ASAPP
AI FOR ENTERPRISE

Personal Credit Algorithm Proposal

June 2019

The Problem

- Customers with good credit record are potentially erroneously flagged as bad customers by the FICO Scores
  - Their loan applications get rejected for reason like lack of many saving accounts, for example

- Borrowers with financial problems are not correctly and timely reflected in the Scores, leading to potential risk of loan delinquencies to the bank

- Constitution of FICO Scores
  - Payment History – 35%
  - Amount Owed – 30%
  - Length of Credit History – 15%
  - Credit Mix or Types of Credit Used – 10%
  - New Credit – 10%

- Merely relying on a FICO credit score falsely caused banks to take on avoidable risk and lose lots of businesses

- How can banks make more precise lending decision?

Data

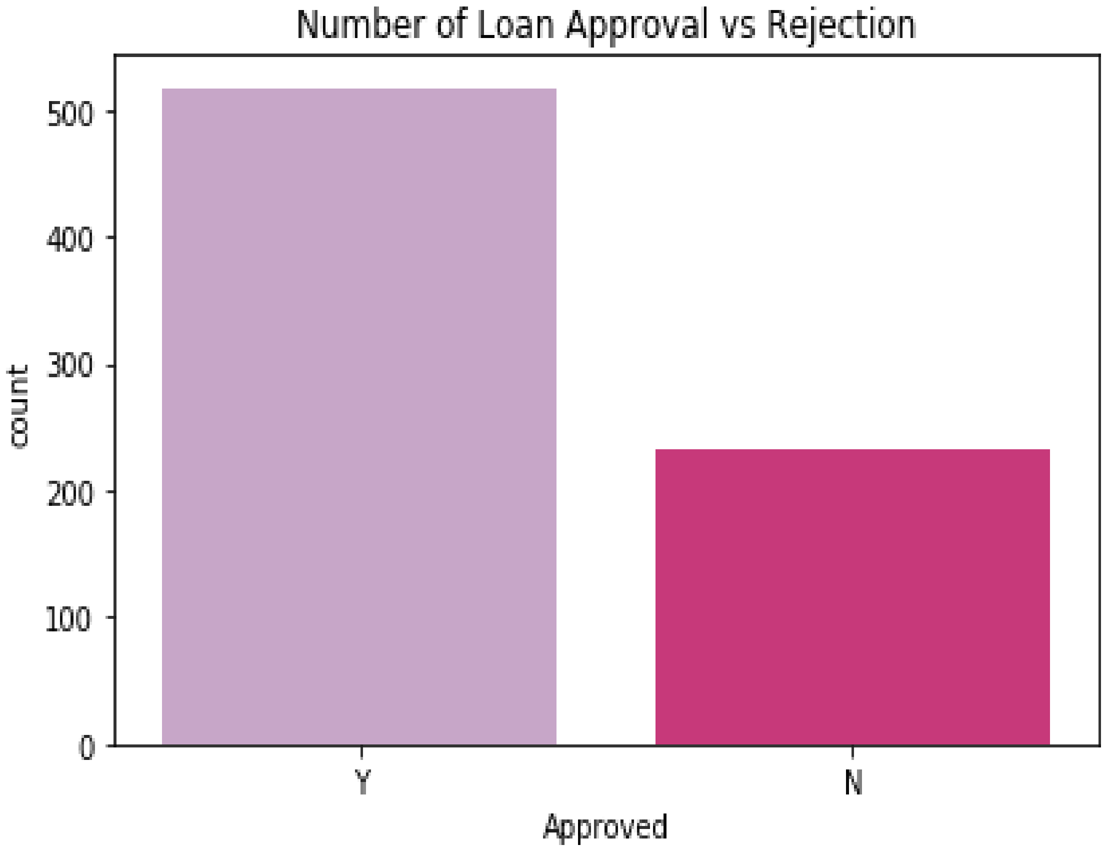
- Banks have a lot more information about their borrowers than what the FICO Scores represents

- Four Datasets

	Borrower	Credit	Application	Result
Features	1. Customer ID 2. Checking Account Balance 3. Savings Account Balance 4. Debts Paid 5. Current Open Loan Applications	1. Customer ID 2. Loan Payoff Period In Months 3. Loan Reason 4. Requested Amount 5. Interest Rate 6. Co-Applicant	1. Customer ID 2. Years At Current Employer 3. Years In Current Residence 4. Age 5. Rent Or Own Home 6. Type Of Current Employment 7. No. Of Dependents	1. Customer ID 2. Was The Loan Approved

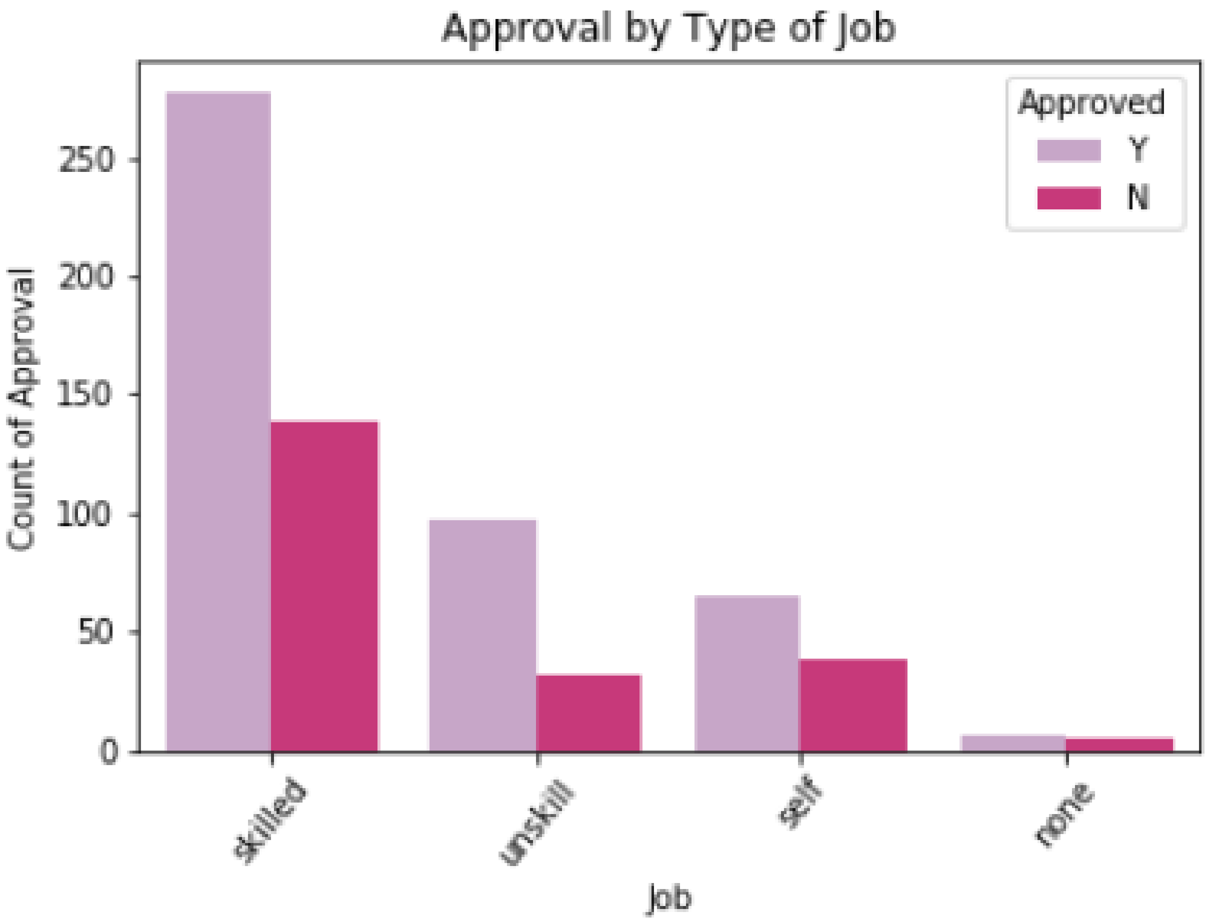
- We will develop an algorithm on the internal data to better segment “good borrowers” and “bad borrowers”

