Lending Club Presentation

An Investor’s Perspective

Agenda

01 Introduction

02 Data Exploration

03 Data Visualization

04 Models

05 Results & Conclusion

Introduction

Team dotFit

Jessica Joy
EDA Lead

Eugene Ng
Business Lead

Thomas Laetsch
Mentor

David Kim
EDA Lead

Darish Sakeesing
Tech Lead

David Wasserman
Tech / Business Lead

Who We Are

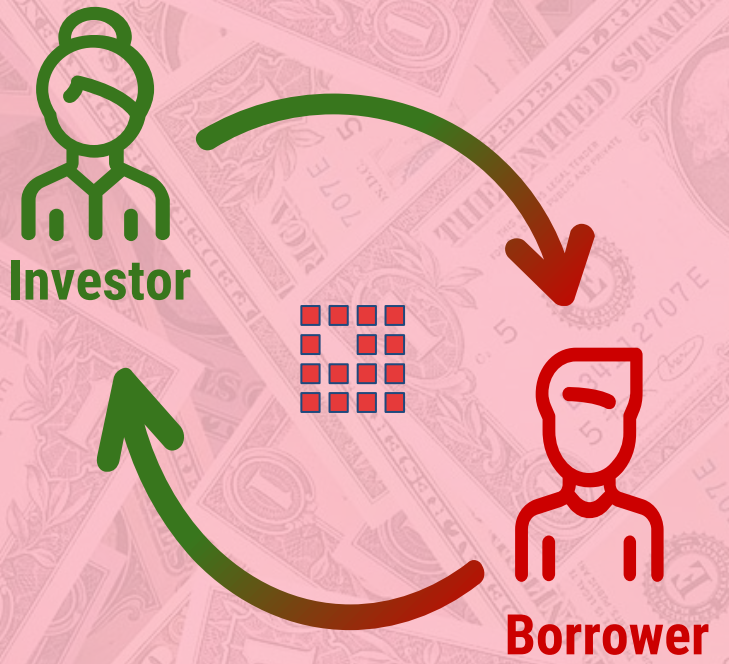
A group of data scientists engaged by a consortium of private investors seeking to maximize their return on investment

Objectives

- Evaluate investment opportunities in Lending Club loans
- Isolate and remove loans likely to default and/or perform poorly

What is LendingClub

- An American peer-to-peer lending company
- Lending Club enables borrowers to create **unsecured** personal loans between $1,000 and $40,000.
- Investors are able to browse the loan listings on the Lending Club website and select loans that they want to invest in based on the information supplied about the borrower, loan amount, loan grade, loan purpose, etc.



LendingClub Timeline

2007

Lending Club launched as a Facebook App

2014

Lending Club IPOs

2016

Troubled waters

Founder and CEO ousted due to conflict of interest and $22M in incorrect loans sold to Jeffries

2017

Lending Club drops grade F & G loans due to high default rates

2020

Lending Club ceases P2P loan market and acquired Radius to become a traditional bank

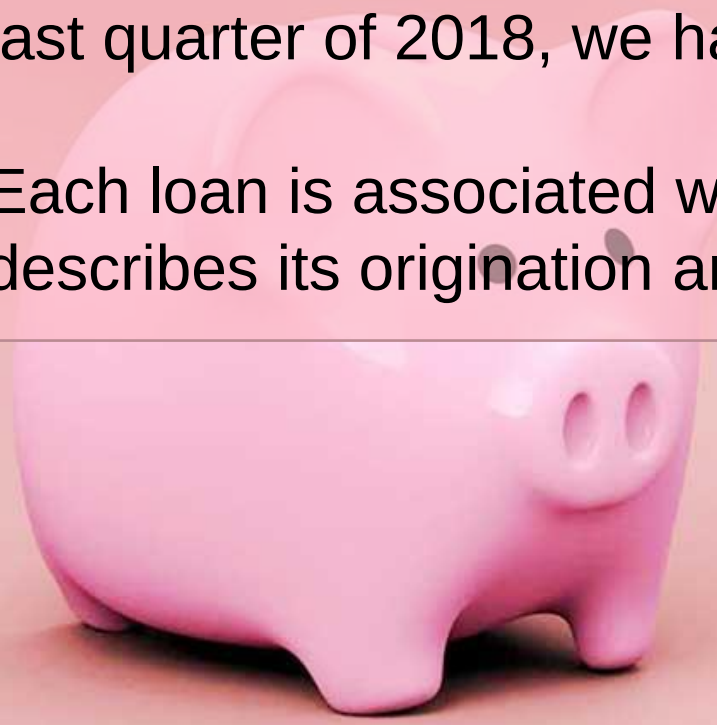
Data Exploration

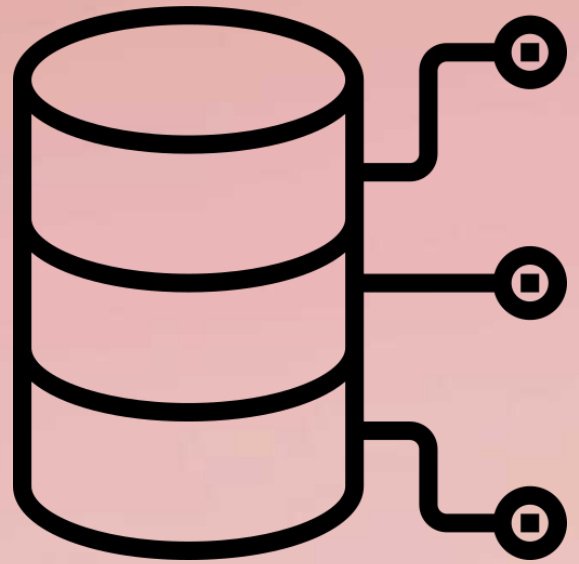
Size of the Data

The data is tabular where each row represents an individual loan.

Since the data contains all the loans issued on the Lending Club platform from the year 2007 up to the last quarter of 2018, we have millions of rows.

Each loan is associated with 151 data points which describes its origination and performance.





