

Lending Club Loan Analysis – Case Study

Exploring Patterns in Loan Defaults and Risk Factors

Prepared by: Naseem I Kesingwala & Omkar Vamsi Krishna Pechetti

Date: November 28th, 2024

Institution: MS - AI & ML Program, IIIT-B and UpGrad

Problem Statement and Objectives

In this case study, we aim to apply Exploratory Data Analysis (EDA) techniques to understand the risk factors associated with loan defaults in the consumer finance industry. Working with historical loan data, the goal is to identify key attributes that influence the likelihood of a borrower defaulting on a loan. This analysis is crucial for minimizing financial losses, as approving loans for risky applicants leads to credit losses, while rejecting loans can result in lost business. By analyzing variables such as income, loan amount, and repayment history, we can uncover patterns that help predict defaults. These insights can guide decisions such as denying loans, adjusting loan amounts, or setting higher interest rates for high-risk borrowers. The ultimate objective is to support the company's risk management and enhance the overall loan approval process.

Objectives

➤ **Analyze Customer and Loan Attributes to Detect Default Trends**

Investigate how various customer characteristics (such as income, credit score, and financial behavior) and loan-specific features (like loan amount, interest rates, and loan purpose) are linked to loan defaults. The objective is to uncover patterns and correlations that reveal which factors most significantly impact default rates, allowing Lending Club to better understand risk factors and improve credit assessments.

➤ **Provide Actionable Insights to Minimize Credit Losses**

Based on the insights derived from the data, provide practical, data-backed recommendations to minimize credit losses. These insights aim to refine Lending Club's lending strategies, optimize credit scoring models, and help identify high-risk borrowers early in the loan approval process, preventing potential defaults and improving overall loan quality.

➤ **Explore Key Variables Driving Loan Defaults Using Exploratory Data Analysis (EDA)**

Use exploratory data analysis (EDA) techniques to identify the key variables contributing to loan defaults. By visualizing and testing various hypotheses within the dataset, the analysis will uncover the primary risk factors, giving Lending Club a deeper understanding of the critical variables that drive loan defaults and enhance its decision-making processes.

Data Cleaning Process

To ensure the dataset was clean, reliable, and ready for in-depth analysis, several crucial data cleaning procedures were undertaken.

- **Fixing Missing Values:** The first priority was to address missing data. Missing values were carefully handled to ensure no bias would affect the analysis. This step was critical in ensuring the dataset's completeness and accuracy for further exploration.
- **Removing Columns with Excessive Missing Data:** Columns containing more than 90% missing values were removed from the dataset. This helped to eliminate irrelevant data and ensure that only the most valuable features were retained, thus maintaining the integrity of the analysis.
- **Dropping Irrelevant Columns:** Additionally, columns that were deemed unnecessary or irrelevant to the objectives of the case study were removed. This allowed for a more focused dataset, streamlining the data and ensuring the analysis was centered on the key attributes most likely to impact loan defaults.
- **Validating Data Consistency:** To ensure data integrity, all columns were validated for correct data types and consistency. This step identified and corrected any inconsistencies or errors in the dataset, ensuring the data was structured accurately for further analysis.