EDA – Key Features



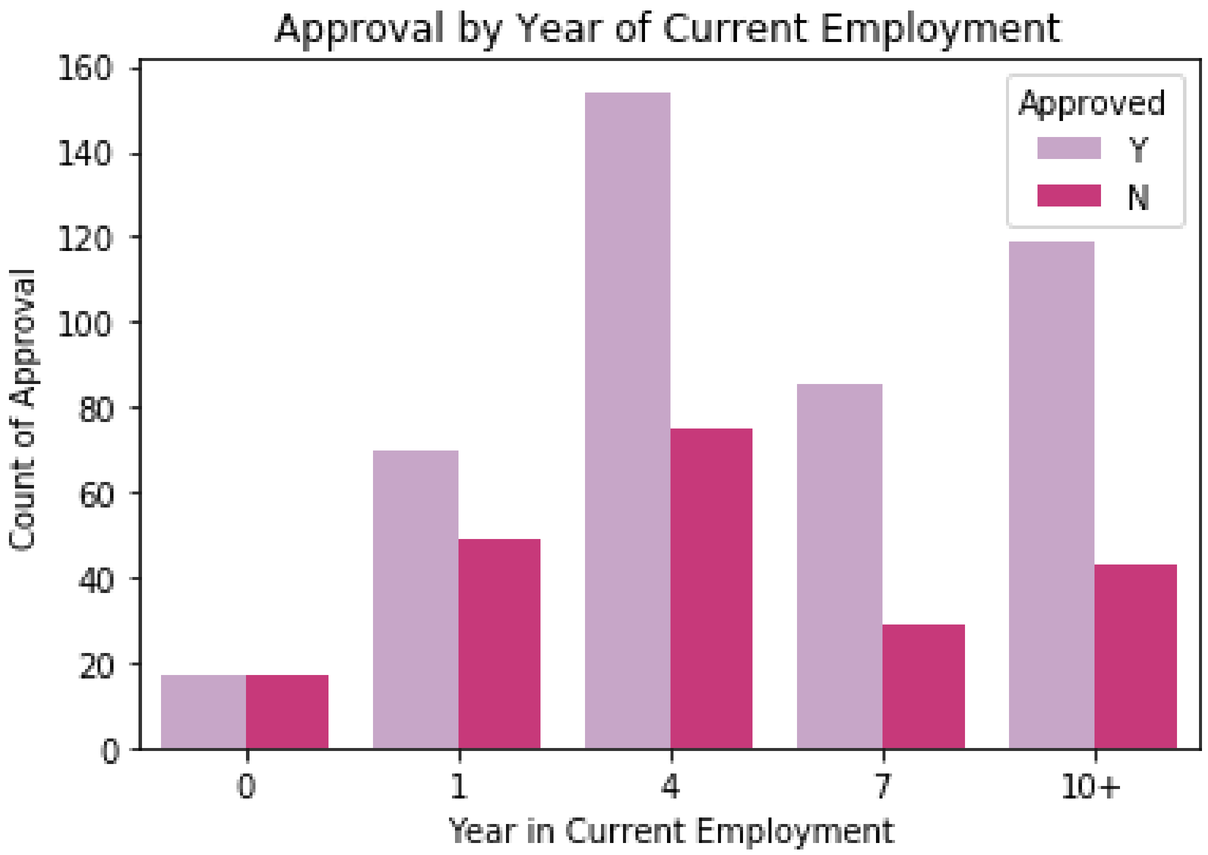
- In this data set, “Approved” is >1x more than “Rejected”

EDA – Key Features



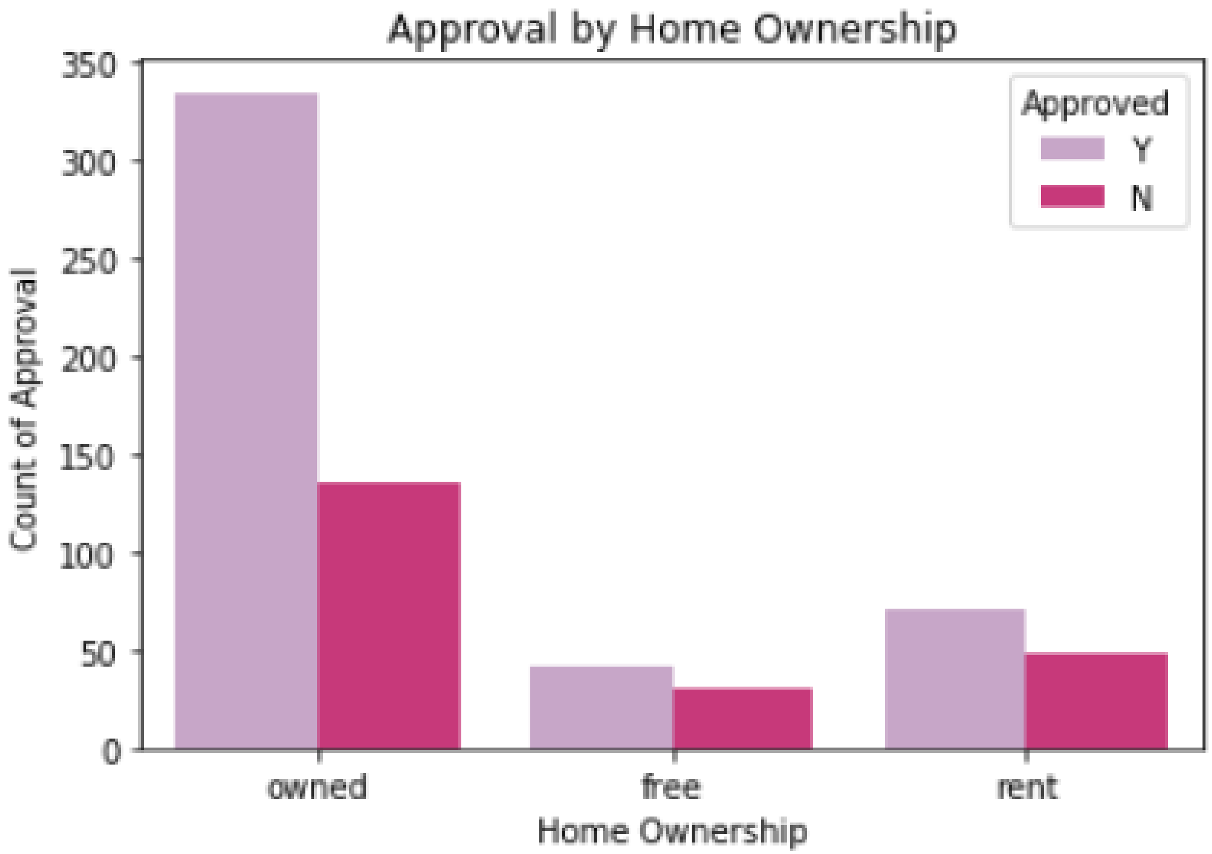
- Banks favored borrowers with a job and regular income
- Self-employed borrowers are less preferred because of the riskier and less stable profit from a business, but they are better than the unemployed

EDA – Key Features



- For the same reason, banks liked to give loans to those with a job tenure equal to or longer than four years
- Explained why the young professionals were less likely to get loans

