

# Problem Identification

What properties of a borrower make a loan more likely to be fully paid? charged off?

- 1 Context: Moneylending is potentially a very lucrative business because many kinds of business like real estate, etc. leverage money from loans as starting capital. However, lending money to borrowers unlikely to pay it back is very costly.
- 2 Criteria for success: When determining whether a loan will be charged off,
  - have a low false positive rate (don't want to deny profitable loans)
  - have a very low false negative rate (want to get paid back)
- 3 Scope of solution space: The focus will be on numerical and categorical data from information about borrowers and their loan requests. We will be focusing on loans through Lending Club.

# Problem Identification

- ④ Constraints within solution space:
  - The data may not be up to date (it is 10 years old)
  - The insights gained may not apply to borrowers in other countries
- ⑤ Stakeholders to provide key insight:
  - Banks
  - Pawnbrokers
- ⑥ Key data sources: data.world dataset about loans issued by lendingclub.com from 2007-2011 with performance data – contains complete loan data for all loans issued through the time period stated, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information.

# Recommendation and key findings

- We are able to determine the four traits most strongly associated with fully paid loans. They are high mean FICO score, high annual income, credit card related purpose, and debt consolidation related purpose.
- We are also able to determine the four traits most strongly associated with charged off loans. They are longer term loans, inquiries within the last six months, credit utilization rate, and small business related purpose.

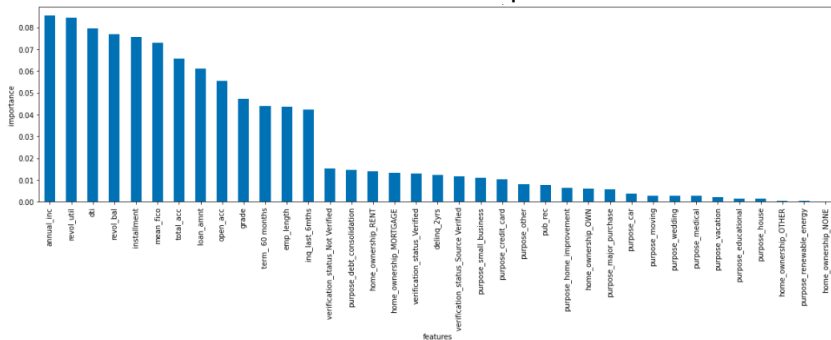
# Modeling results and analysis

Logistic regression coefficients for fully paid loans:

mean_fico	0.378308
annual_income	0.288602
purpose_credit_card	0.179052
purpose_debt_consolidation	0.129990
purpose_car	0.083845
⋮	⋮
loan_amnt	-0.147008
purpose_small_business	-0.153739
revol_util	-0.155017
inq_last_6mths	-0.234819
term_ 60 months	-0.383156

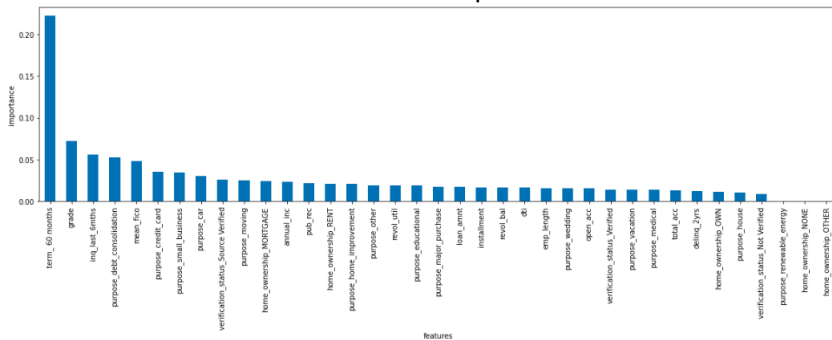
# Modeling results and analysis

## Random forest feature importances:



# Modeling results and analysis

## XGBoost feature importances:



# Summary and conclusion

- The most creditworthy features are high mean FICO score, high annual income, credit card related purpose, and debt consolidation related purpose.
- The most subprime features are longer term loans, inquiries within the last six months, credit utilization rate, and small business related purpose.
- Other important features determining loan status are monthly debt to income ratio, total credit balance, and loan grade.