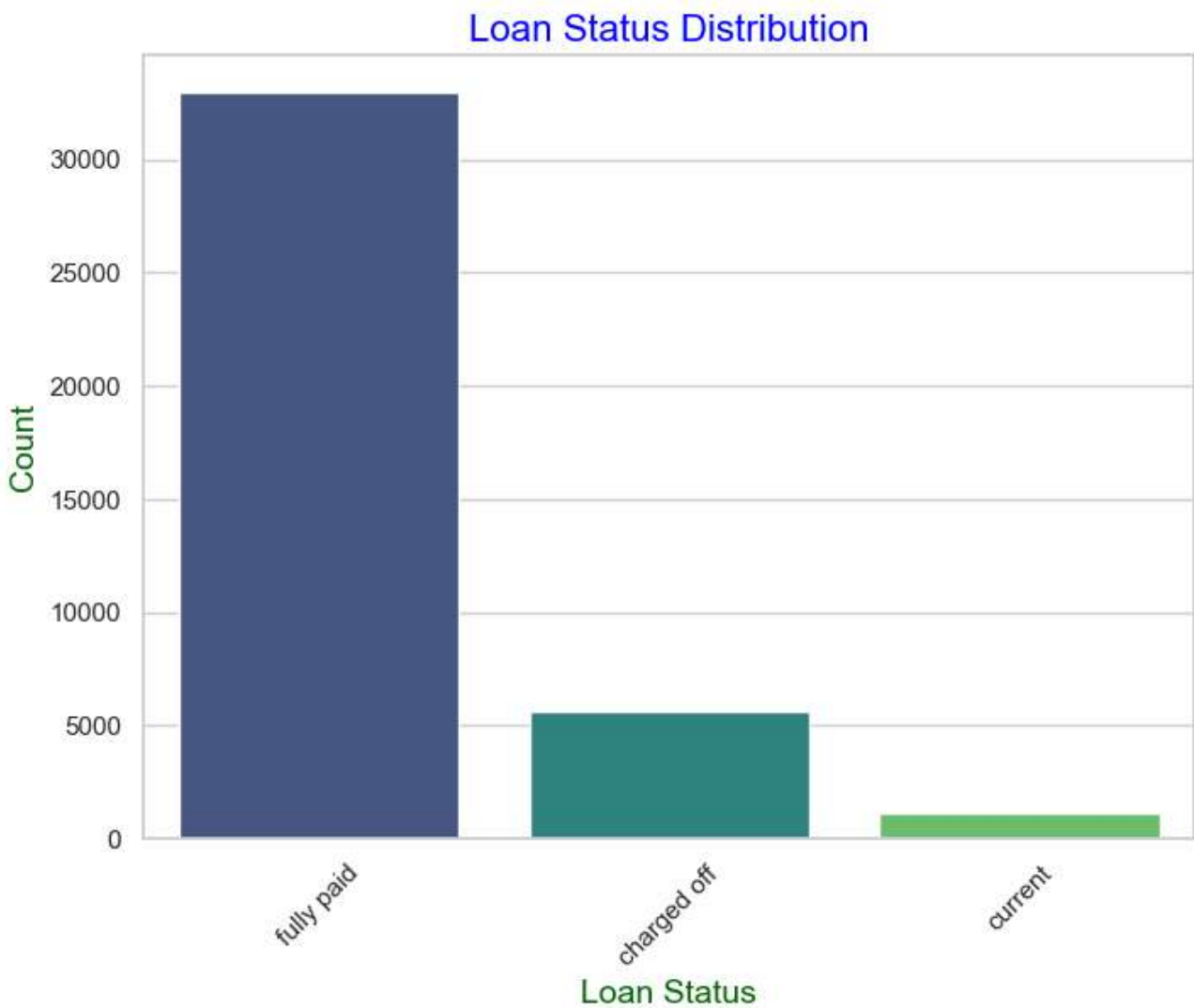
These data cleaning procedures helped prepare a clean, consistent dataset, primed for detailed exploratory analysis and enabling more accurate insights into the factors driving loan defaults.

Exploratory Data Analysis (EDA)