EDA – Key Features

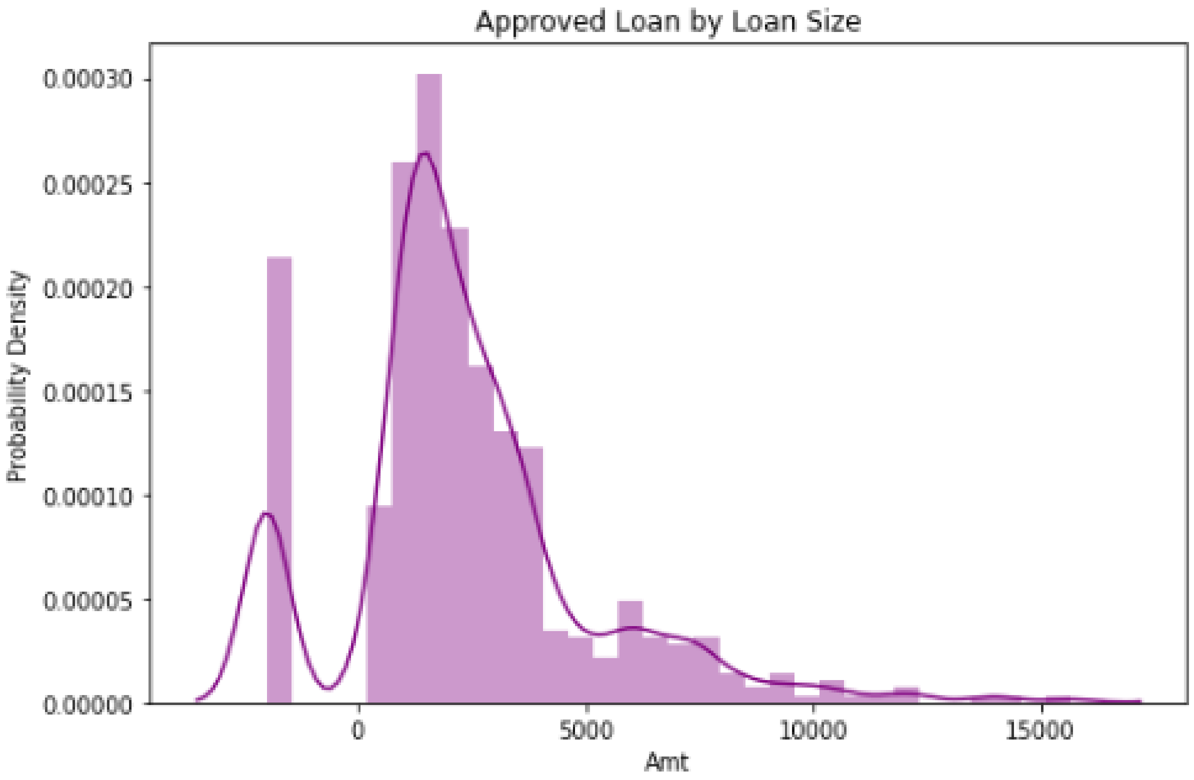


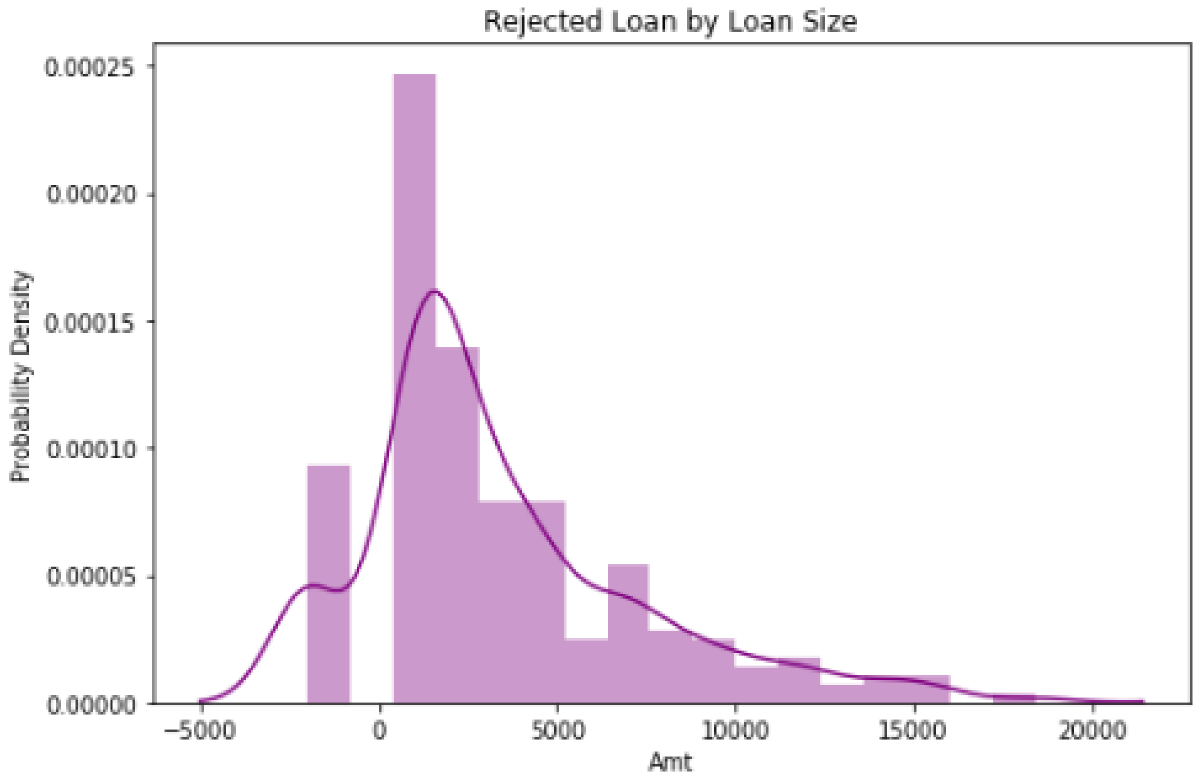
- Banks were more likely to grant loans to home owner
- Real estate is a collateral

EDA – Key Features

- Interestingly, the chance of getting approved was higher for the group of customers who delayed their loans, comparing to the group that paid off their debts
- Not all late payers are bad customers. The opposite of bad, they are gems to the banks, as long as they have the ability to repay
- Banks can charge penalty interest and late fees
- Late payment undermines the FICO Scores
- Hence, relying simply on the Scores would undermines banks' business growth

EDA – Key Features

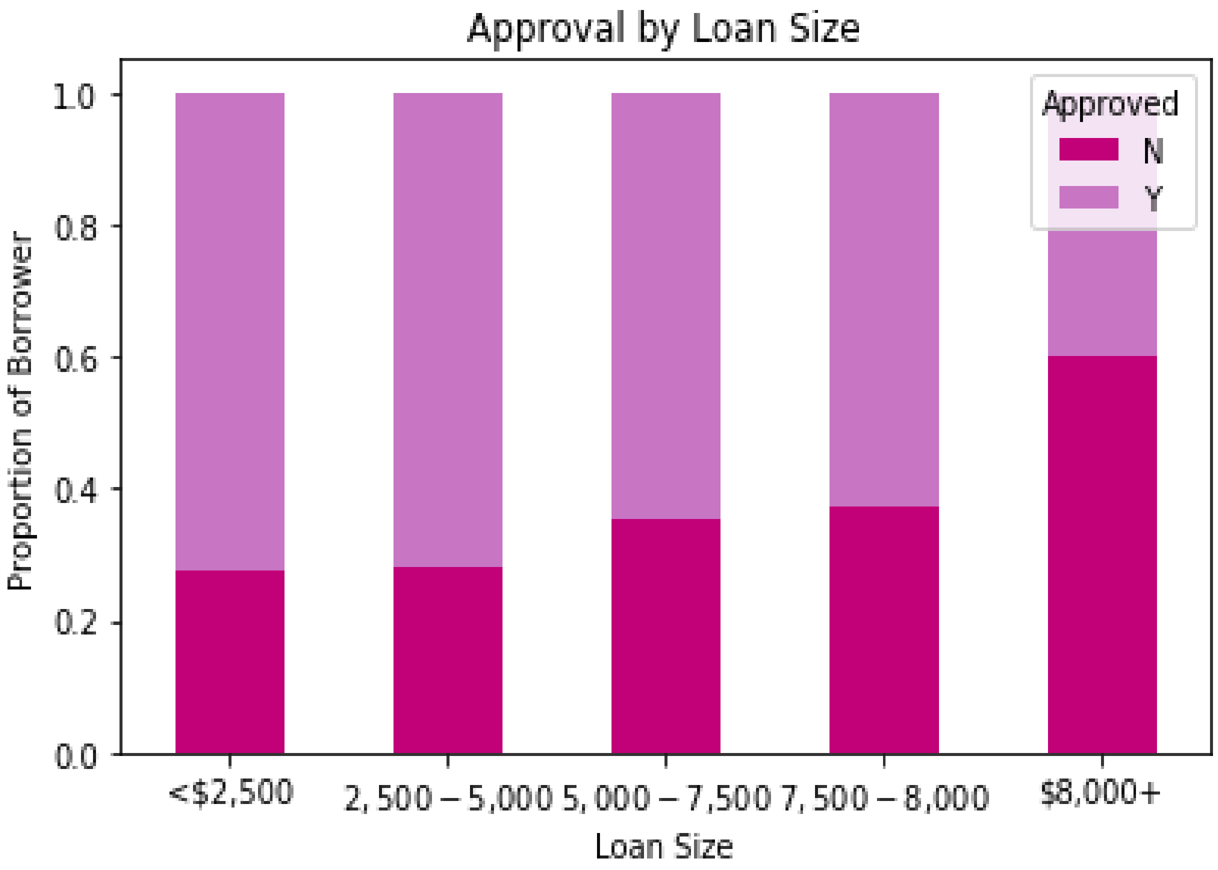




- Most of the loans that the banks gave out were relatively small under $3,000
- Size of rejected loans went beyond that of approved loans, implying that banks were reluctant to underwrite ultra large loan to individuals

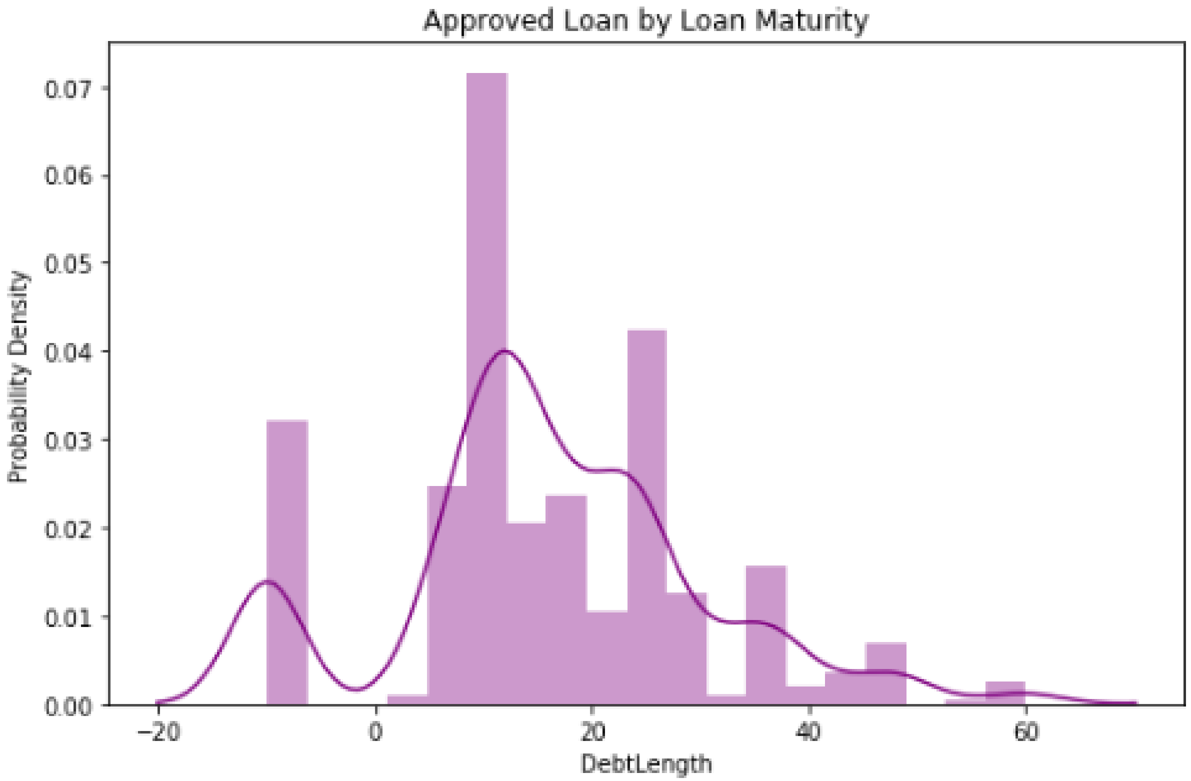
**We replaced n/a with -2000, the bar to the very left.

