

A blurred background image of a business meeting. In the foreground, a person's hand holds a black pen, pointing at a document with various charts and graphs. Another hand is visible, holding a silver pen. The documents feature pie charts, bar charts, and tables. The overall scene is out of focus, emphasizing the text overlay.

Lending Club Case Study

C71 – AI/ML - Huy Vo & Zueb Murthuza

Executive Summary

- This analysis aims to minimize credit losses by identifying risky loan applicants through a detailed examination of borrower and loan attributes. By leveraging data insights, the company can make informed decisions on loan approvals, such as:
 - Rejecting high-risk applications
 - Modifying loan amounts to mitigate exposure
 - Adjusting interest rates for specific borrower profiles
- The focus is on "Charged Off" loans, which represent applicants who defaulted, causing financial losses. The findings and recommendations provided aim to strengthen risk assessment processes and improve portfolio health.

Objectives

- Identify key factors contributing to loan defaults.
- Leverage insights to:
 - Target low-risk applicants.
 - Reduce financial exposure by refining loan approval criteria.
- Establish a data-driven approach to enhance decision-making and improve overall lending practices.





Analysis Approach

Data Understanding:

- Loaded and reviewed the dataset to understand its structure and key variables.
- Referenced the data dictionary for context and relevance.

Data Cleaning:

- **Missing Values:** Dropped columns with >80% missing data and imputed critical fields (e.g., emp_length) using income buckets.
- **Standardization:** Converted variables to appropriate formats (e.g., int_rate to float, term to integer). Extracted key information from columns like issue_d into issue_month and issue_year.
- **Redundancy Removal:** Removed irrelevant or duplicate columns (e.g., id, url, policy_code).
- **Outliers:** Applied the IQR method to handle extreme values in annual_inc.
- **Data Filtering:** Retained relevant records by excluding "Current" loan statuses.
- **Consistency Checks:** Standardized values (e.g., merged Source Verified into Verified) and ensured uniform categories.