1. Overview on Loan Status Distribution

- **Overview on Loan Status Distribution**

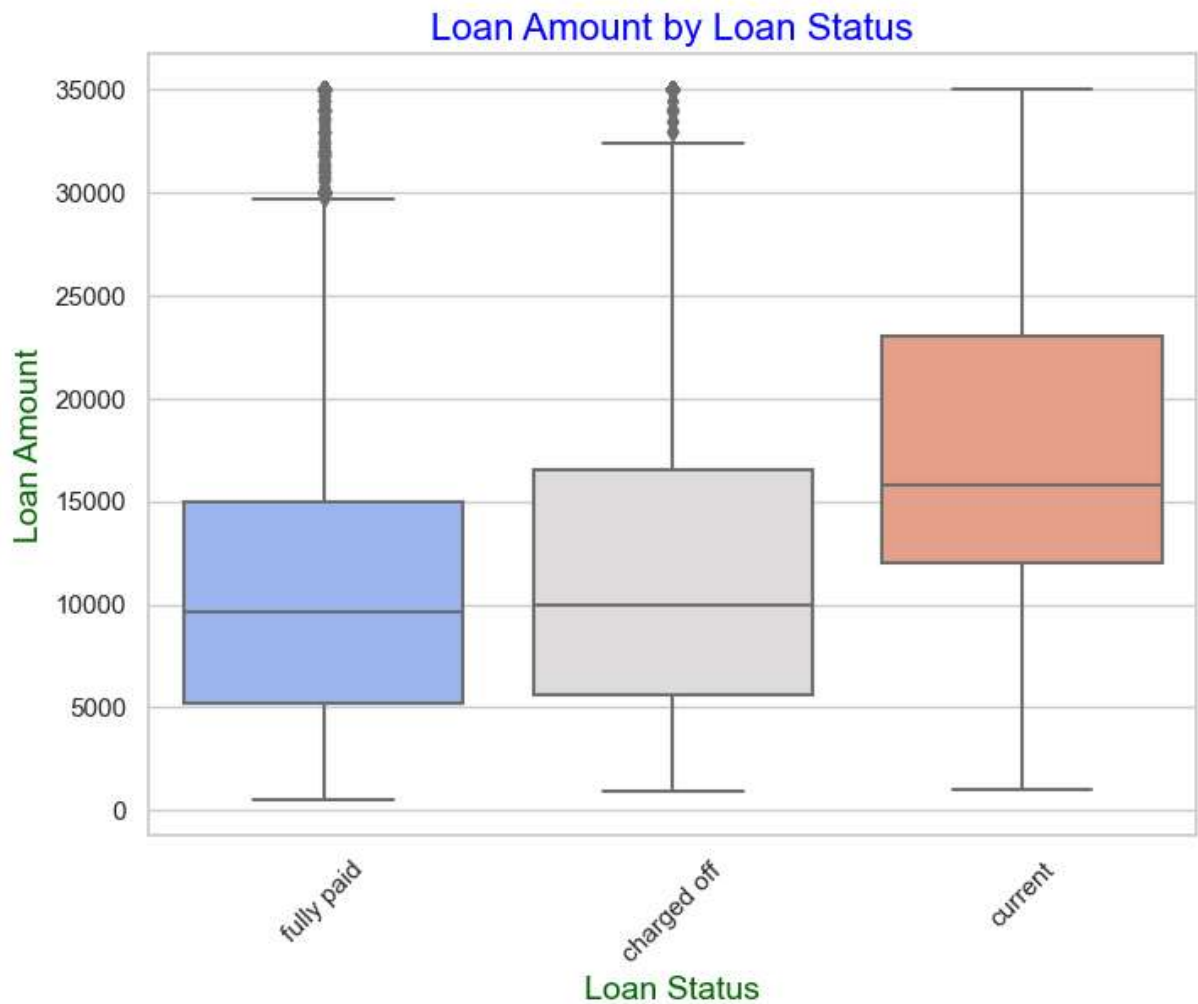
The distribution of loan statuses provides a clear understanding of the proportion of loans that are fully paid, charged off, or currently being repaid. This helps in assessing the overall risk associated with the loan portfolio and identifying trends in loan repayment behavior. By analyzing these distributions, Lending Club can better tailor its risk management strategies.