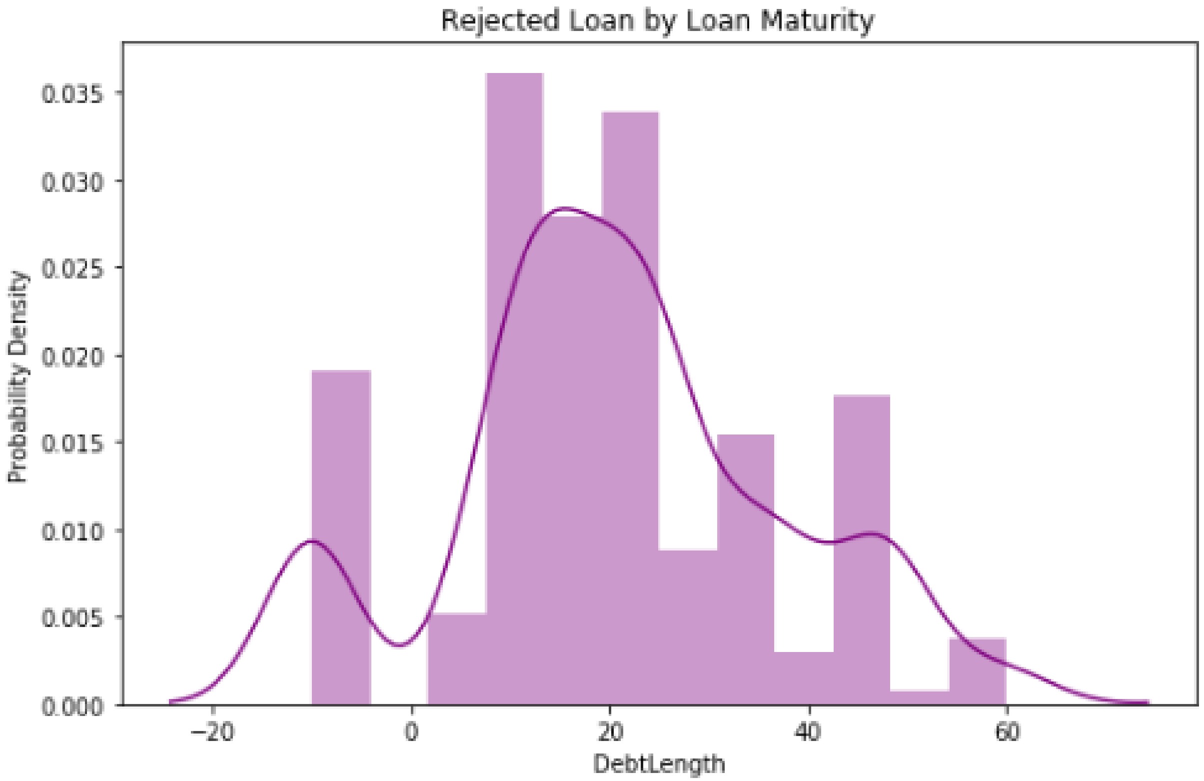
EDA – Key Features



- It is clear that banks wanted smaller size loans
- Despite underwriting a large volume of higher risk loans, banks diversified the risk by managing the size of the loans

EDA – Key Features

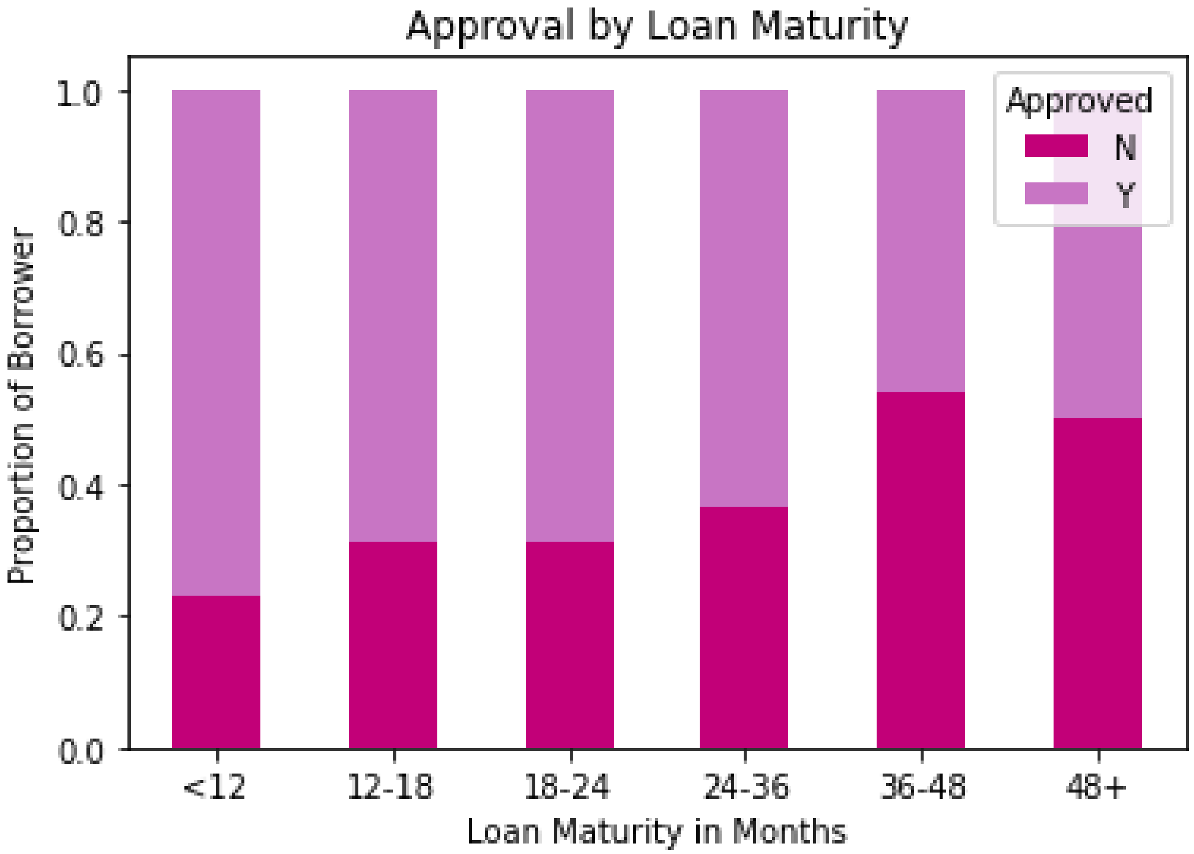




- The distributions are fat, indicating that length of debt might not be a very good positive indicator of loan approval. Let's drill deeper.

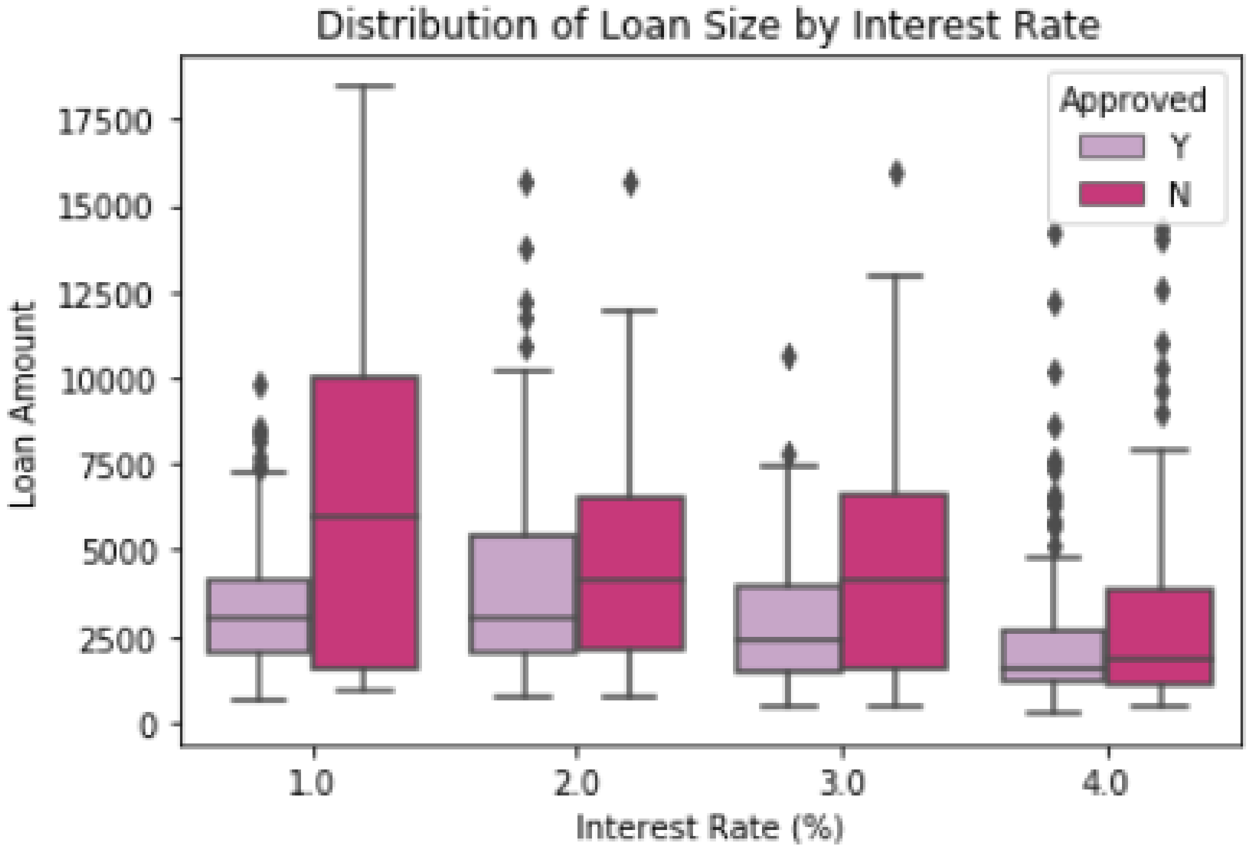
**n/a was filled with -10, the bar to the very left.

