Prediction of Default Rates for Peer-to-Peer Loaning Companies

By Group 10: Sammi Chang, Junguo He, Xueting Li, Jialu Li

PAUL MERAGE

Focus on Innovation

PRODUCED
BY GROUP 10

Meet The Team

Jialu(Molly) Li

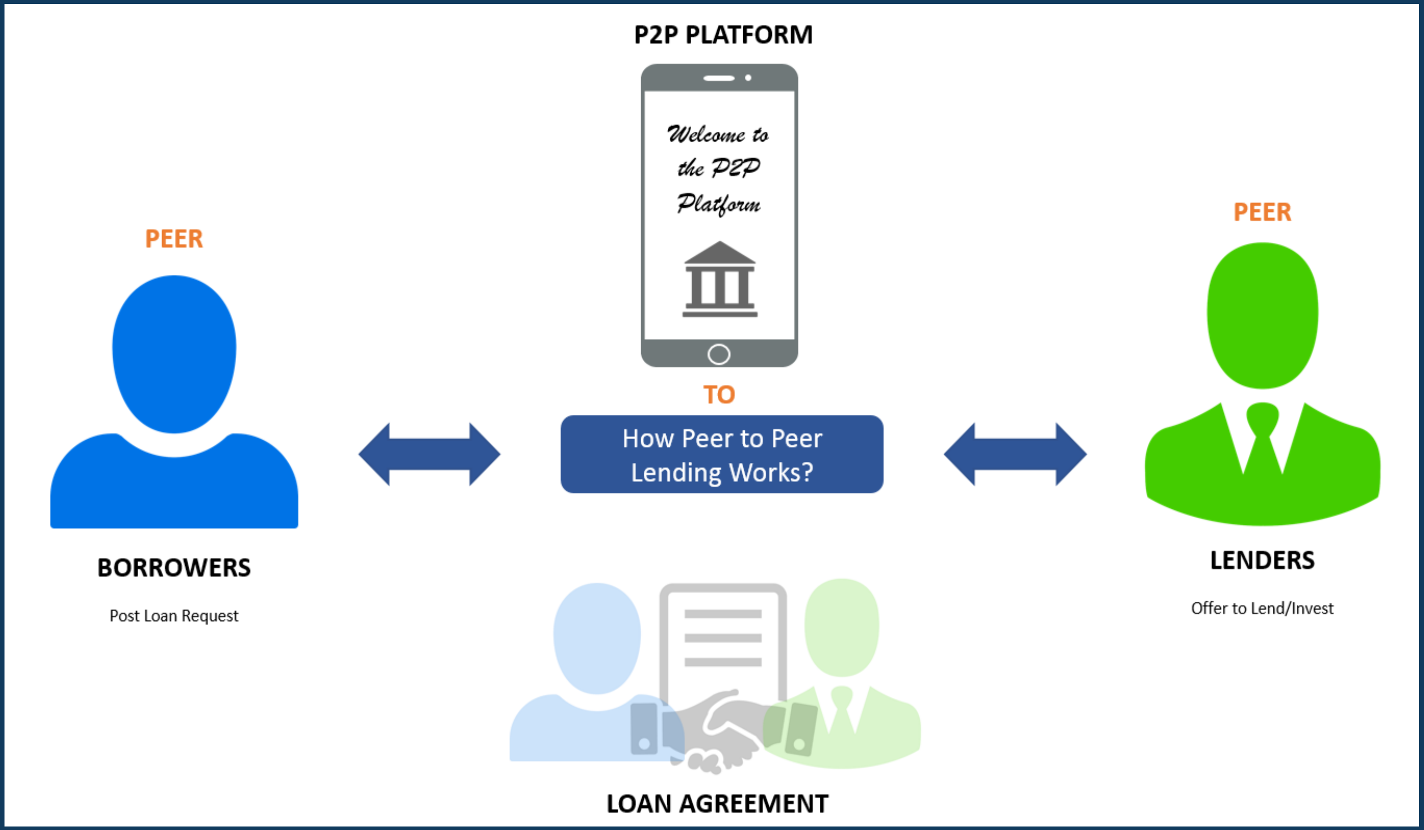
Junguo(Nick) He

Xueting(Lexie) Li

Sammi (Lishan)
Chang

Background

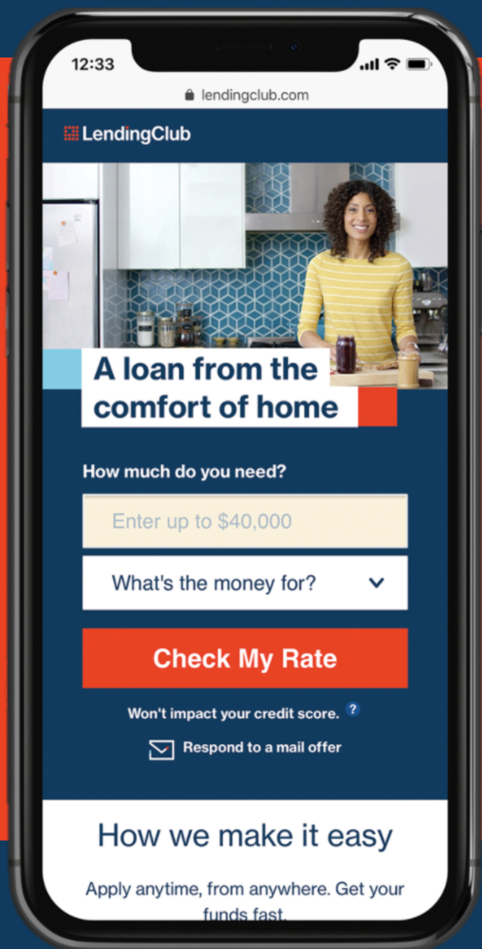
What is a P2P company?