2. Analysis on Historic Loans Data

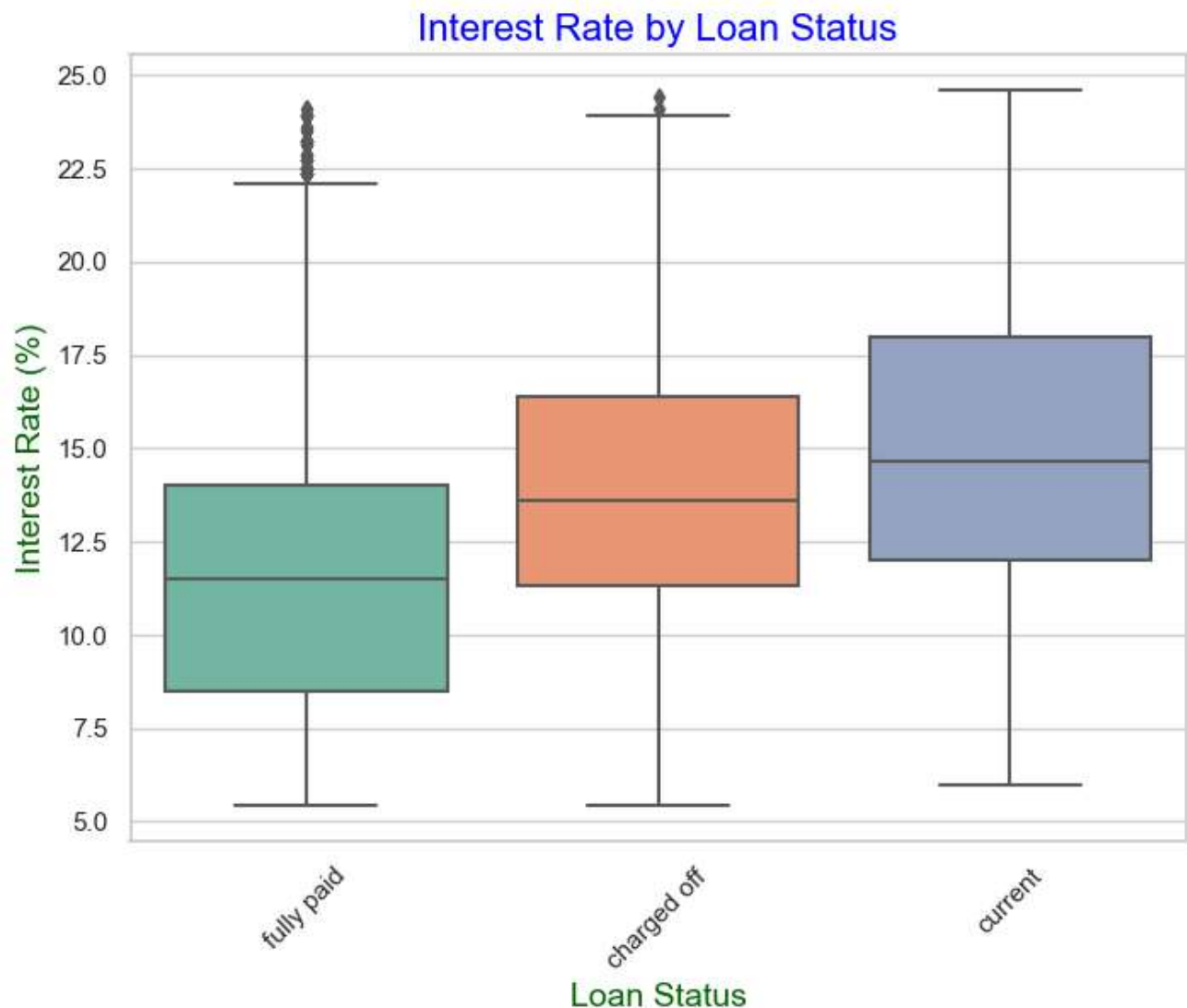
- **Loan Amount vs. Loan Status**

The relationship between loan amount and loan status is analyzed to determine whether larger loan amounts correlate with an increased likelihood of defaults. This analysis suggests that higher loan amounts may be linked to a higher risk of non-repayment. As a result, this finding emphasizes the importance of implementing stricter evaluations or setting more stringent limits on large loans. By doing so, Lending Club can effectively mitigate the financial risks associated with larger loans and better manage their overall credit exposure.



