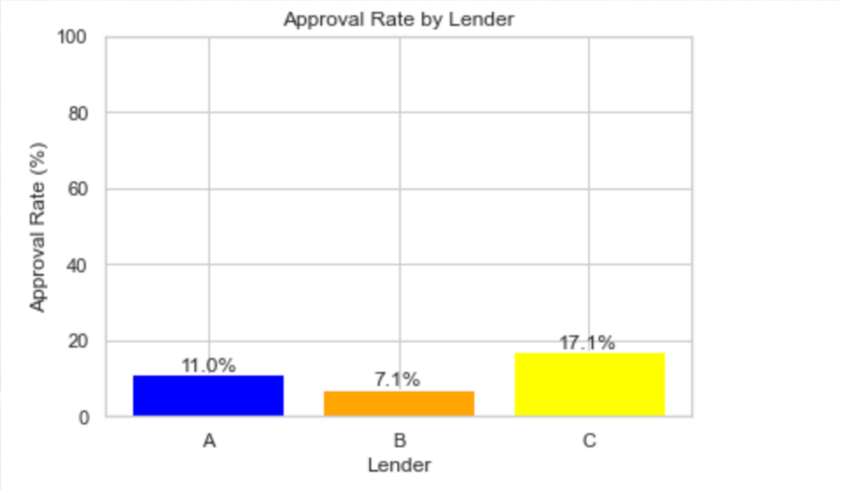
Bankrate (RV interview)

Letao (Julietta) Zhu

Agenda

- Executive Summary
- Data and Methods
- Findings and Impacts

Bankrate stands to add $2M to its commission revenue with a better matching model



Low approval rates among banks

$2.6M

commission revenue

$4.5M

Potential commission revenue*

* represents revenue growth potential of >.7x; key drivers include value recovered from miss-matching

Executive Summary - With Optimized Matching, Bankrate gains $2M

Background

Bankrate, a 3rd party agency helping consumers select their best loan option, want to improve commission revenue with a better and smarter matching system.

3 Questions

- What feature matter the most to determine loan approval?
- Any differences in who gets approved across lenders?
- Other ways to monetize?

3 Methods

- Random forest binary classifier
- feature importance to determine differences in three lenders
- Expected value from the product of approval probability and unit revenue

Recs

- Use exiting data/info about consumers to generate differential recommendations of lenders, leveraging a machine learning model for accurate assignment.
- Gather additional consumers information to continuously enhance the accuracy and effectiveness of the prediction model.

Agenda

Executive Summary

Data and Methods

Findings and Impacts

We leveraged leading data science methods for targeted analysis

Analysis Part	Variables, Packages, and Process
Binary Classification & Feature Importance Analysis	• RandomForestClassifier in the sklearn package in python • Construct 2 new variables: housing_ratio & loan_income_ratio • Build a general approval model and separate models for each lender using a random forest approach. • Analyze feature importance to identify significant predictors.
Differential Approval and Revenue Prediction	• Investigate top 4 features and examine differences in approval criteria among lenders. • Predict approval probabilities for each lender on a test set. • expected revenue per customer analysis
Optimizing Revenue through Lender Recommendation	• Recommend the lender with the highest expected revenue for each customer. • Compare the predicted revenue with current revenue

Our analysis can be further reinforced if key assumptions and limitations are overcome

Assumptions

- Loan approval failures of an otherwise creditworthy user occur solely due to incorrect lender recommendation
- There's only one uniform loan product for all lenders
- Fixed commission for all loans
- Model trustworthiness

