Loan Default Predictor using Machine Learning

Francis Muthoni

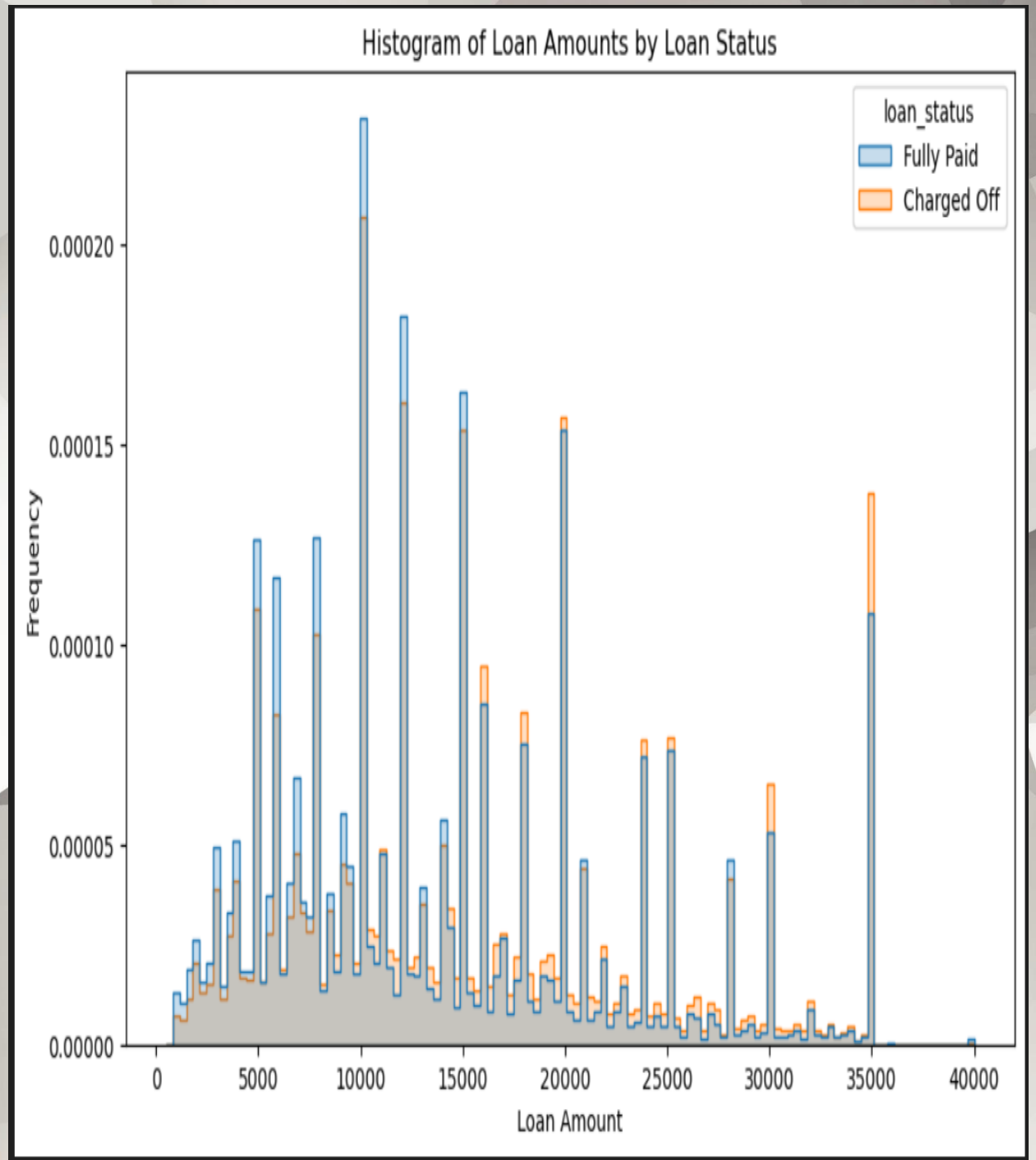
Problem Statement

LendingClub, a leading peer-to-peer lending platform, faces the challenge of minimizing the risk associated with lending money to borrowers. The company needs to assess the creditworthiness of potential borrowers to reduce the likelihood of loan defaults. The goal is to build a predictive model that helps identify high-risk borrowers, thereby improving the quality of the loan portfolio and enhancing overall profitability

