

LENDING CLUB'S LOAN DATA & ML MODELING



By Affad, Ali, & Marie

RESEARCH QUESTION

A black and white graphic of a human head profile facing left. Inside the head, a detailed white circuit board is visible, representing the brain or mind.

What features
contributes to
Lending Club's
loan grade?



1st & Largest P2P lender

Charges fees for loans

& Lends \$\$\$

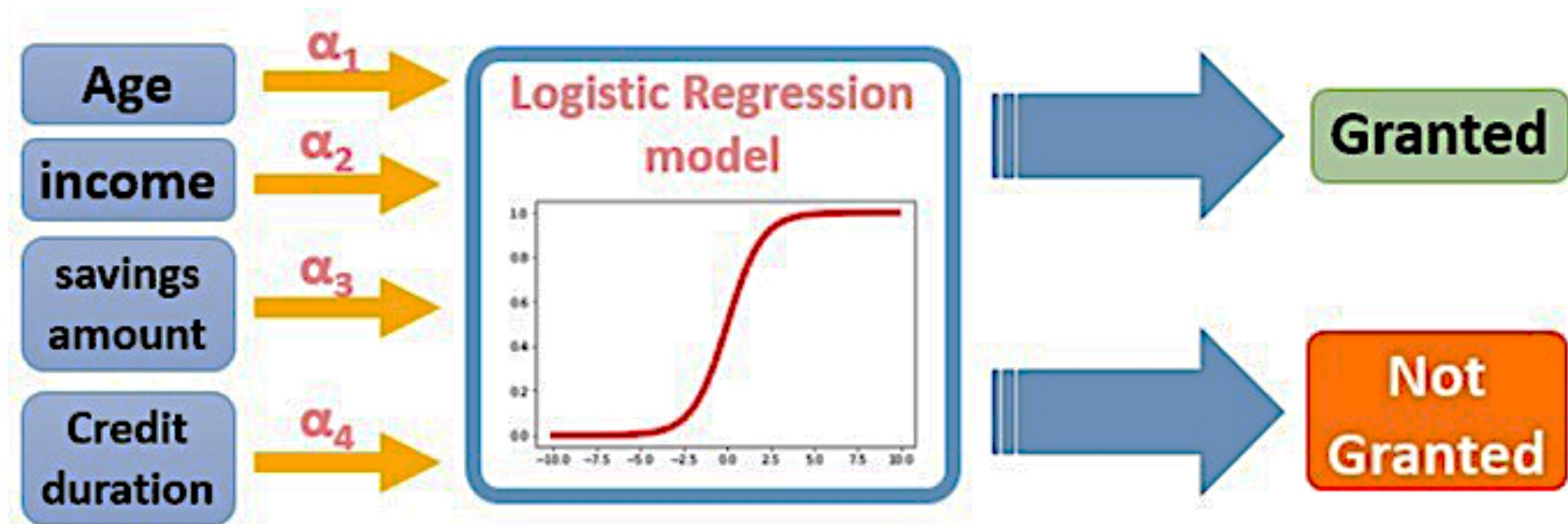
Youssef Fenjiro

“Machine Learning for

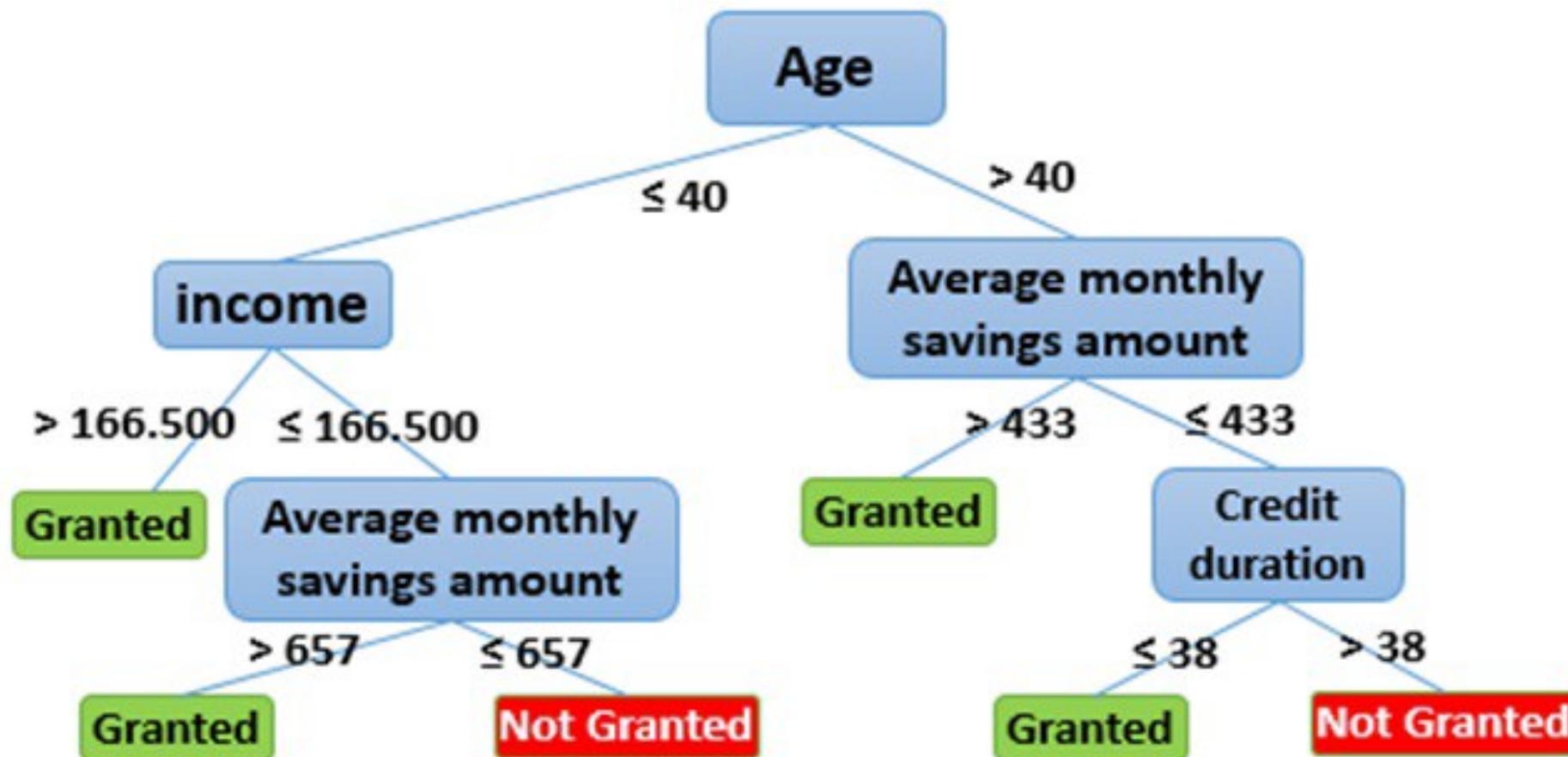
Banking: Loan approval

use case,” July 24, 2018

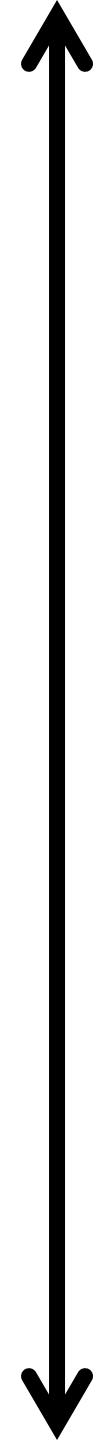
Source: Fenjiro's "Machine Learning for Banking: Loan Approval," 2018



Source: Fenjiro's "Machine Learning for Banking: Loan Approval," 2018



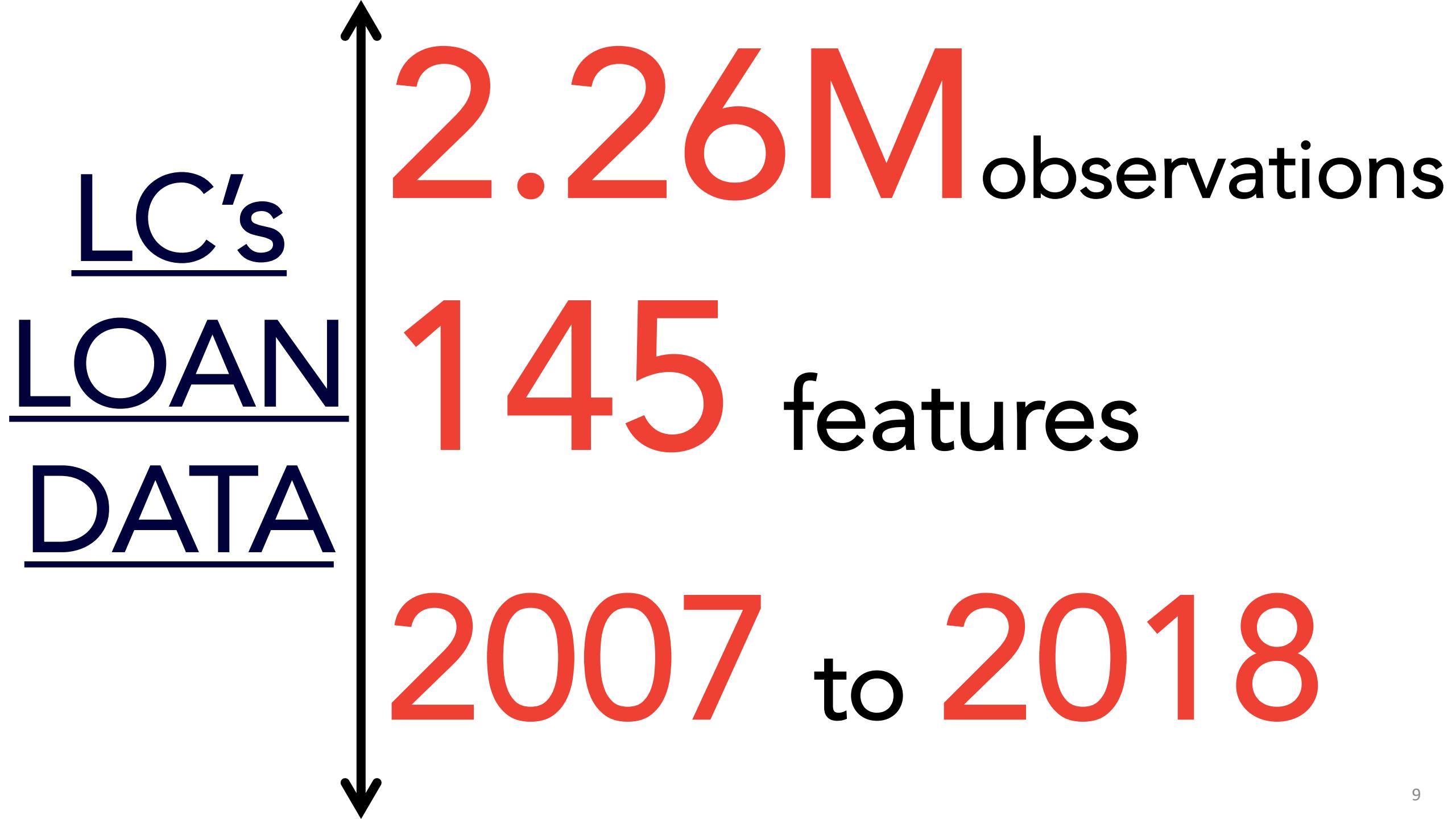
WHAT IS
THE LOAN
GRADE?