Peer-to-peer (P2P) lenders are secondary marketplaces for loans that connect individual borrowers and lenders over an online platform. Lenders and borrowers never directly interact with each other; the P2P platform acts instead as a middleman. The most common type of P2P loan is a personal or business loan.

LendingClub

How LendingClub works



1

Apply

Apply on LC App or website based on user questionnaire

2

Choose a Loan Offer

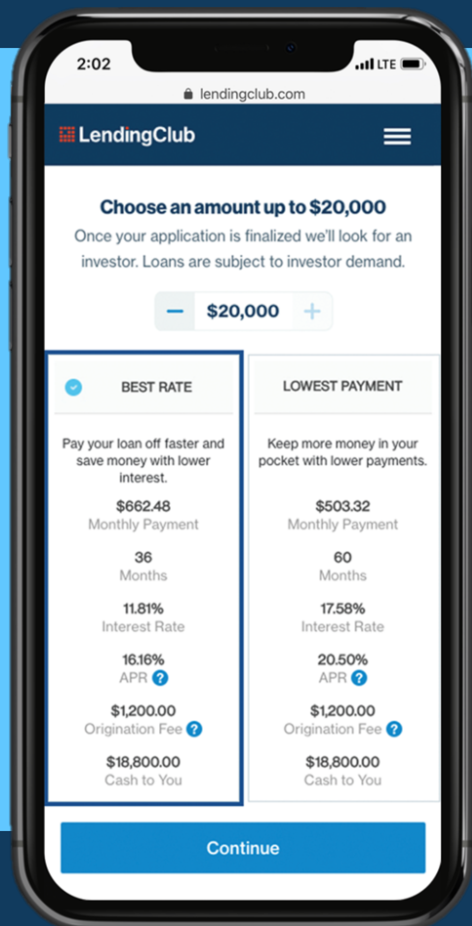
3

Get Funded

How LendingClub works

1

Apply



2

Choose a Loan Offer

Select rate, term, and payment options

3

Get Funded

How LendingClub works

1

Apply

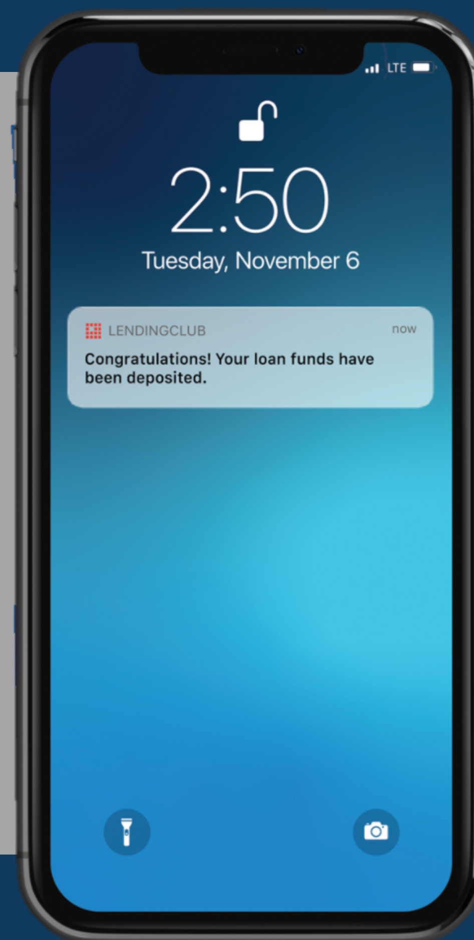
2

Choose a Loan Offer

3

Get Funded

After evaluation and matching an investor, loan can be sent directly to the borrower's account