Business Undertanding

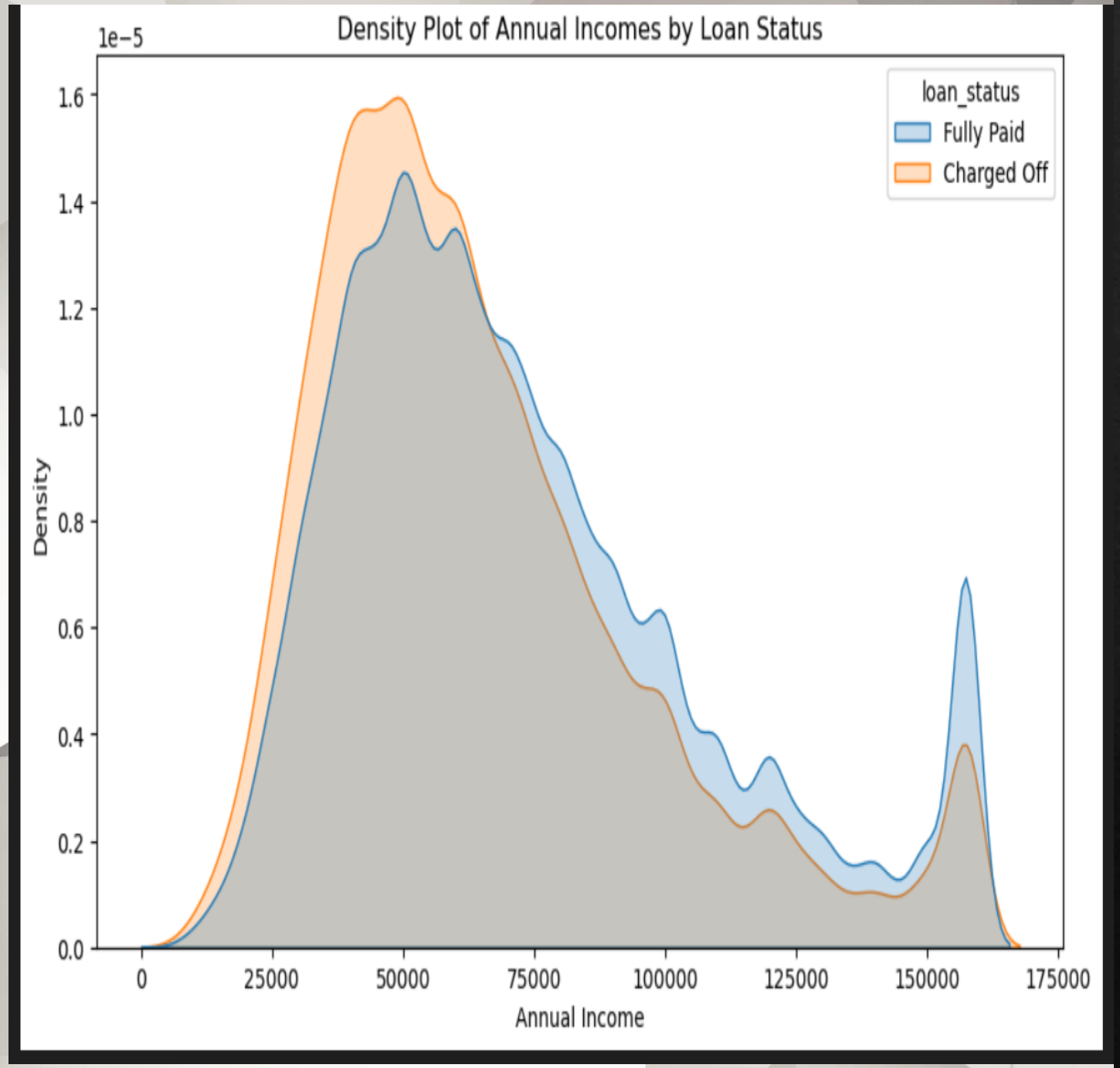
Company Overview: LendingClub specializes in offering loans to urban customers
Loan Approval Decision: When receiving a loan application, LendingClub must decide whether to approve it based on the applicant's profile.

