# Home Credit Data Set

Matthew Daly

# Predicting Loan Default Risk

## The Company

–

Home Credit: lender who provides loans to populations unable to use traditional credit services.

## The Problem

–

how to lower loan default risk by identifying patterns from within historical data.

## The Method

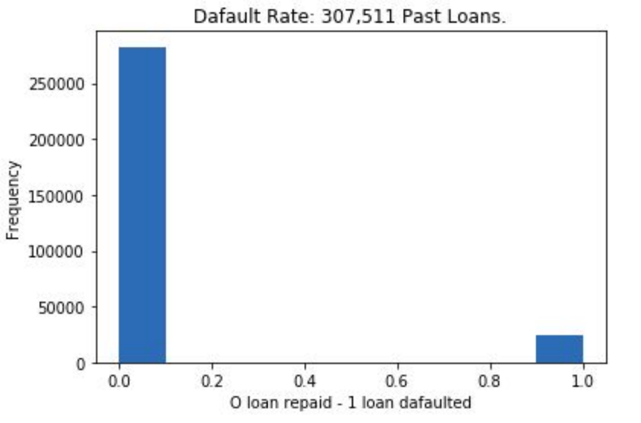
–

a combination of Exploratory Data Analysis, Principal Component Analysis, and various Machine learning Algorithms to identify the most predictive data points.

# The Data

307,511 loans

122 features - including age, employment, past credit, region…



Default Rate

~

8.8%

# The Potential

Even a small decrease in the default rate would translate to a significant reduction in losses and an increase in capital available for further loans.

# TOP THREE DATA FEATURES

## Outside Credit Agency Scores

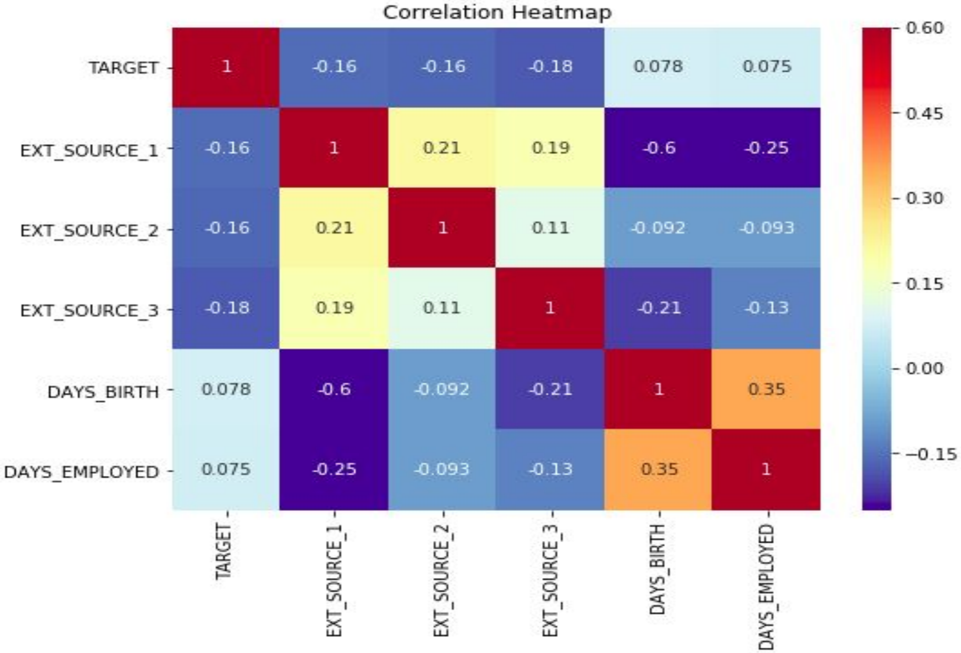
(EXT_SOURCE_[1-3)

## Length of Employment

(DAYS_EMPLOYED)

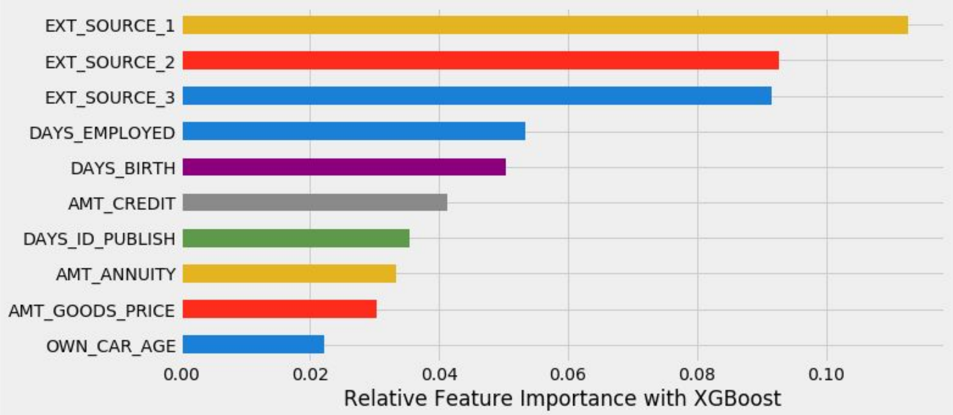
## Age of Borrower

(DAYS_BIRTH)



