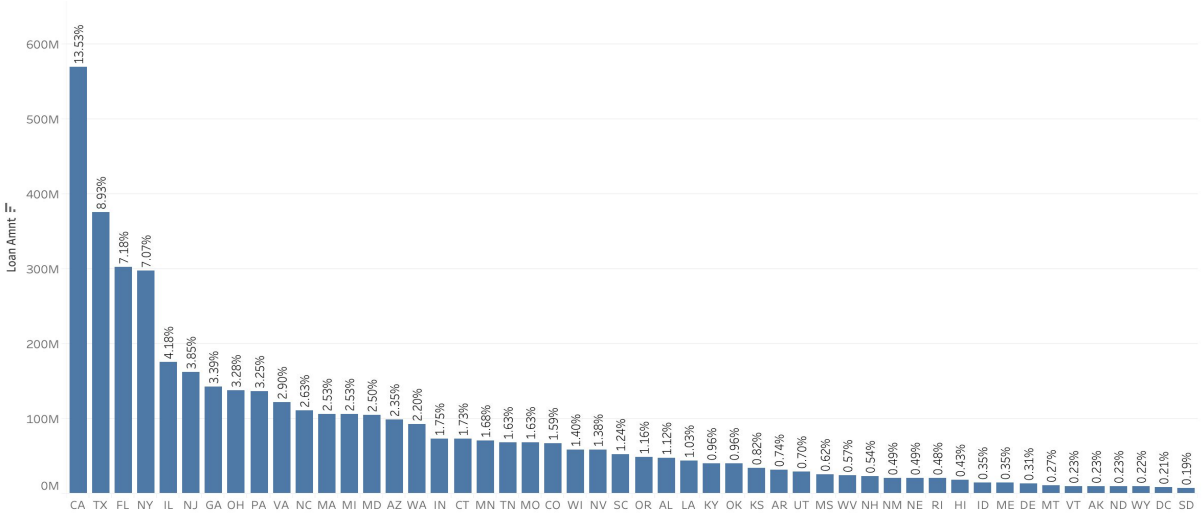
- Open source data made available by LC
- 3 most recent quarters: 2020 Q1, 2019 Q3, 2019 Q3
- 376,305 Rows x 107 columns
- 254,792 rows of cleaned data, 4.2 Bln in total transactions