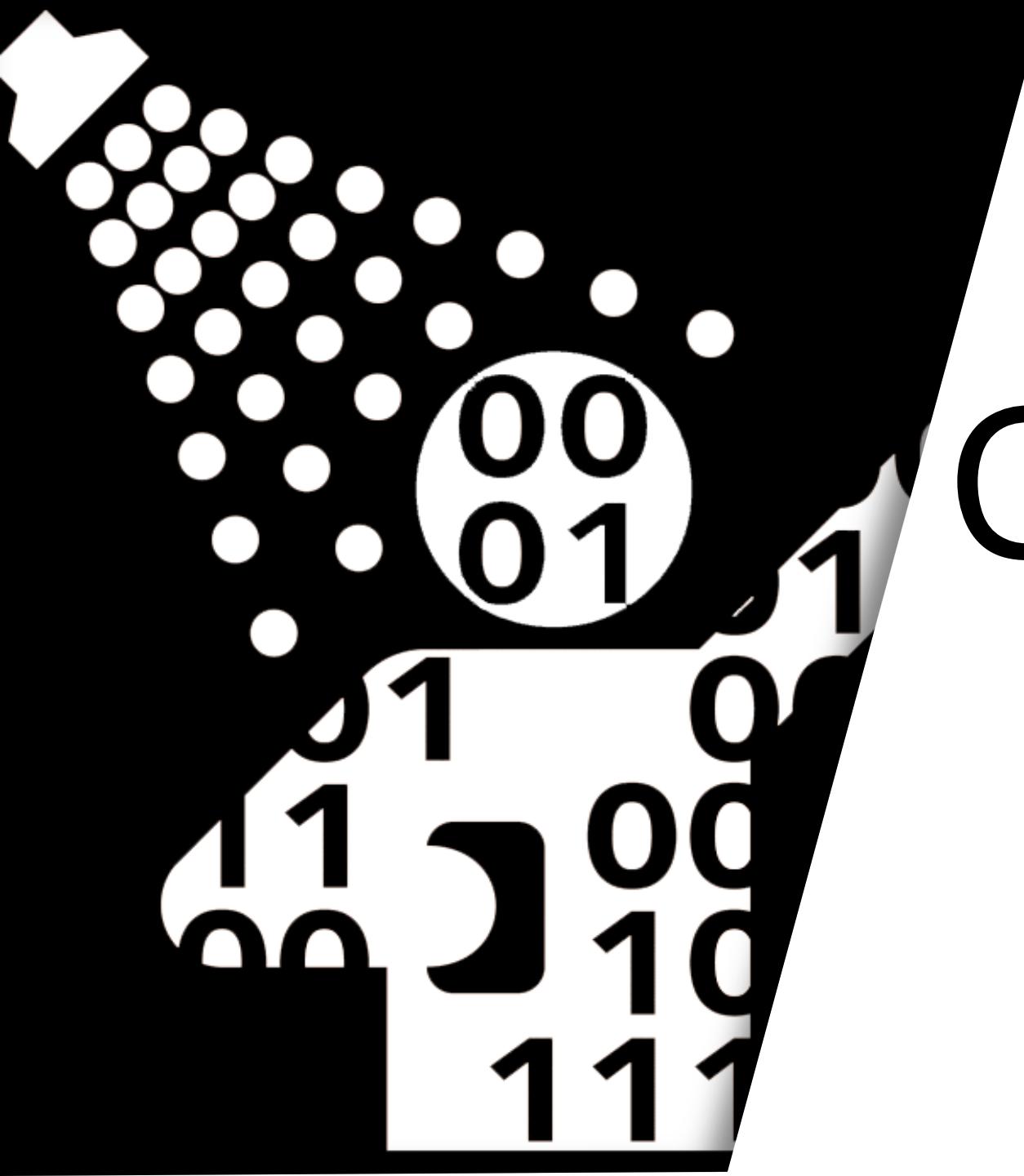


Rate information
Lending Club's interest rates take into account credit risk and market conditions. The final interest rate for each loan grade is the result of the following equation:

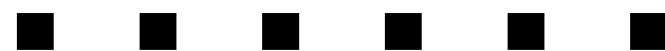
Lending Club Account

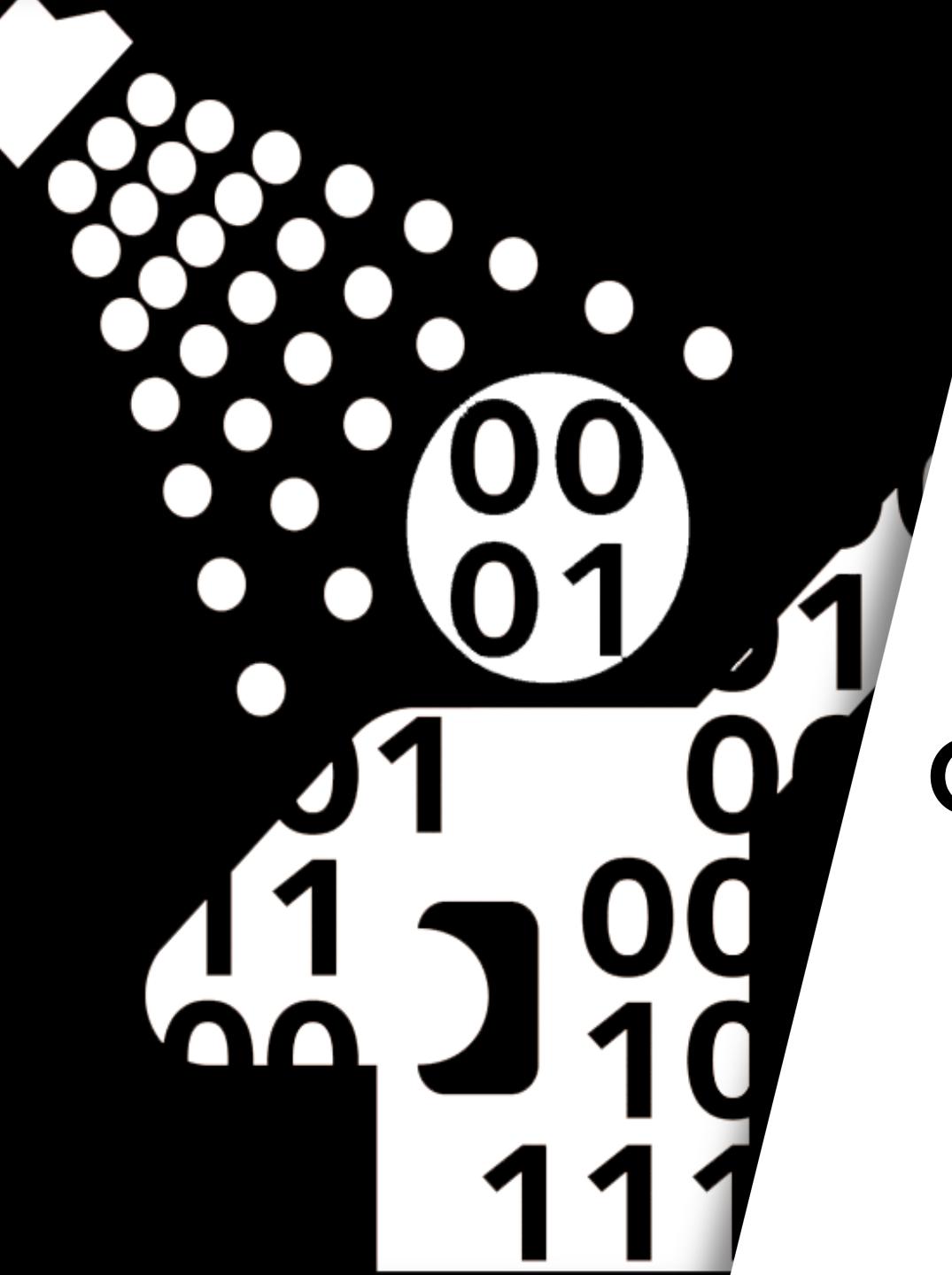
LOAN GRADE	SUB-GRADE	LENDING CLUB BASE RATE	ADJUSTMENT FOR RISK & VOLATILITY	INTEREST RATE
A	1	5.05%	1.41%	6.46%
	2	5.05%	1.97%	7.02%
	3	5.05%	2.51%	7.56%
	4	5.05%	3.14%	8.19%
	5	5.05%	3.76%	8.81%
B	A	3	5.05%	2.51% 7.56%
		4	5.05%	3.14% 8.19%
		5	5.05%	3.76% 8.81%
	B	1	5.05%	5.28% 10.33%
		2	5.05%	5.97% 11.02%
C	B	3	5.05%	6.66% 11.71%
		4	5.05%	7.25% 12.69% 17.74%
	1	5.05%		12.92% 17.97%
	2	5.05%		14.95% 20.00%
	3	5.05%		17.45% 22.50%
D	4	5.05%		19.95% 25.00%
	5	5.05%		23.75% 28.80%
	1	5.05%		23.85% 28.90%
	2	5.05%		23.85% 28.90%
	3	5.05%		23.87% 28.92%
E	4	5.05%		23.90% 28.95%
	5	5.05%		23.92% 28.97%
	1	5.05%		23.95% 29.00%
	2	5.05%		
	3	5.05%		





AFTER
CLEANING





2016/2018

~500,000

observations (each)

& 65 features

TRENDS

IN THE

DATA?



