



LOAN DEFAULT PREDICTION

Joshua Edelstein

AGENDA



Address
Stakeholder



Data Background



Predictive Model



Conclusion/
Recommendations



Next Steps

BUSINESS QUESTION



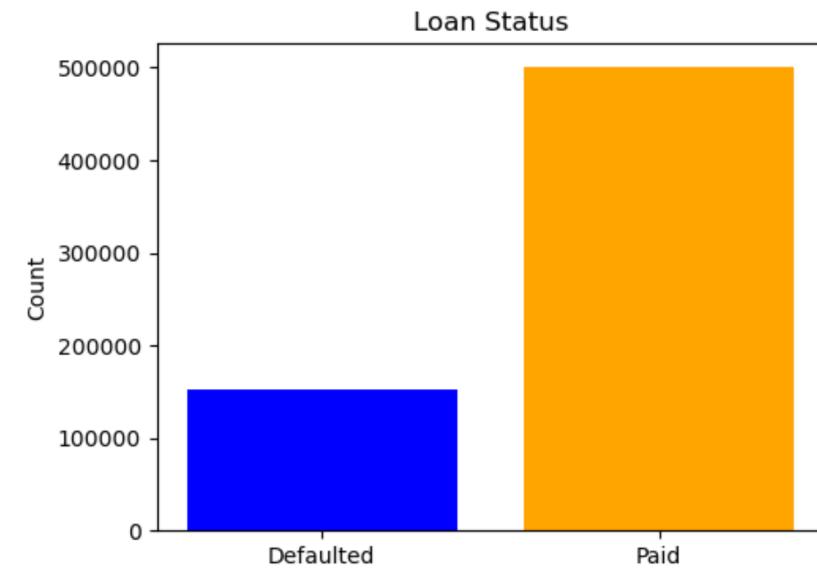
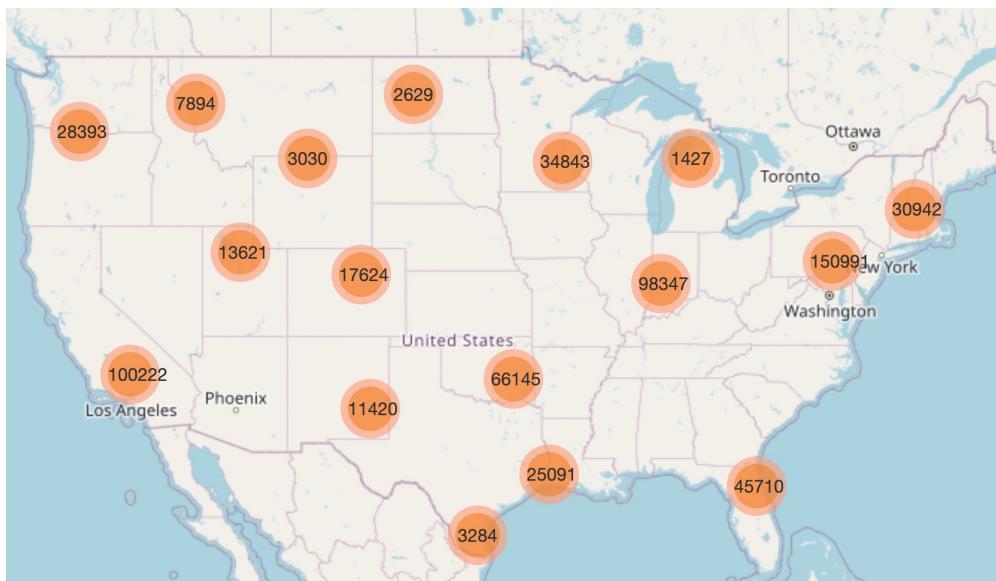
RECENT INCREASE IN DEFAULT
RATES FOR BUSINESS LOANS.