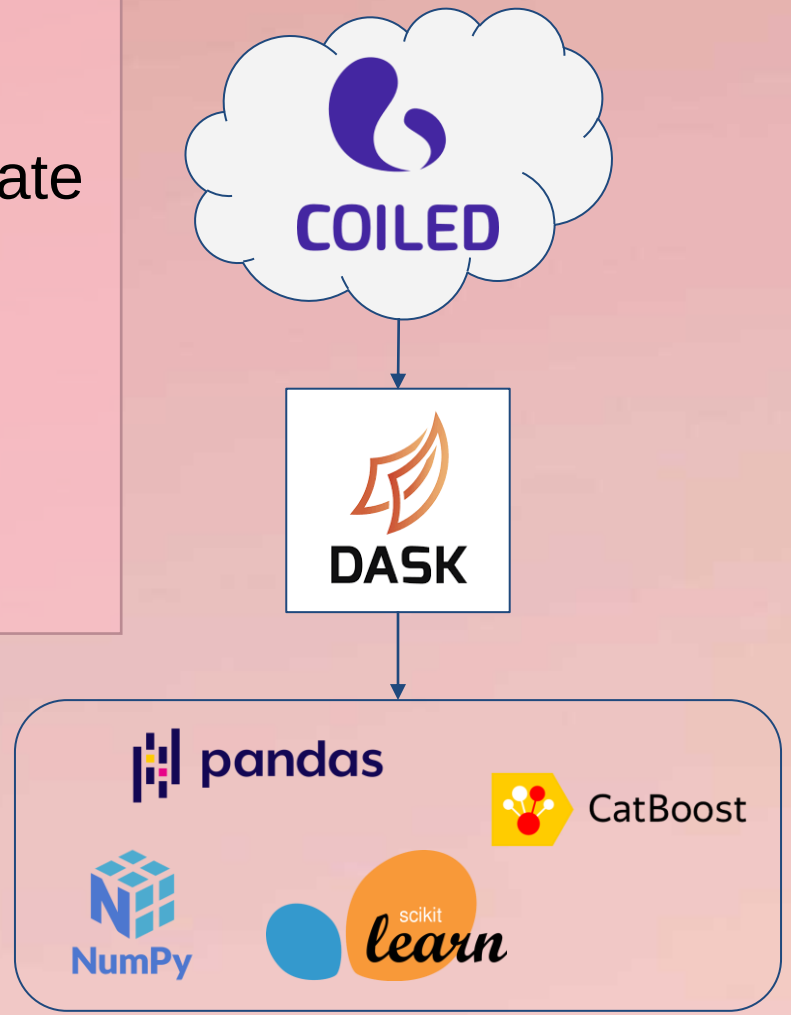
2.26 million rows by 151 columns (1.6 GB)

Tools Used

The sheer size of the data posed an immense challenge to load it and perform exploratory analysis.

We had to adopt cutting edge technologies to manipulate the data.

We decided on Dask and Coiled since they work well with popular frameworks like pandas and sklearn.

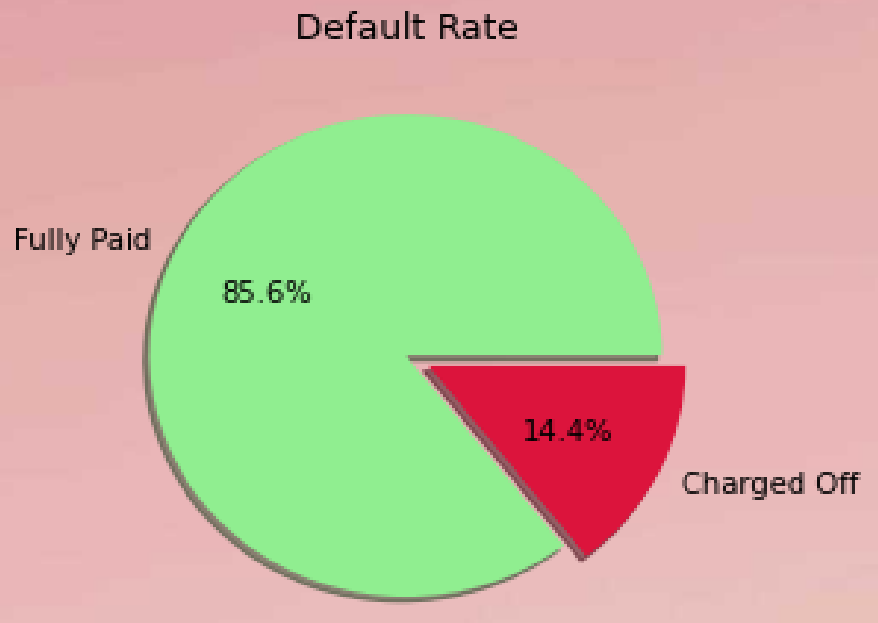


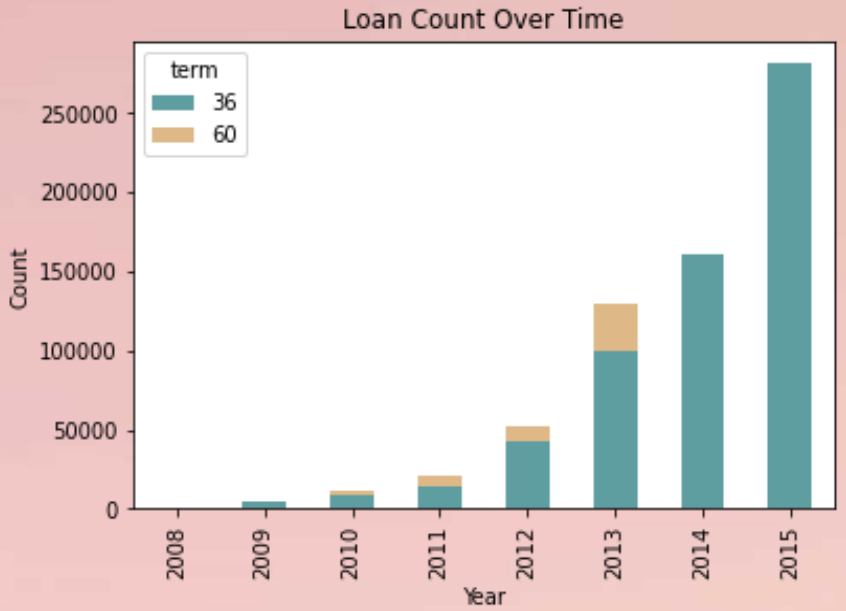
Imbalanced Data

The dataset was imbalanced as expected. Only a minority of loans ended up default.

Considerations to be taken:

- Kept proportions when taking samples, splitting, et al.
- Adjusted models to perform well even with imbalanced datasets (upsampling, changing priors, etc.)





Determining What Data was Available at Loan Origination

The dataset held information about loan origination and performance.