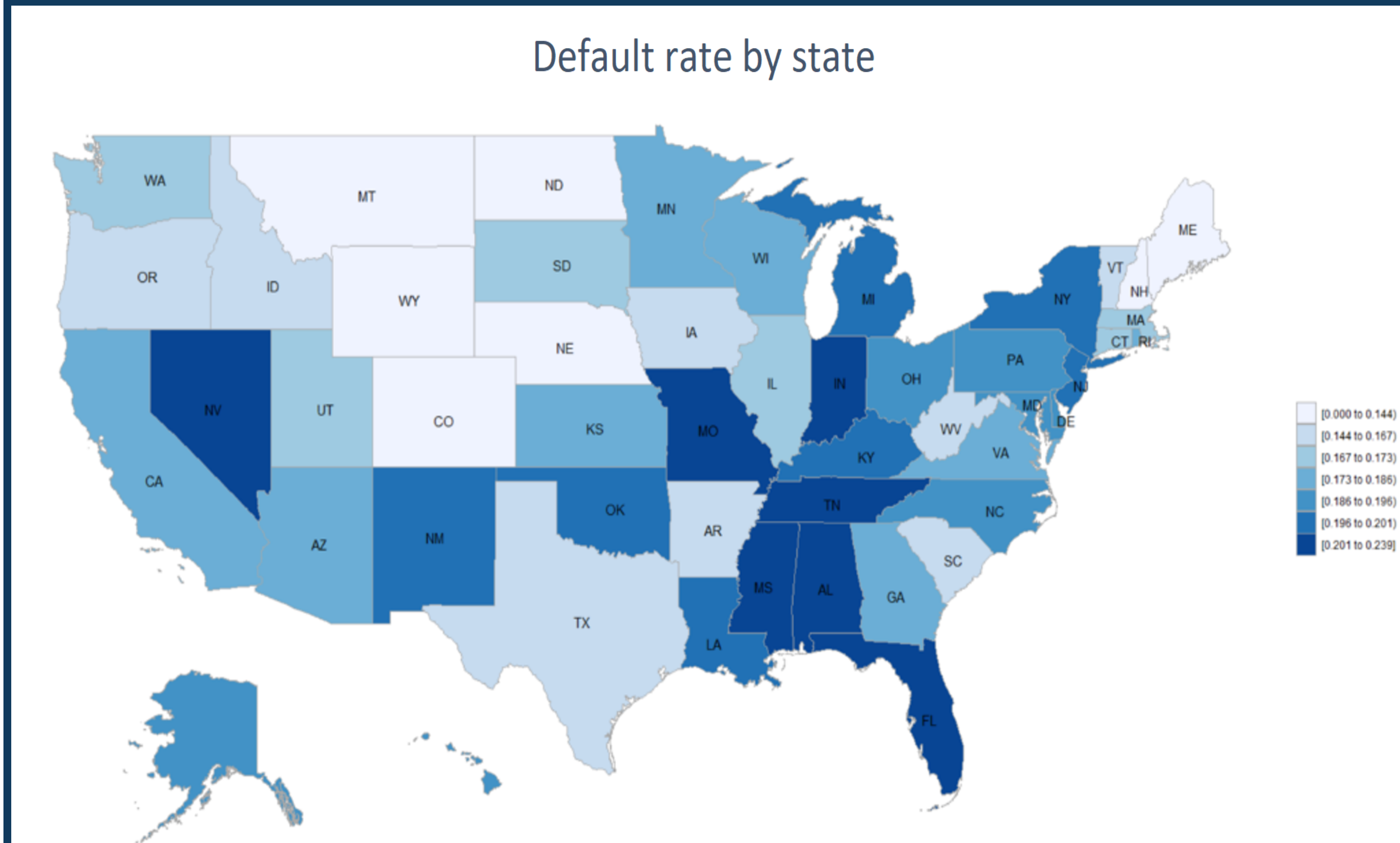
Why predicting default rate?

- Higher default rate compared to bank

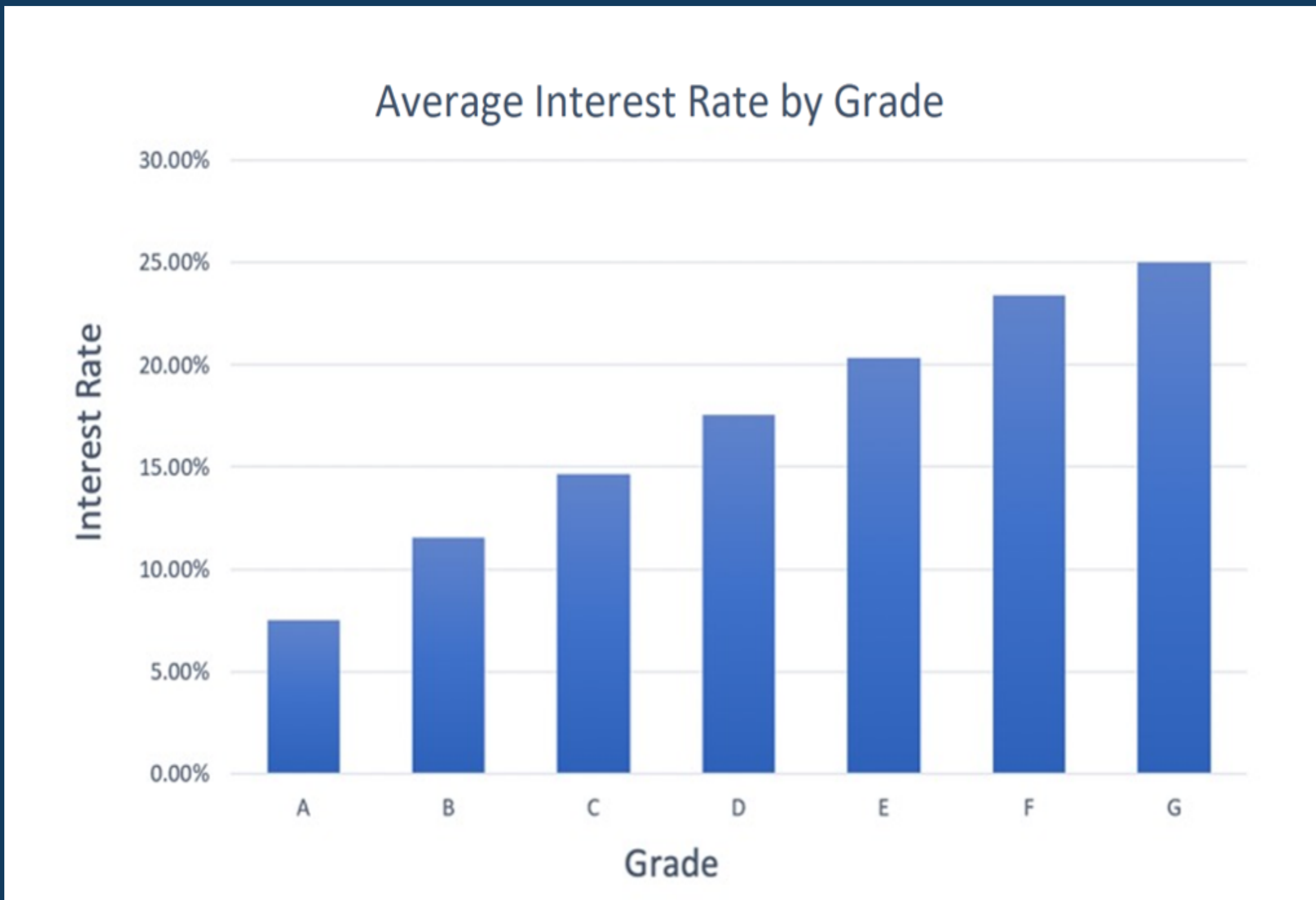
- As a prediction model for P2P companies