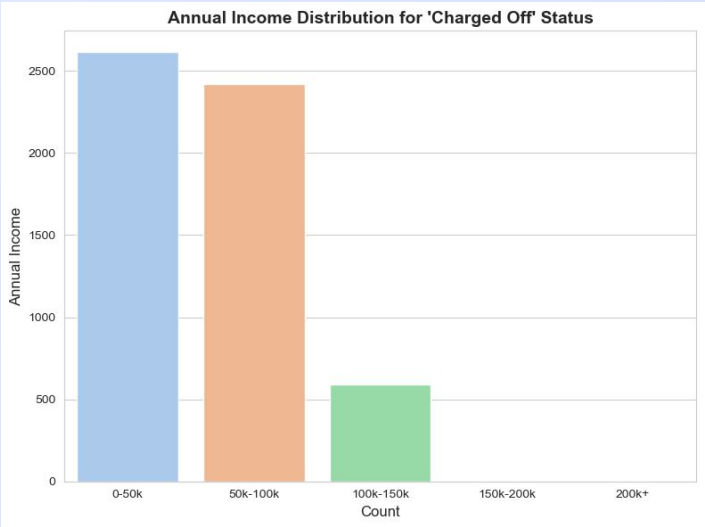
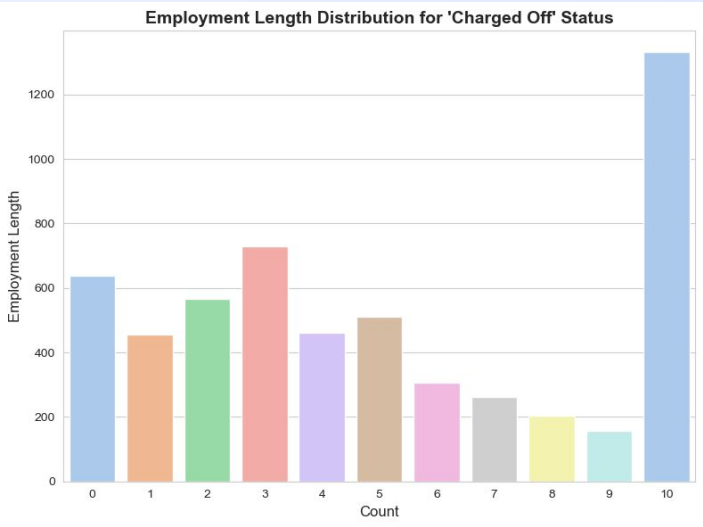
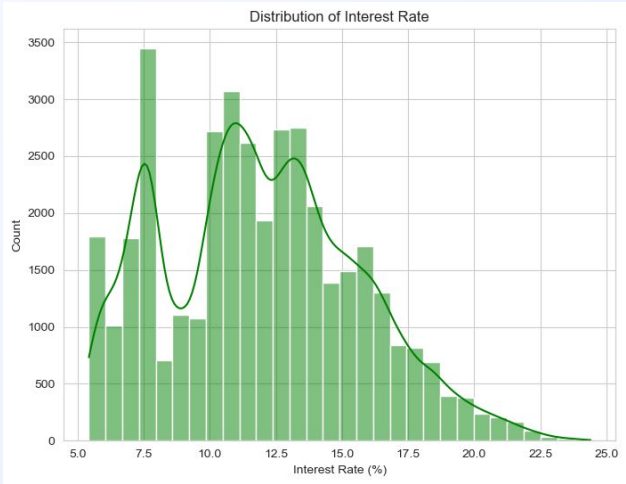
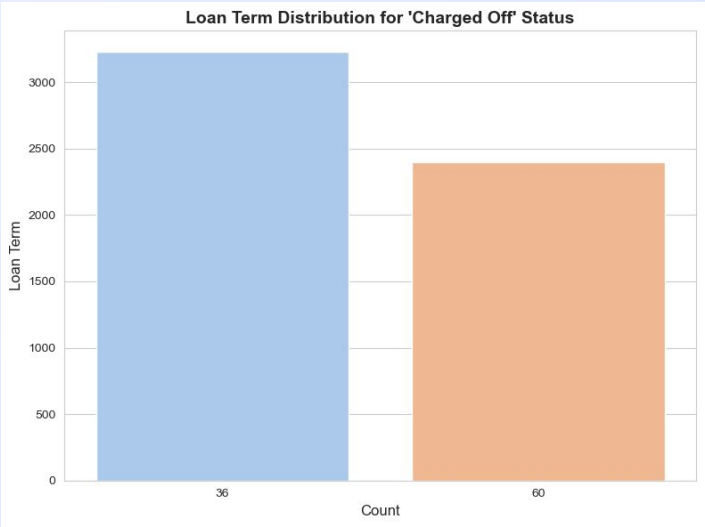
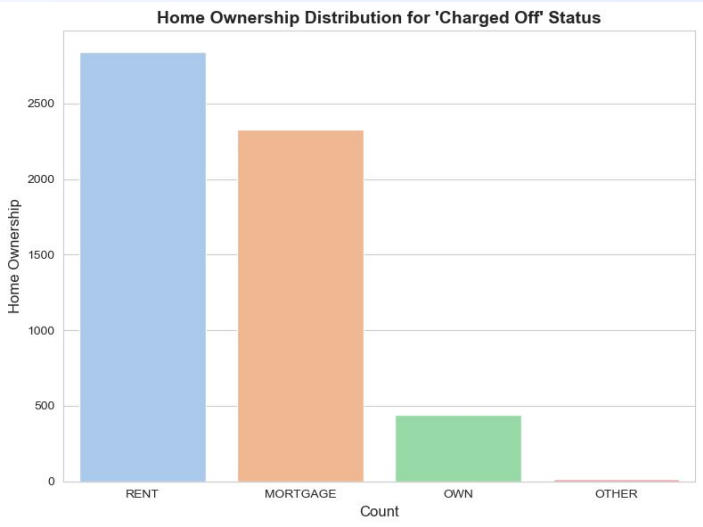
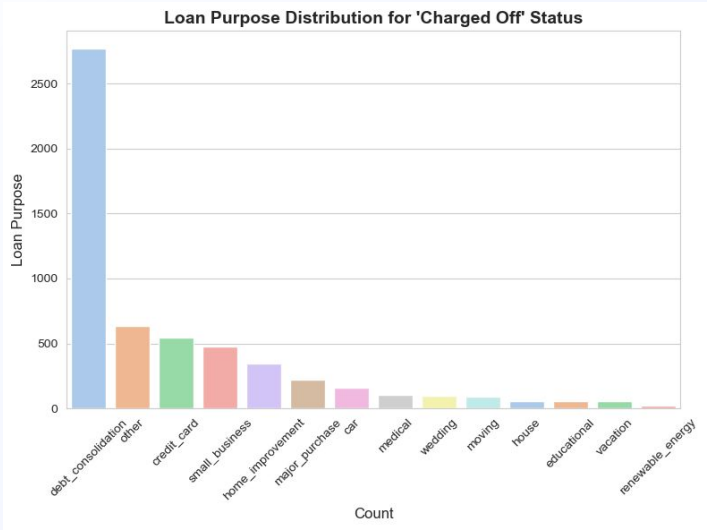
Exploratory Data Analysis (EDA):

- Conducted univariate and bivariate analyses to uncover patterns and relationships.
- Visualized trends using bar charts, boxplots, and correlation heatmaps.