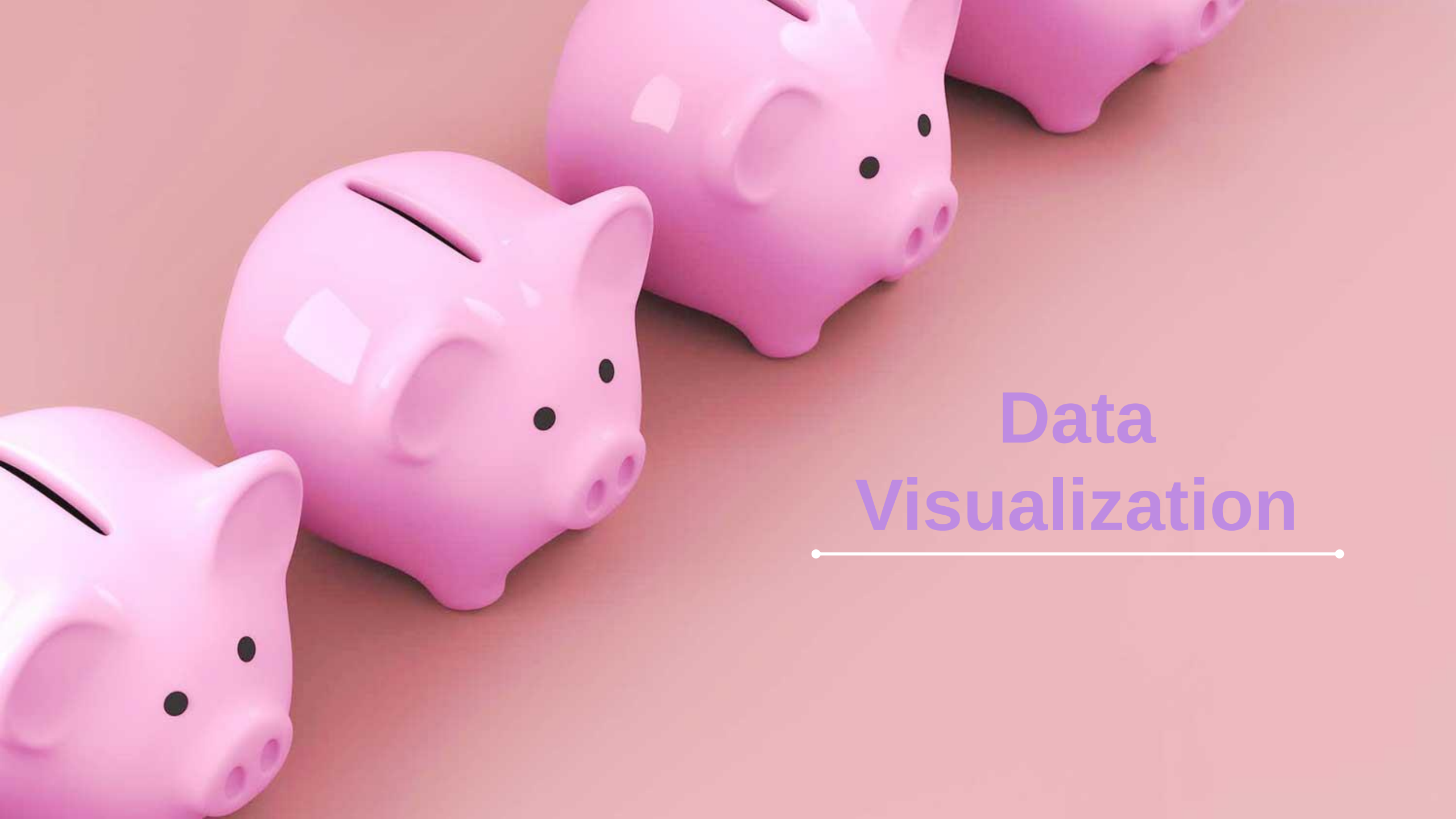
It was important to separate what was available to investors at the time of loan origination and only use those features in our models.

Another consideration was that Lending Club would often pull the credit report of borrowers after loan origination and update fields like how many open accounts, etc. Therefore, we could not use those datapoints.

Examples:

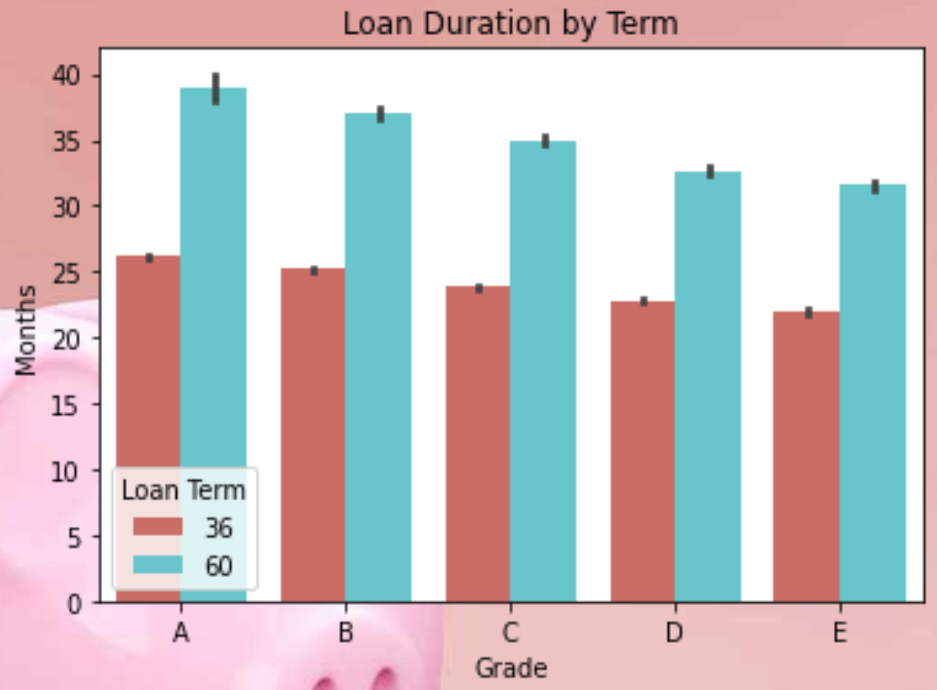
dti:	Debt to Income ratio
addr_state:	The state for the address of the borrower
purpose:	Category provided by borrower
term:	Number of payments on the loans

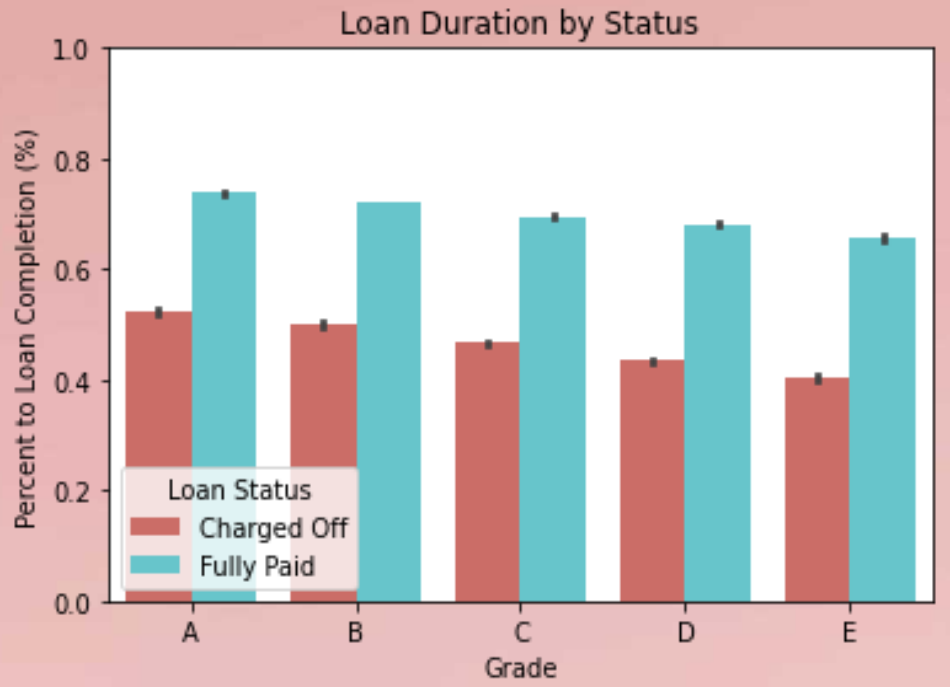
We ended up with 30 features.



