Lending and Underwriting Practices

An Analysis of Lenders Underwriting Considerations Through the Context of Digital Humanities

DIGHUM 101 Individual Project

By: Jack Peterson

Table of Contents

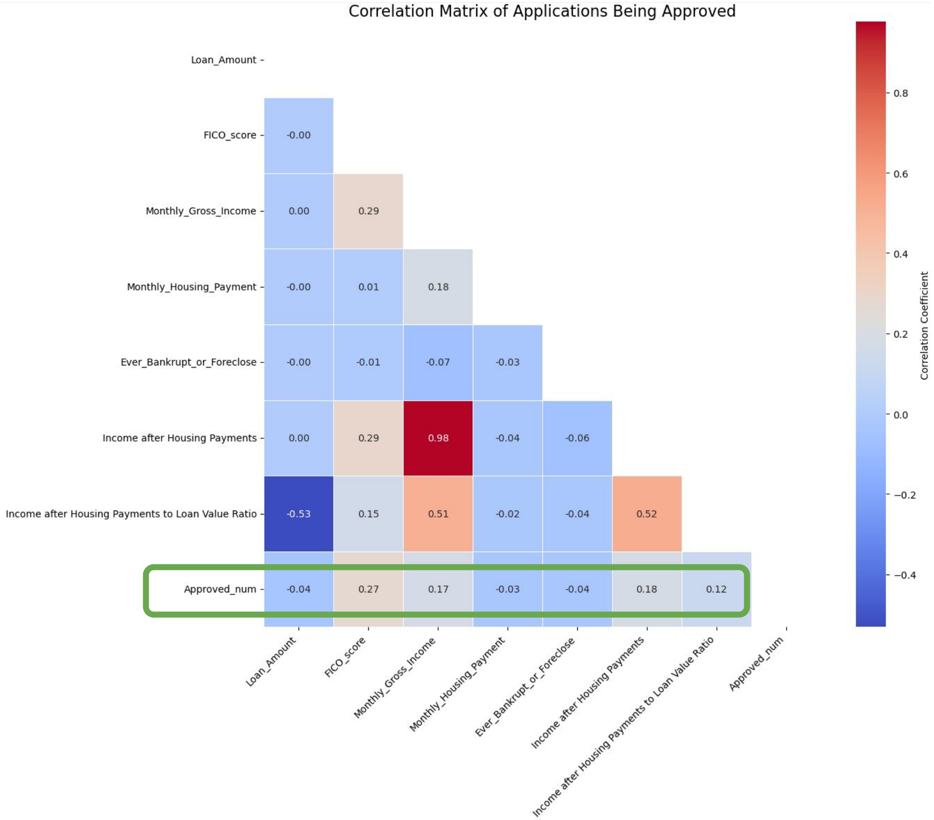
Overview

In this project I will be leveraging a dataset of loan applications from three major lenders (real names redacted) in order to best understand the differences in their underwriting considerations when it comes to job type and employment sectors in order to understand why applicants are being denied and then see how we can benefit applicants through different online tools in order to improve their approval rates as well as determine which lender has the highest approval rates in order consider which lender is the most socially equitable. Furthermore, I will consider if there are any potential government sponsored online tools or policies that would be appropriate to increase the working class's approval rates.

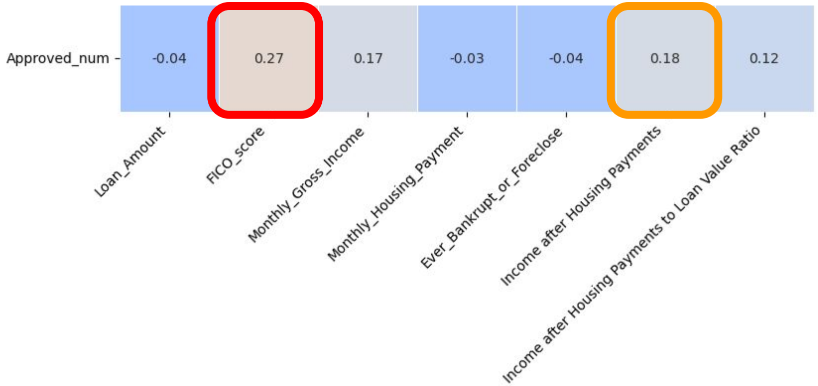
Leveraging Correlation Matrices and Regression Models

- I will be using correlation matrices to show the linear association between the various numerical factors and the likelihood of an application being approved.
- I will then be using a logistic regression model to further consider the non-linear relationship between the various factors and the likelihood of being approved. This is critical as many lenders tend to have particular baseline figures they look for in order to make approval decisions as the decision itself is binary.
- This dual approach allows us to thoroughly investigate both linear and non-linear patterns in the data, guiding more informed and effective lending decisions.
- Using a logistic regression model is fairly precise in predicting applications that are not going to be approved but have relatively low precision in predicting approved applications. Thus, our correlation matrices can assist us in predicting these approved applications.