- Help customers to understand their chance of getting loan

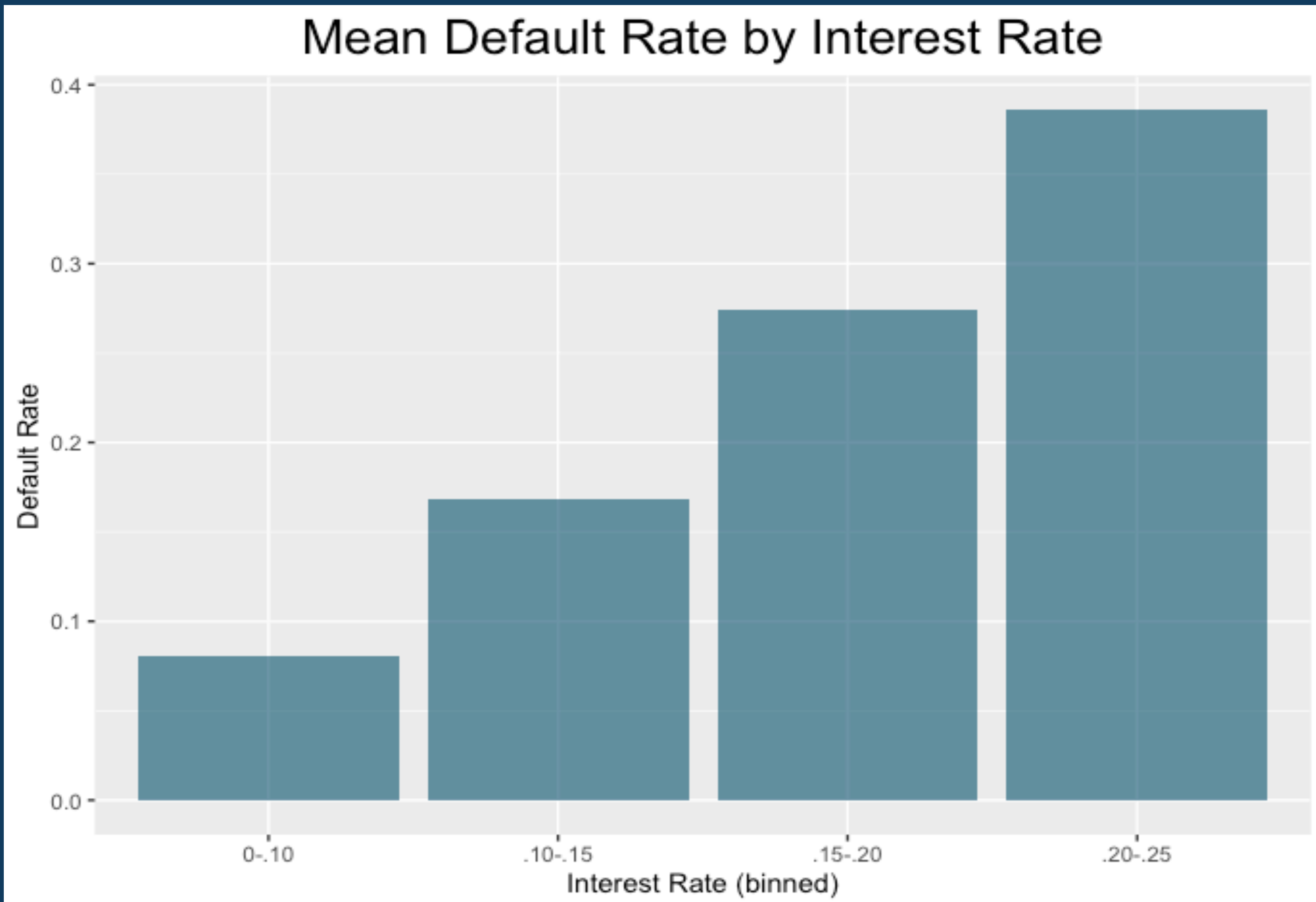
Data Visualization

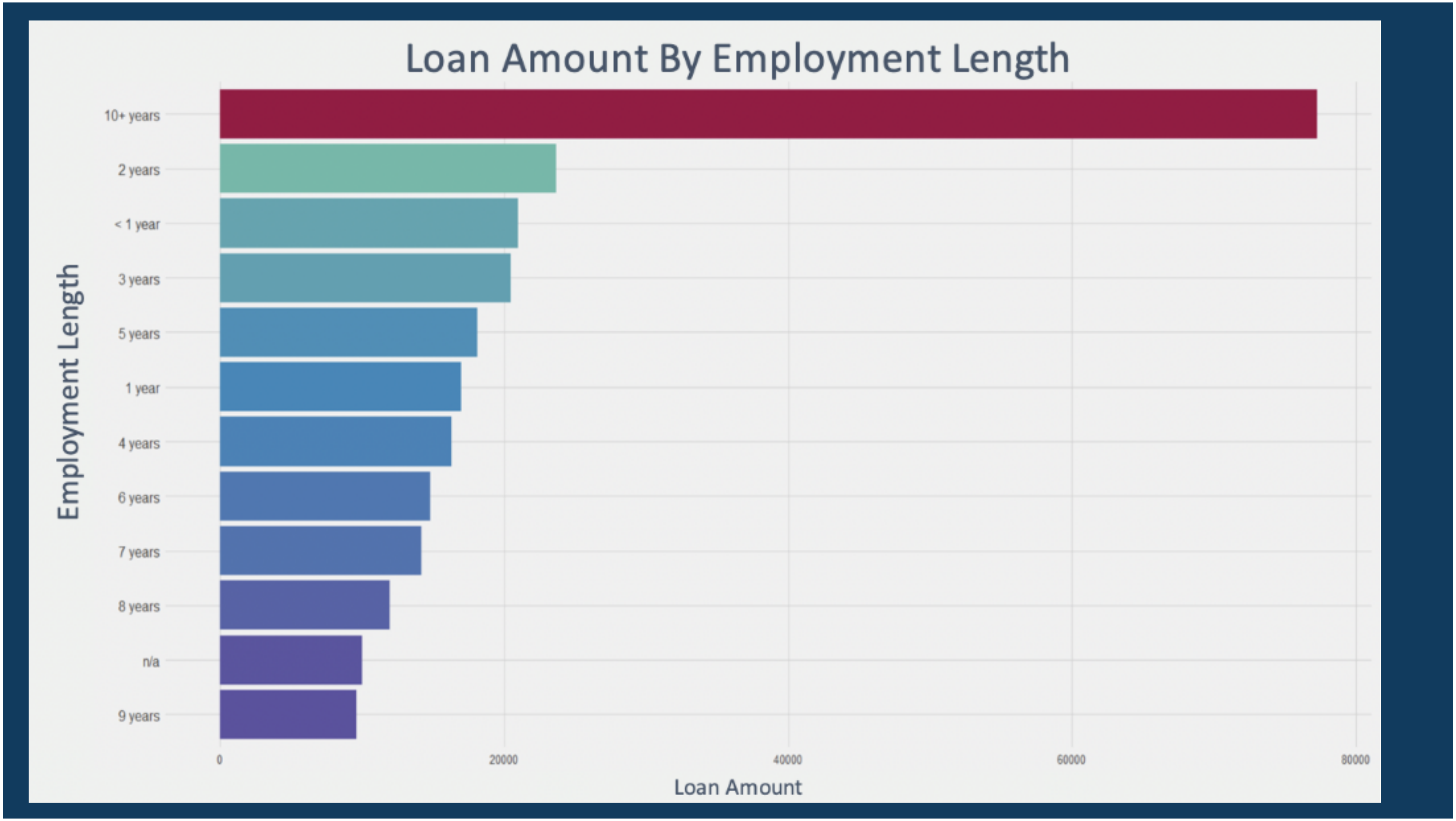


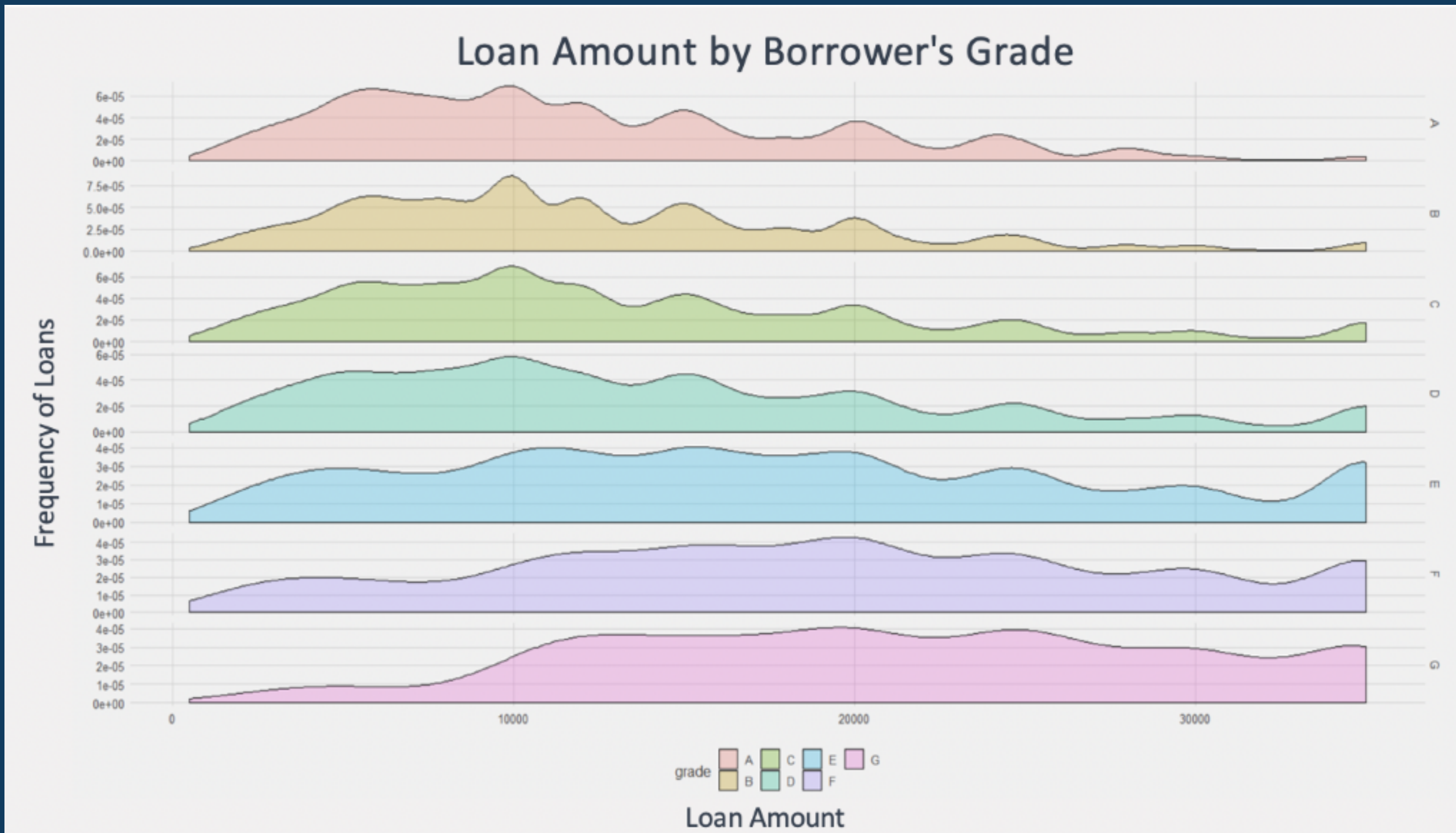
This chart shows the default rate per state. Darker colors indicate a higher average default rate. California sits in the middle, with Nevada, Florida, Alabama, Missouri, Indiana, Tennessee, and Mississippi holding the biggest default rates.