Grade

A

433,027

B

663,557

C

650,053

D

324,424

E

135,639

F

41,800

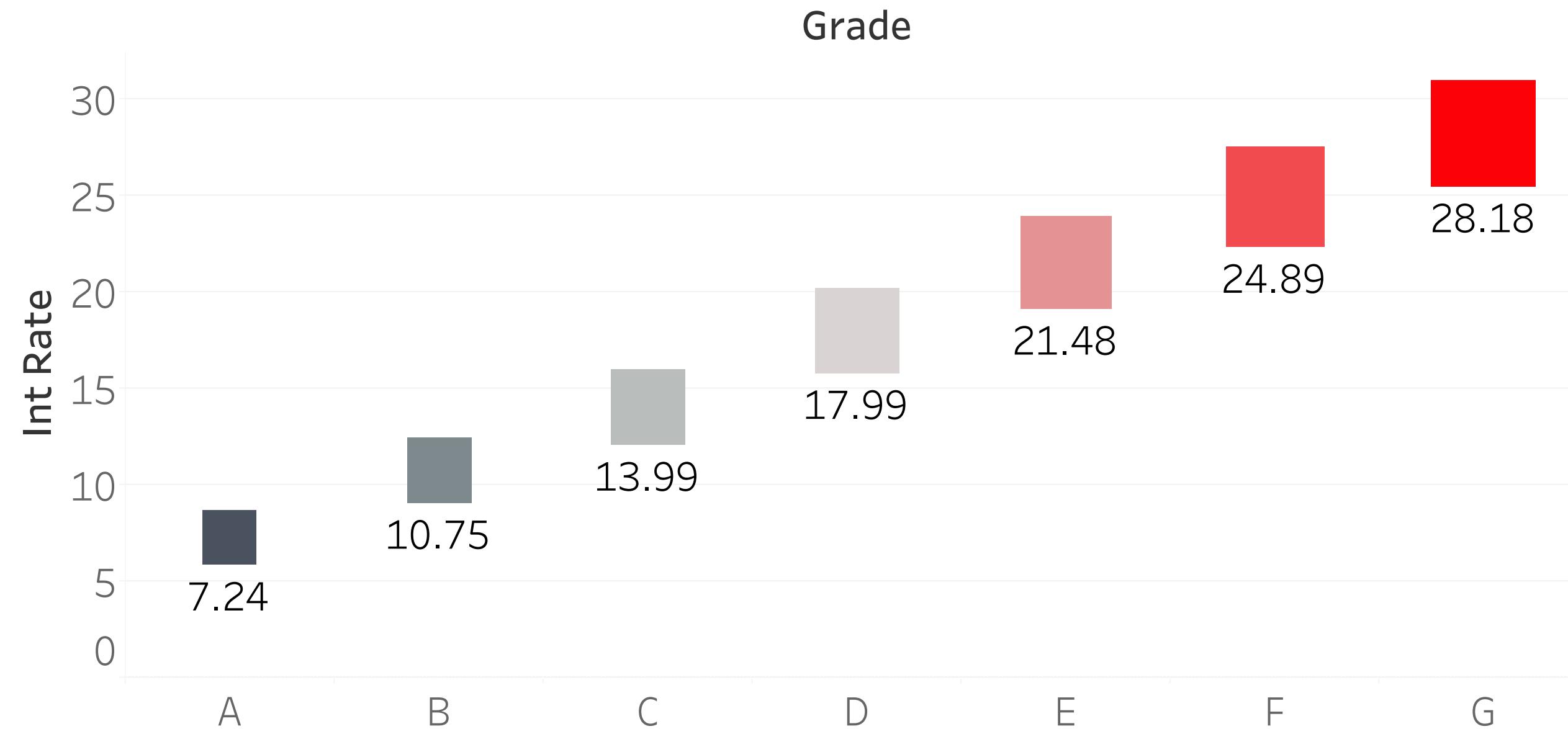
G

12,168

AMOUNT
OF LOANS
PER EACH
GRADE

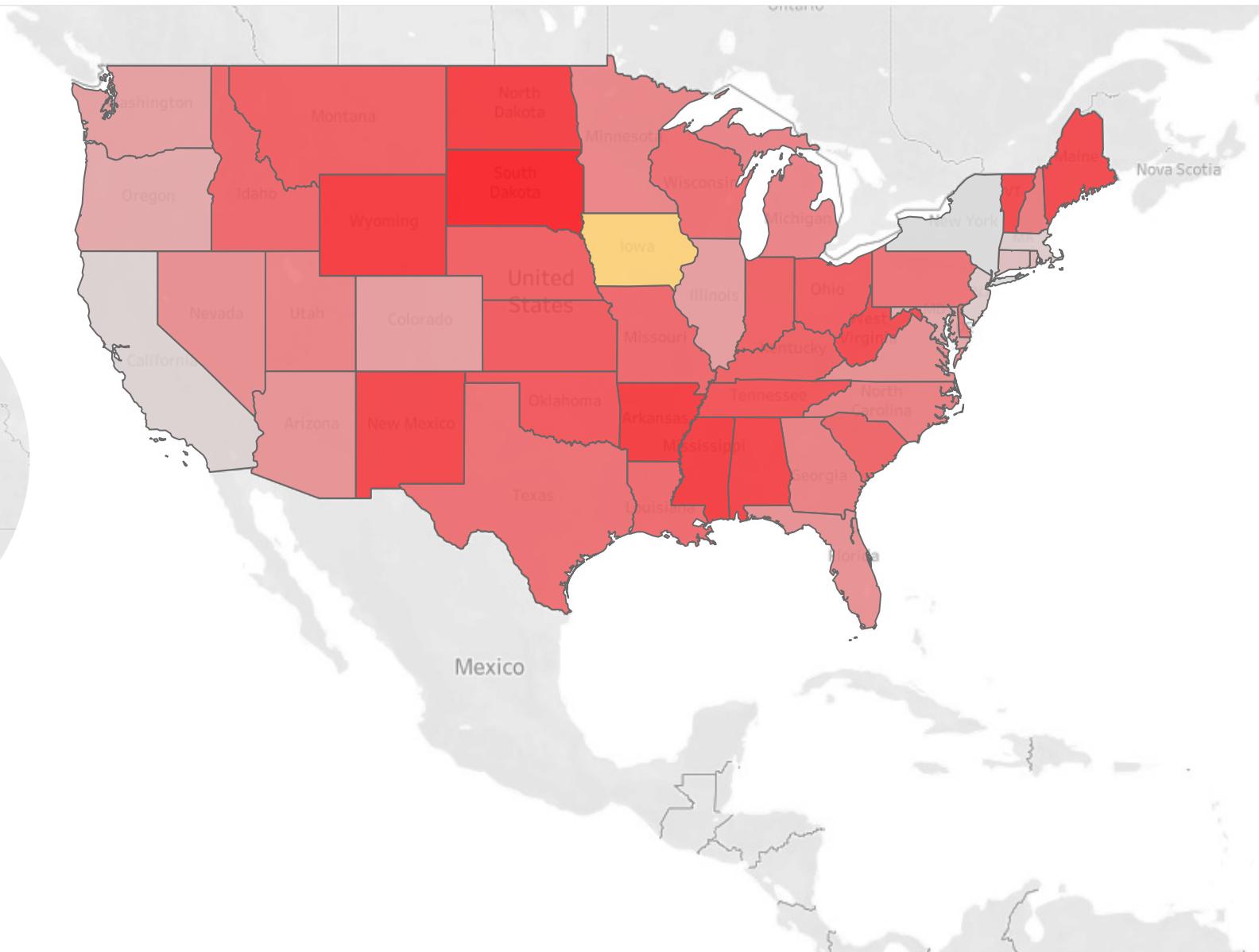
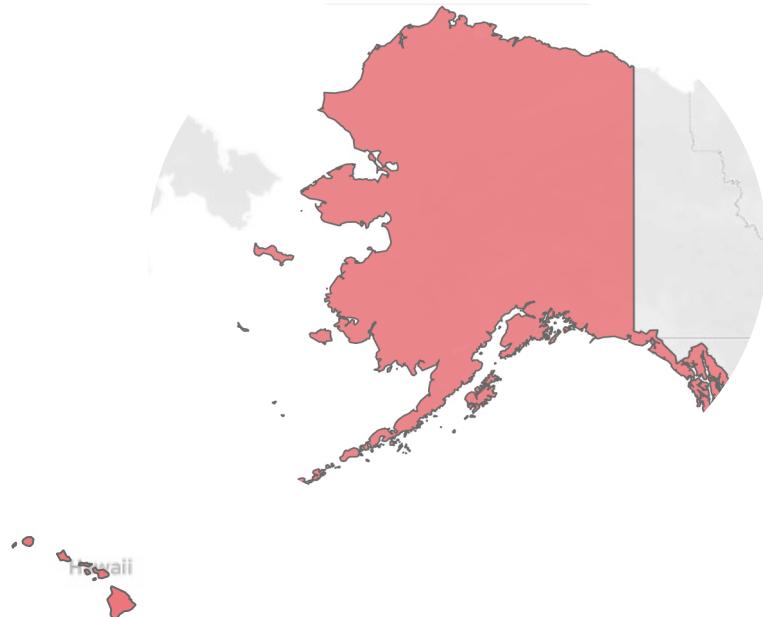
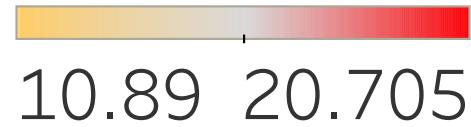