Correlation Matrix for All Lenders



Correlation Coefficients Impacting Approval

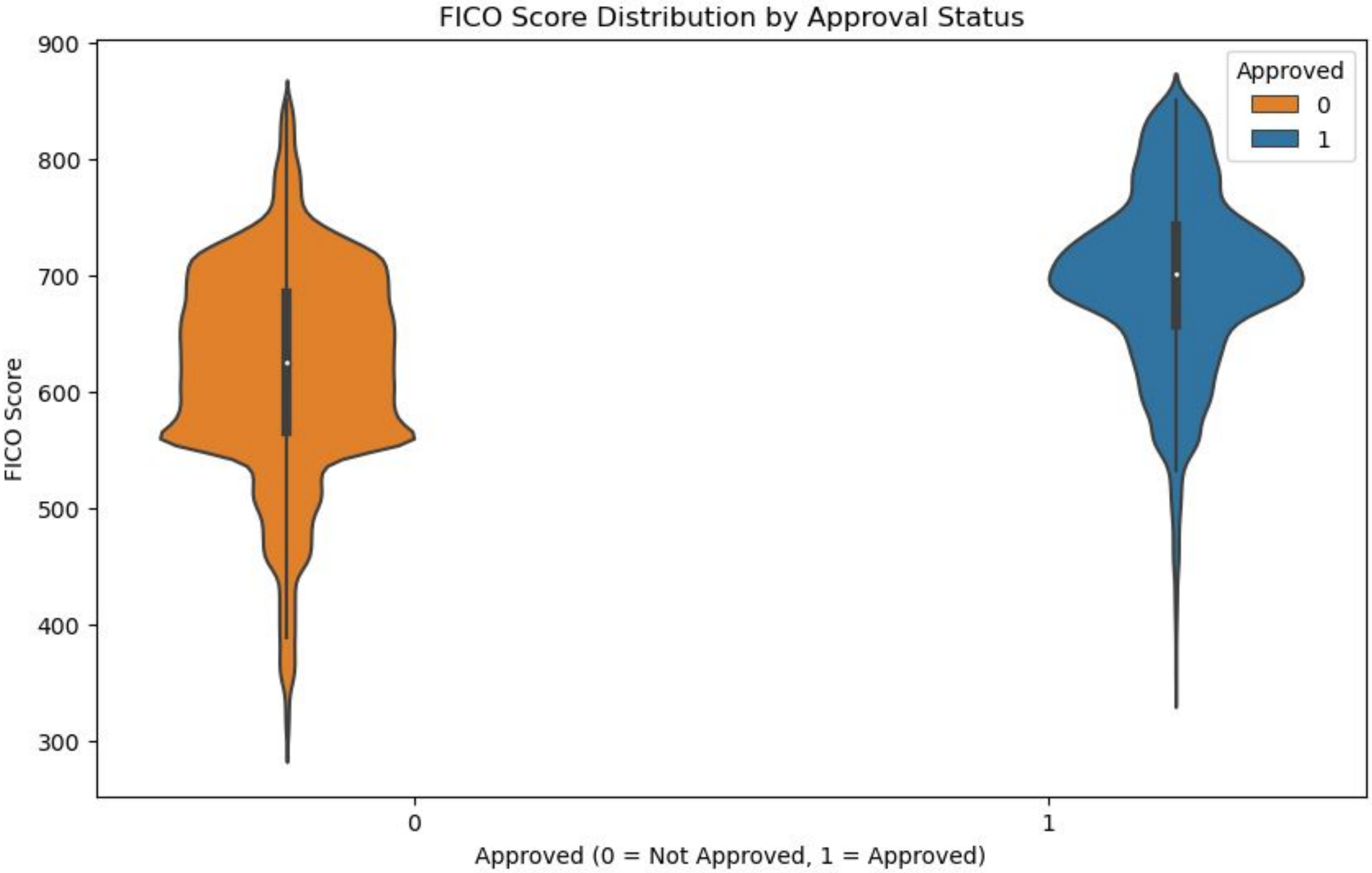


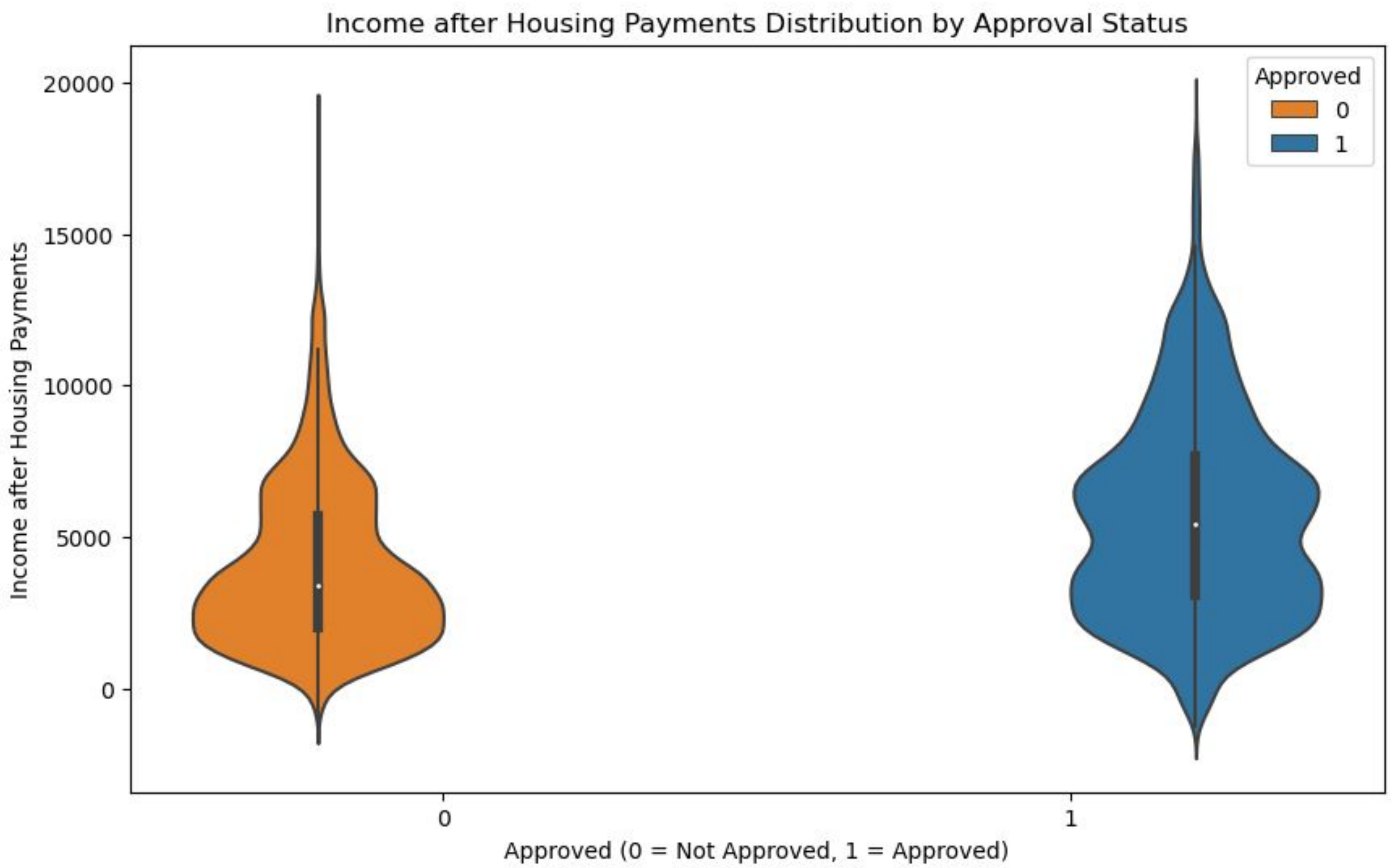
- FICO score has the largest correlation coefficient, followed by monthly income after housing payments.
- I was able to create a higher correlation coefficient than monthly gross income by considering said income after monthly housing payments are subtracted.