EDA – Key Features



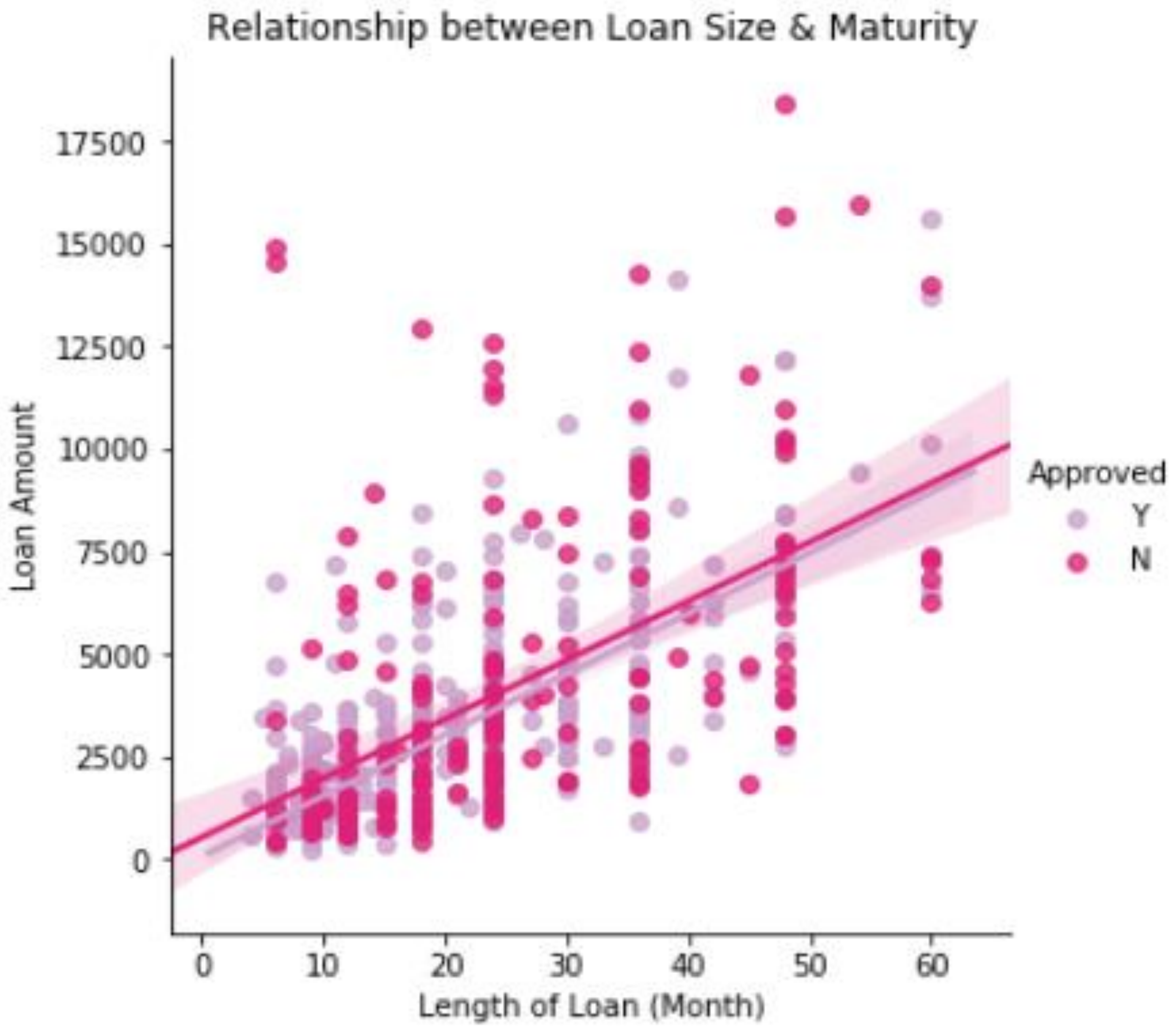
- However, the chart shows that there is a clear trend of increasing chance of rejection with time of loan

EDA – Interaction Effects



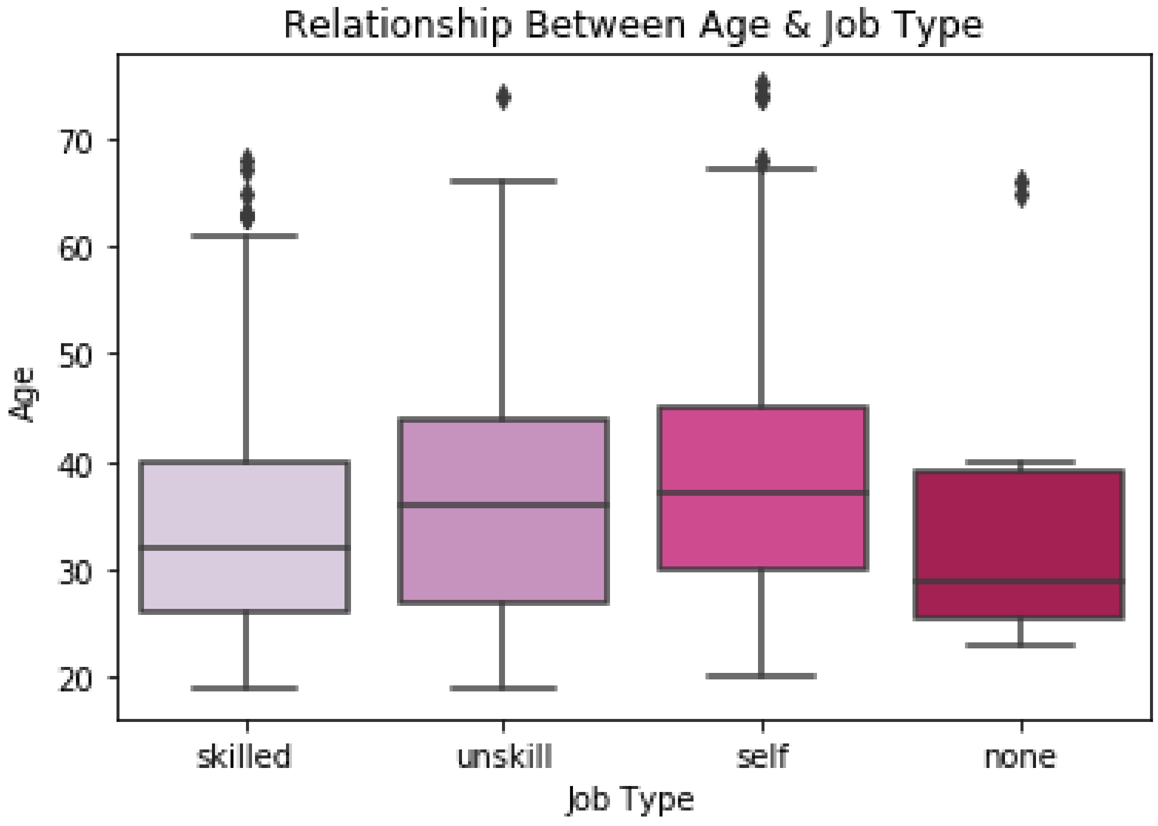
- At 4% interest rate, median loan size was around $2,000 only
- The dispersion of the amount is also smaller than that of other rates, represented by the lower IQRs