Grade to Median Interest Rate - All Loan Data

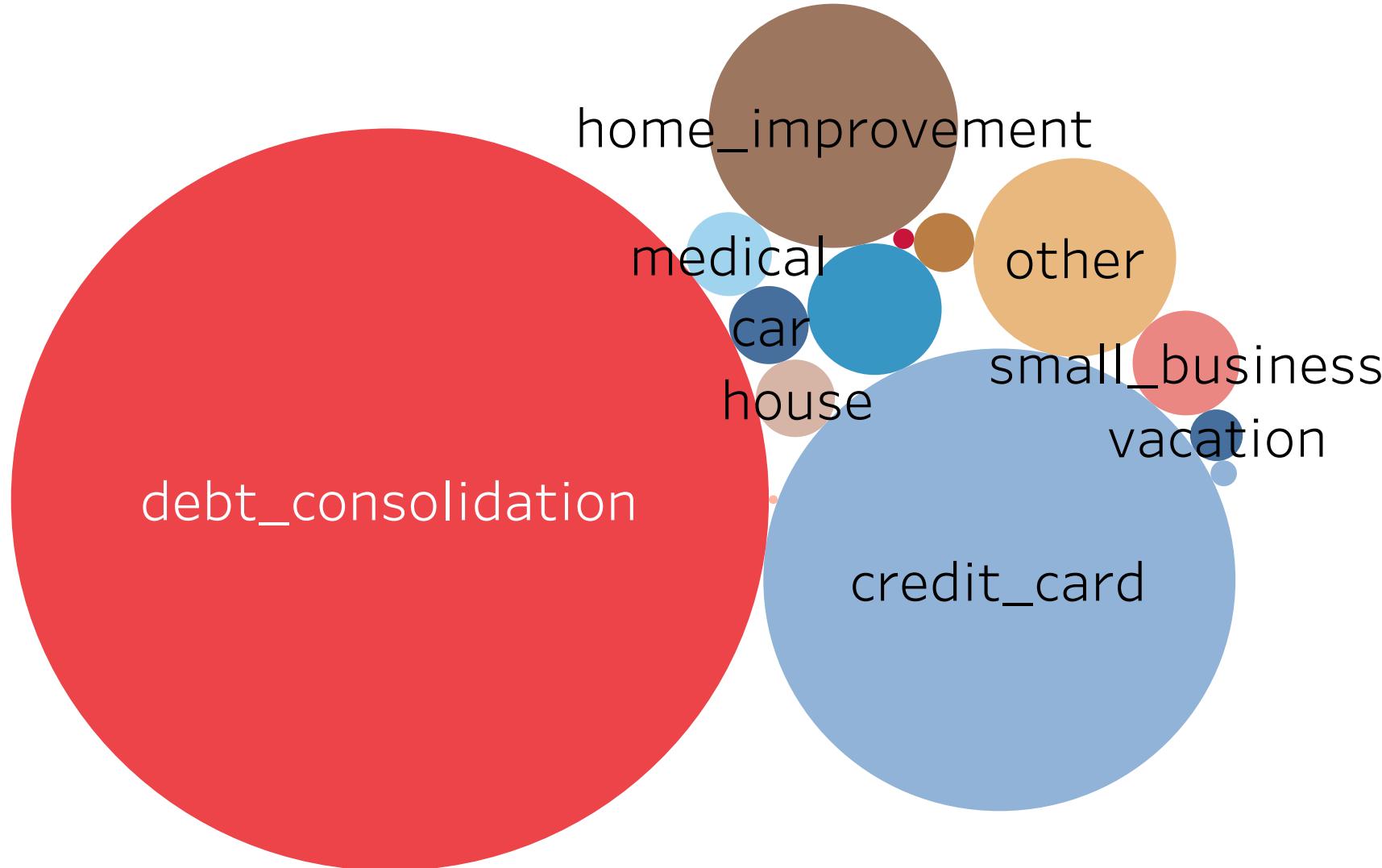


Median DTI by US State - All

Dti

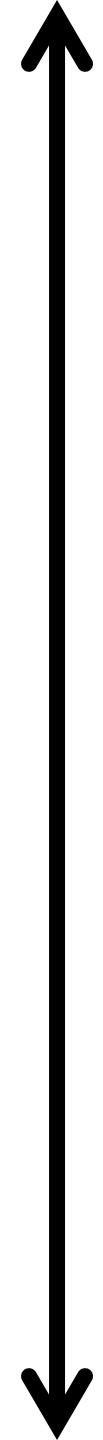


Amount of Loans in Each Category - All

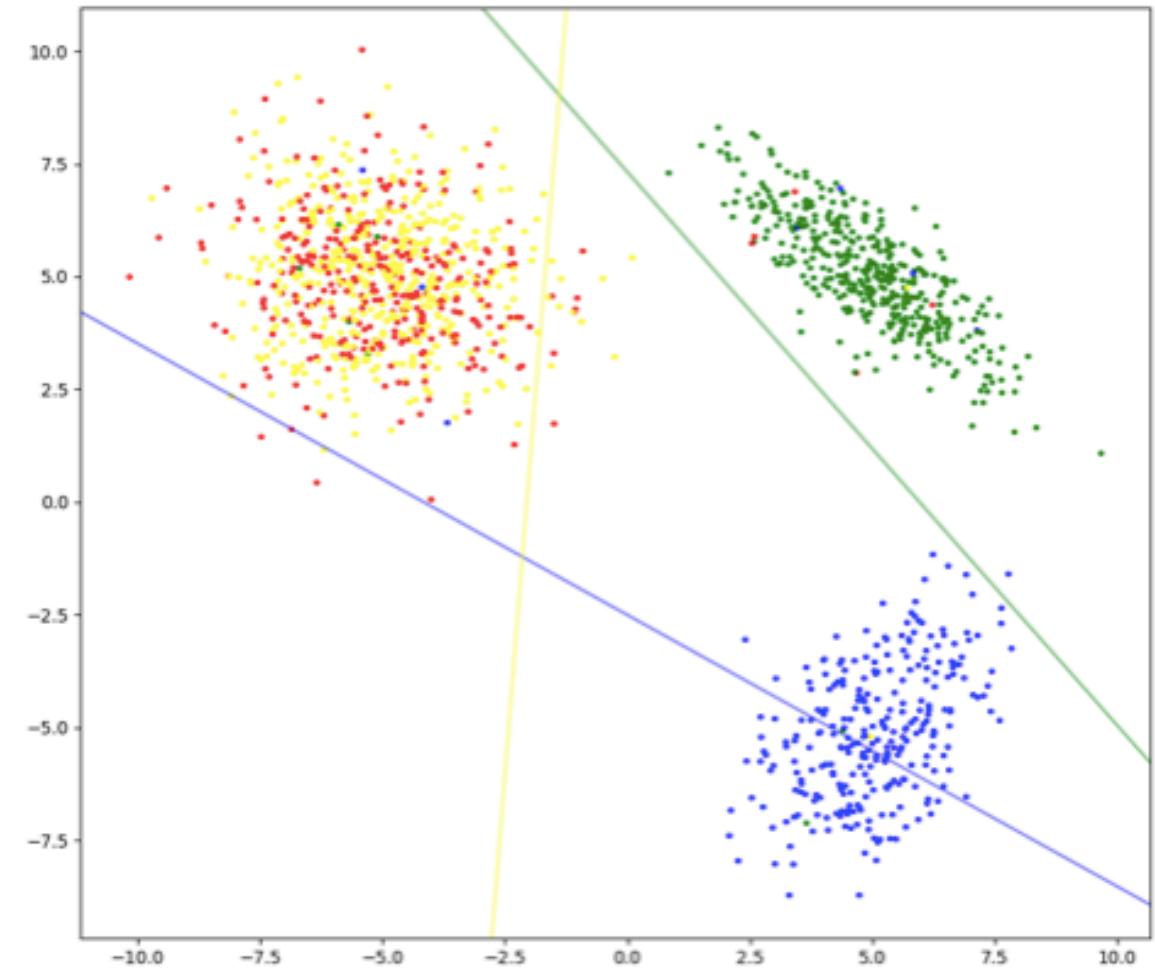
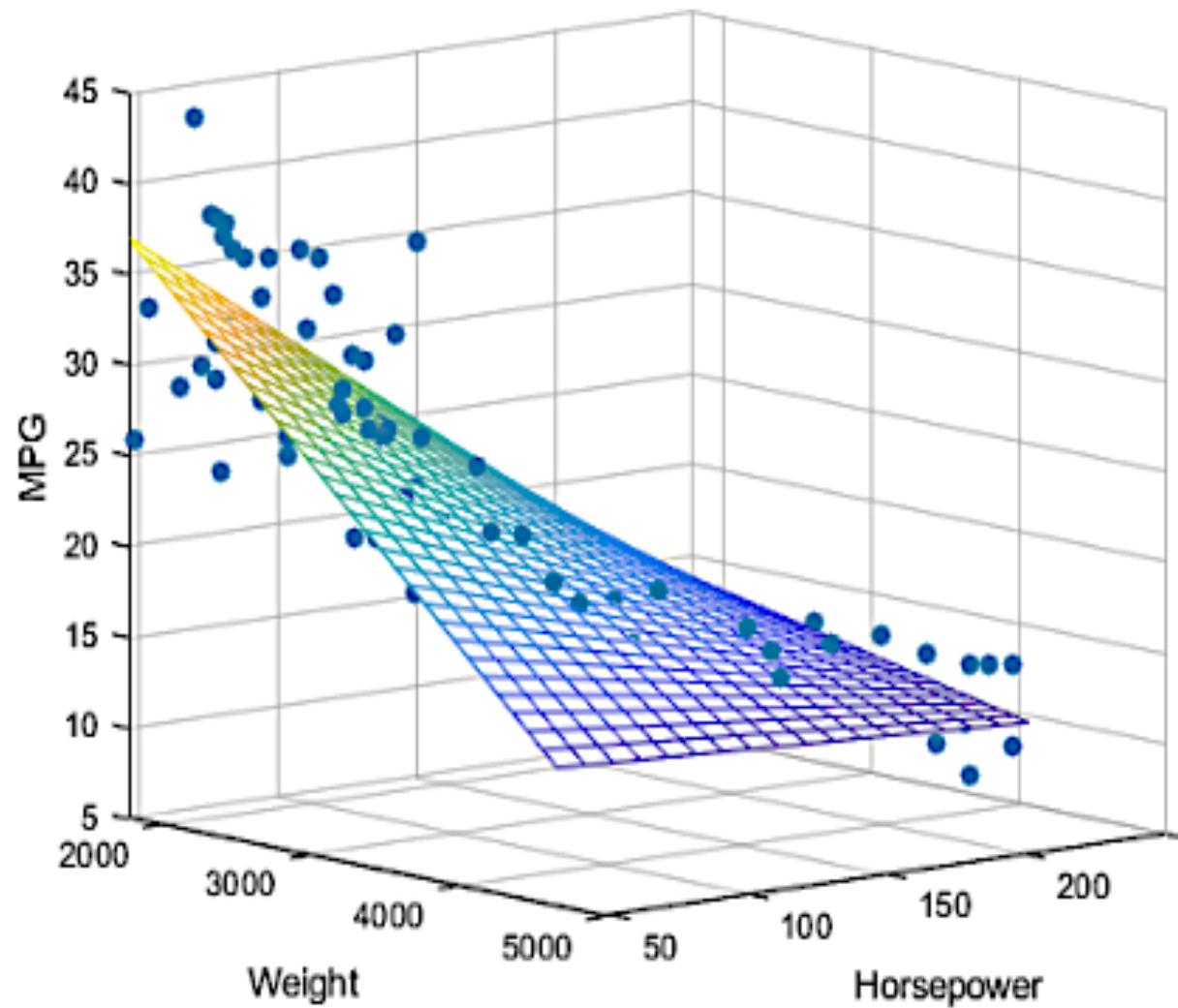