Logistic Regression Model Output for All Lenders

```
Classification Report:
              precision    recall  f1-score   support

           0       0.90      0.99      0.94     26750
           1       0.48      0.06      0.11      3250

    accuracy                           0.89     30000
   macro avg       0.69      0.53      0.53     30000
weighted avg       0.85      0.89      0.85     30000

Confusion Matrix:
[[26537   213]
 [ 3052   198]]
                                                  Coefficient
Loan_Amount                                         -0.000006
FICO_score                                           0.011115
Monthly_Gross_Income                                -0.000021
Monthly_Housing_Payment                             -0.000126
Ever_Bankrupt_or_Foreclose                          -0.989967
Income after Housing Payments                        0.000105
Income after Housing Payments to Loan Value Ratio   -0.084631
Debt to Income Ratio                                -0.255156
```

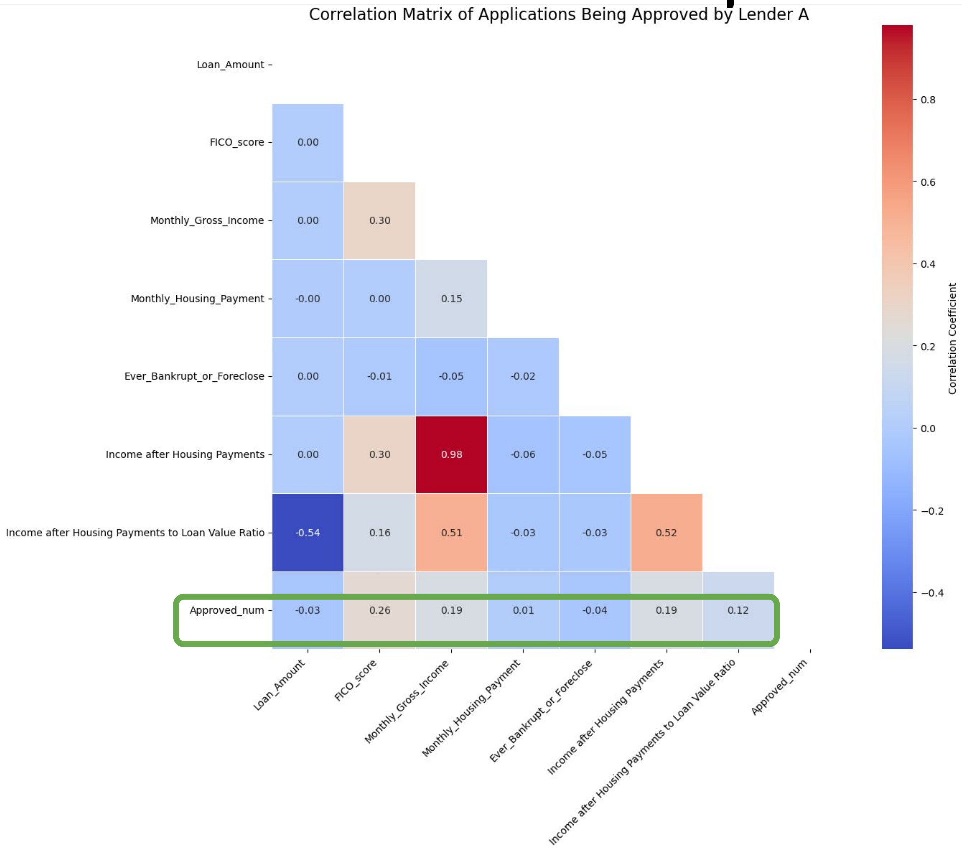
- Using a logistic regression model is 90% precise in predicting applications that are not approved but only 48% precise predicting approved applications.
- In this model our most impactful factor on an application appears to be whether or not they have been foreclosed on or gone bankrupt, with a coefficient close to -1.
- Following this, the most important factor is their Debt to Income Ratio, then their Net Income to Loan Ratio, and then their FICO score.


FICO Score Distribution by Approval Status
900
800
700
600
500
400
300
FICO Score
Approved
0
1
0
1
Approved (0 = Not Approved, 1 = Approved)