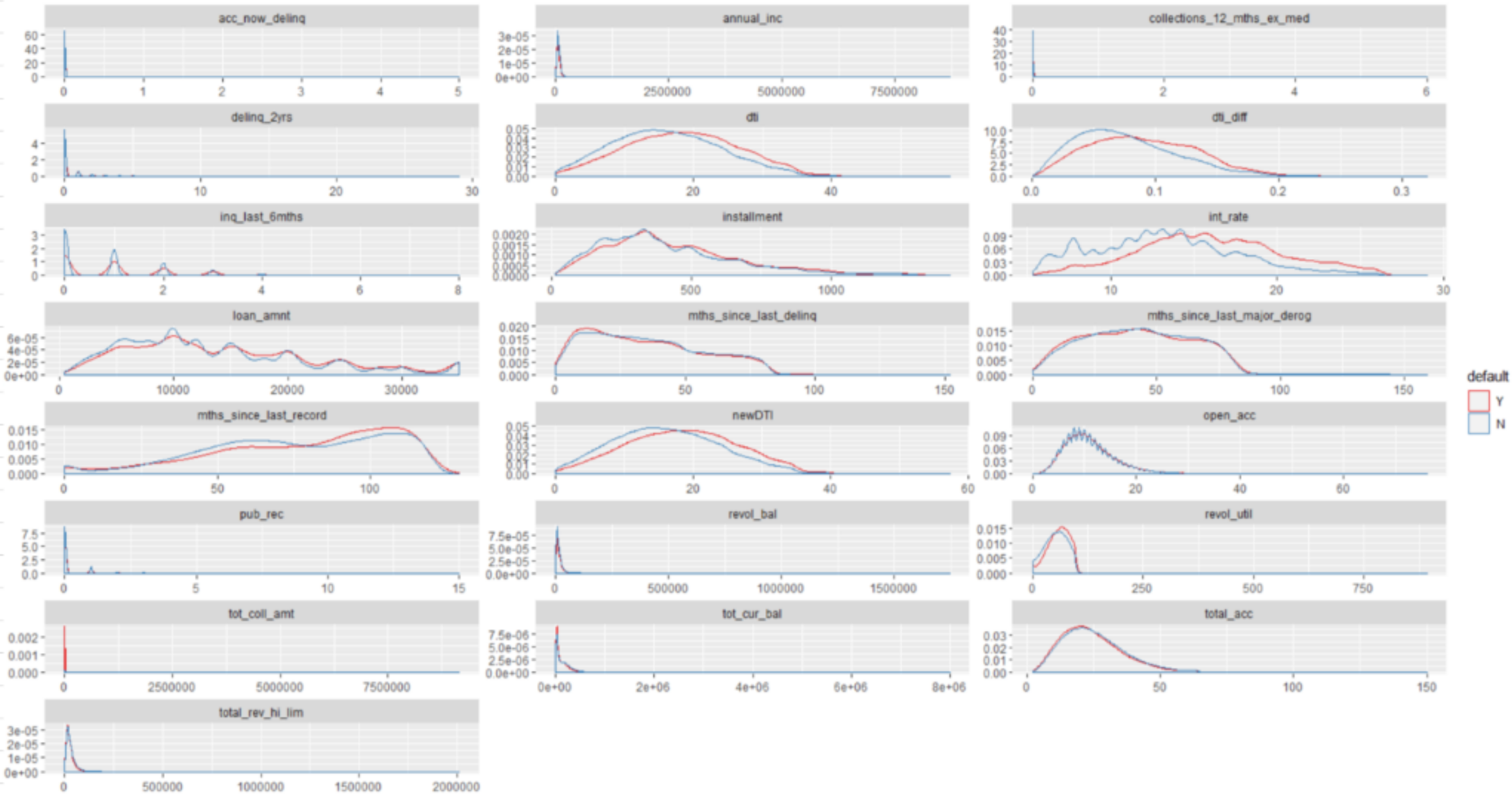
Average Interest Rate by Grade
30.00%
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
Interest Rate
A
B
C
D
E
F
G
Grade


Mean Default Rate by Interest Rate
Default Rate
0.4
0.3
0.2
0.1
0.0
0-.10
.10-.15
.15-.20
.20-.25
Interest Rate (binned)


Loan Amount By Employment Length
Employment Length
10+ years
2 years
< 1 year
3 years
5 years
1 year
4 years
6 years
7 years
8 years
n/a
9 years
0
20000
40000
60000
80000
Loan Amount


Loan Amount by Borrower's Grade
Frequency of Loans
Loan Amount
grade
A
B
C
D
E
F
G
0
10000
20000
30000