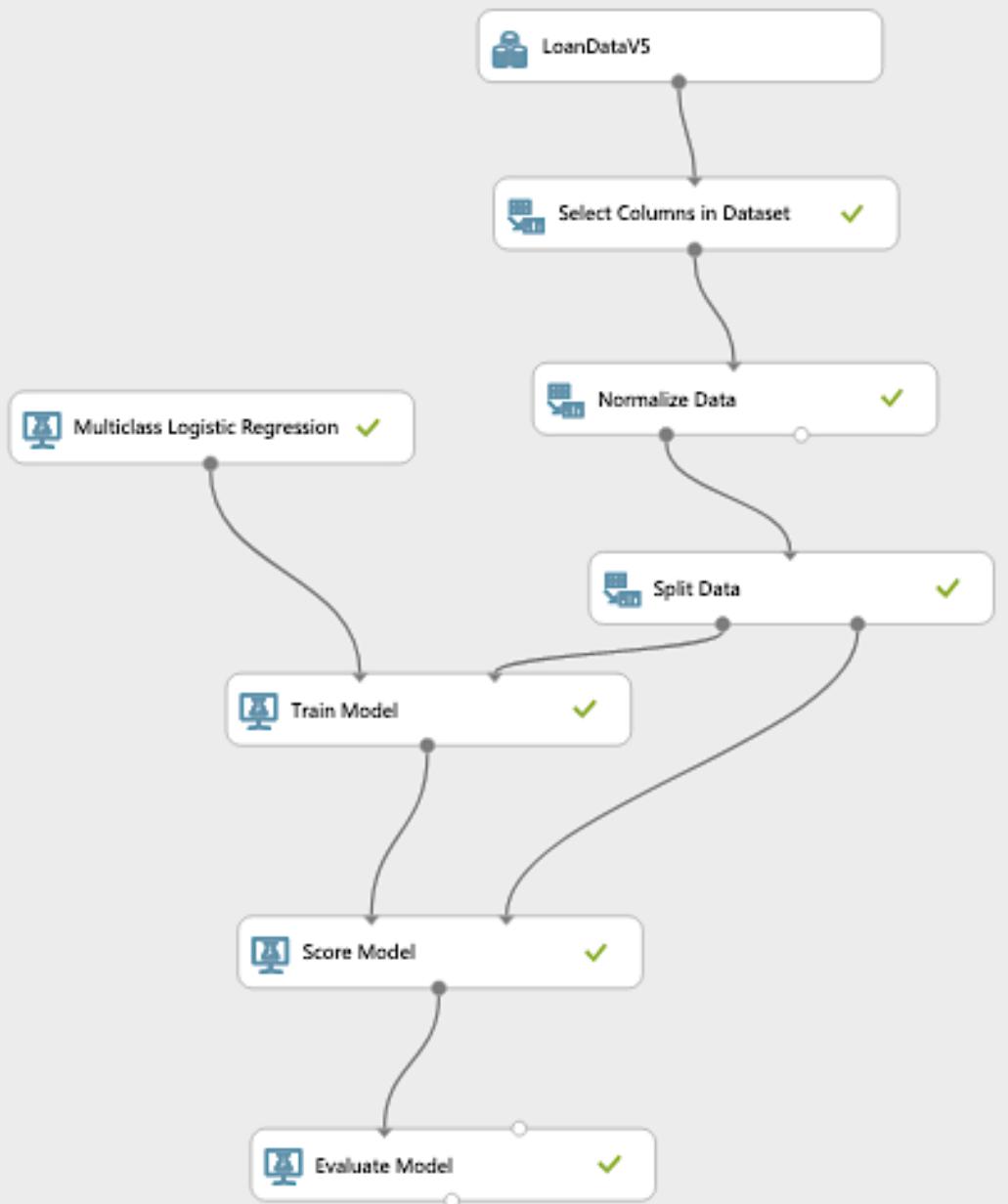
- Purpose**
- debt_consolidation
 - educational
 - credit_card
 - house
 - car
 - major_purchase
 - medical
 - home_improvement
 - renewable_energy
 - moving
 - other
 - small_business
 - vacation
 - wedding