AMERICAN BANKERS
ASSOCIATION SEEKS SOLUTIONS

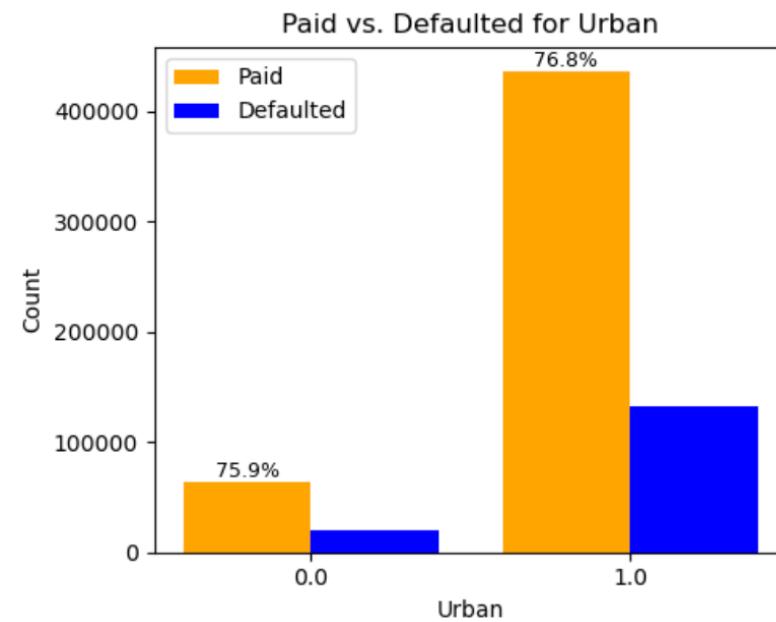
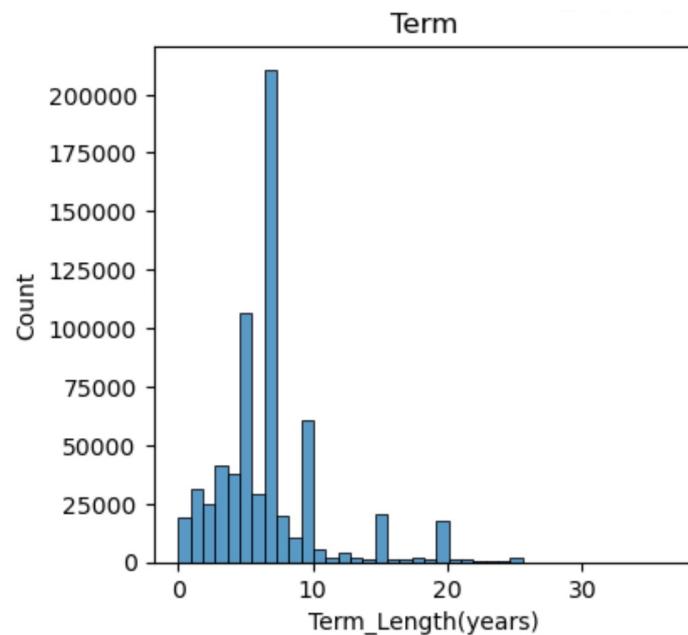
DATA UNDERSTANDING

- ~650,000 loans from across the 50 states
- From loans offered from 1969-2014, and due between 1973-2035
- Default rate of 23%



DATA UNDERSTANDING

- Numerical Data:
 - Term length, Number of Employees, Amount
- Categorical Data:
 - State of Business, Bank State, Year, Urban or Rural, Category of Business