Data Visualization

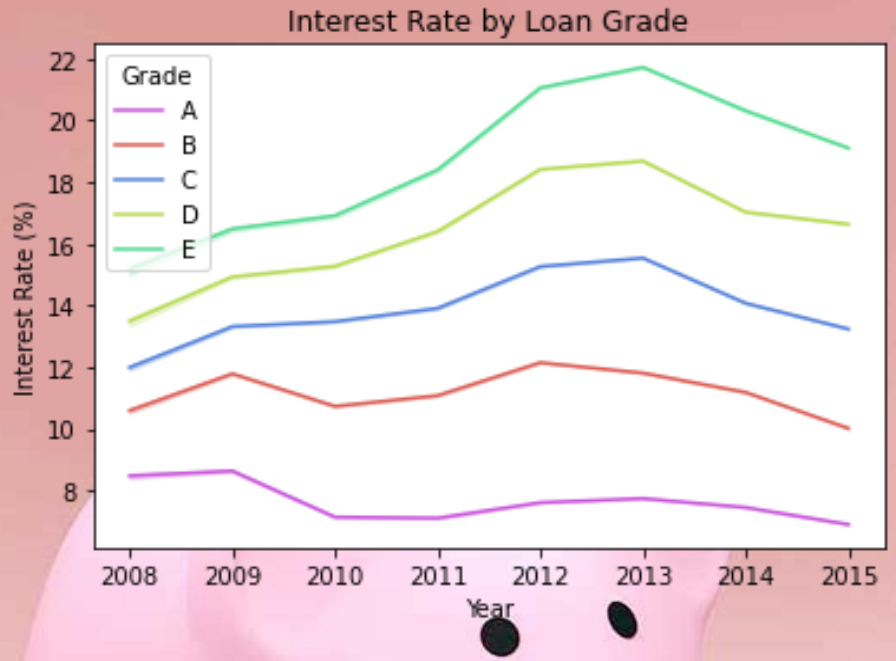
Loan Duration

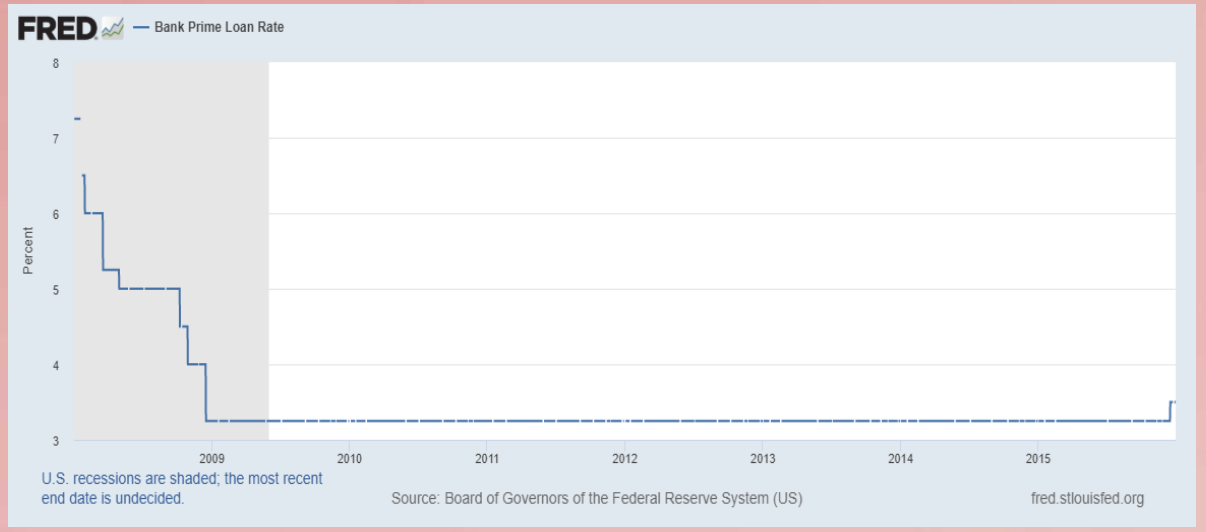




- As the grade worsens, loan duration decreases

Prime vs. Lending Club Interest Rates

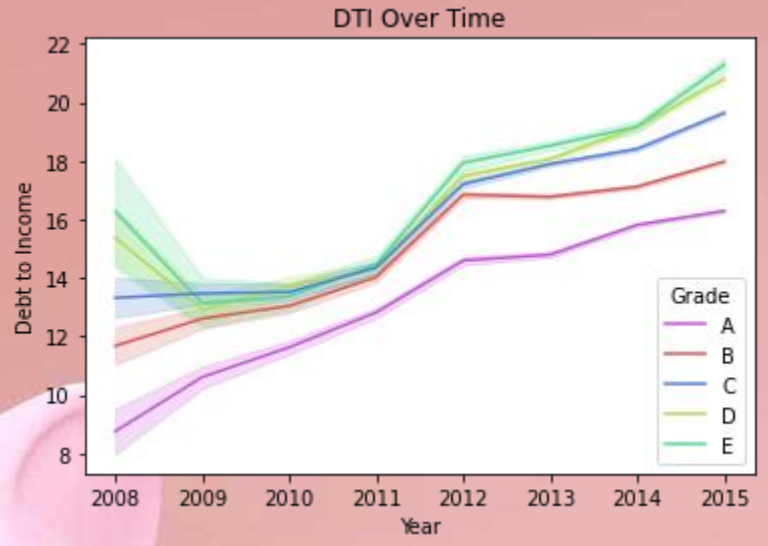


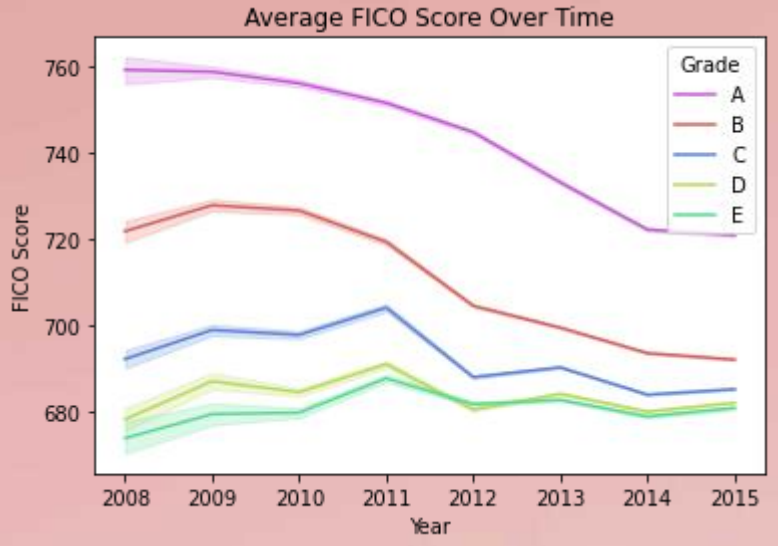


What is the prime rate?