MODELS



MULTICLASS LOGISTIC REGRESSION





2018

MULTICLASS

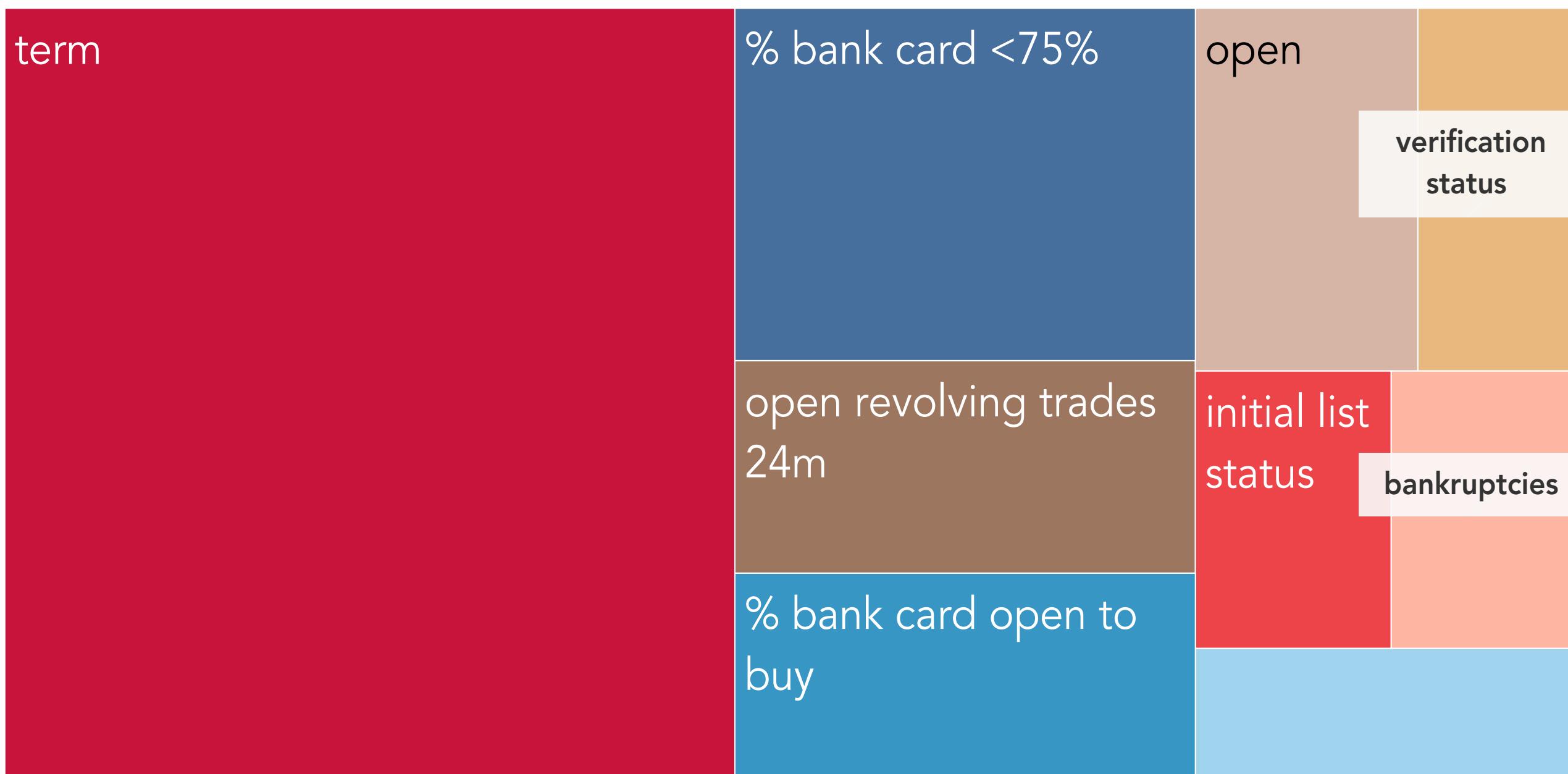
LOGISTIC

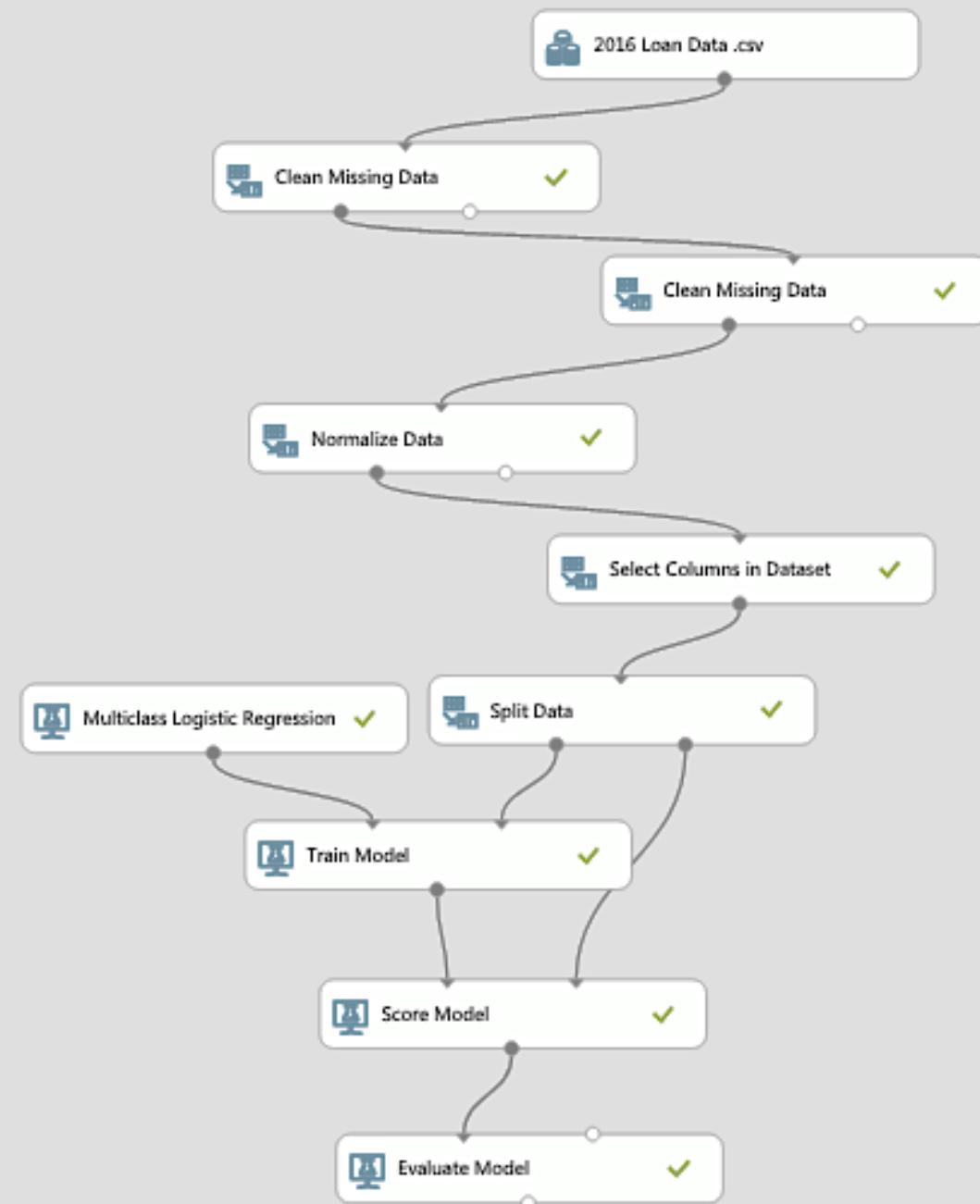
REGRESSION

2018 RESULTS

		Predicted Class						
		0	1	2	3	4	5	G
Actual Class	0	72.2%	20.4%	7.2%	0.1%	0.0%		
	1	28.4%	42.1%	28.3%	1.1%	0.2%		
	2	11.7%	29.4%	53.4%	4.9%	0.6%		
	3	6.8%	18.9%	59.8%	13.5%	1.0%		
	4	4.9%	12.0%	60.8%	19.2%	3.1%		
	5	2.6%	6.1%	51.3%	39.9%	0.1%		
	G	6.2%	9.3%	50.9%	33.5%			

