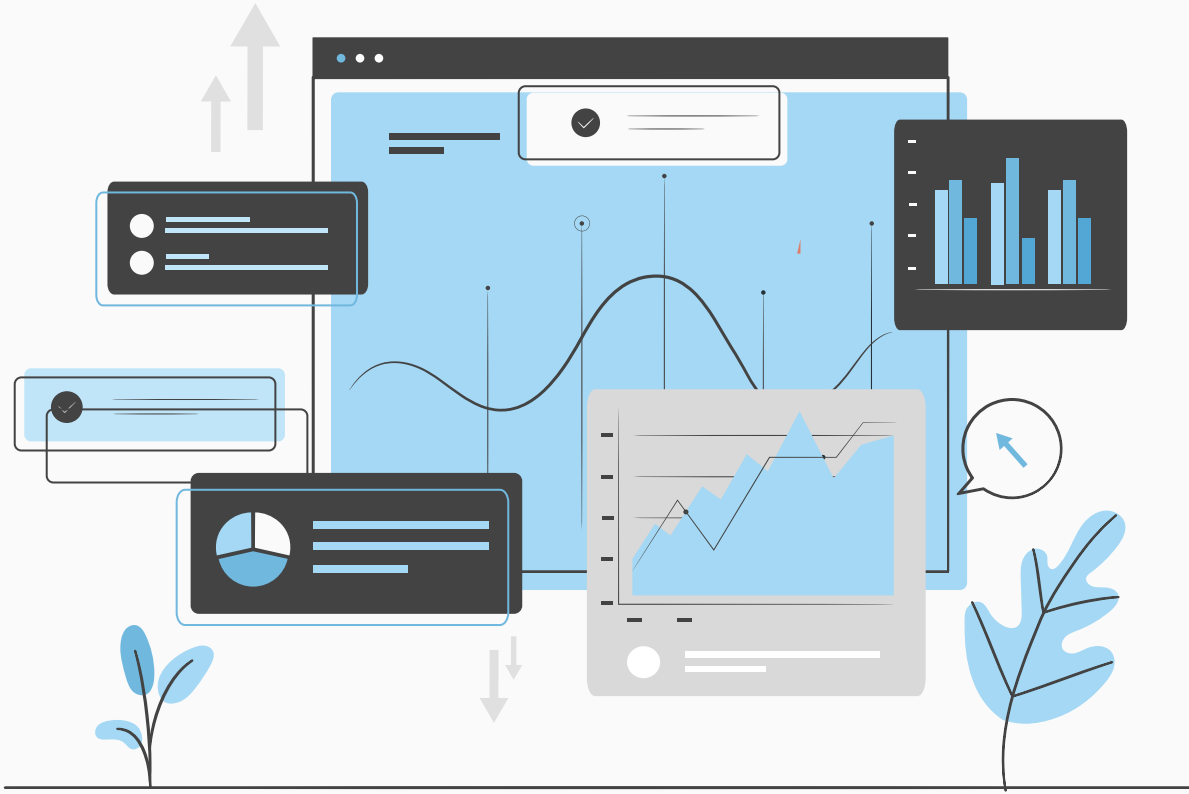
SBA LOANS

Predicting defaults



OVERVIEW

BACKGROUND:

- Created in 1953, the SBA (Small Business Administration) provides resources, both informational and financial, to small businesses and entrepreneurs across the country.

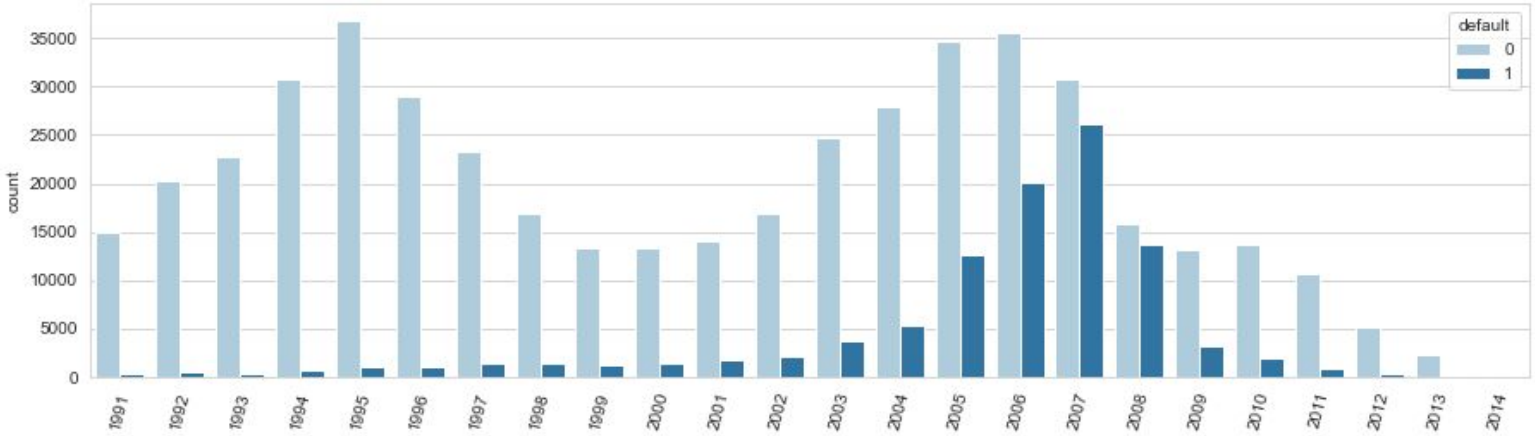
GOAL:

- Build a model to predict SBA loan defaults for loans issued in 2007.

DATA:

- Initial dataset contained over 850,000 loan records issued between 1968 and 2014.
- Imbalanced classes in a ratio of about 6 to 1.

OVERVIEW



DEFAULTS PEAKED IN THE DATASET IN 2007 DUE TO THE GREAT RECESSION.

METHODOLOGY

TRAINING SET:

- Choose loans who had run their course before 2007 (i.e. had either paid in full or defaulted before the end of 2006).

FEATURE ENGINEERING:

- Added features including: **state default rate, sector default rate, and real estate collateral.**

MODEL AND MEASUREMENT:

- Measuring success on an F-Beta score weighted towards recall, as well as general accuracy of the model.
- Measuring how much money the model would save/lose compared to the test set.
- A **gradient boosted tree model** provided the best performance.

RESULTS

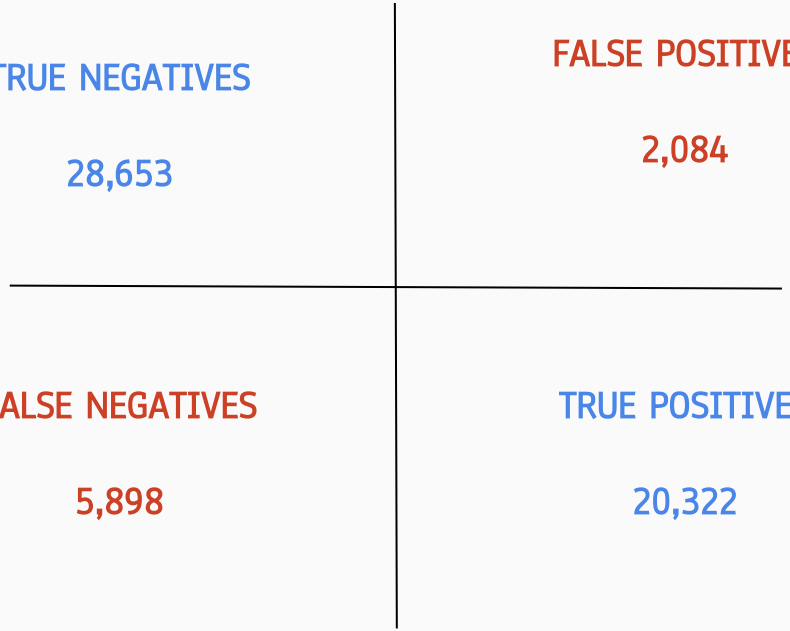
TRUE NEGATIVES
+ amortized amount of the loan at 6% interest

FALSE POSITIVES
+ $0.00
(loan not issued)

FALSE NEGATIVES
- Amount Owed

TRUE POSITIVES
+ $0.00
(loan not issued)

RESULTS


TRUE NEGATIVES
28,653
FALSE POSITIVES
2,084
FALSE NEGATIVES
5,898
TRUE POSITIVES
20,322

RESULTS

DID THE MODEL HELP?:

- The value of the test set originally was... -$1,200,000,000!!!
- Using the model, banks would have lost a collective $200,000,000
- That represents a savings of around $1,000,000,000 for 2007

CONCLUSION:

- Using a generalized model can help banks make wiser decisions in their lending practices to maximize earnings.
- There are additional factors to consider when issuing loans such as the true risk to a bank versus the SBA, de-risking by selling portions of the loans on secondary markets, etc.

BUILDING AN APPLICATION WITH A GENERALIZED MODEL...