- The prime rate is the interest rate that commercial banks charge their most creditworthy corporate customers.
- The rates for mortgages, small business loans, and personal loans are based on prime.

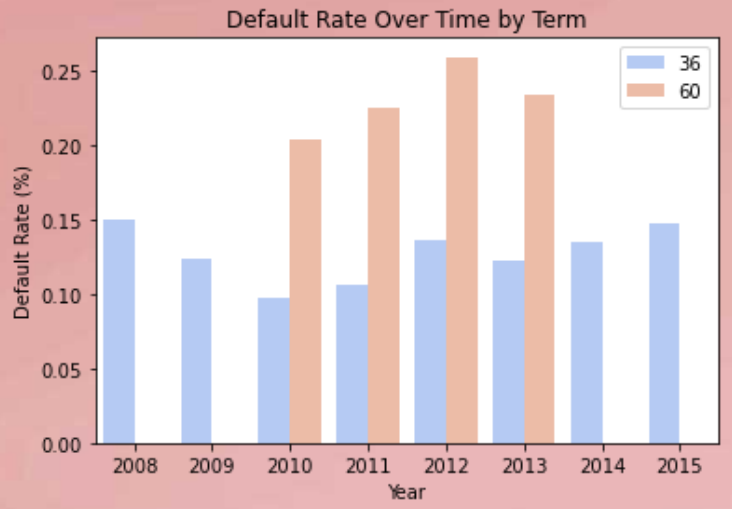
Loan Grade Factors

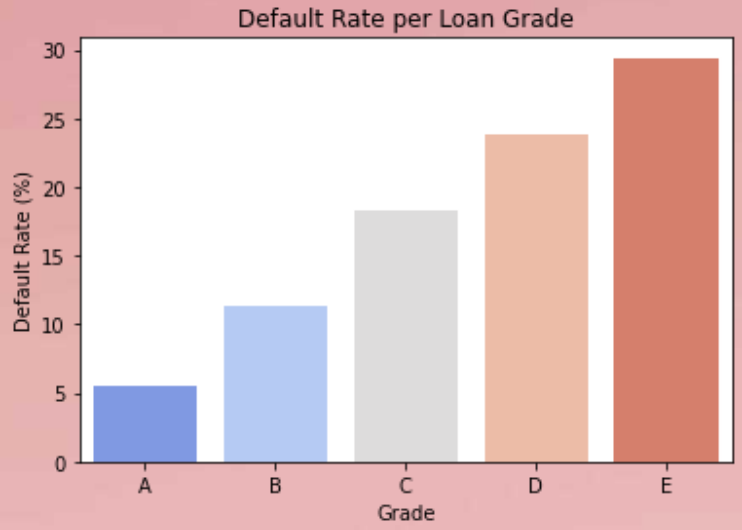


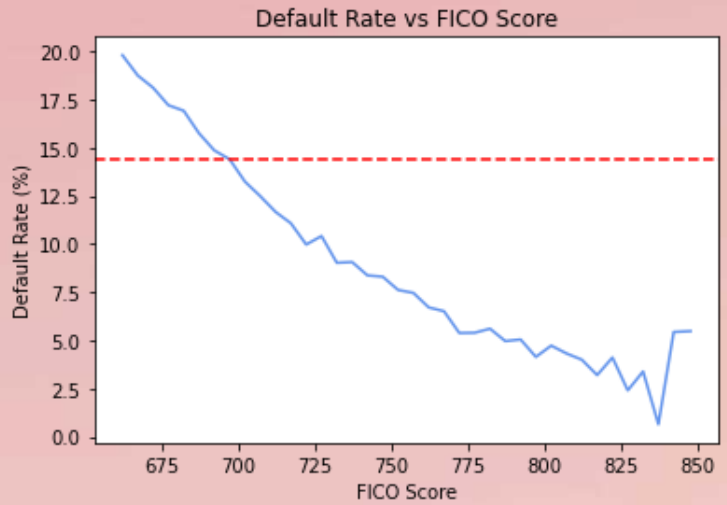


- **DTI**: Debt-to-income ratio, sum of outstanding debt divided by income
- **FICO Score**: Credit score ranging from 300 to 850