2018 MCLR Features of Significance





2016

MULTICLASS

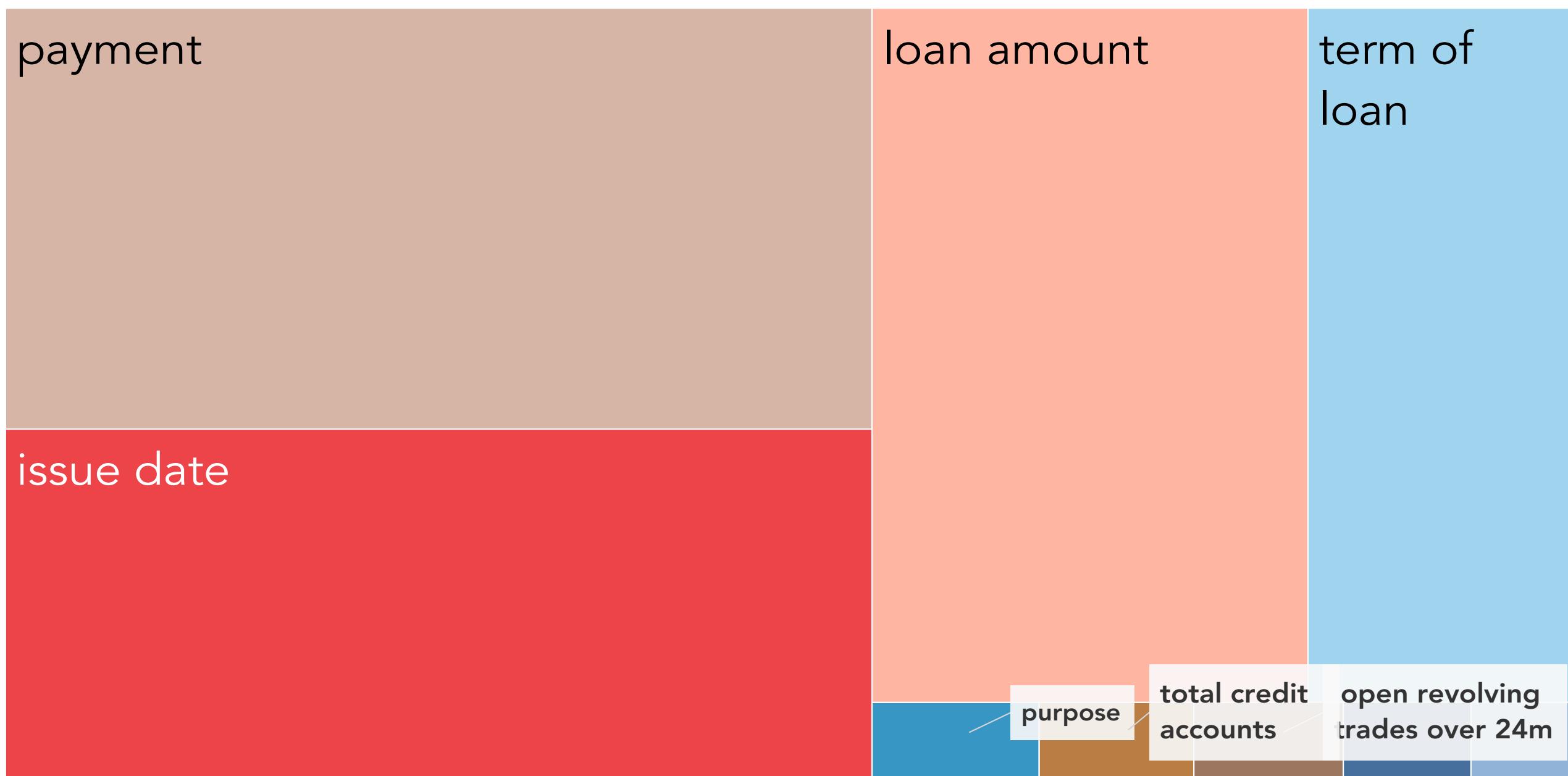
LOGISTIC

REGRESSION

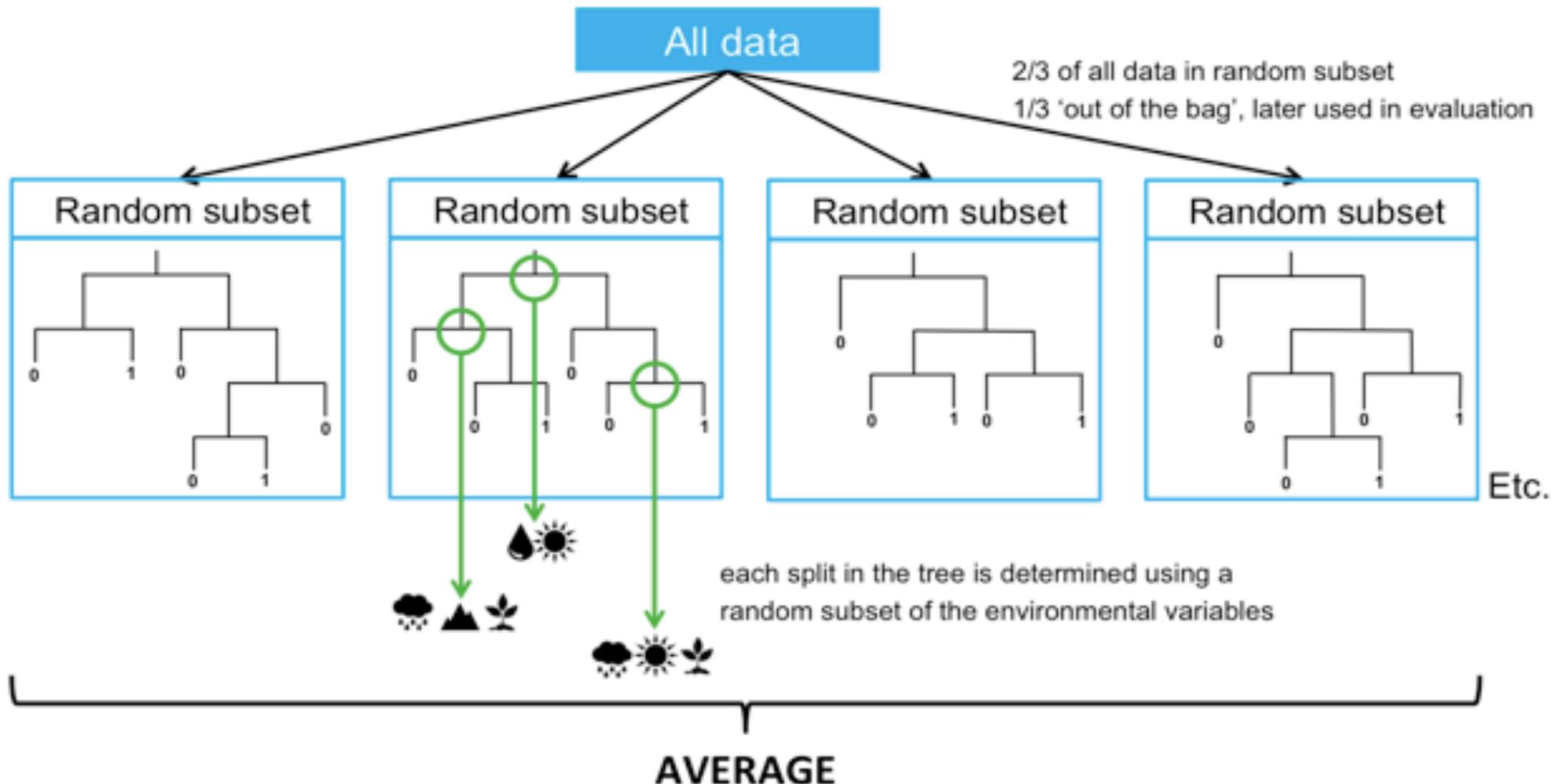
2016 RESULTS

		Predicted Class						
		A	B	C	D	E	F	G
Actual Class	A	62.2%	35.3%	2.5%	0.0%	0.0%		
	B	10.6%	66.5%	22.6%	0.2%	0.0%		
	C	1.1%	31.4%	60.7%	5.8%	1.0%		
	D	0.2%	8.0%	55.6%	29.4%	6.8%	0.0%	
	E	0.0%	2.2%	26.4%	37.7%	32.4%	1.2%	0.0%
	F	0.1%	0.9%	8.3%	21.3%	58.9%	10.3%	0.2%
	G		0.8%	3.6%	10.0%	42.3%	42.8%	0.5%

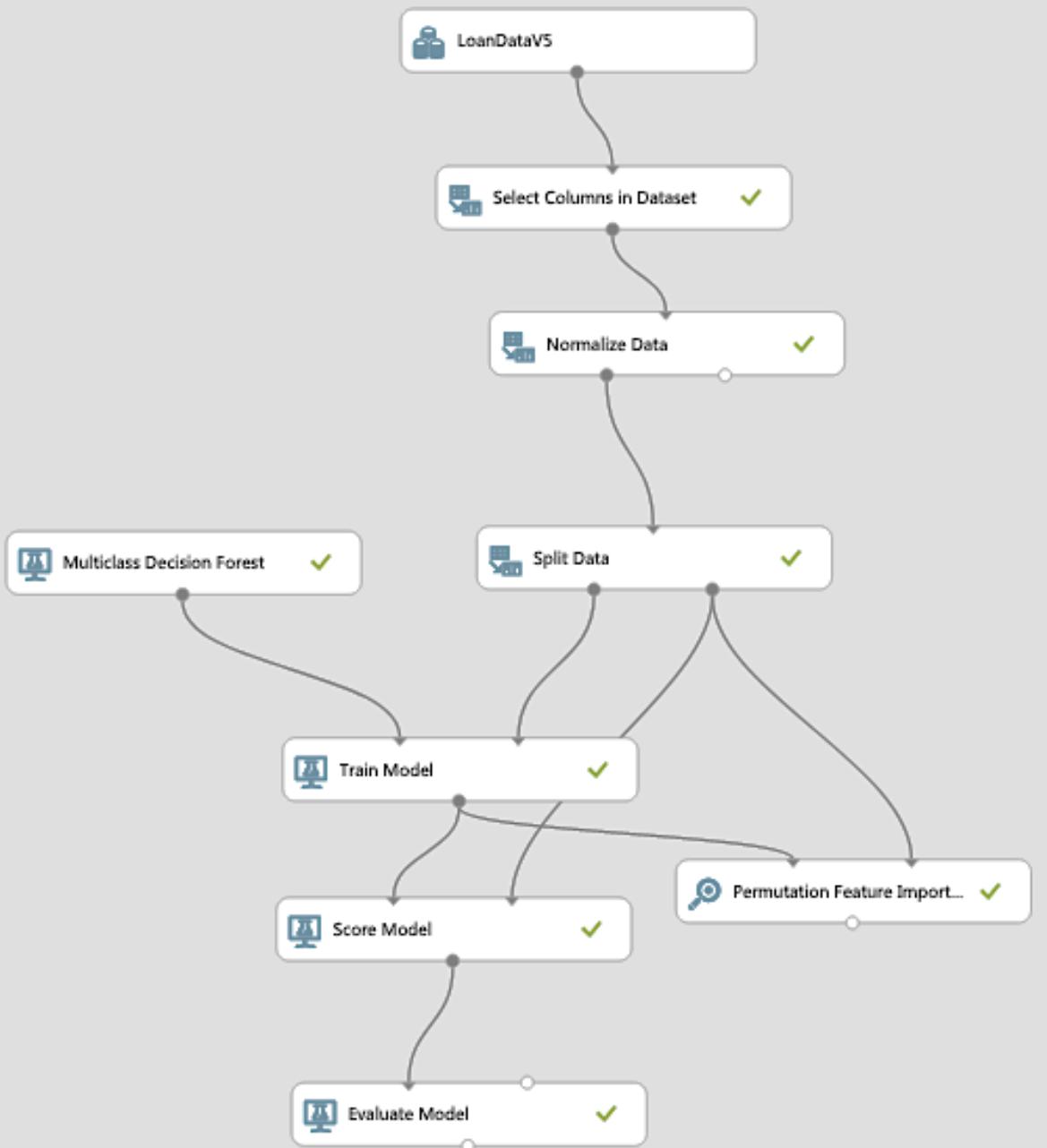
2016 MCLR Features of Significance



MULTICLASS DECISION FOREST



> find the set of predictor variables that produce the strongest classification model