EDA – Interaction Effects



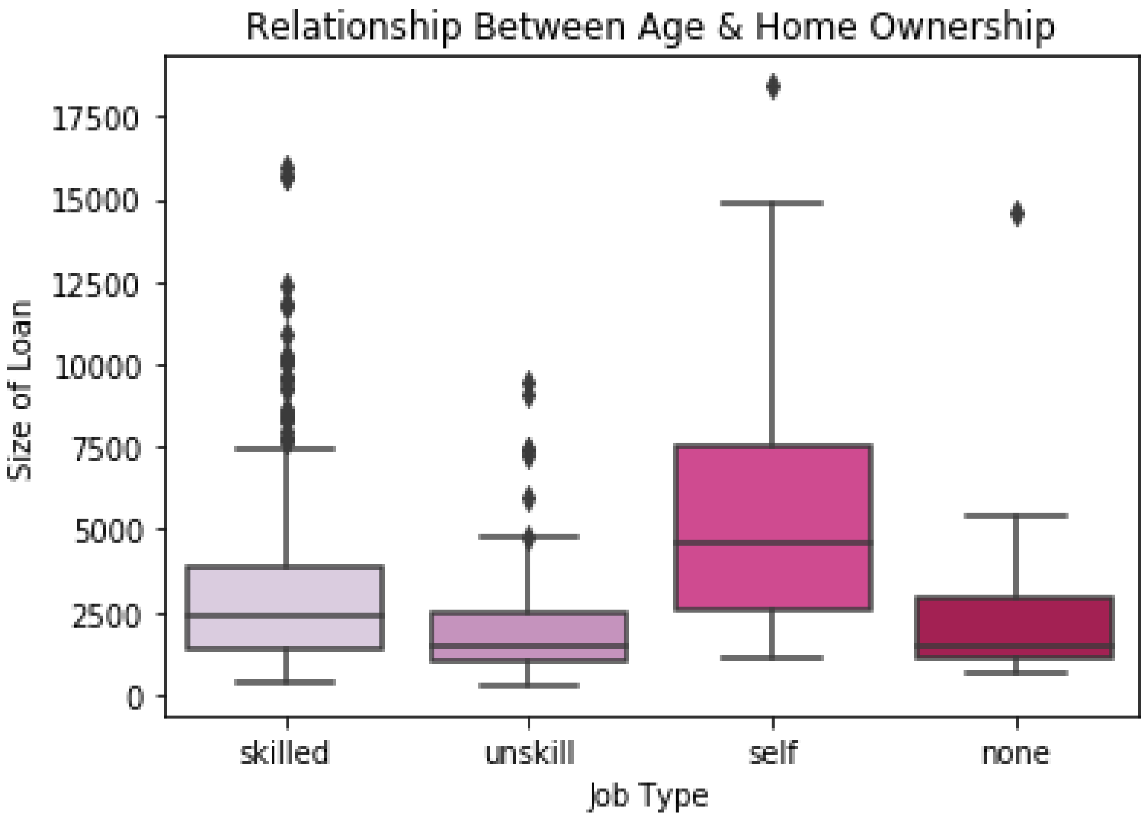
- There was a weak positive correlation of 0.63 between the size and maturity of loan
- Also reinforced our previous observation that loan application above $5,500 tended to have a higher chance of being rejected, as indicated by the red dots on upper part of the chart

EDA – Interaction Effects



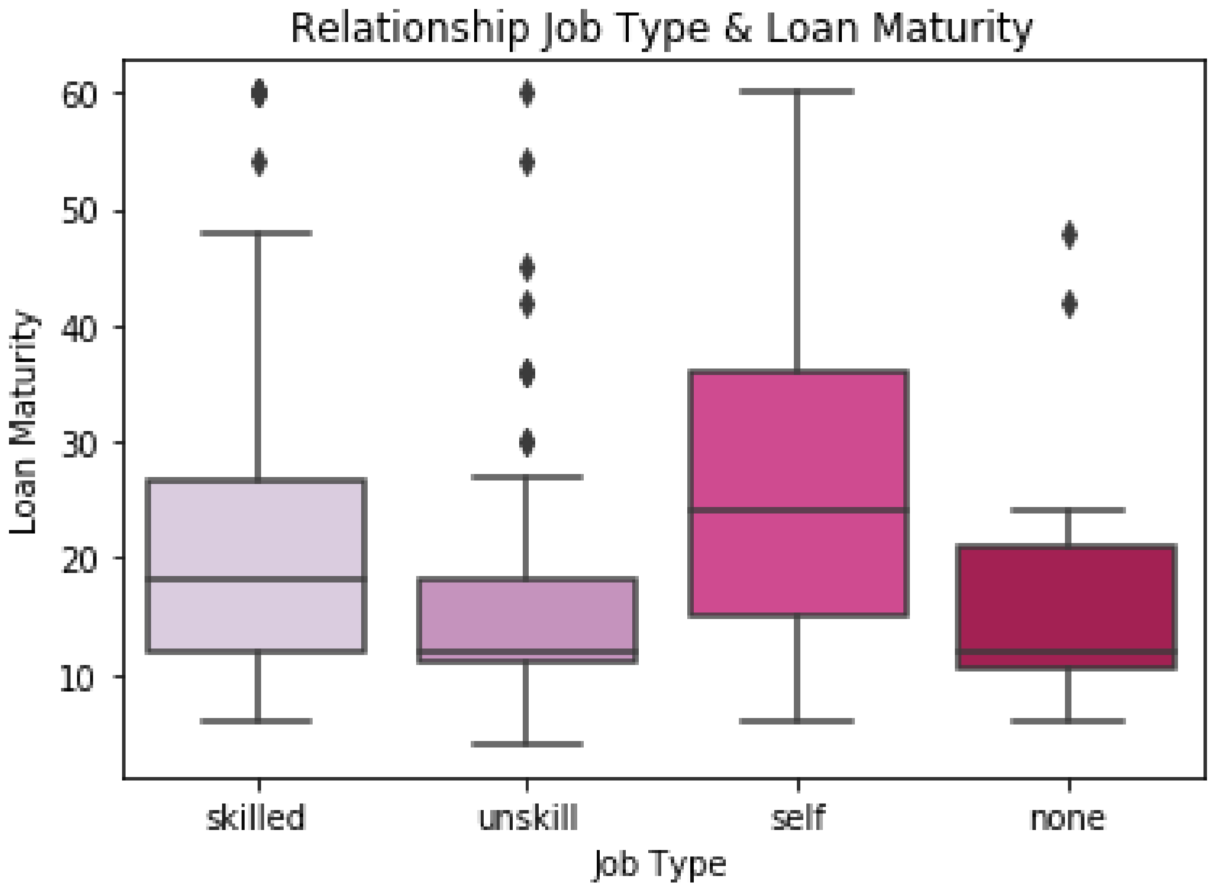
- The distributions across “skilled”, “unskill”, and “self” are pretty similar, suggesting that the age of borrowers might be independent of the three job types
- But it distinguishes the unemployed

EDA – Interaction Effects



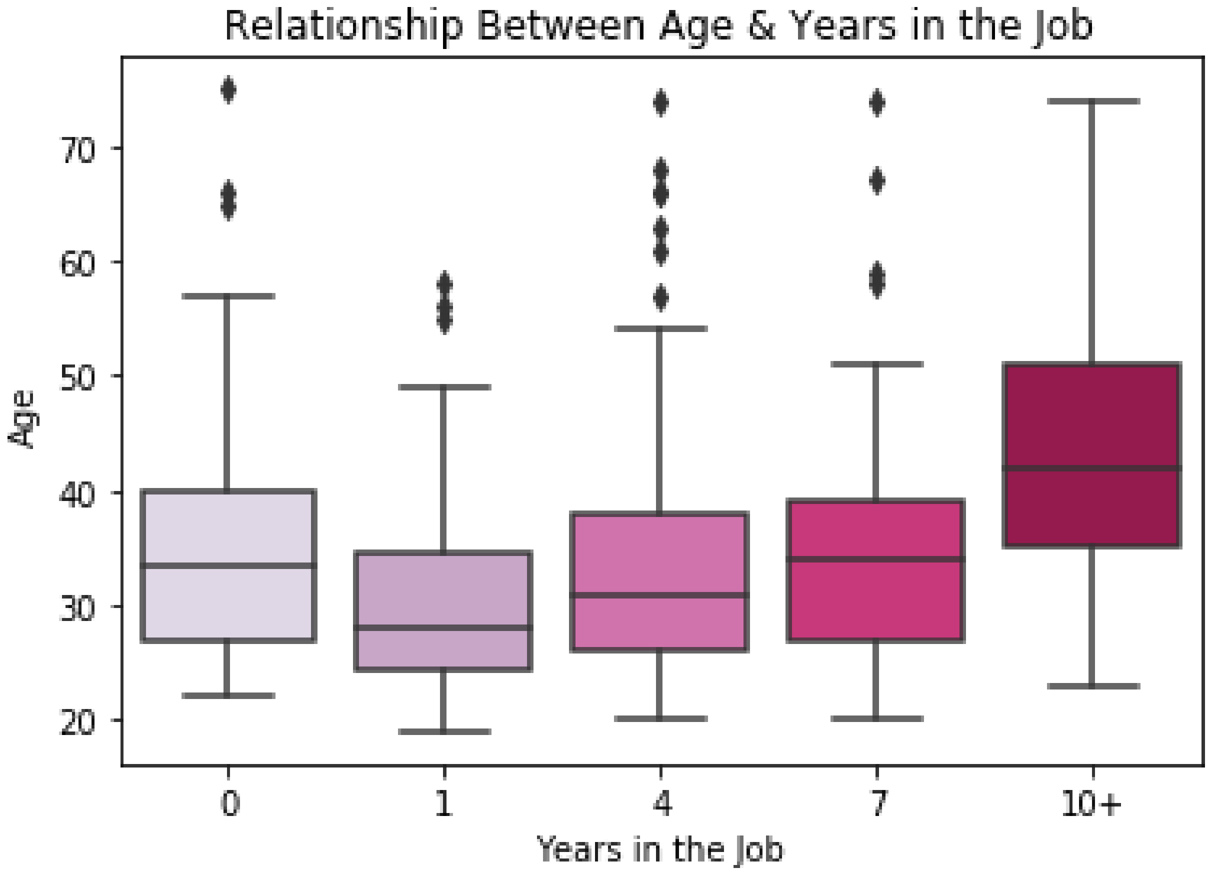
- Self-employed inclined to borrow larger loan with median size of $5,000. The borrow may go to finance their companies