Default Rate Analysis

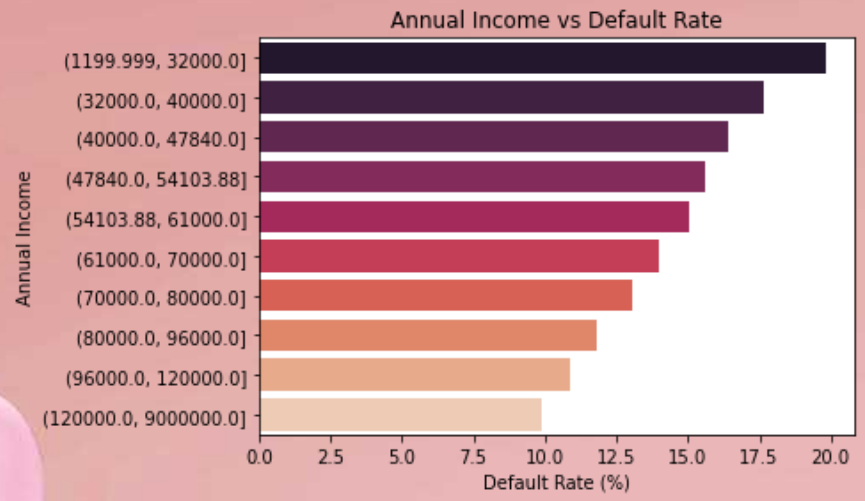


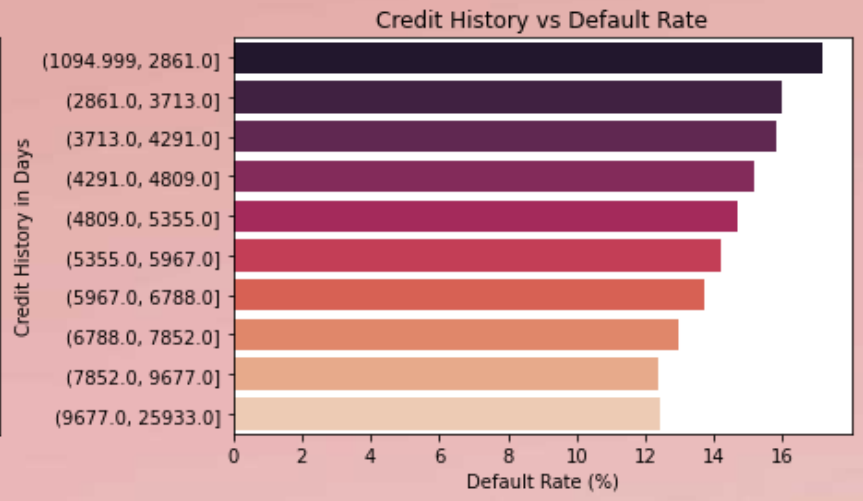


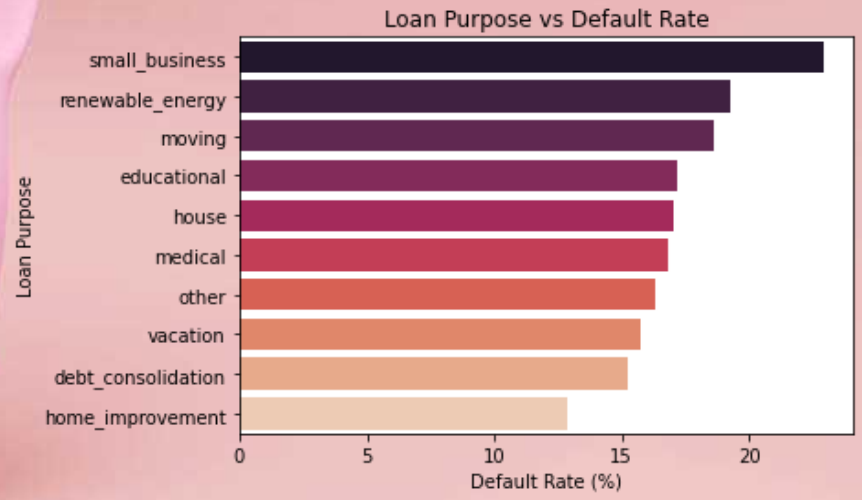


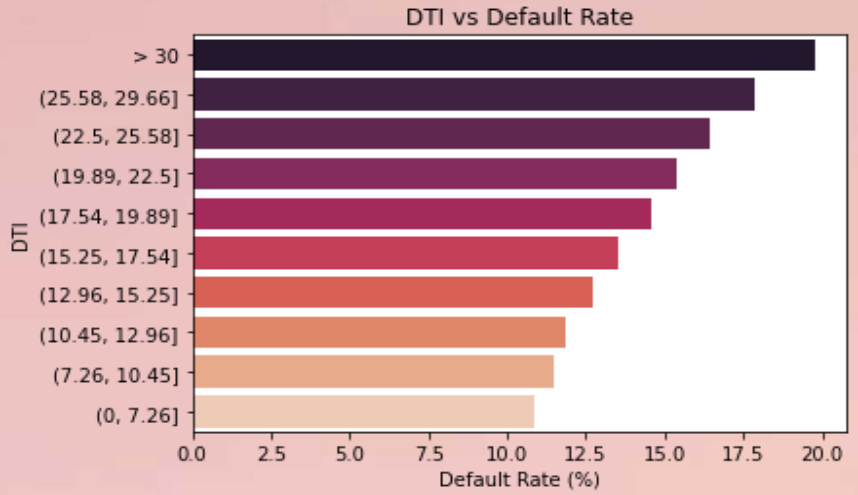
$$\text{Default rate} = \frac{\text{\# of loans charged off}}{\text{total \# of loans issued}}$$

Default Rate vs. Key Features

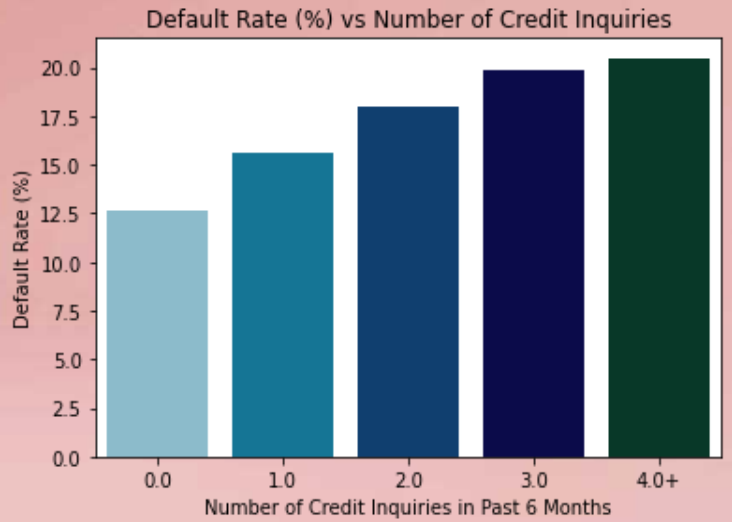


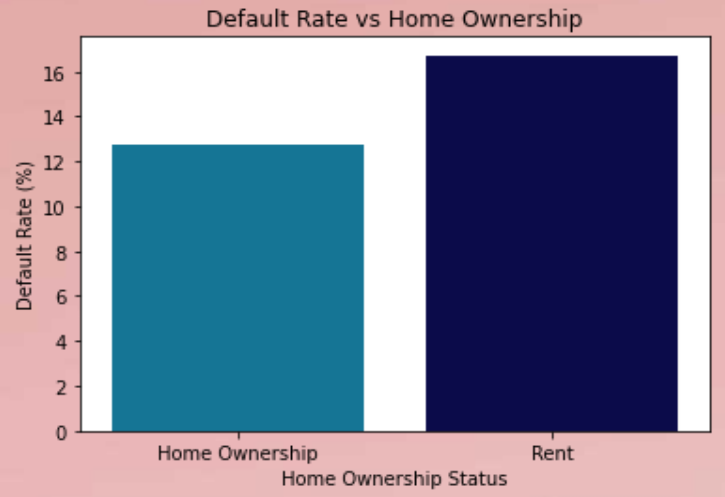


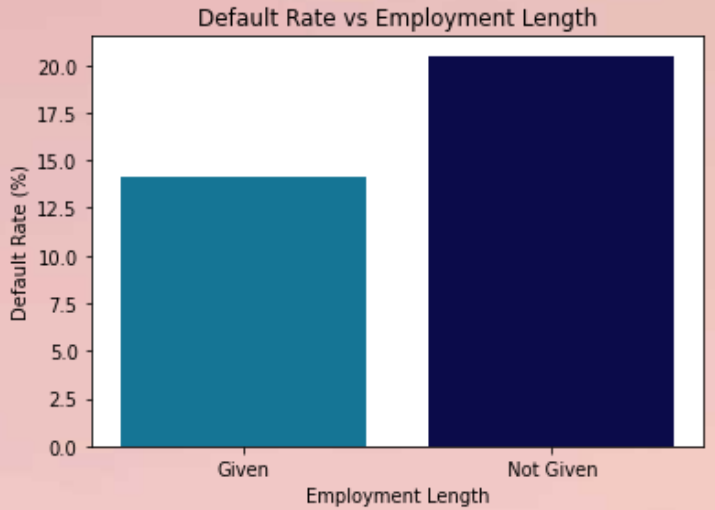




Default Rate vs. Key Features







Models

Logistic Regression

	Pred: Charged Off	Pred: Fully Paid
True: Charged Off	59,207	36,287
True: Fully Paid	213,450	351,959