Obstacles to Analysis and Modeling

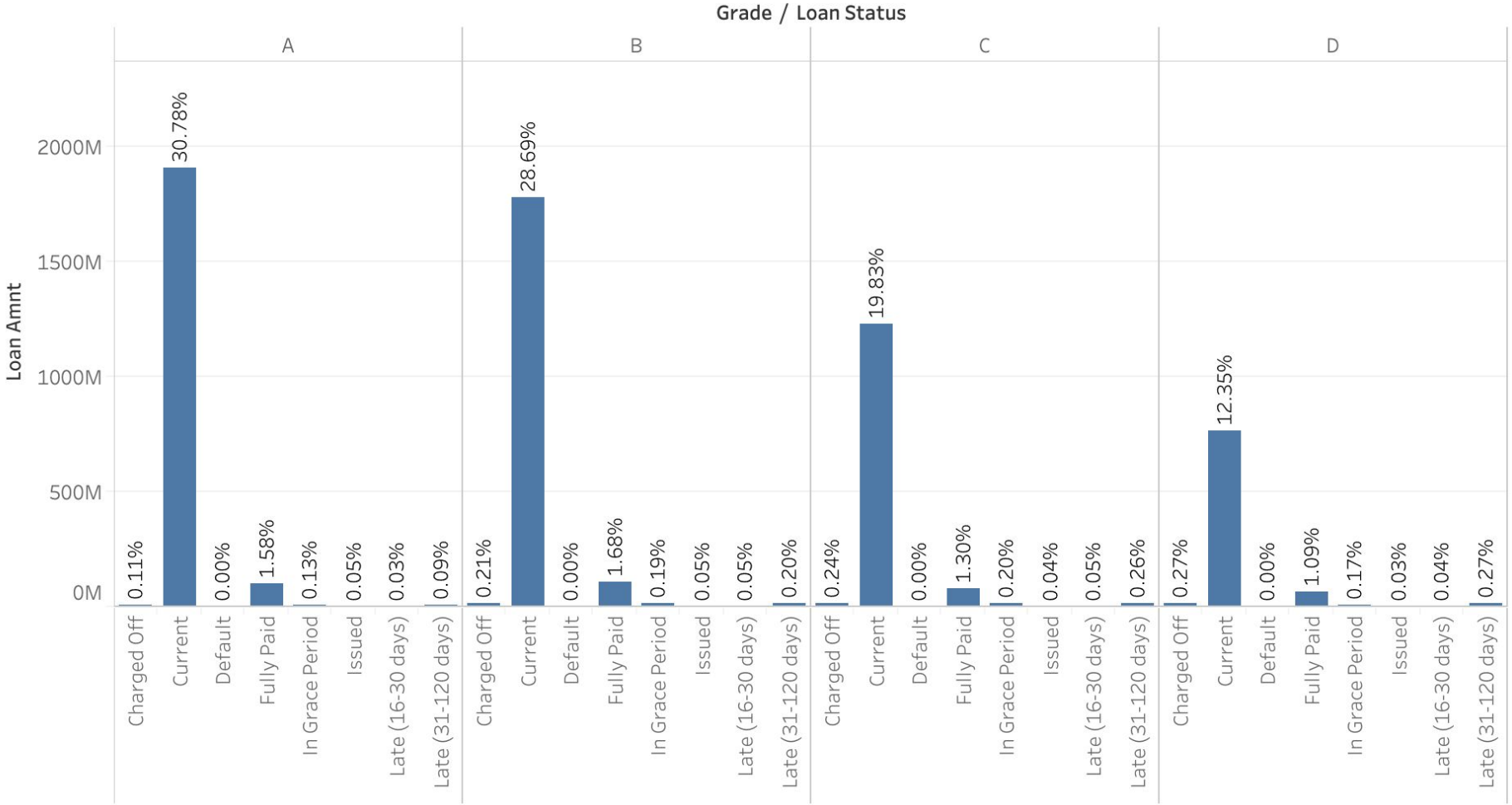
- Significant class imbalance in certain classes
- Similarities within some other classes
- Number of variables and extensive data dictionary
- Weak correlations
- Challenges of a real life dataset, data cleaning and modeling



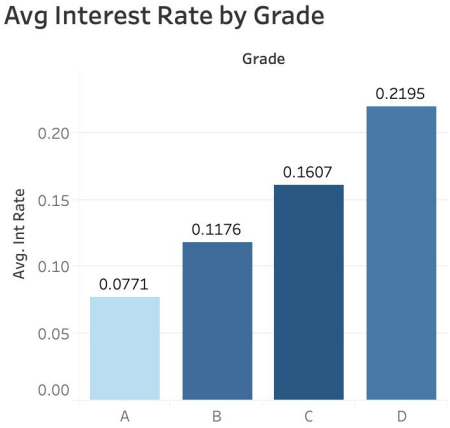


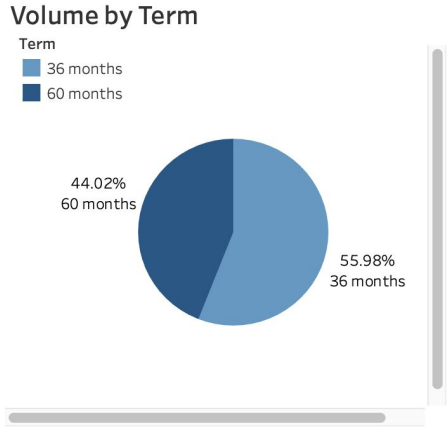
Loan Status	% of Total Loan Amnt along Lo..	Loan Amnt
Current	91.47%	3,849,160,050
Fully Paid	5.98%	251,817,400
Charged Off	0.77%	32,448,475
Late (31-120 days)	0.79%	33,097,575
In Grace Period	0.65%	27,501,125
Late (16-30 days)	0.15%	6,394,475
Issued	0.17%	7,325,375
Default	0.01%	364,100