- **Risk Types**:
- **Loss of Business**: Not approving a loan for an applicant who is likely to repay results in a missed opportunity and loss of business.
- **Financial Loss**: Approving a loan for an applicant who is likely to default could lead to financial losses.
- **Data Utilization**: The available data contains historical information about loan applicants and their default status.
- **Objective**: Identify patterns indicating the likelihood of an applicant defaulting, to inform decisions such as denying the loan, adjusting loan amounts, or setting higher interest rates for riskier applicants.



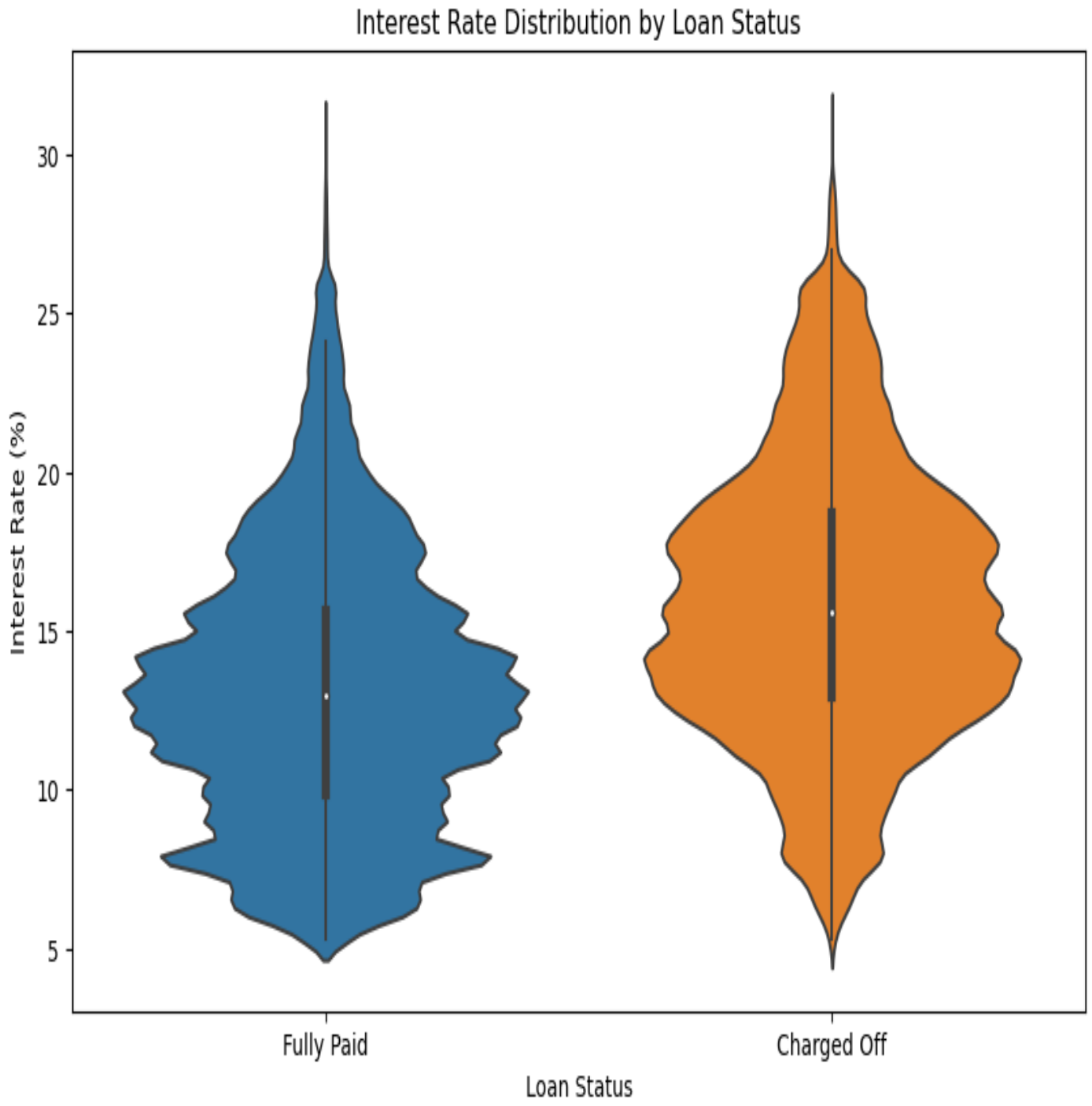
ANALYSING RELATIONSHIP BETWEEN LOAN AMOUNTS AND LOAN STATUS

- from the histogram it is clearly evident that as the loan amount increase the chances of default also increase since below 15000 most loans are fully paid but above the more and more loans are being charged off



