
Lending Club Historical Loan Data Analysis

Executive Summary

Our primary objective was to analyze a historical dataset of loan disbursements, enriched with borrower financial demographics and specific loan details, to uncover patterns in credit profile behaviors.

This analysis helped us identify trends, pinpoint areas of oversight, and devise potential strategies to:

Reduce the percentage of loan defaults.

Refine marketing and lending strategies.

Key Findings

Debt-to-Income Ratio (DTI):

Higher DTI ratios are strongly associated with loan defaults. Charged-off loans generally have higher DTI ratios compared to fully paid loans.

Annual Income:

Higher income levels correlate with lower loan default rates. As annual income increases, the number of charged-off loans decreases.

Recommendations

Focus on High DTI Borrowers:

Implement stricter lending criteria for borrowers with high DTI ratios. Consider setting a maximum DTI threshold below 40 for approval or offering higher interest rates to compensate for the increased risk.

Targeted Marketing Campaigns:

Develop targeted marketing campaigns aimed at higher-income brackets (\$191,951 and above), as they exhibit a lower likelihood of loan defaults. Highlight the benefits and competitive rates available to these demographics.

Income Verification and Support Programs:

Strengthen income verification processes to ensure accurate income reporting. Additionally, offer financial literacy programs and debt management support to borrowers in lower-income brackets (\$0-\$47,150) to help them manage their finances better.

Dynamic Risk Assessment Models:

Use dynamic risk assessment models that continuously update borrower risk profiles based on real-time financial behavior and credit usage. This can help identify potential defaulters early and implement preemptive measures.

Detailed Analysis:

Loan Status vs DTI

Descriptive Analysis:

The box plot shows that the median DTI for charged-off loans is higher than for fully paid-off loans.

The interquartile range (IQR) for charged-off loans is higher, indicating greater variability in DTI.

Inferential Analysis:

Mann-Whitney U Test

Null Hypothesis: There is no difference in DTI distributions between charged-off and fully paid loans.

Result: A near-zero p-value suggests the observed difference in DTI distributions is unlikely to have occurred by chance, rejecting the null hypothesis.

Effect Size:*Cliff's Delta*

Consistently negative values (around -0.18) indicate a moderate effect size, suggesting a greater likelihood of higher DTI values in the charged-off group.

*Loan Status vs Annual Income***Descriptive Analysis:**

The count plot indicates that as annual income increases, the number of charged-off loans decreases.

Inferential Analysis:*Linear Regression:*

Model: Loan default status as a function of annual income.

Result: The model shows a relationship between income level and the number of charged-off loans, although the p-value suggests this relationship might be due to chance, necessitating further testing.

*Loan Status vs DTI & Annual Income***Descriptive Analysis:**

The box plot shows that the IQR of DTI is consistently higher across almost all income brackets for charged-off loans compared to fully paid loans, suggesting higher DTI ratios for charged-off loans.

Inferential Analysis:*Logistic Regression:*

Model: Loan status as a function of DTI and income brackets.

Result: DTI and income brackets were significant predictors of loan status. Higher income brackets were associated with a lower likelihood of charge-offs, while higher DTI values increased the likelihood.

Effect Size:

Odds Ratios:

The logistic regression coefficients and odds ratios indicate that a one-unit increase in DTI is associated with a ~3.8% increase in the odds of a loan being charged off. Each movement up an income bracket corresponds to a ~16.88% decrease in the odds of loans being charged off.