EDA – Interaction Effects



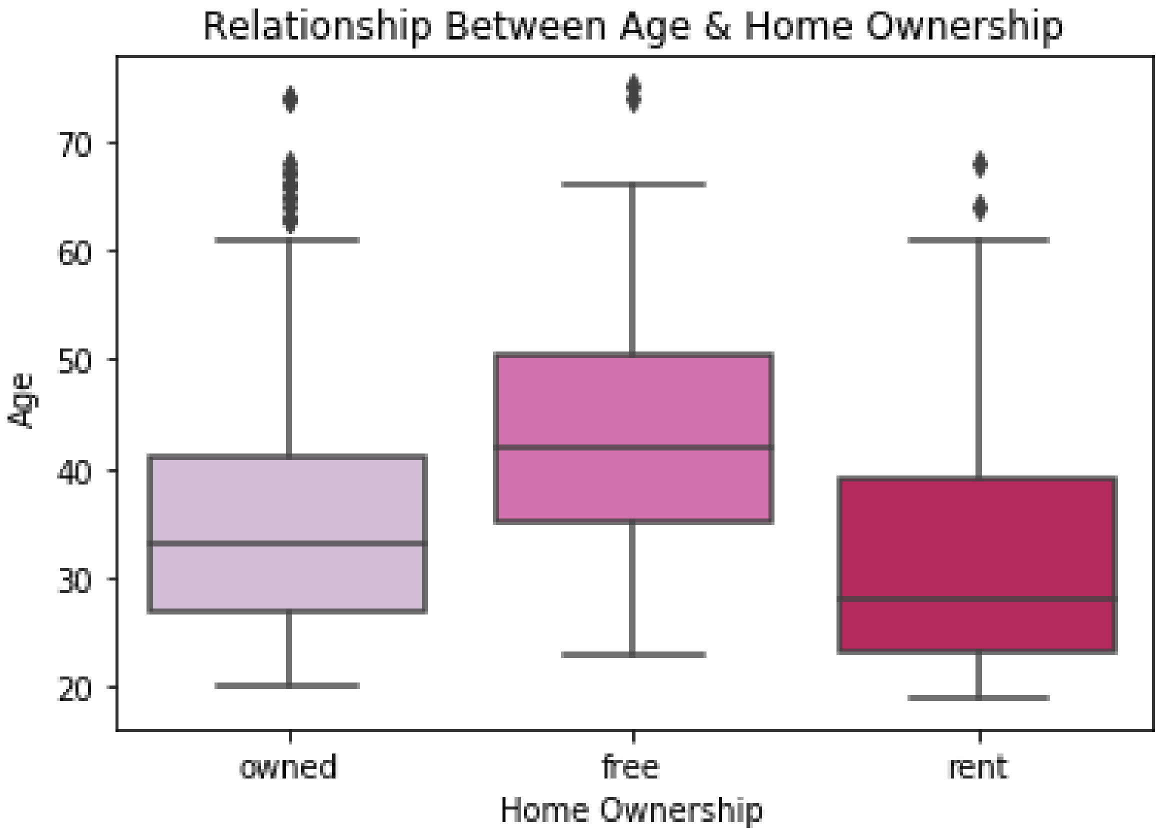
- Perhaps for the same reason, self-employed inclined to keep the loan longer with median of 2 years
- Liquidity is always critical to a business

EDA – Interaction Effects



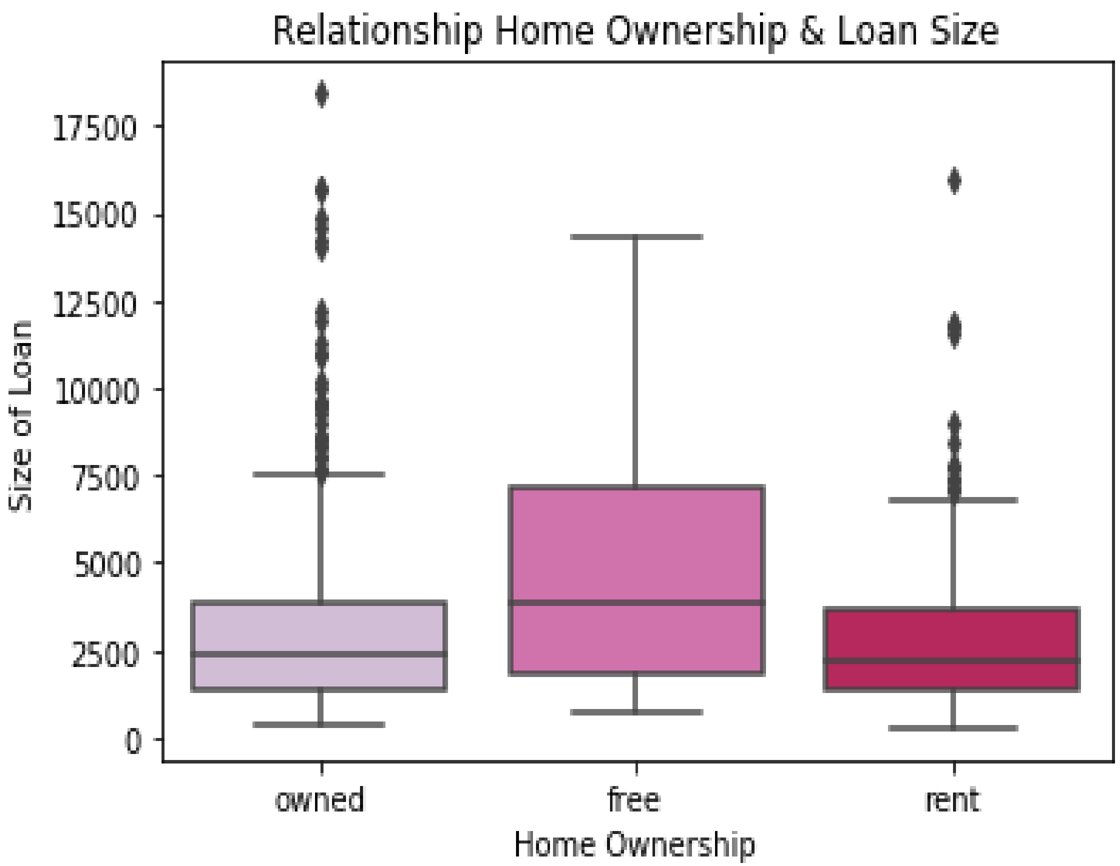
- The age of borrowers seems to be independent of the years of jobs under 10, too. Nevertheless, it distinguishes those stay in existing company for over 10 years. The median age moved up for the “10+” group too
- Career movement tends to slow down from 40s

EDA – Interaction Effects



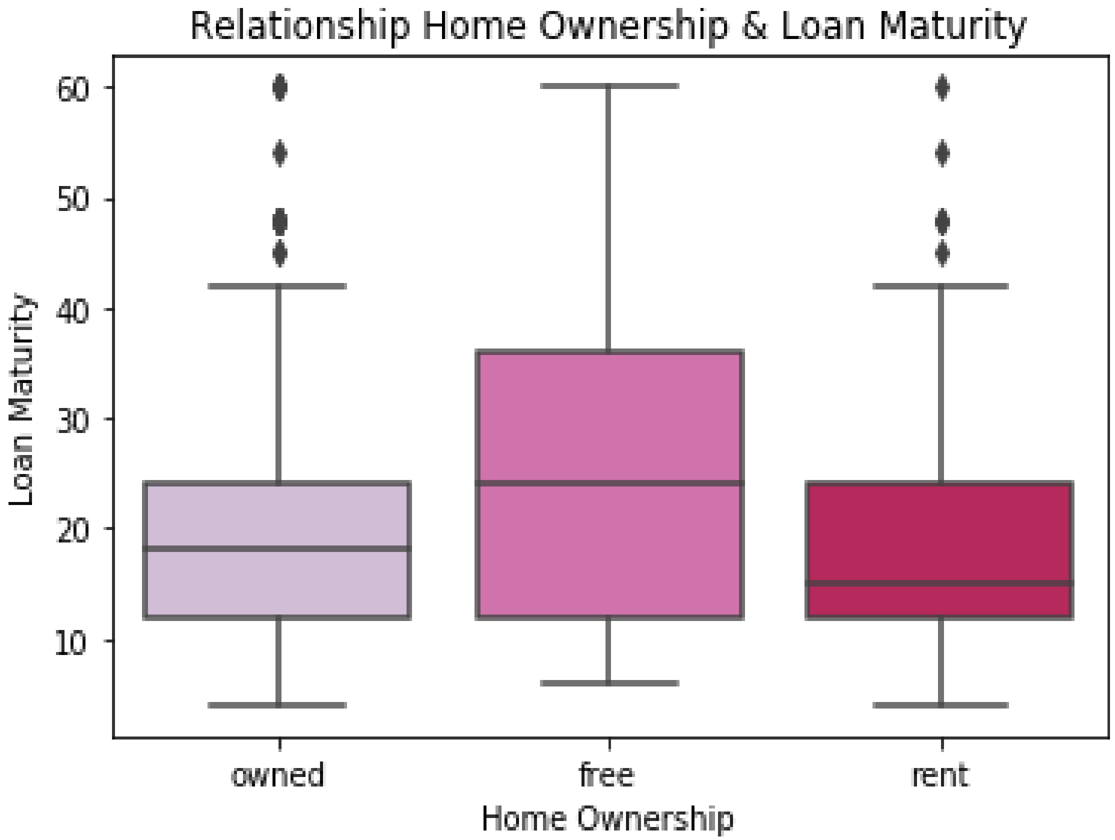
- Although the distributions across types of home ownership are similar, the median age of those live in free housing is relatively high
- Senior people with little savings are easier to obtain free housing