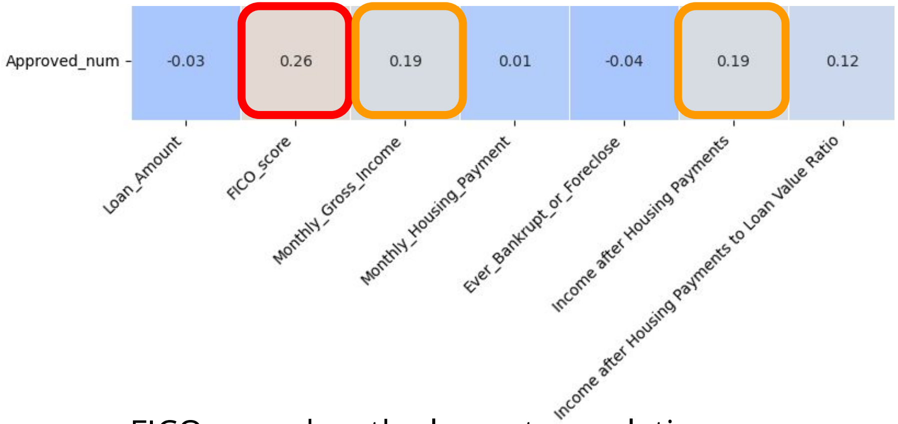


Income after Housing Payments Distribution by Approval Status
Income after Housing Payments
20000
15000
10000
5000
0
Approved
0
1
0
1
Approved (0 = Not Approved, 1 = Approved)

Correlation Matrix for Lender A



Correlation Coefficients Impacting Approval



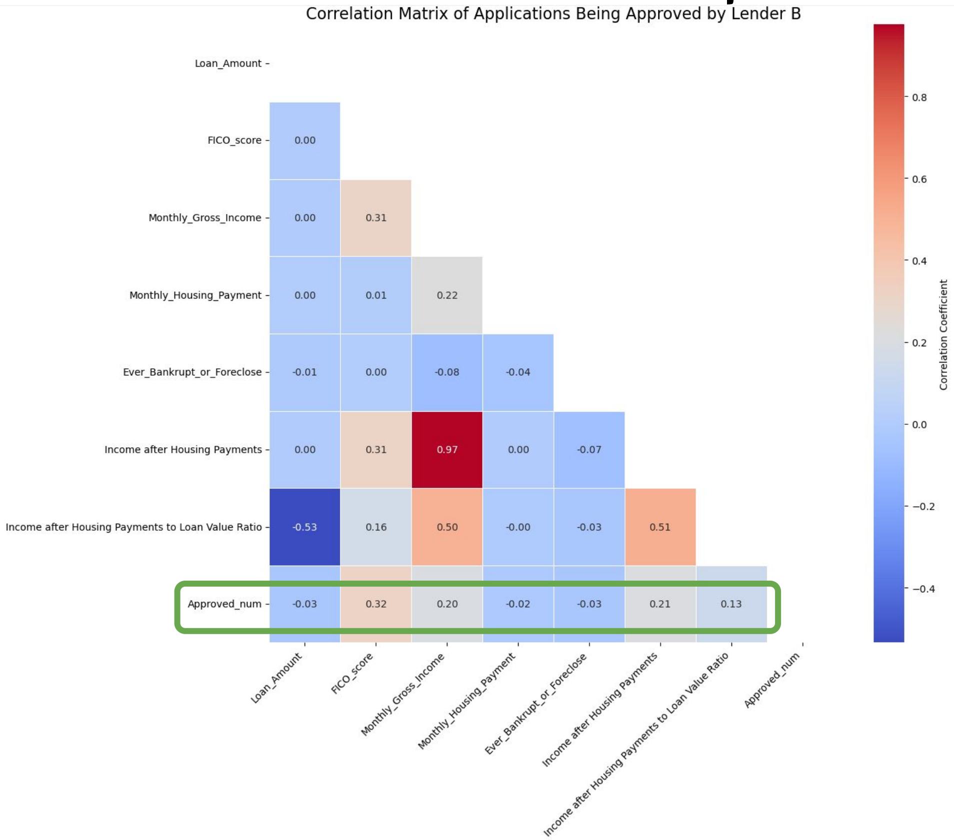
- FICO score has the largest correlation coefficient, while monthly gross income and monthly income after housing payments share the same correlation coefficient.
- This suggests that Lender A views the importance of gross income and income after housing payments equally. Furthermore, this lender values Income after Housing Payments to Loan Value ratio as the next most relevant coefficient.