Correlation Heatmap

| | TARGET | EXT_SOURCE_1 | EXT_SOURCE_2 | EXT_SOURCE_3 | DAYS_BIRTH | DAYS_EMPLOYED |
|---|---|---|---|---|---|---|
| TARGET | 1 | -0.16 | -0.16 | -0.18 | 0.078 | 0.075 |
| EXT_SOURCE_1 | -0.16 | 1 | 0.21 | 0.19 | -0.6 | -0.25 |
| EXT_SOURCE_2 | -0.16 | 0.21 | 1 | 0.11 | -0.092 | -0.093 |
| EXT_SOURCE_3 | -0.18 | 0.19 | 0.11 | 1 | -0.21 | -0.13 |
| DAYS_BIRTH | 0.078 | -0.6 | -0.092 | -0.21 | 1 | 0.35 |
| DAYS_EMPLOYED | 0.075 | -0.25 | -0.093 | -0.13 | 0.35 | 1 |



## Feature Selection via XGboost

Training Accuracy: 92.01%

Testing Accuracy: 91.99%

# Recommendation - 1

## Trust and Emulate the Credit Agencies

The top three indicators of repayment correlate to credit scores assigned by outside agencies.

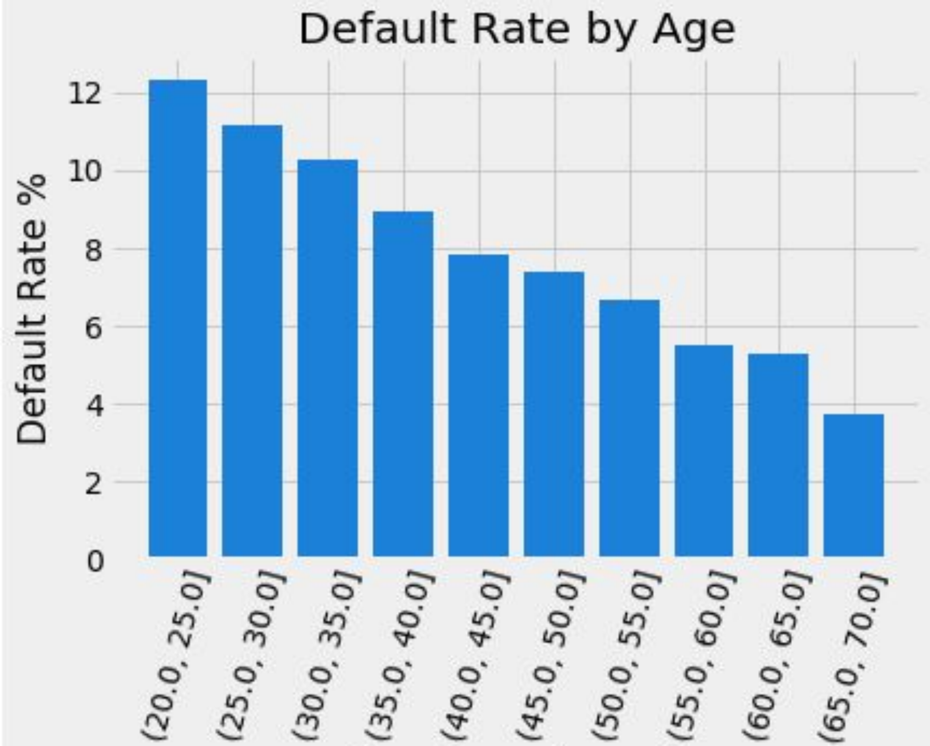
# Recommendation - 2
# Investigate Employment Anomaly

Anomaly: Max Days Employed 365243

Anomaly Default Rate: 5.4%

Potential Gain 3.26%

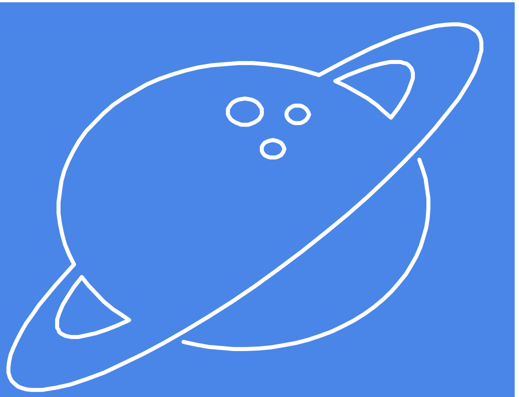
Non-Anomaly Default Rate: 8.66%



# Recommendation - 3

Aggressive debt counseling for younger segments .

# Future Work



- fully Explore the Employment Anomaly for Insights
- engineer Domain Specific Data Features
- research the Best Practices credit agencies are currently employing and apply them to the data set

Full credit and appreciation to Will Koehrsen for his Kaggle guides to this data set. All hail Will!