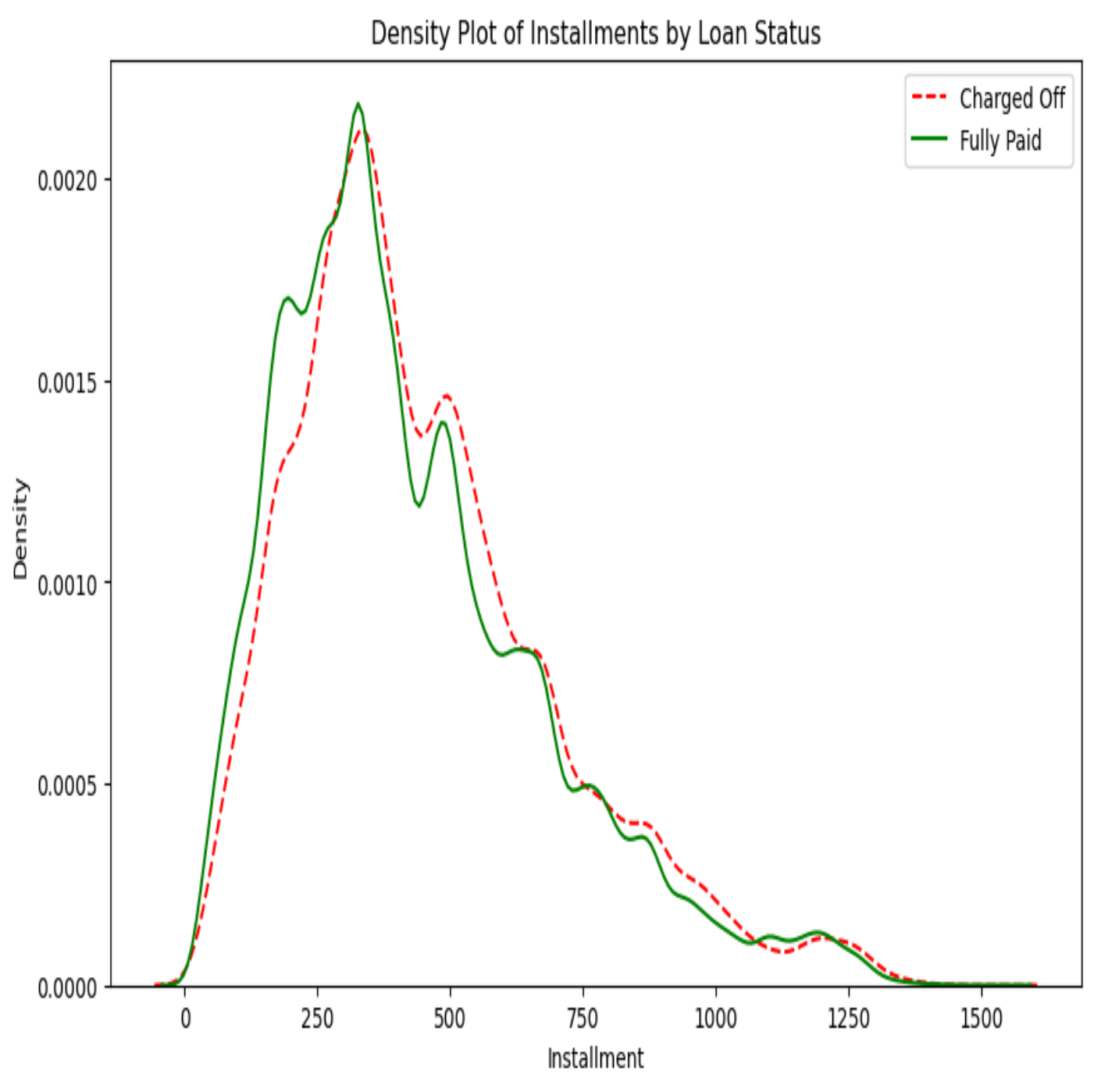
ANALYSING RELATIONSHIP BETWEEN ANNUAL INCOME AND LOAN STATUS

- it is clearly evident people with a lesser annual income tend to not repay their loans compared to the shift on the right side where annual income of above 75000 tend to repay the loans better