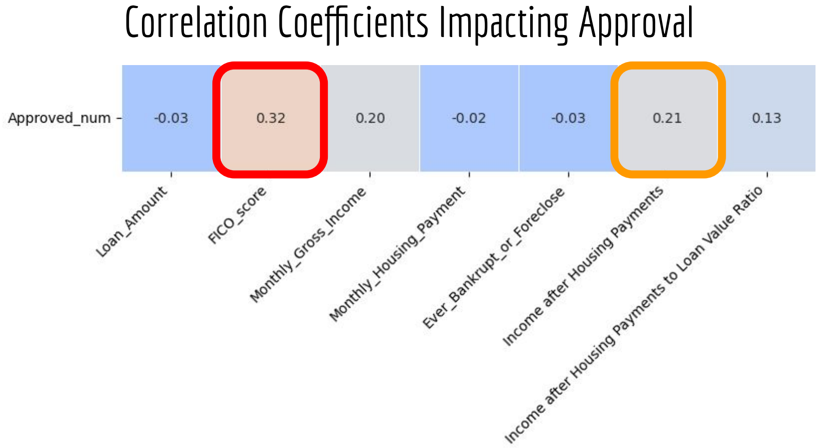
Logistic Regression Model Output for Lender A

```
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.99      0.94     14686
           1       0.46      0.06      0.10      1814

    accuracy                           0.89     16500
   macro avg       0.68      0.52      0.52     16500
weighted avg       0.85      0.89      0.85     16500

Confusion Matrix:
[[14567   119]
 [ 1712   102]]
                                                    Coefficient
Loan_Amount                                           -0.000004
FICO_score                                             0.010742
Monthly_Gross_Income                                   0.000018
Monthly_Housing_Payment                               -0.000105
Ever_Bankrupt_or_Foreclose                            -1.364505
Income after Housing Payments                          0.000123
Income after Housing Payments to Loan Value Ratio     -0.008081
Debt to Income Ratio                                   1.157463
```

- **Key Coefficients:**
  - **FICO Score:** Positive impact from a high score.
  - **Bankrupt/Foreclosure:** A negative coefficient larger than the overall suggests that Lender A weighs this heavier than other lenders on average.
  - **Debt to Income Ratio:** Unusually high positive coefficient, however this does not take existing instances of debt into account.

Correlation Matrix for Lender B





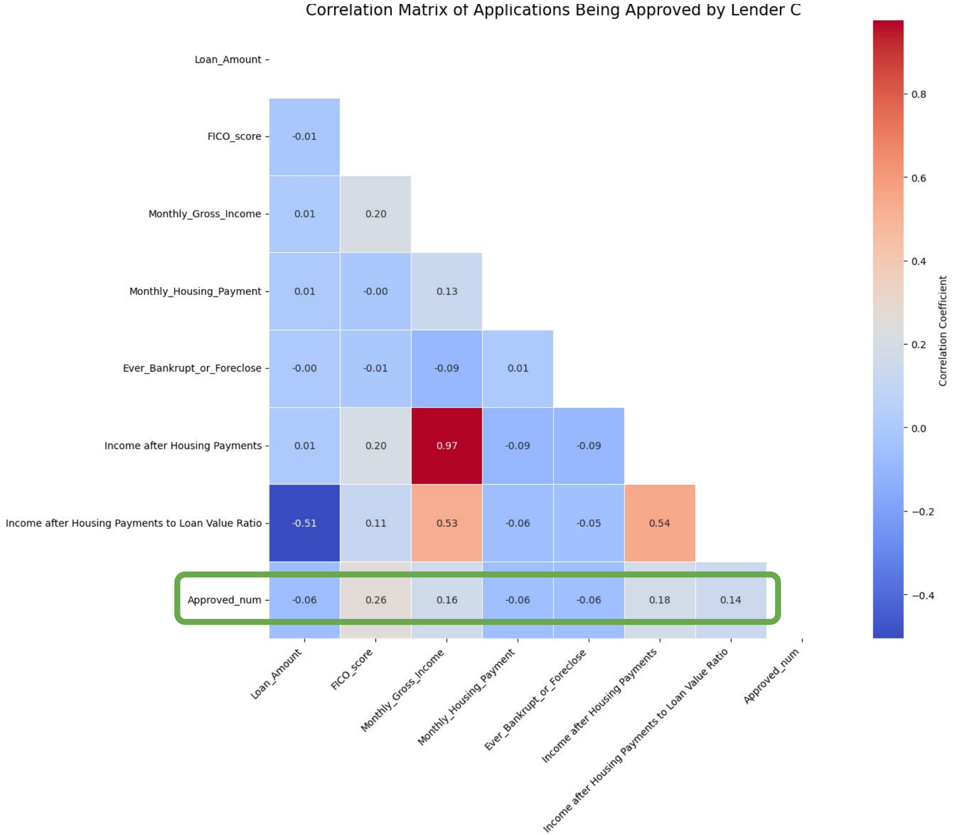
- Lender B values a high FICO score more than either of the other two lenders.
- Lender B also values FICO Score, Monthly Gross Income, and Monthly Income after Housing Payments more than the other two lenders, valuing the latter more so.