PREDICTIVE MODEL

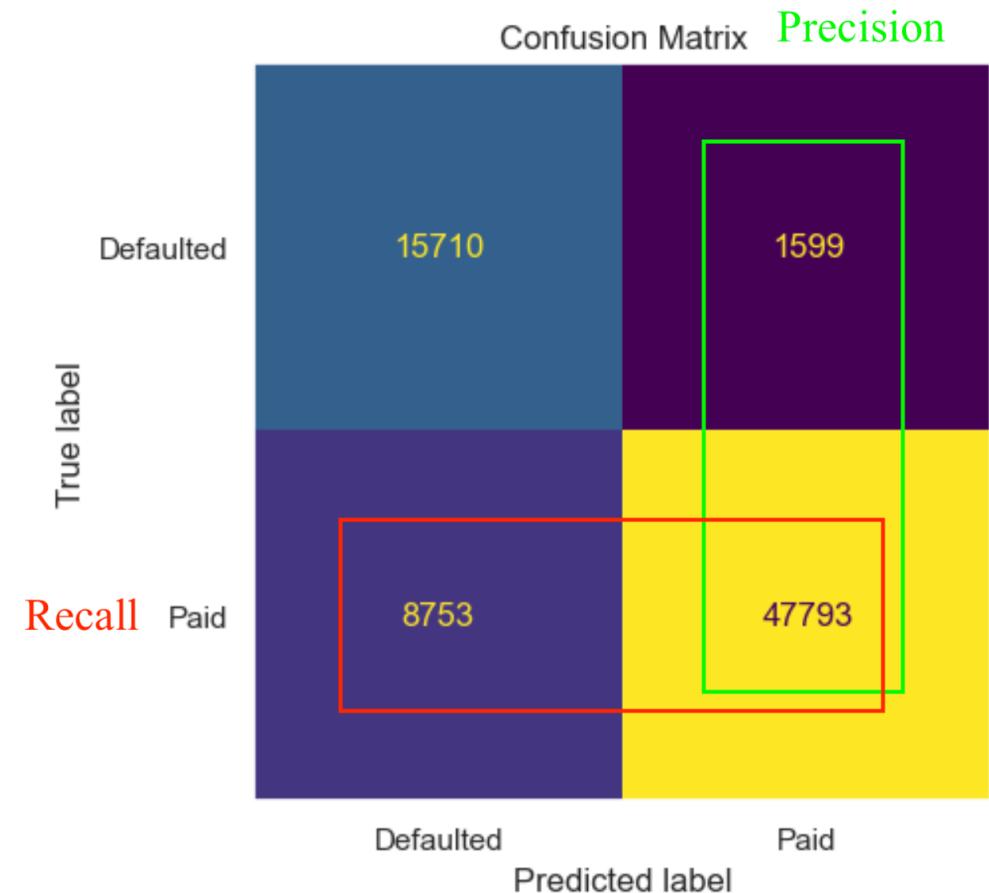
PRECISION VS RECALL

- False Positive- offer loan to a business that will default
 - Precision
- False Negative- reject loan application when client will pay
 - Recall



MODEL METRICS

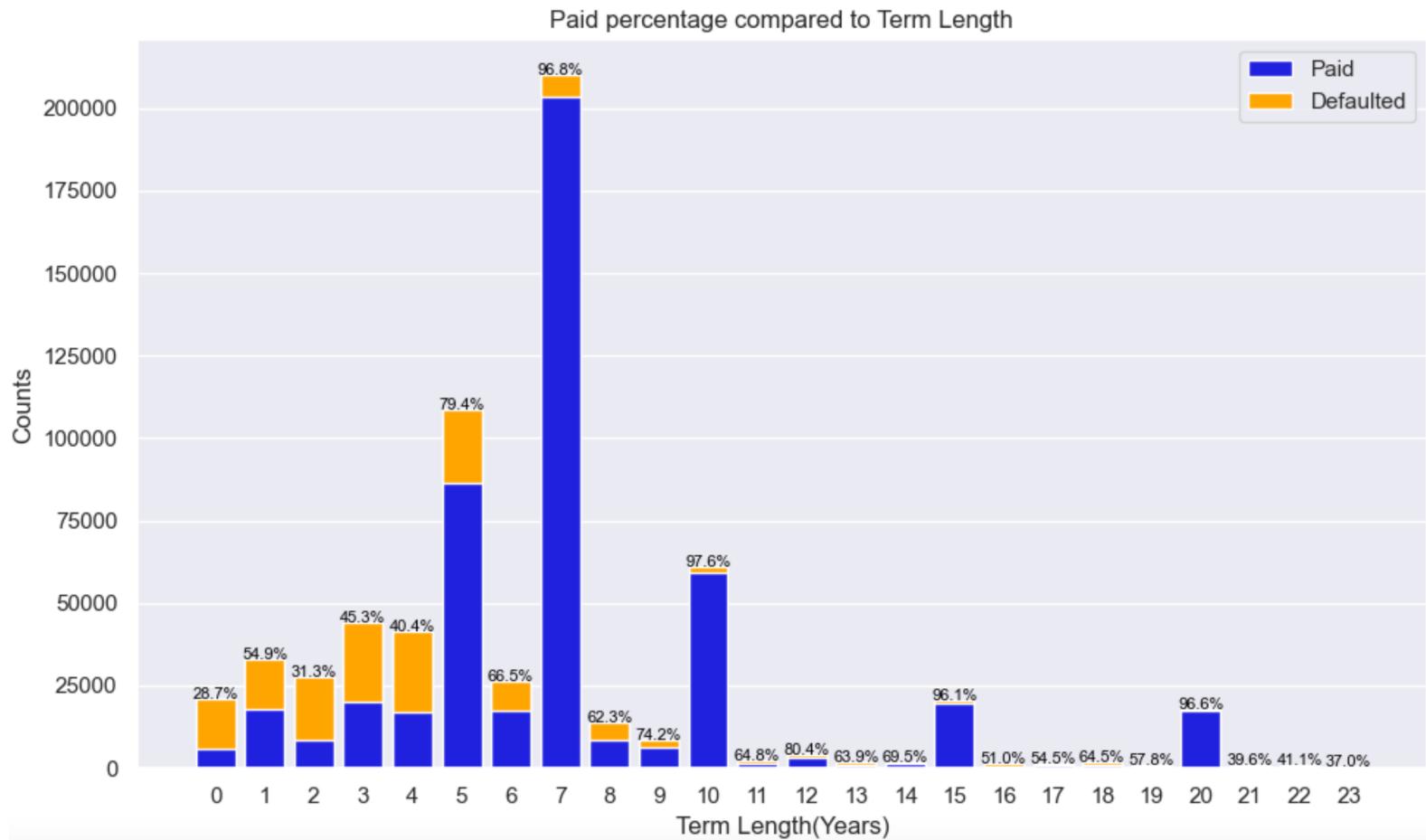
- XGBoost Model- combination of many smaller models
- Tested on 73,855 unseen business loans
 - Accuracy: 86%
 - Recall: 85%
 - Precision: 97%
 - F1 score: .90