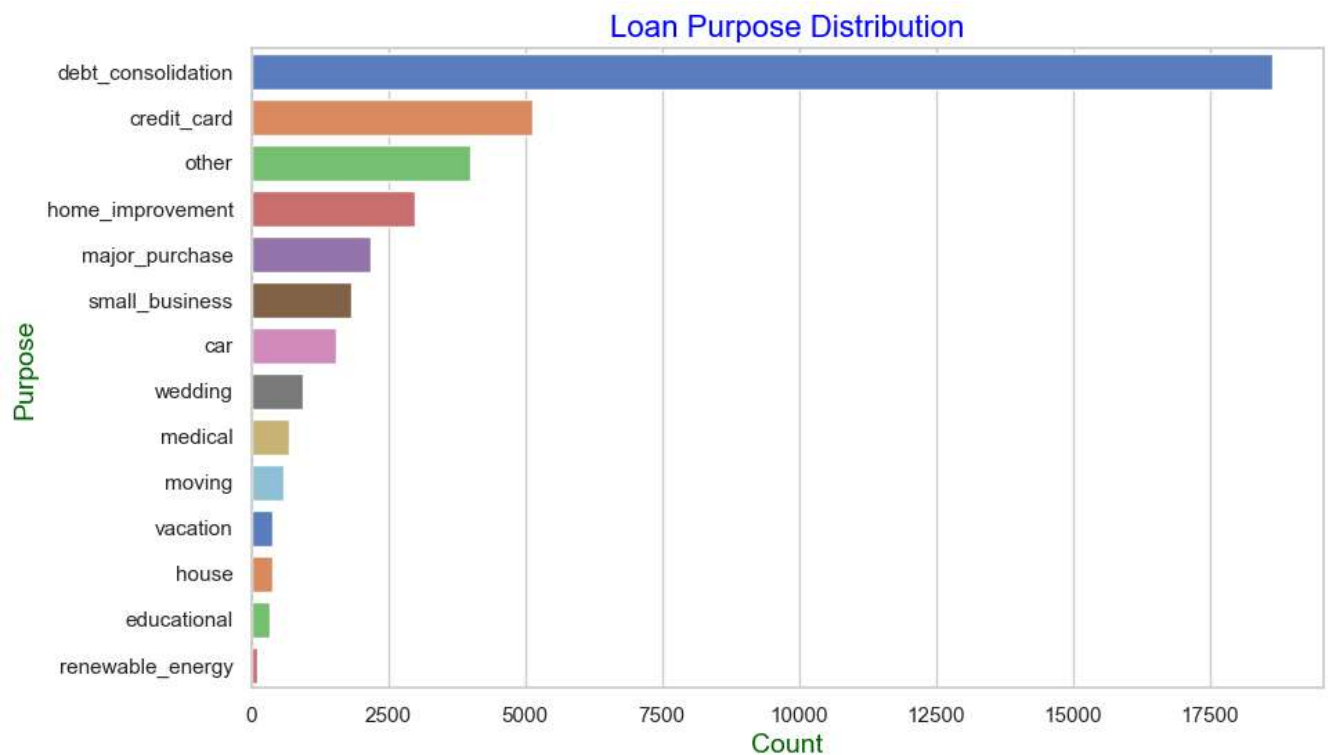
- **Interest Rates by Loan Status**

In this analysis, interest rates are compared with loan status to uncover potential patterns that show how loans with higher interest rates might be more prone to defaults. If higher interest rates indeed correlate with an increase in defaults, it could signal that borrowers with greater financial risk are more likely to opt for loans with higher rates. This finding provides crucial insights for Lending Club to reassess and refine its pricing strategy, ensuring that the risk profiles of borrowers are considered more effectively when setting loan interest rates.



- **Loan Purpose Distribution**

This section investigates the various reasons for which loans are granted, such as debt consolidation, home improvement, or other personal needs, and examines how different loan purposes may relate to default rates. By understanding which loan purposes are more likely to lead to defaults, Lending Club can make informed decisions about adjusting their loan eligibility criteria. For instance, stricter checks could be implemented for loans related to high-risk purposes like debt consolidation, helping to reduce overall default rates and maintain the financial stability of the lending portfolio.



Key Observations and Business Takeaways

- **Impact of Loan Amount on Defaults**

Larger loans tend to carry a higher risk of default, suggesting that Lending Club may need to implement more stringent approval processes for higher-value loans. Introducing more detailed evaluations or capping larger loans could be a way to reduce exposure to defaults and better manage financial risks associated with big-ticket loans.

- **Interest Rates and Their Impact on Defaults**

Loans with higher interest rates are correlated with higher default rates. This indicates that these loans are often taken on by higher-risk borrowers. Lending Club may need to reconsider its interest rate policies or adjust its risk assessment models to ensure that higher interest rates are not leading to unsustainable lending practices. A more granular risk analysis could be implemented to account for the borrower's repayment capability in relation to the loan's interest rate.

- **Loan Purpose and Default Rates**

A significant proportion of defaults come from loans categorized under “debt consolidation.” This suggests that Lending Club may want to re-evaluate the loan approval process for this category. Strengthening eligibility requirements for borrowers seeking debt consolidation loans or diversifying the loan portfolio could be effective in reducing defaults associated with this loan type.

- **Income Levels and Default Patterns**

Borrowers with annual incomes below \$50,000 tend to default more frequently. This observation suggests that lower-income borrowers may face greater financial challenges, which impacts their ability to repay loans. To mitigate this risk, Lending Club could refine its credit scoring algorithms or adjust its lending criteria to better assess the repayment capacity of individuals in this income group. Enhanced financial assessments could reduce defaults and improve the risk management process.

References

1. UpGrad. (2024). **Machine Learning and AI: Lending Club Case Studies** [Videos]. Available on UpGrad App.
 2. UpGrad. (2024). **Loan.csv dataset**: Historical loan data for analysis. Provided for the Lending Club case study.
 3. Python Software Foundation. (2024). **Python Documentation**. Retrieved from: <https://docs.python.org/>
 4. eBooks on Python. (2024). **Python Programming Resources**. Various eBooks for Python learning and development.
-

Thank You!

Thank you for your time and attention.
Feel free to reach out with any questions or feedback.

Contact:

Naseem I Kesingwala

Email: naseem.kesingwala@gmail.com

Omkar Vamsi Krishna Pechetti

Email: opechetti@gmail.com