EDA – Interaction Effects



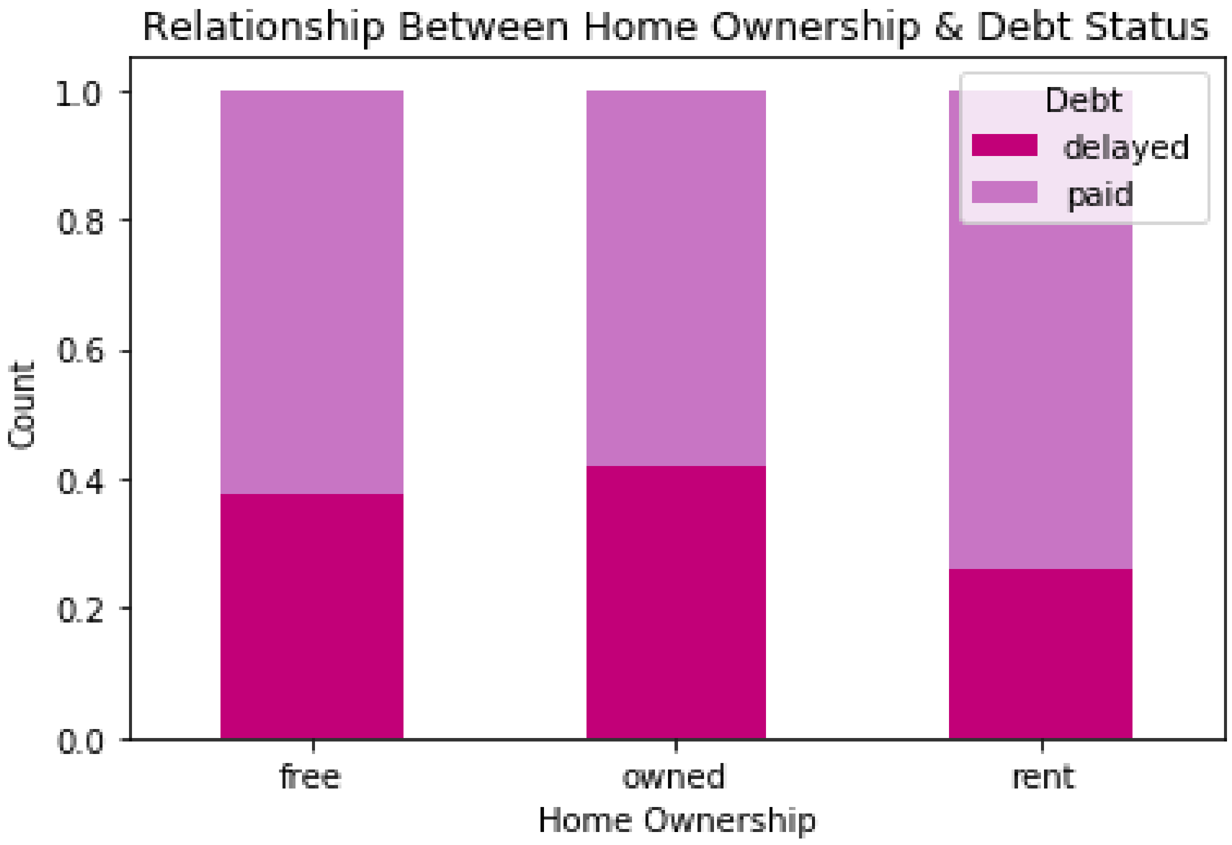
- Since we suspect that older people needed to live in free housing because of their financial needs, it might also explain why they needed larger loans than the average

EDA – Interaction Effects



- And they wanted to pay back slowly

EDA – Interaction Effects



- Borrowers who had extra cash outflow from rent were more punctual in payment
- The large rent expenses might cause borrowers to watch their finance more closely

Solution: Logistic Regression Model

- **The Model**
  - Classical model that provides confidence intervals for coefficients, high inferences
- **Limitations**
  - Assumes linearity of independent variables and log odds
- **Began with the baseline model, we removed the insignificant independent variables. The refined model achieves an AUC of only 0.65. The algorithm classifies 73% correctly**

Solution: XGBoost Model

- **The Model**
  - Tree-based model that reduces residuals of prior result and attaches regularization

- **Special Features**
  - High performance, usually delivers excellent predictive result, less overfitting

- **Limitations**
  - Depends largely on finding the right combination of the tuning parameters

- **AUC is 0.73, the algorithm classifies 71% correctly**

Solution: Random Forest Model

- **The Model**
  - Tree-based model that reduces residuals of prior result and attaches regularization

- **Special Features**
  - High predictive power, less overfitting, provides feature importance

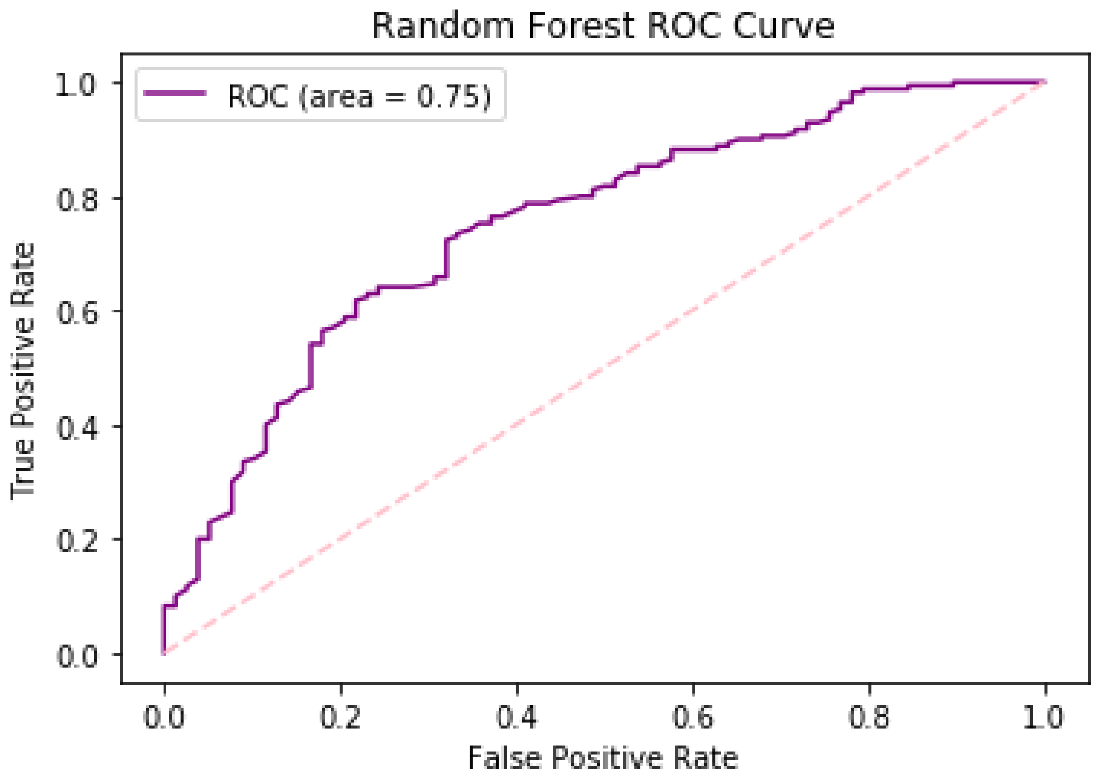
- **Limitations**
  - Slower running time
  - Considered a black box algorithm

- **AUC is 0.75, Random Forest classifies 73% correctly**

Results

	Logistic Regression	XGBoost	Random Forest
AUC	0.6536	0.7331	0.7453
Accuracy	0.7275	0.7130	0.7270

- Random Forest model performed the best with cross validated AUC of 0.75
- 73% (true-positive) and 59% (true-negative) accuracy rates, respectively
- Looking beyond accuracy rate, false positive is equally important as an indicator of bad loans to the lending business
- Hence, instead of maximizing the accuracy, we pick the model with the highest ROC score.
- Bank should choose the optimal threshold according to its risk profile



Conclusion & Next Steps

- **We suggest to incorporate additional available data**

  For example:
  - Sex (Male/Female)
  - Checking and saving accounts balance against debt
  - Repeated borrow rate with the bank

- **We provide**
  - High precision (AUC > 0.85) and performance algorithm
  - Statistical reliability tests
  - Training
  - World class customer service

Conclusion & Next Steps

- **Machine learning models are in many ways improving personal lending businesses**
  - From better fraud detection or more credit underwriting, to effective customer acquisition
- **ASAPP is one of the best artificial intelligence platform globally**
  - Applying ASAPP’s best model to the bank’s exclusive and valuable internal customer information would allow the bank to release its highest potential
- **We put our customers first. We look forward to provide our customer with best quality products**

ASAPP
AI FOR ENTERPRISE

Thank You

info@asapp.com

www.asapp.com