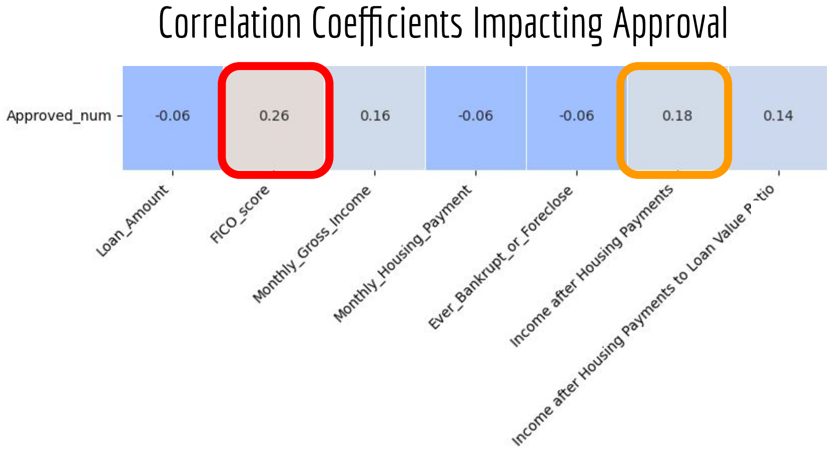
Logistic Regression Model Output for Lender B

```
Classification Report:
              precision    recall  f1-score   support

           0       0.94      0.99      0.96      7669
           1       0.50      0.15      0.23       581

    accuracy                           0.93      8250
   macro avg       0.72      0.57      0.60      8250
weighted avg       0.91      0.93      0.91      8250

Confusion Matrix:
[[7580   89]
 [ 492   89]]
                                                Coefficient
Loan_Amount                                       -0.000005
FICO_score                                         0.019184
Monthly_Gross_Income                               0.000067
Monthly_Housing_Payment                            0.000127
Ever_Bankrupt_or_Foreclose                        -1.889800
Income after Housing Payments                     -0.000061
Income after Housing Payments to Loan Value Ratio  0.079255
Debt to Income Ratio                              -2.451117
```

- **Key Coefficients:**
  - **FICO Score:** Lender B has the highest coefficient for FICO Score compared to the other lenders.
  - **Bankrupt/Foreclosure:** A negative coefficient larger than all others suggests this could be close to a dealbreaker for Lender B.
  - **Debt to Income Ratio:** Strong negative coefficient for Debt to Income Ratio, suggesting that Lender B strongly prefers customers to have disposable income more than anything else.

Correlation Matrix for Lender C





- While Lender C values FICO score the highest, the following three largest coefficients are weighted more equally as each other by Lender C than by the other lenders.