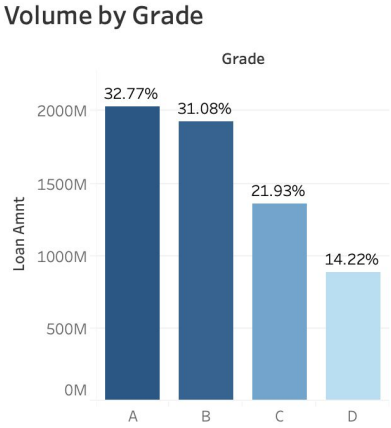
Status by Loan Grade

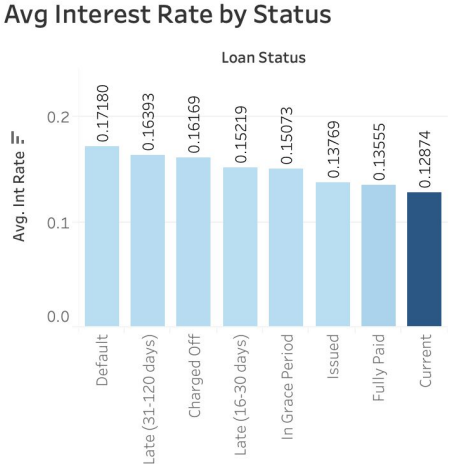


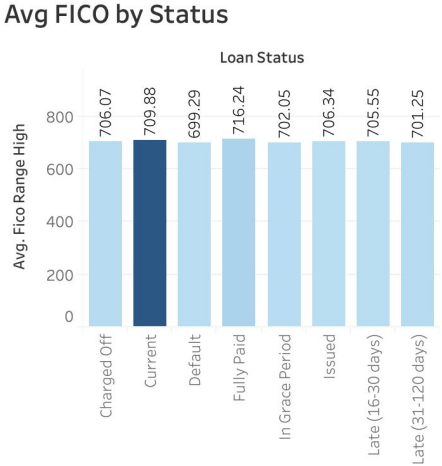
Exploring the Variables

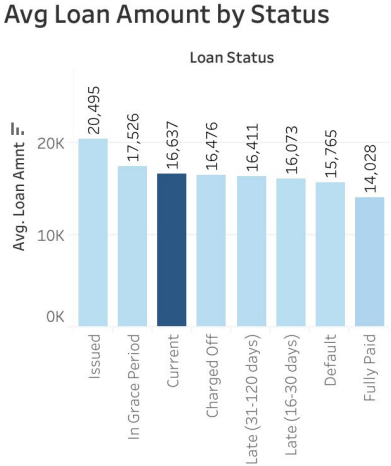






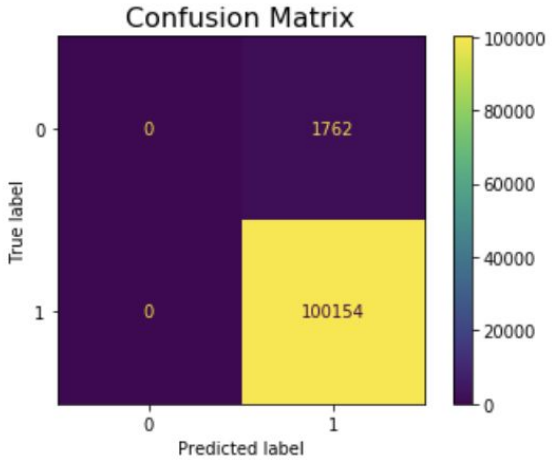




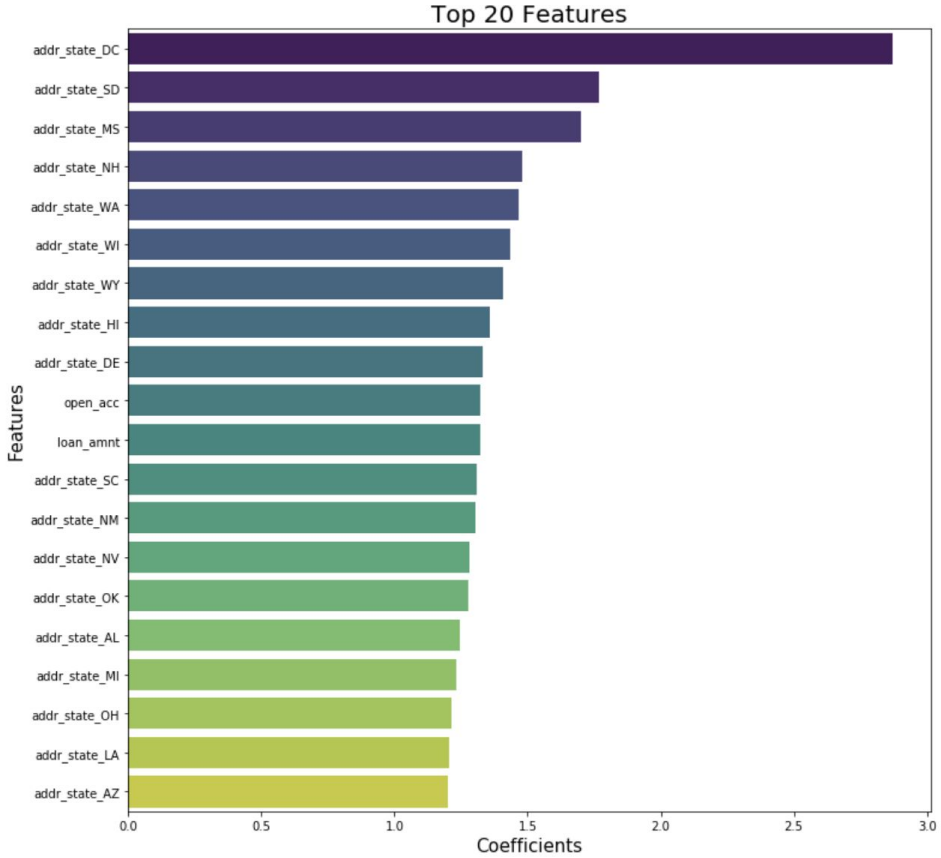


Modeling

- Logistic Regression
  - Goal: find factors that affect longevity of a loan
  - Current loan status: 1, Charge off + late + default: 0
  - Data scaling in dataset with continuous and categorical variables
  - Accuracy score 98.27% (due to baseline!)
  - Confusion Matrix: 1.7% FP
- Random Forest
  - Accuracy score 98.29%