Limitations

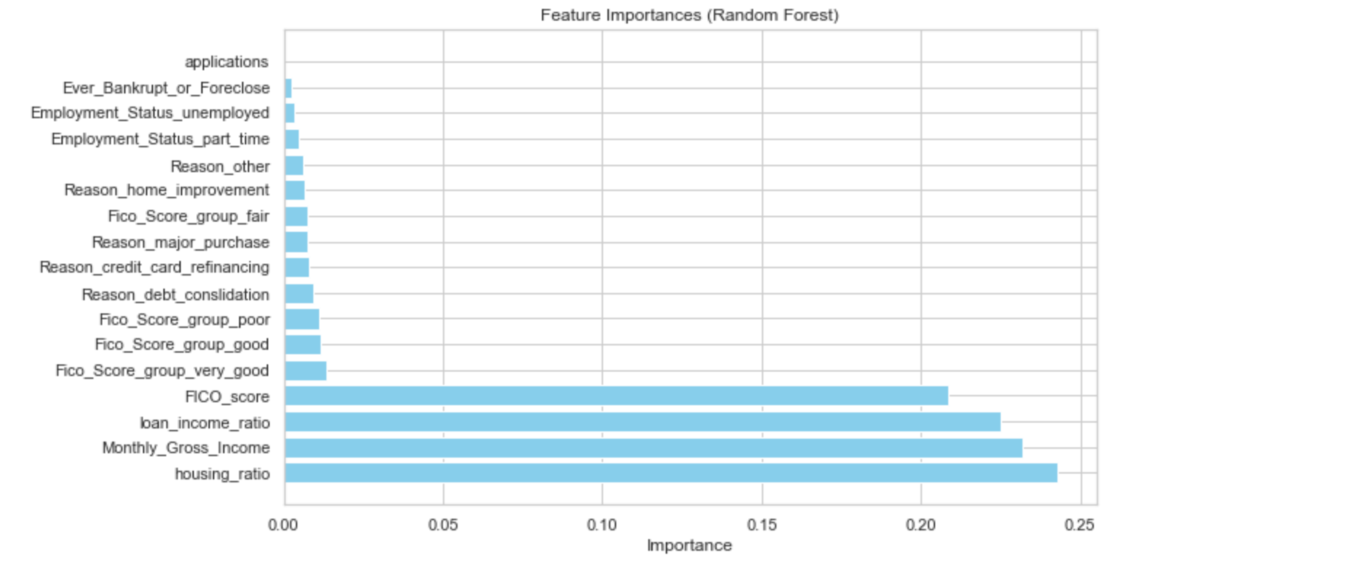
- Lack of information about loan products such as interest rate
- Lack of cost information (consumer acquisition expenses, operation costs, etc)
- Lack of information about other ways of monetization (e.g. from consumers instead of lenders)

Agenda

- Executive Summary
- Data and Methods
- Findings and Impacts

1

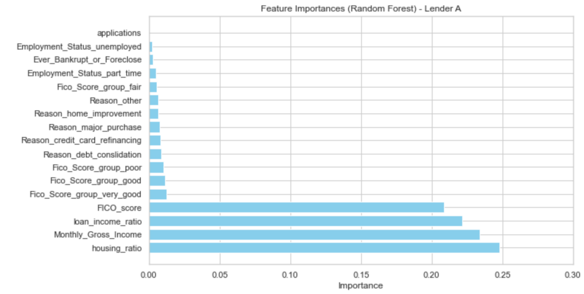
Analyze feature importances to identify significant predictors



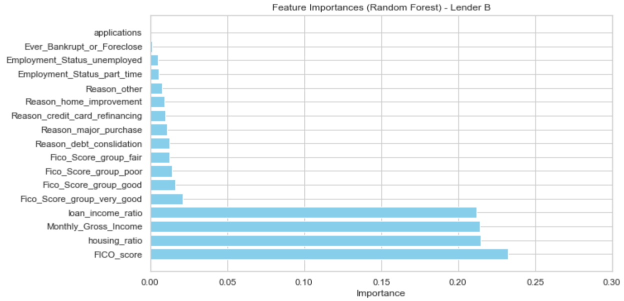
- The model have a **88%** accuracy score on test set.