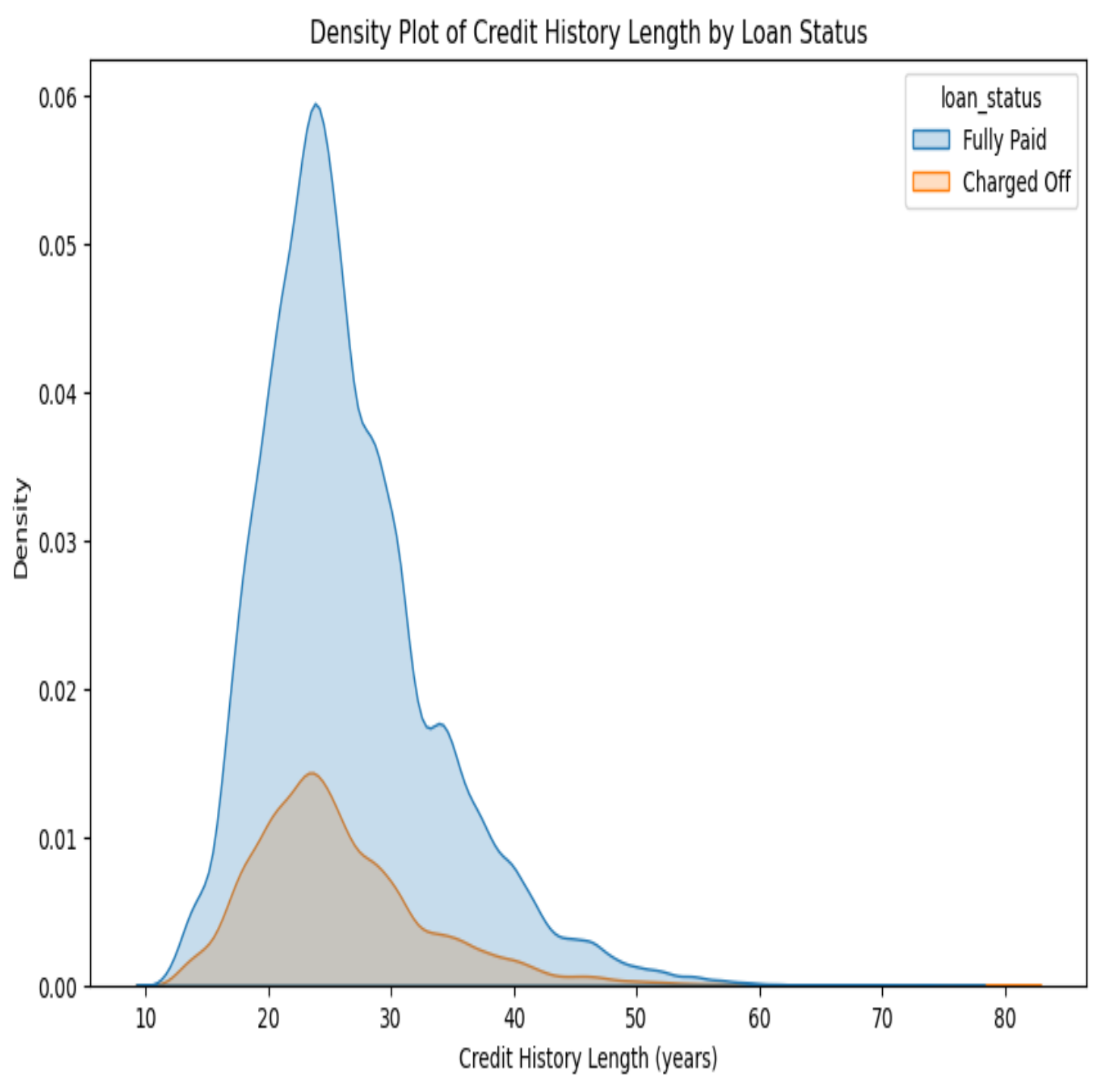
ANALYSING RELATIONSHIP BETWEEN INTEREST RATES AND LOAN STATUS

- It is clearly evident loans with a lesser interest rate are plone to be paid and despite there being almost equal rates of default in the 10-15 rates the default is still more consistent moreover default increase above the 20% rate.