2018

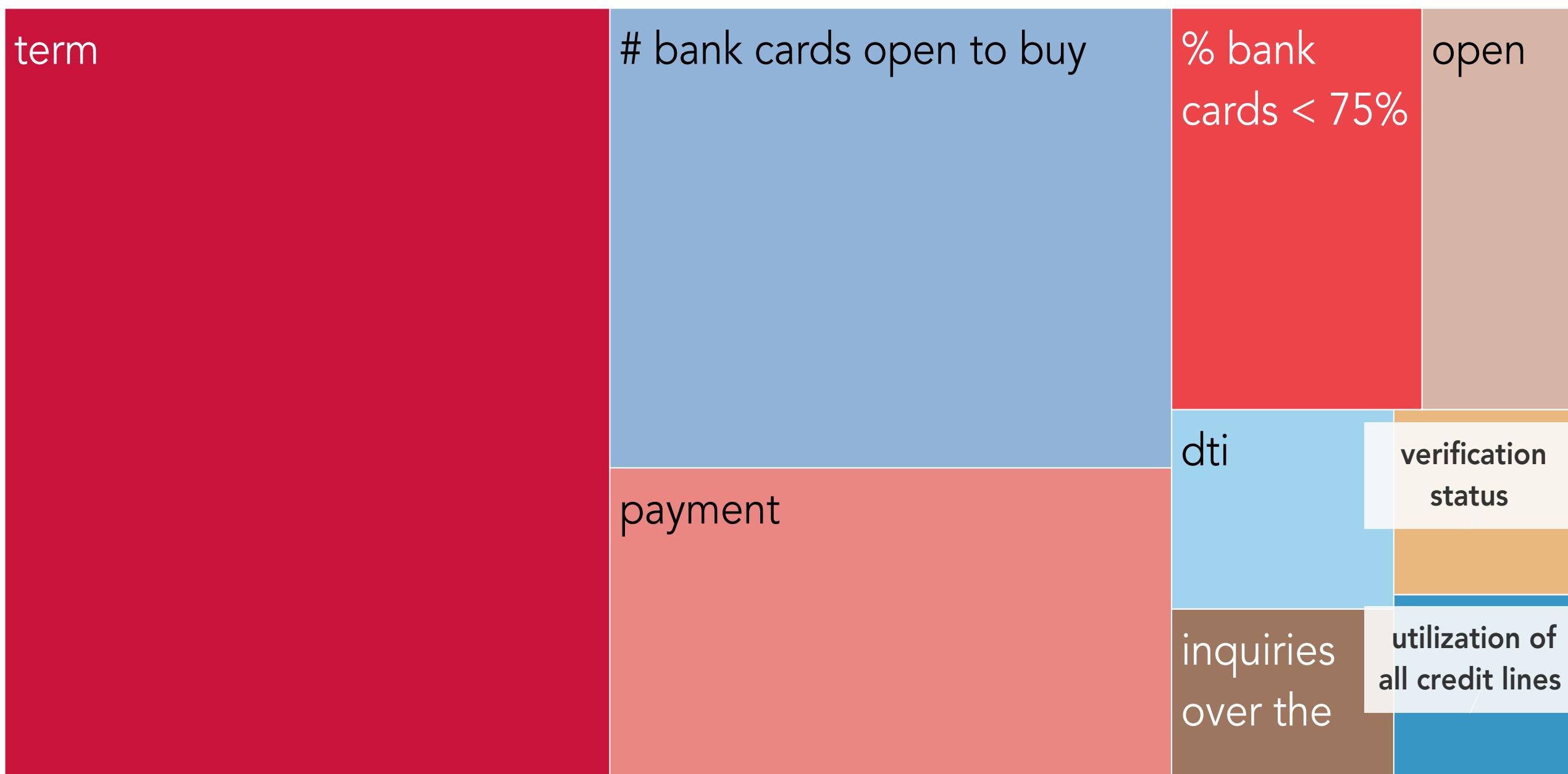
MULTICLASS DECISION FOREST

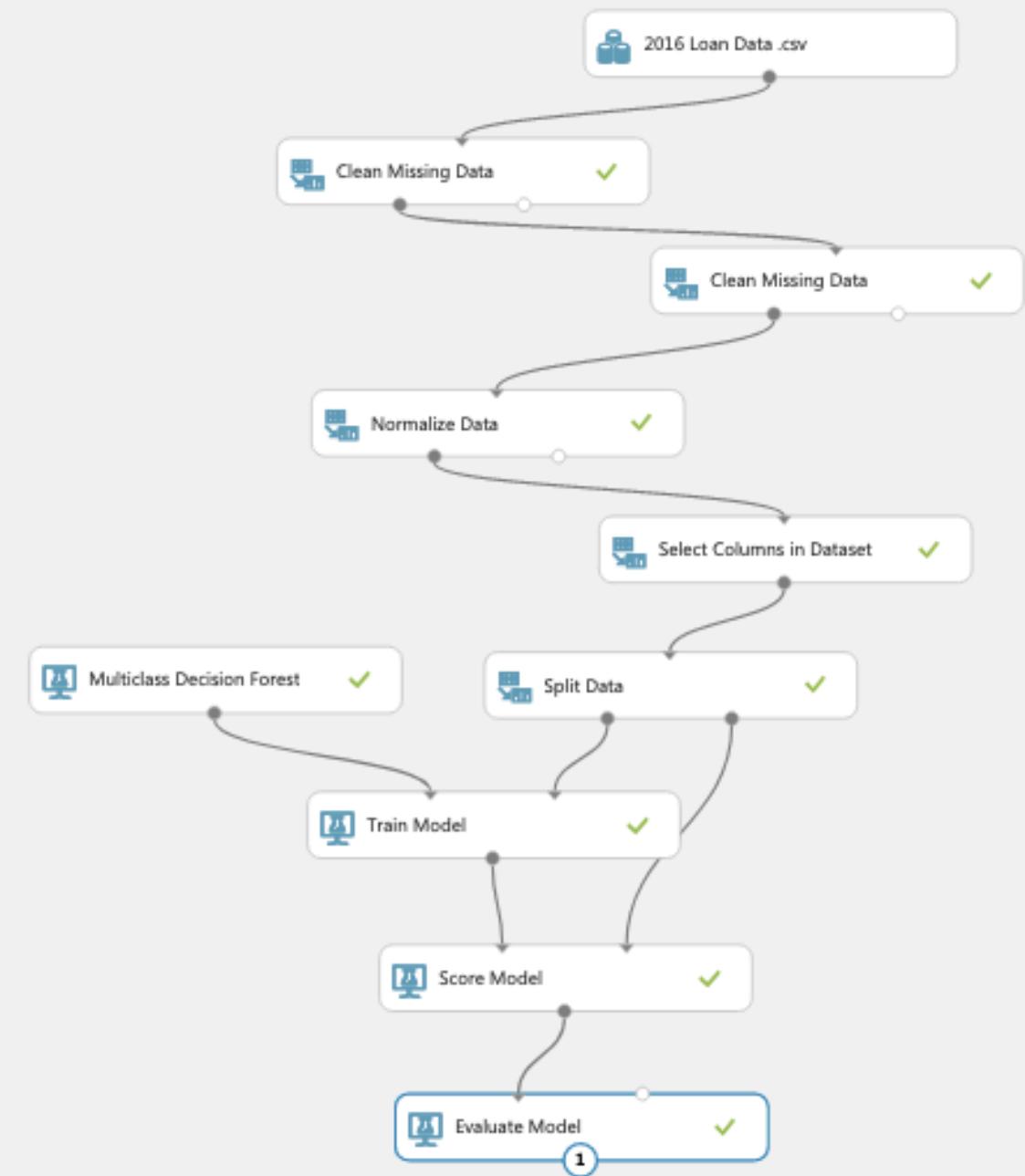
2018 RESULTS

Predicted Class

	0	1	2	3	4	5	G
0	75.3%	18.6%	5.3%	0.7%	0.1%		
1	28.7%	44.5%	22.2%	4.0%	0.5%	0.0%	0.0%
2	12.1%	32.0%	42.7%	11.4%	1.6%	0.1%	0.0%
3	7.0%	24.6%	42.3%	23.0%	2.9%	0.2%	0.0%
4	2.9%	15.6%	40.9%	31.6%	8.6%	0.4%	0.0%
5	1.1%	9.5%	38.7%	40.0%	7.8%	2.4%	0.4%
G	2.2%	14.4%	30.9%	38.8%	7.9%	3.6%	2.2%

2018 Decision Forest Features of Significance





2016

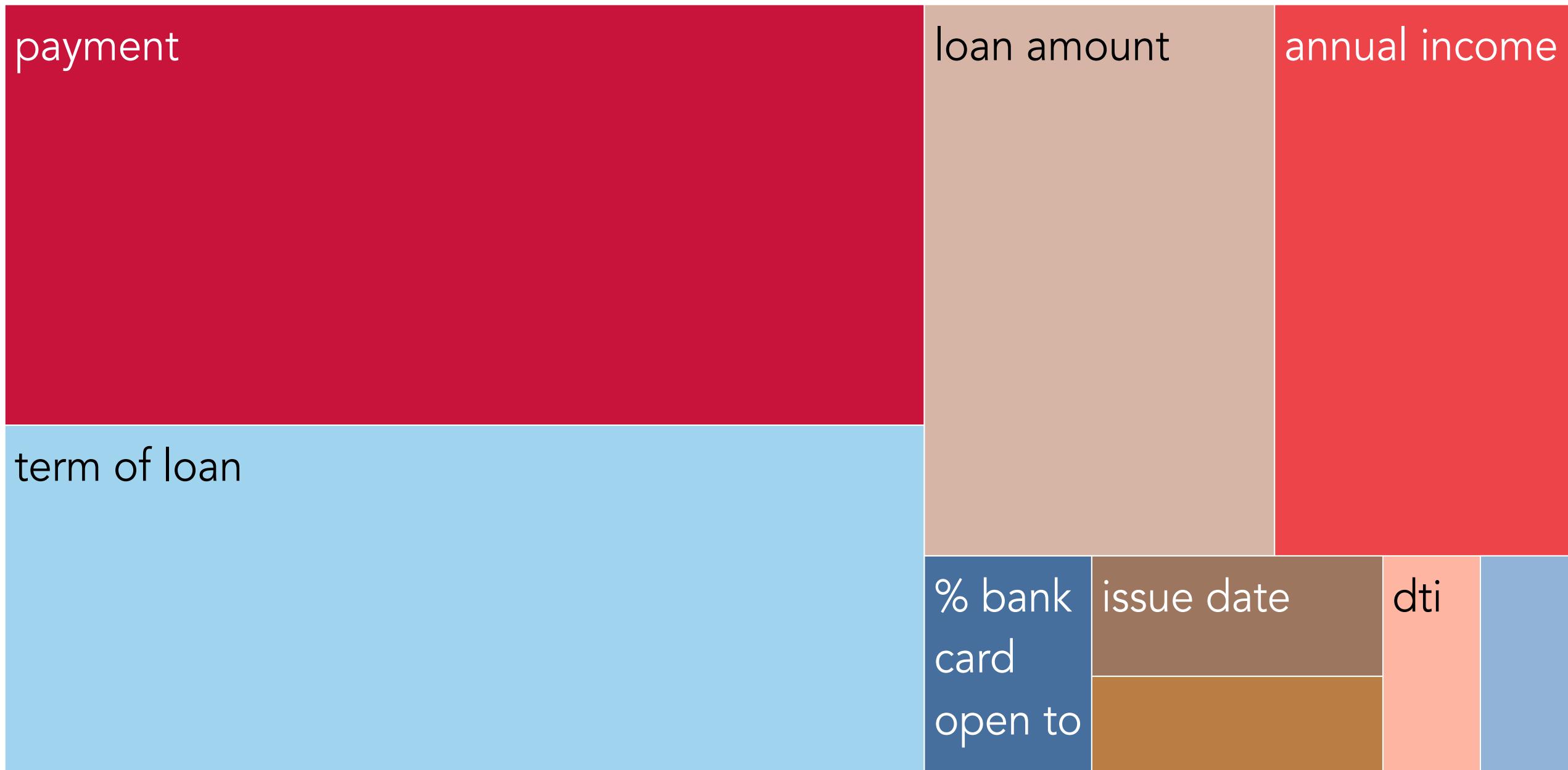
MULTICLASS DECISION FOREST

2016 RESULTS

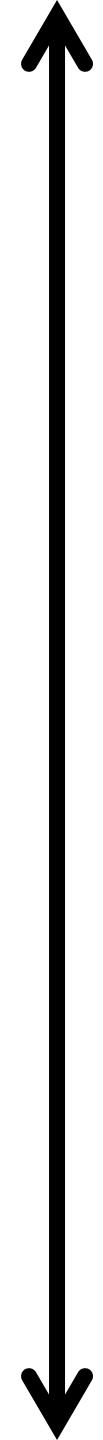
Predicted Class

	A	B	C	D	E	F	G
A	62.2%	32.0%	5.4%	0.4%	0.0%		
B	16.3%	63.8%	18.0%	1.8%	0.1%	0.0%	
C	5.4%	29.3%	60.5%	4.3%	0.4%	0.0%	
D	2.5%	18.4%	30.7%	45.3%	3.1%	0.1%	0.0%
E	1.0%	8.4%	15.5%	22.0%	50.6%	2.6%	0.0%
F	0.6%	4.7%	10.2%	10.2%	31.0%	40.3%	3.0%
G		2.7%	7.0%	8.1%	10.8%	43.6%	27.9%

2016 Decision Forest Features of Significance



SUMMARY



	2016 MCLR	2018 MCLR	2016 Decision Forest	2018 Decision Forest
Overall Accuracy	55%	47%	58%	47%
Average Accuracy	87%	84%	88%	84%

	2016 MCLR	2018 MCLR	2016 Decision Forest	2018 Decision Forest
Top Three Features	Payment	Term of Loan	Payment	Term of Loan
	Issue Date	% BC < 75%	Term of Loan	# BC open to buy
	Loan Amount	Trades over 24m	Loan Amount	Payment



THANK
YOU