M O S T I M P O R T A N T F E A T U R E S (1)

Term

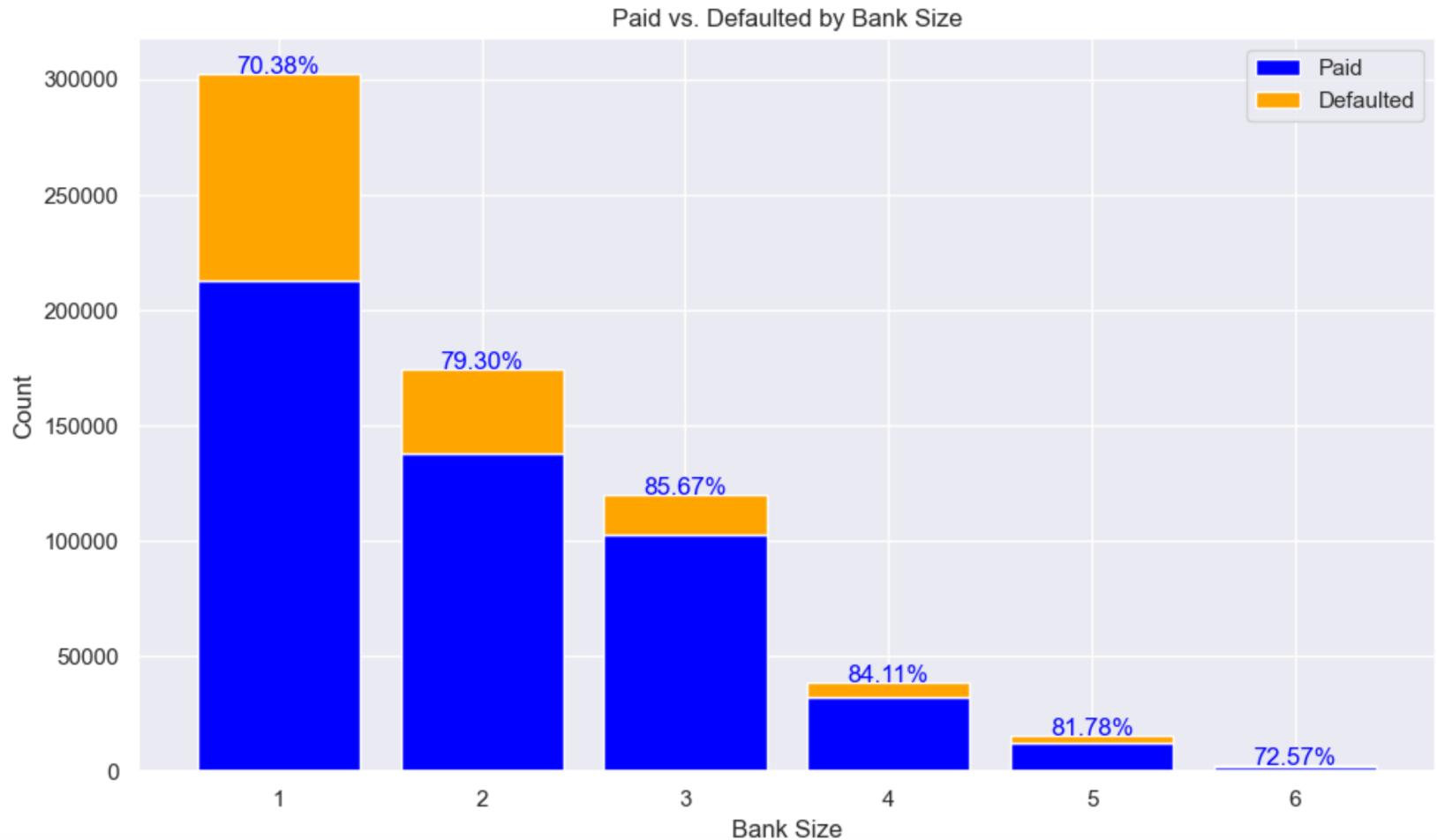
- Longer term loans have higher paid rate



MOST IMPORTANT FEATURES (2)

Bank

- Size of the bank chain. Medium size chains have the highest paid rate



RECOMMENDATIONS



IMPLEMENT MODEL



ADVISE BANKS TO BE CAUTIOUS
WITH SHORTER TERM LOANS



NOTIFY LARGER BANKS THAT
DEFAULT RATE IS AN ISSUE

NEXT STEPS

- Collect more historic and monetary data about the business asking for the loan
- Further research relationship between term and chain size with default rate



TH A N K S

Joshua Edelstein

joshuaedelstein15@gmail.com

github.com/joshuaedelstein15

www.linkedin.com/in/joshua--edelstein/