	LogR Return	NULL Return
Total Return:	8.41%	9.34%
Compounded Annual Return*:	4.12%	4.57%

Accuracy	62.21%
Loan Approval Rate	58.74%
Precision for 'Pred: Fully Paid'	90.65%
Recall	62.00%
Specificity	62.25%

* assumes 24 month average life of the loan

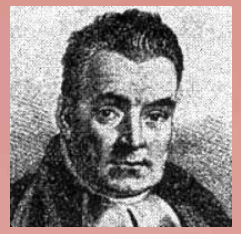
Support Vector Machine – Linear SVC

	Pred: Charged Off	Pred: Fully Paid
True: Charged Off	250	95244
True: Fully Paid	565	564844

	SVM Return	NULL Return
Total Return:	9.28%	9.34%
Compounded Annual Return*:	4.54%	4.57%

Accuracy	85.50%
Loan Approval Rate	99.87%
Precision for 'Pred: Fully Paid'	85.57%
Recall	0.26%
Specificity	99.90%

* assumes 24 month average life of the loan



Linear Discriminant Analysis

	Pred: Charged Off	Pred: Fully Paid
True: Charged Off	55,733	39,761
True: Fully Paid	211,823	353,586

	LDA Return	NULL Return
Total Return:	9.28%	9.34%
Compounded Annual Return*:	4.54%	4.57%

* assumes 24 month average life of the loan

Accuracy	61.93%
Loan Approval Rate	59.52%
Precision for 'Pred: Fully Paid'	89.89%
Recall	58.36%
Specificity	62.54%

Random Forest

	Pred: Charged Off	Pred: Fully Paid
True: Charged Off	65,165	30,329
True: Fully Paid	197,240	368,169

	RF Return	NULL Return
Total Return:	9.64%	9.34%
Compounded Annual Return*:	4.71%	4.57%

Accuracy	65.56%
Loan Approval Rate	60.30%
Precision for 'Pred: Fully Paid'	92.39%
Recall	68.24%
Specificity	65.12%

* assumes 24 month average life of the loan

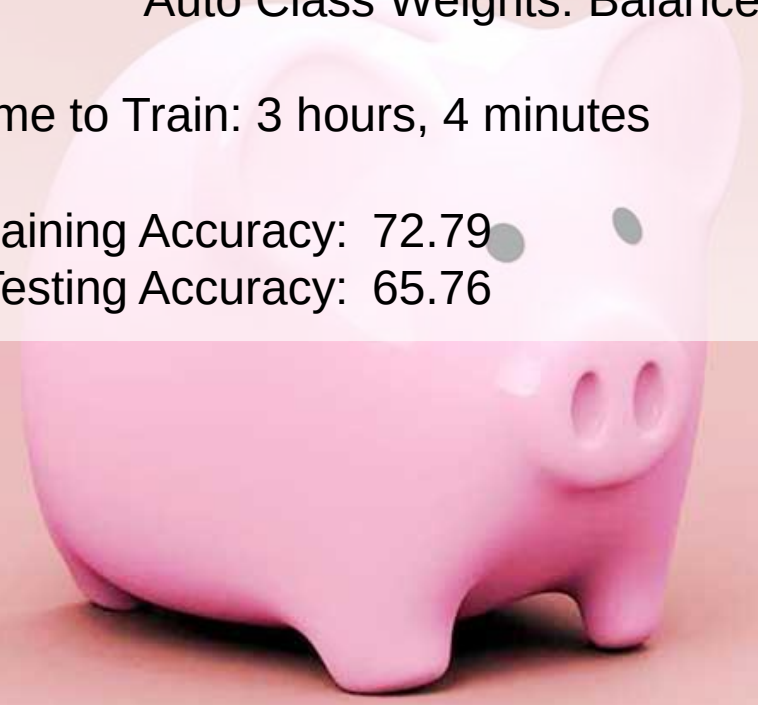
CatBoost

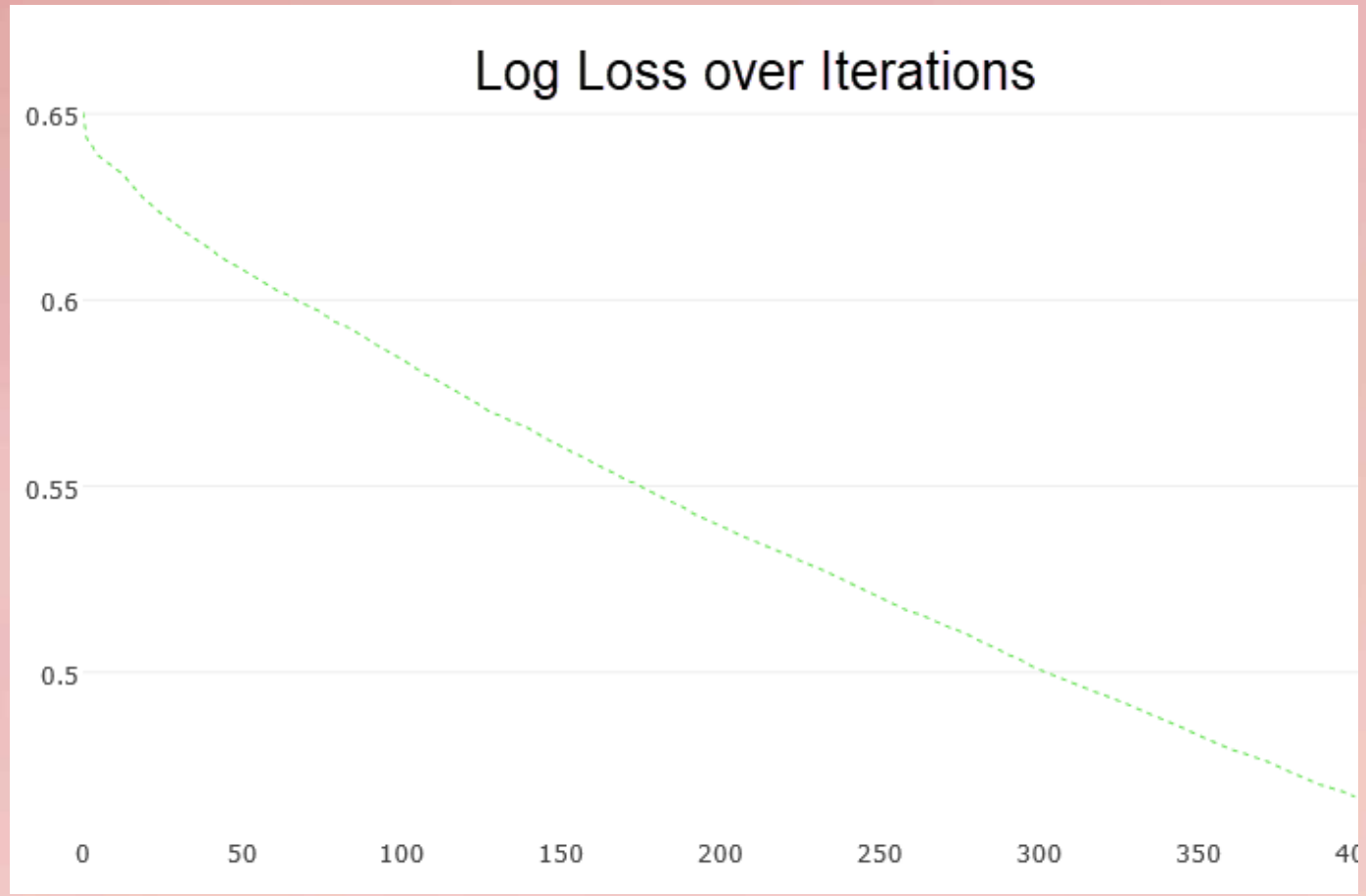
Best parameters:

- Maximum Tree Depth: 8
- Iterations: 400
- L2 Leaf Regularization: 4
- Learning Rate: 1
- Auto Class Weights: Balanced

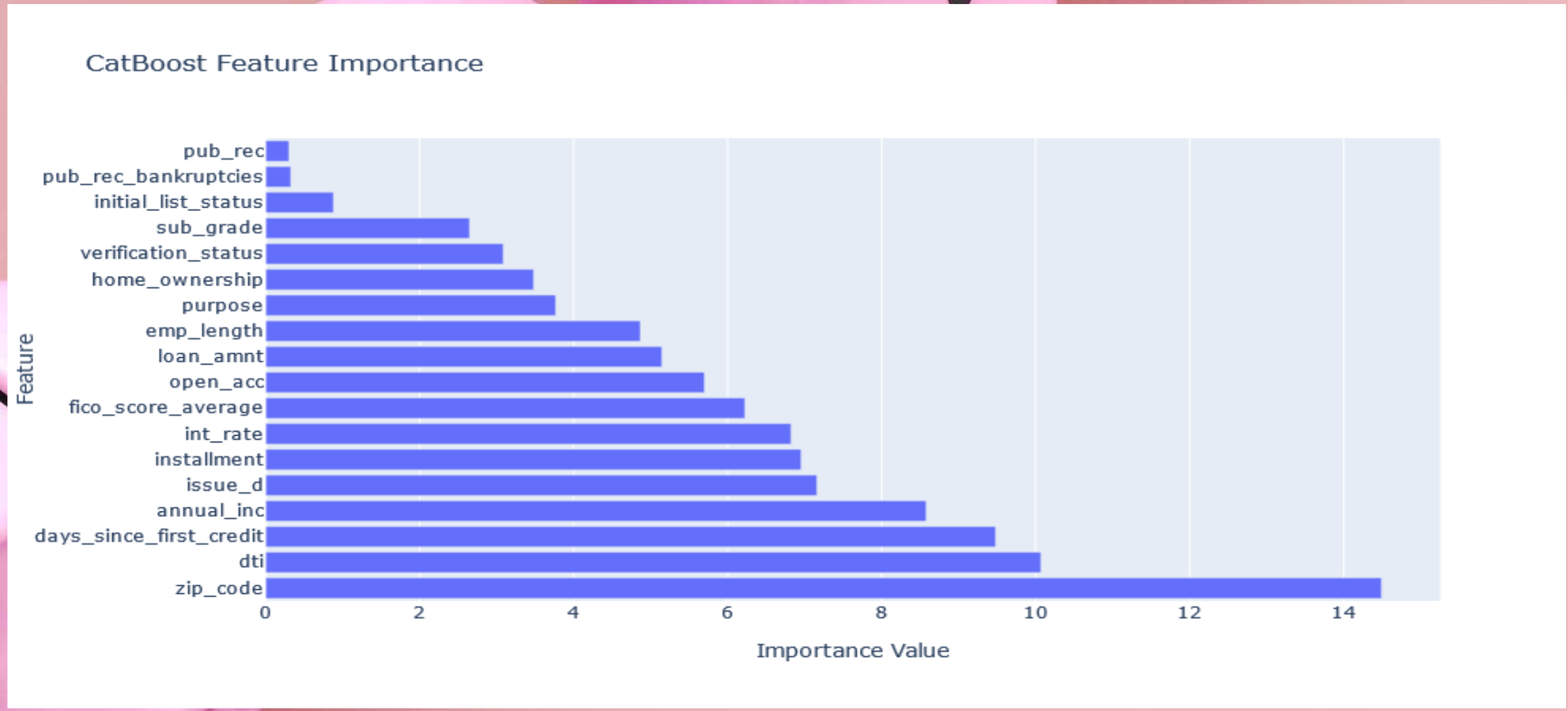
Time to Train: 3 hours, 4 minutes

Training Accuracy: 72.79
Testing Accuracy: 65.76





CatBoost


CatBoost Feature Importance
pub_rec
pub_rec_bankruptcies
initial_list_status
sub_grade
verification_status
home_ownership
purpose
emp_length
loan_amnt
open_acc
fico_score_average
int_rate
installment
issue_d
annual_inc
days_since_first_credit
dti
zip_code
Feature
0
2
4
6
8
10
12
14
Importance Value

CatBoost

	Pred: Charged Off	Pred: Fully Paid
True: Charged Off	63,911	31,583
True: Fully Paid	157,372	408,037

	CatBoost Return	NULL Return	Idealized Return**
Total Return:	12.44%	9.34%	16.90%
Compounded Annual Return*:	6.04%	4.57%	8.12%

Accuracy	71.41%
Loan Approval Rate	66.52%
Precision for 'Pred: Fully Paid'	92.82%
Recall	66.93%
Specificity	72.17%

* assumes 24 month average life of the loan

** assumes 100% accuracy in picking loans

CatBoost

Manual Portfolio Optimization

Process:

- Exclude small business loans
- Exclude employment length where N/A
- Exclude renters
- Exclude Nevada residents
- Exclude annual incomes less than $42,000

Result:

- ROI went from 12.44% to 12.65%

Results & Conclusions

Is Lending Club the Best Investment Option?

Null Model	4.57%
CatBoost	**6.04%**
Lending Club Ideal	8.12%

VS

10-Year Treasury Notes*	2.8%
AAA Bonds**	5.9%
Speculative-Grade Bonds**	8.0%
Real Estate***	10.5%
S&P 500 Average Annual Return****	10.7%

Lending Club Stock Price History

* https://finance.yahoo.com/quote/%5ETNX/
** https://www.mindfullyinvesting.com/historical-returns-of-corporate-bonds/
*** https://www.investopedia.com/ask/answers/060415/what-average-annual-return-typical-long-term-investment-real-estate-sector.asp
**** https://www.fool.com/investing/how-to-invest/stocks/average-stock-market-return/

Lessons Learned and Next Steps

- More feature engineering to possibly improve performance of low performing models
- Understand why borrowers’ credit reports were monitored after origination
- Utilize neural networks
- Model other targets: PnL, Lending Club's approval criteria

Questions & Answers

THANK YOU