Density plot for Default



Features With The Most Information Gain

=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

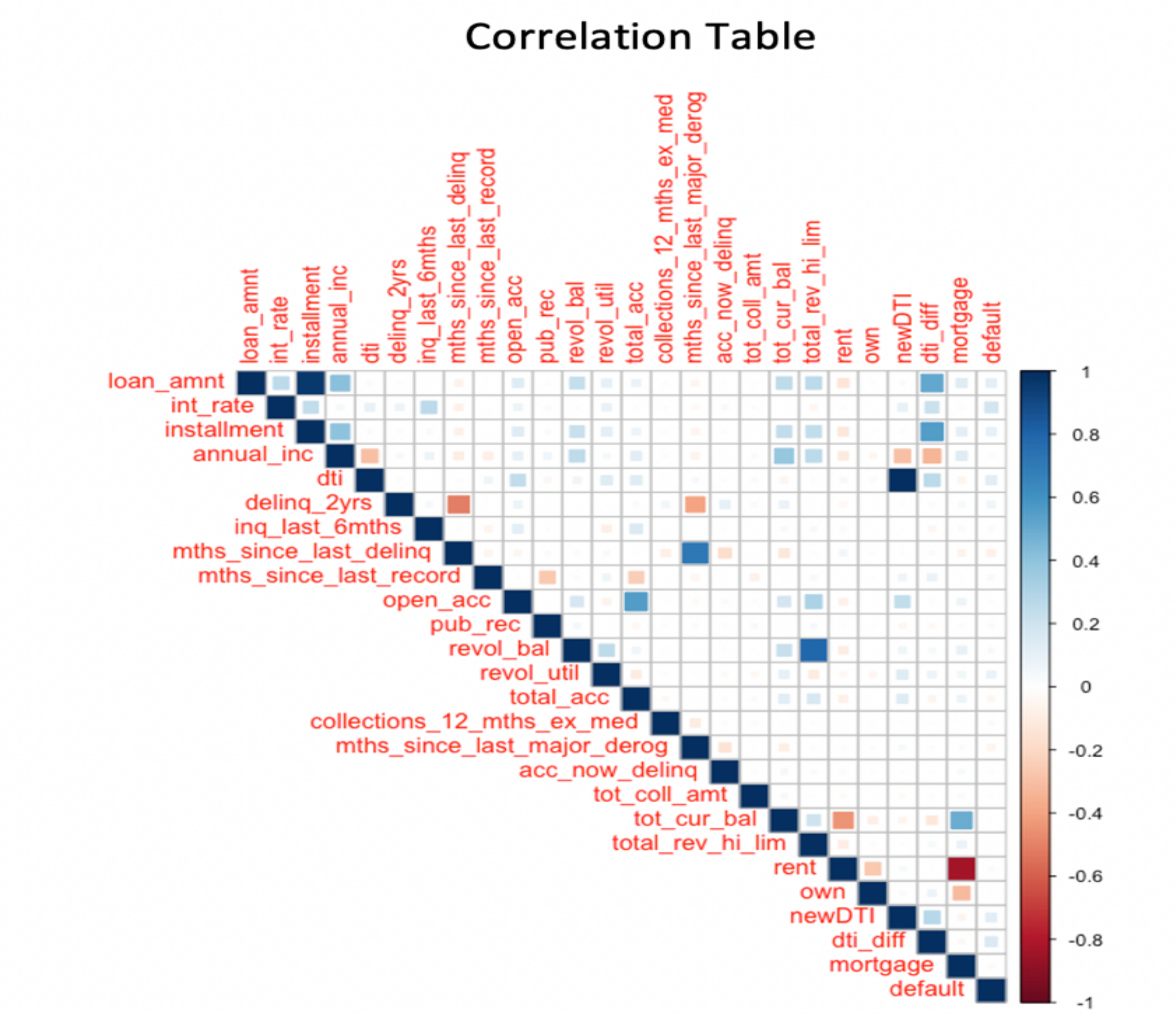
average merit	average rank	attribute
0.05 +- 0	1 +- 0	2 int_rate
0.017 +- 0	2.2 +- 0.6	29 grade_a
0.016 +- 0	3.3 +- 0.64	57 dti_diff
0.013 +- 0	6 +- 0	6 dti
0.011 +- 0	7 +- 0	52 thirtysix_mo
0.009 +- 0	8 +- 0	30 grade_b
0.008 +- 0	9 +- 0	32 grade_e
0.007 +- 0	10.4 +- 0.49	12 revol_util
0.007 +- 0	10.6 +- 0.49	4 annual_inc
0.006 +- 0	12.2 +- 0.4	17 tot_cur_bal
0.006 +- 0	12.9 +- 0.54	33 grade_f
0.005 +- 0	13.9 +- 0.3	1 loan_amnt
0.005 +- 0	15 +- 0	5 verification_status
0.004 +- 0	16.5 +- 0.5	22 mortgage _
0.004 +- 0	18.1 +- 0.54	20 rent
0.004 +- 0	18.5 +- 1.02	3 installment
0.003 +- 0	20 +- 0	18 total_rev_hi_lim
0.002 +- 0	21 +- 0	8 inq_last_6mths
0.002 +- 0	22 +- 0	34 grade_g
0.001 +- 0	23 +- 0	48 credit_age
0.001 +- 0	24 +- 0	13 total_acc
0.001 +- 0	25 +- 0	51 application_type_dummy
0.001 +- 0	26.4 +- 0.49	53 Last_Record
0.001 +- 0	27.2 +- 0.87	37 Small_Business
0.001 +- 0	27.4 +- 0.66	27 emp_length_10plus
0.001 +- 0	29 +- 0	50 Derog
0.001 +- 0	30.2 +- 0.4	49 Delinq

These features are kept in the final analysis.