Logistic Regression Model Output for Lender C

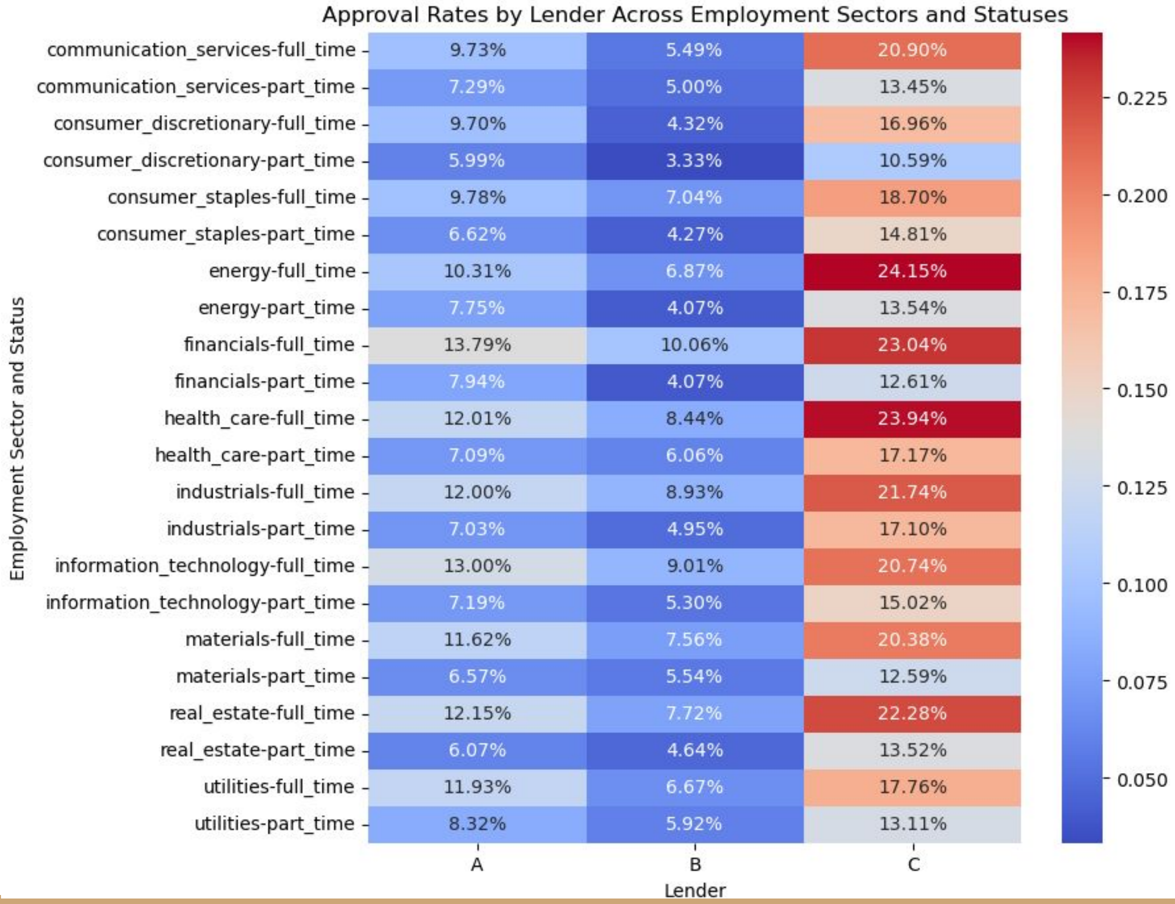
```
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.99      0.91      4362
           1       0.59      0.07      0.12       888

    accuracy                           0.83      5250
   macro avg       0.71      0.53      0.52      5250
weighted avg       0.80      0.83      0.78      5250

Confusion Matrix:
[[4319   43]
 [ 827   61]]
                                              Coefficient
Loan_Amount                                     -0.000006
FICO_score                                       0.008822
Monthly_Gross_Income                             0.000075
Monthly_Housing_Payment                          0.000174
Ever_Bankrupt_or_Foreclose                      -0.887685
Income after Housing Payments                   -0.000099
Income after Housing Payments to Loan Value Ratio  -0.012827
Debt to Income Ratio                            -2.974494
```

- **Key Coefficients:**
  - **FICO Score:** Lender C has a slightly lower coefficient for FICO Score compared to other lenders.
  - **Bankrupt/Foreclosure:** Suggests this is not preferred but more bearable than it is viewed by others.
  - **Debt to Income Ratio:** Negative coefficient for Debt to Income Ratio, suggesting that Lender C prefers customers to have disposable income more so than the other lenders.

Lender Approval Rates

Overall	10.98%
Lender A	10.97%
Lender B	7.13%
Lender C	17.06%

- By segmenting customers into these employment categories we can increase overall approval rates.
- Lender C has substantially higher approval rates across the board whereas Lender B has the lowest.



Identifying Key Metrics for Approval

- **Key Metrics:** The top metrics that were the most relevant in predicting loan approval were FICO score, bankruptcy or foreclosure history, monthly gross income, and net income after monthly housing payments.

- **Identifying Key Metrics:** Due to the linearity of FICO scores and income levels, these metrics are best expressed through the correlation matrices whereas the relevance of bankruptcy or foreclosure history can be clearly seen through the logistic regression models.

Insights and Recommendations

- **Highest Approval Rates:** Lender C appears to have substantially higher approval rates across the board, in some sectors approving a higher percentage of applicants than Lender A and B combined.

- **Social Equity and Financial Inclusion:** These higher approval rates would signify that Lender C has substantially more equitable underwriting practices than with Lender A or B. With these being 3 of the country's top lenders, Lender C likely has a key role to play in the future of financial inclusion for those who are currently underserved by the banking industry.

- **Increasing Individual Approval Rates:** the government should consider supporting the development of online tools for individuals to both improve their current metrics that lenders use, as well as direct applicants to the lender that is most likely to approve them. While this currently seems to be Lender C across the board, this could change in the future.