- This graph shows us which features are the most important when determine approval
- Top 4 predictors: housing_ratio, Monthly_Gross_Income, FICO_score, loan_income_ratio

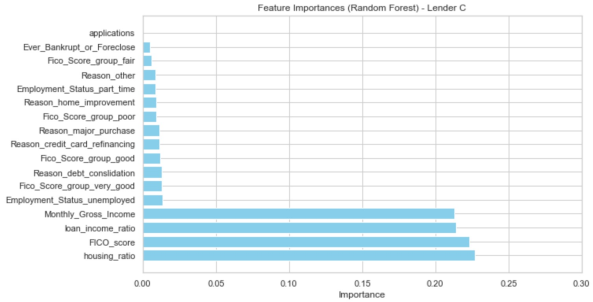
2 Investigate differences in what matters the most for each lender



Accuracy for Lender A: 0.9451



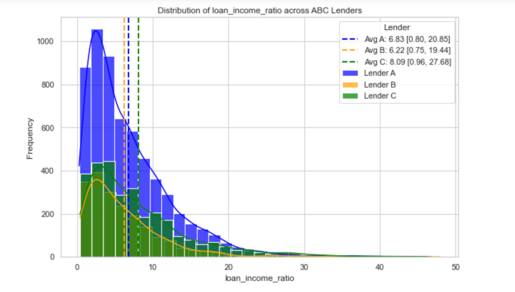
Accuracy for Lender B: 0.90945

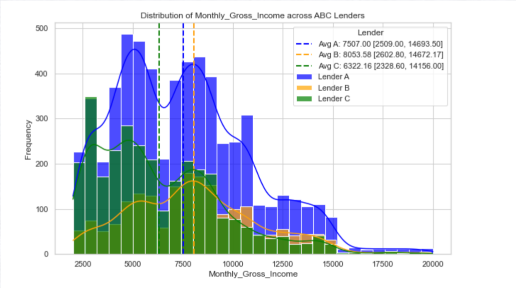


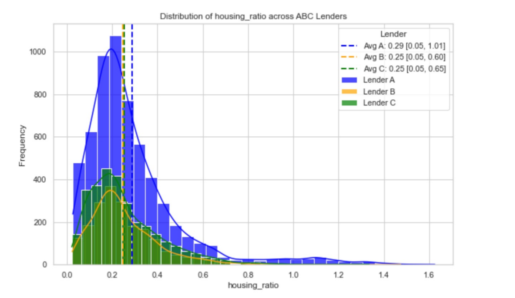
Accuracy for Lender C: 0.90255

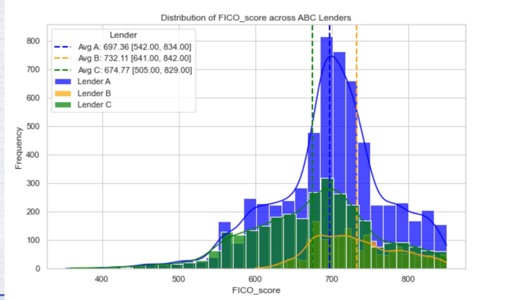
- The top 4 features are consistent across group
- **There are differences in the order of the variable importance. To see if the differences in order are significant, We can look at these 4 features separately (next page)**

2 Investigate top 4 features and examine differences in approval criteria among lenders.









- Lender B requires higher monthly income and FICO score; Lender C have relatively lower requirement to approve a loan
- In the next step, we can use machine learning predictive model to handle complex non-linear patterns that's not shown on these graphs

3 Recommend the lender with the highest expected revenue for each customer

	Expected_Revenue_Best	Recommended_Lender	Expected_Revenue_Recommend_A	Probability_of_approval_A	Expected_Revenue_Recommend_B
0	32.5	A	32.5	0.13	7.0
1	25.0	A	25.0	0.10	0.0
2	28.0	B	5.0	0.02	28.0
3	19.5	C	10.0	0.04	0.0
4	52.5	A	52.5	0.21	14.0

```
The percentage revenue growth will potentially be around 0.70
The predicted revenue with the new matching model is: 4,483,836
```

- Use separate models to predict a probability of approval for each lender for a held-out test set
- Calculate expected revenue and determine the lender with the highest expected revenue
- We will get a 70 percent revenue growth compared to the initial earning hence estimated $2M increase in revenue