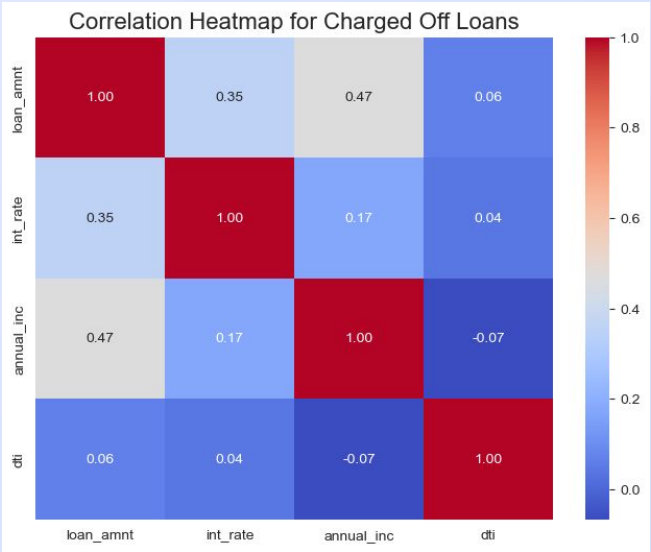
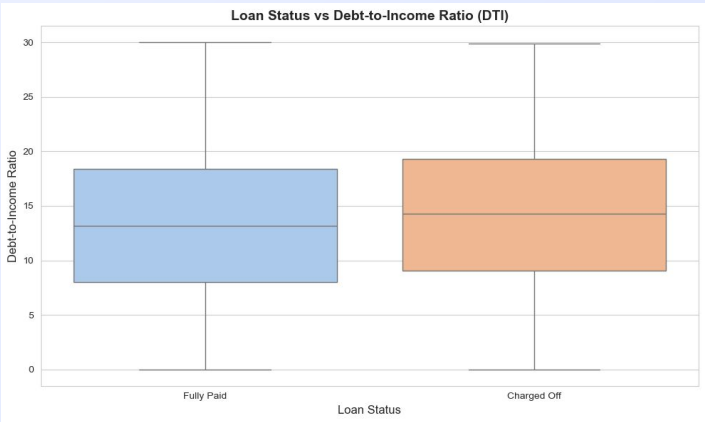
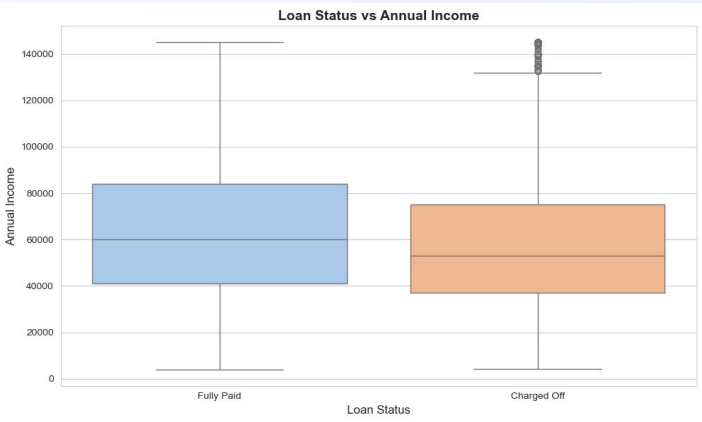
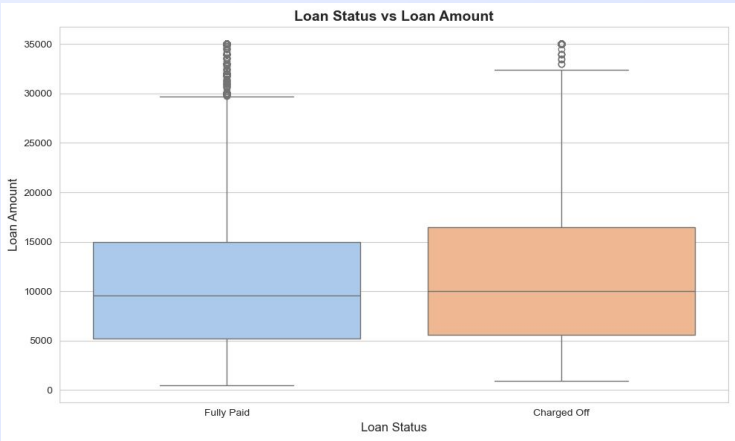
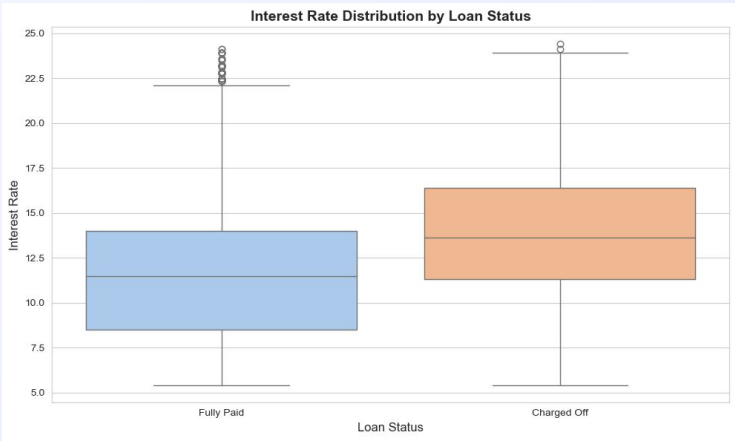
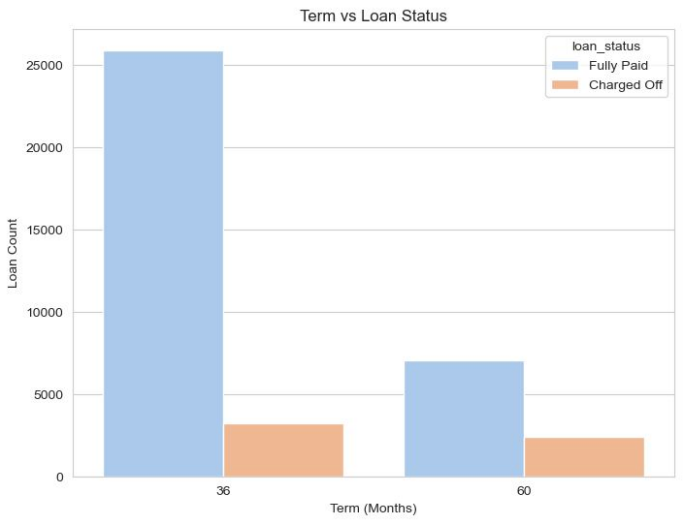
Insights and Recommendations:

- Identified key variables influencing defaults and proposed actionable steps to mitigate risk.

Key Visuals - Univariate



Key Visuals – Bivariate & Multivariate



Key findings - Univariate

Loan Status

Fully Paid: Represents **85.4%** of loans, indicating effective credit policies.

Charged Off: Accounts for **14.6%**, highlighting areas for improvement in borrower assessment.

Interest Rate Distribution

Peaks observed at **10%-15%** for charged-off loans, suggesting interest rates are a significant risk factor.

Loan Purpose

Majority of charged-off loans are for **debt consolidation**, with smaller portions attributed to credit card and small business loans.

14.6%
Charged off
Loans

10-15%
Peak interest rates
with most
defaults

**Debt
Consolidation**



Key findings – Univariate contd.

Home Ownership

Borrowers with **"RENT"** and **"MORTGAGE"** home ownership statuses dominate the charged-off category, likely due to financial pressures from rent or mortgage payments.

Annual Income

The majority of borrowers with charged-off loans earn less than **\$100K** annually, indicating income as a critical determinant of default risk.

Employment Length

Borrowers with **10+ years** of employment form a significant portion of defaults, showing that employment stability alone does not mitigate risk.

Rent &
Mortgage

100K
And less

10+ Years



Key findings – Bivariate & Multivariate

Interest Rate vs. Loan Status

Borrowers with interest rates above **10%** show a higher likelihood of default.

Loan Amount vs. Loan Status

Charged-off loans tend to have slightly higher **loan amounts** on average compared to fully paid loans.

DTI vs. Loan Status

Marginally **higher DTI** observed for charged-off loans, although its correlation with other variables is low.

Correlation Heatmap

Moderate correlations between Loan Amount and Annual Income (**0.47**) and Loan Amount and Interest Rate (**0.35**).

Weak or negligible correlations for DTI, suggesting it acts independently in assessing default risk.

Loan Term

Charged-off loans are more common for **36-month** terms, indicating shorter-term loans may pose higher risks.



Recommendations

- **Dynamic Risk-Based Pricing:**
 - Adjust interest rates for borrowers in the 10%-15% range to better balance risk and profitability.
- **Enhanced Screening for Debt Consolidation Loans:**
 - Strengthen risk assessment protocols for borrowers applying for debt consolidation, which accounts for the majority of charged-off loans.
- **Income-Driven Assessments:**
 - Incorporate robust income-based criteria to identify high-risk borrowers, particularly those earning below \$100K.
- **Refined Loan Structuring:**
 - Introduce caps on debt-to-income ratios to minimize default risks.
 - Offer tailored loan amounts and terms based on borrower profiles and repayment history.
- **Employment Tenure Evaluation:**
 - Reassess reliance on employment length as a primary stability factor and consider additional financial metrics.
- **Portfolio Monitoring and Policy Adjustments:**
 - Continuously analyze trends in loan purposes, borrower demographics, and repayment behaviors.
 - Utilize predictive modeling to proactively refine risk management strategies.



Thank you