ANALYSING RELATIONSHIP BETWEEN INSTALMENT AND LOAN STATUS

- It is clearly evident the lesser the instalment the better repayment there is and as the installment increase the chances of defaukt that is associated with the increase also is directly related



VISUALING THE RELATIONSHIP BETWEEN CREDIT HISTORY AND LOAN STATUS

- from the density plot above shows credit history has no major influence on the outcome of the loan status

Recommended Model

Balanced Random Forest Classifier Modelling

Outputs explanation

1. Recall for 'Charged Off': 64%

Effectively identifies 64% of defaulting loans.

2. Precision for 'Charged Off': 27%

Only 27% of predicted defaults are actual defaults.

3. F1-Score for 'Charged Off': 38%

Moderate score reflects trade-off between recall and precision.

4. Recall for 'Fully Paid': 58%

Fails to identify 42% of fully paid loans correctly.

5. Precision for 'Fully Paid': 87%

Correctly predicts 87% of fully paid loans.

6. F1-Score for 'Fully Paid': 70%

Higher than 'Charged Off', but impacted by lower recall.

7. Overall Accuracy: 59%

Lower overall accuracy due to focus on detecting more defaults.

Business Conclusions

1. ***Effective Risk Detection:***

- The model effectively identifies potential loan defaults with a high recall rate for "Charged Off" loans. This capability is crucial for minimizing financial risks and managing credit effectively.

2. ***Model Consistency and Reliability:***

- Cross-validation shows that the model is consistent across different data segments with an average accuracy of about 76%. This consistency underscores the model’s reliability, although its accuracy in predicting actual defaults is lower (59%)

3. ***Operational Efficiency Concerns:***

- The model's low precision for predicting "Charged Off" loans leads to operational challenges. Incorrect classifications consume resources and may negatively affect relationships with borrowers who are wrongly targeted as high risk.

4. ***Impact on Customer Relationships:***

- Misclassifications, particularly of "Fully Paid" loans as likely to default, risk harming customer trust and satisfaction. This could deter borrowers from seeking additional services due to perceived unfair treatment.

5. ***Need for Model Optimization:***

- Given the trade-offs between identifying defaults and maintaining customer relationships, further model refinement is necessary to balance accuracy and customer impact effectively.
- These points succinctly capture the model's strengths in risk management and its drawbacks in terms of operational efficiency and customer relations, while noting the consistent performance demonstrated through cross-validation.

Business Recommendations:

1. *Update Credit Policies:*

- Review and adjust credit scoring and approval processes to better balance risk management with efficient operations and positive borrower experiences.

2. *Boost Data Analytics:*

- Enhance data analytics capabilities by incorporating broader data sets, such as economic trends and personal spending behaviors, to improve prediction accuracy.

3. *Utilize Advanced Modeling Techniques:*

- Use a mix of different modeling techniques, like ensemble methods, to increase the accuracy and fairness of predictions, especially in imbalanced datasets.

4. *Continuously Update Models:*

- Regularly updating the predictive models based on new data and changing market conditions to keep the predictions relevant and accurate.

5. *Educate Stakeholders:*

- Train stakeholders, particularly those interacting with customers, on understanding and using model predictions effectively to ensure informed decision-making.

6. *Perform Cost-Benefit Analysis:*

- Analyze the costs associated with incorrect predictions and adjust strategies to mitigate the most financially impactful errors.

7. *Ensure Compliance and Fairness:*

- Make sure all model applications comply with legal standards and ethical practices, particularly concerning fairness across different groups of borrowers.

- These recommendations aim to streamline the application of predictive models in your business, ensuring they not only improve financial outcomes but also support sustainable and ethical

• Strategic Conclusion

Risk vs. Reward:

1. The decision to accept this trade-off depends on the specific costs associated with false positives (wrongly predicting a loan will default) versus false negatives (failing to predict a default). If the cost of missing a "Charged Off" loan (false negative) is higher than the cost of incorrectly predicting a loan will default (false positive), this trade-off might be acceptable.

2. In lending, missing a default could mean a significant financial loss, whereas incorrectly predicting a default might lead to lost opportunities but potentially less direct financial harm.

Area of improvement

- Balancing these factors is key to optimizing the model's performance according to the specific business context and the relative costs of different types of errors.