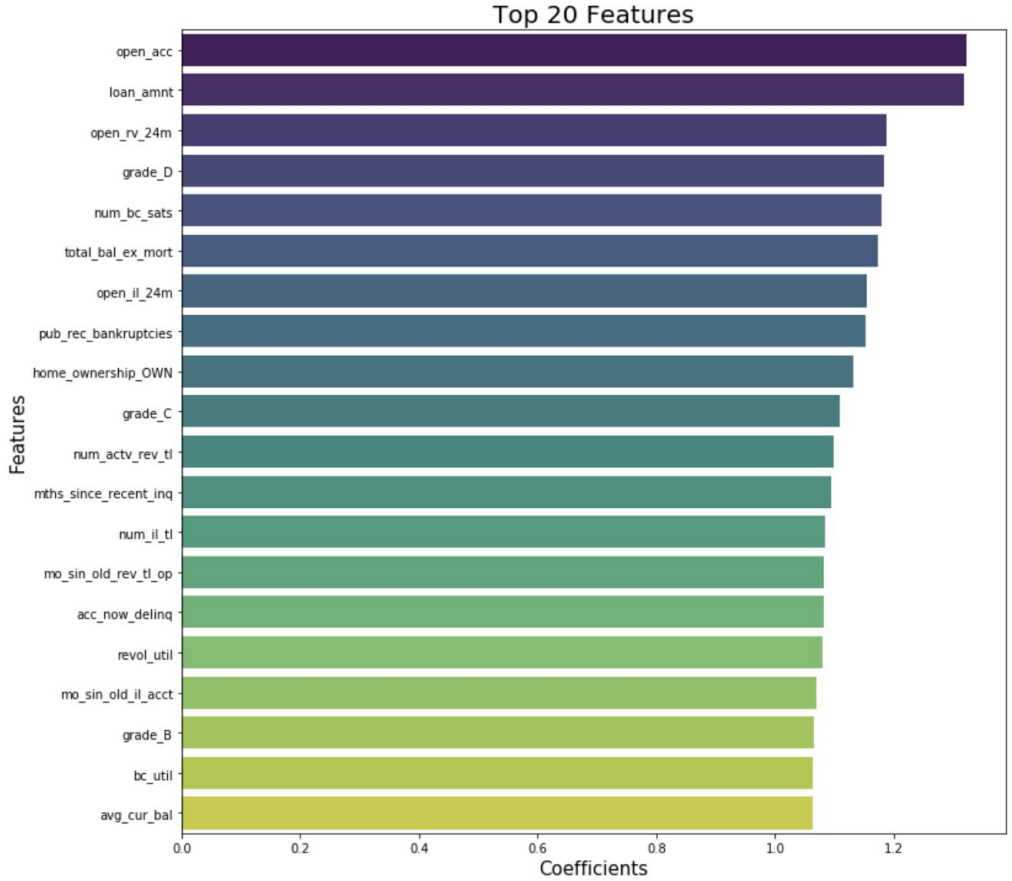


States and Top Features



Top Features

coef	features
1.321741	open_acc
1.317992	loan_amnt
1.188337	open_rv_24m
1.184430	grade_D
1.179015	num_bc_sats
1.172784	total_bal_ex_mort
1.154523	open_il_24m
1.152750	pub_rec_bankruptcies
1.130929	home_ownership_OWN
1.108127	grade_C
1.099443	num_actv_rev_tl
1.095369	mths_since_recent_inq
1.083892	num_il_tl
1.082897	mo_sin_old_rev_tl_op
1.082192	acc_now_delinq
1.079527	revol_util
1.068861	mo_sin_old_il_acct
1.064434	grade_B
1.064199	bc_util
1.063277	avg_cur_bal



LendingClub

Top Features affecting good vs bad loans

1. **open_acc:** The number of open credit lines in the borrower's credit file
2. **Loan_amnt:** Loan amount
3. **Open_rv_24m:** Number of revolving trades opened in past 24 months
4. **Grade:** LC assigned loan grade
5. **num_bc_sats:** Number of satisfactory bankcard accounts
6. **total_bal_ex_mort:** Total credit balance excluding mortgage
7. **open_il_24m:** Number of installment accounts opened in past 24 months
8. **pub_red_bankruptcies:** Number of public record bankruptcies
9. **home_ownership:** The home ownership status provided by the borrower during registration or obtained from the credit report.
10. **num_actv_rev_tl:** Number of currently active revolving trades

Conclusions

- Is it a good idea to invest through LendingClub?
  - 91% of loans are current, 97% have a positive status (incl 5.98% full paid off)
- At origination, what are the top factors that would affect the longevity of the loan?
  - The model identified Top 20 variables
- How well can we predict the possibility of a loan fallout?
  - 1.7% error
  - High accuracy scores due to baseline
  - Overall 97% positive loan status, plus embedded possibility to diversify

LendingClub

Thank you!