Features With The Least Information Gain

0 +- 0	31 +- 0.63	35 Credit_Card
0 +- 0	31.8 +- 0.4	11 revol_bal
0 +- 0	33.2 +- 0.4	16 tot_coll_amt
0 +- 0	34 +- 0.63	7 delinq_2yrs
0 +- 0	34.8 +- 0.4	41 Home_Improvement
0 +- 0	36.8 +- 0.98	9 open_acc
0 +- 0	37 +- 1	36 Car
0 +- 0	38.1 +- 0.83	31 grade_c
0 +- 0	38.1 +- 0.94	14 collections_12_mths_ex_med
0 +- 0	40 +- 0	38 Purpose_Other
0 +- 0	41.2 +- 0.6	44 Moving
0 +- 0	42.5 +- 0.92	43 Medical
0 +- 0	43.1 +- 0.83	10 pub_rec
0 +- 0	43.2 +- 0.75	24 emp_length_lessthan1
0 +- 0	45.2 +- 0.4	25 emp_length_1to5
0 +- 0	45.8 +- 0.4	42 Major_Purchase
0 +- 0	49 +- 0.63	47 Renewable
0 +- 0	49.9 +- 2.66	21 own
0 +- 0	50.4 +- 3.38	46 House
0 +- 0	50.8 +- 2.09	26 emp_length_6to9
0 +- 0	51.2 +- 2.75	40 Debt_Consolidation
0 +- 0	52.2 +- 1.78	45 Vacation
0 +- 0	52.5 +- 3.64	28 emp_length_is_NA
0 +- 0	52.8 +- 3.16	15 acc_now_delinq
0 +- 0	53 +- 2.37	39 Wedding
0 +- 0	53.2 +- 0.87	23 other

These features are removed in the final analysis.

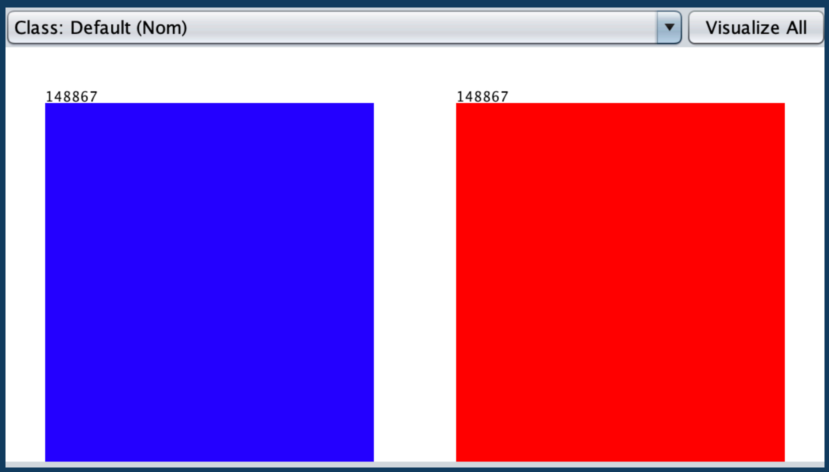


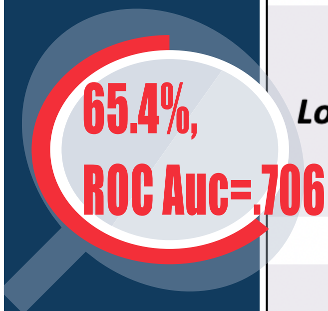
Interesting correlations between variables

- Dti (debt-to-income ratio)
- **Dti_diff (=difference between dti and the new dti)**
- **Int_rate**
- Loan_amnt
- Rent and tot_cur_bal (total current balance of all accounts)

Models

Oversampling





		Oversampling				No resampling	
		With *dti_diff*		Without *dti_diff*			
		Class: 1	Class: 0	Class: 1	Class: 0	Class:1	Class: 0
Logistic Regression	**precision:**	0.348	0.864	0.345	0.865	0.544	0.786
	recall:	0.637	0.659	0.647	0.649	0.058	0.986
	weighted accuracy:	65.41%		64.85%		78.00%	
	ROC Area	0.706		0.705			
J48 Decision Tree	**precision:**	0.333	0.865	0.332	0.864	0.583	0.784
	recall:	0.658	0.624	0.657	0.622	0.042	0.991
	weighted accuracy:	63.13%		63.00%		0.78%	
	ROC Area	0.684		0.684		0.641	
Random Forest	**precision:**	0.367	0.856	0.869	0.687	0.78	0.997
	recall:	0.577	0.716	0.605	0.687	0.692	0.0187
	weighted accuracy:	68.50%		66.90%		78%	
	ROC Area	0.396		0.412		0.3	
Naïve Bayes	**precision:**	0.338	0.865	0.333	0.866	0.835	0.384
	recall:	0.651	0.636	0.666	0.619	0.793	0.451
	weighted accuracy:	63.90%		62.90%		0.74%	
	ROC Area	0.695		0.693		0.69	

Conclusion

- Random Forest is the best model, followed by Logistic Regression.

- Recommendation for investors: buy loans which are in Grade A and have smaller loan amounts and from borrowers who have a low dti (debt-to-income ratio).

THANK YOU

PAUL MERAGE

Focus on Innovation

PRODUCED
BY GROUP 10

How to find us

Xueting(Lexie) Li	Sammi Chang	Jialu(Molly) Li	Junguo(Nick) He
			
xuetinl5@uci.edu	sammic@uci.edu	jialul14@uci.edu	junguoh@uci.edu
https://www.linkedin.com/in/xuetinglio330/	https://www.linkedin.com/in/sammichang/	https://www.linkedin.com/in/jialul/	https://www.linkedin.com/in/junguo-he/