Lending Club Case Study

RAJEEV SRIKANTH

RAJIV GABA

upGrad and IIITB Machine Learning & AI Program (September 2024 cohort)

CASE STUDY OVERVIEW

SECTION 1

Problem Statement and Business Objective

PROBLEM STATEMENT

1. The company must decide on loan approvals based on applicant profiles, balancing the risk of business loss (rejecting creditworthy applicants) and financial loss (approving risky applicants).
2. The objective is to use Exploratory Data Analysis (EDA) to analyze past loan data and identify patterns and variables influencing loan defaults.

BUSINESS OBJECTIVE

1. Reduce credit losses by identifying key factors driving loan defaults and refining the risk assessment process.
2. Maintain accessibility to lower-interest loans while improving decision-making for loan approvals to optimize the company's portfolio and minimize defaults.

Analysis Approach

1. Data cleaning
    1. Removing redundant columns
    2. Data type issue fixing
    3. Null value handling
2. Outlier handling
3. Data analysis
    1. Calculating derived metrics
    2. Data visualisation for insights

1. All redundant columns – columns with either all null values or all rows having same value – were removed
2. Data type issue was fixed for all object data types by converting them to proper format – integer/float/date – using string manipulation techniques
3. Columns with more than 50% null values were dropped completely while null value imputation for columns with few null values were done – through mean/median/0 imputation – based on the data distribution
4. Outlier handling for key continuous variables was done using boxplot analysis
5. Derived metrics for further use in the EDA section were calculated based on existing columns
6. Multiple insights were obtained using EDA and data visualisation techniques

UNIVARIATE ANALYSIS

SECTION 2

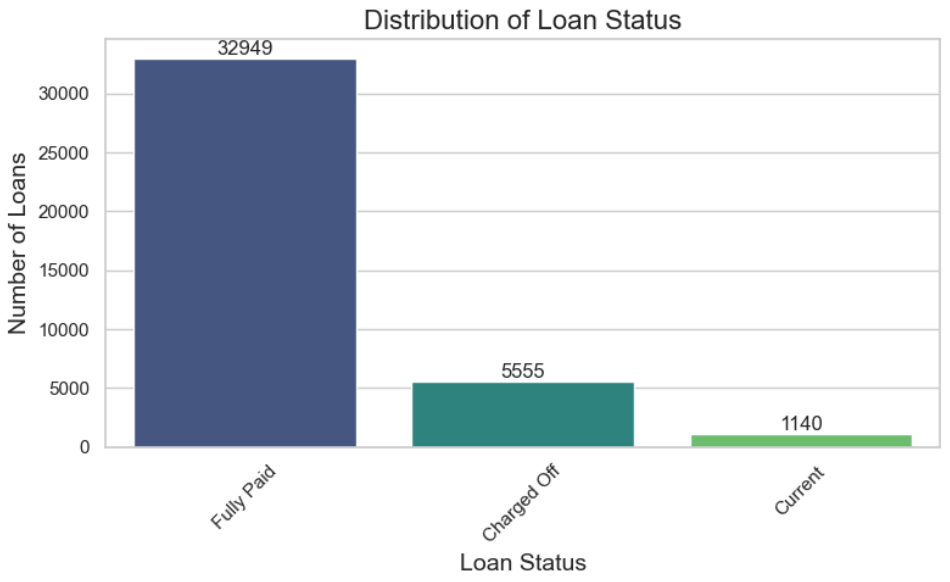
Loan Status Distribution Highlights Default Risk

Analysis

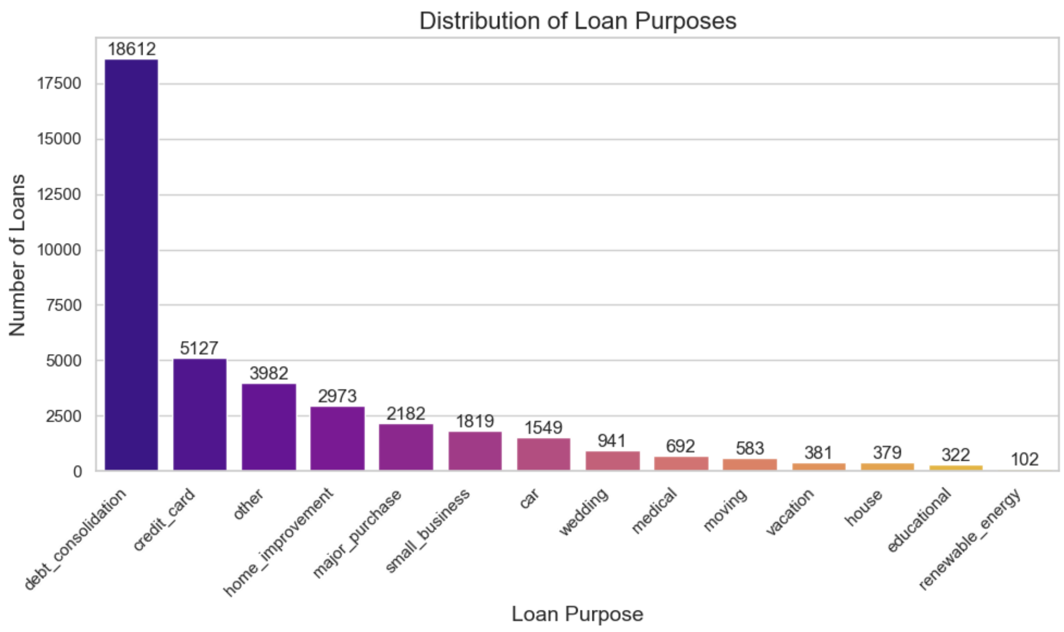
1. **Most loans are successfully repaid:** Over 32,000 loans are fully paid, indicating strong repayment trends among borrowers.
2. **Default rate is significant:** With 5,555 loans charged-off, default risk poses a notable challenge for the lender.
3. **Minimal active loans:** Only 1,140 loans are currently being repaid, suggesting the dataset predominantly contains completed loan records.

Business Implications

1. **Key focus on default reduction:** Charged-off loans represent a critical area for intervention to minimize credit losses.
2. **Leverage repayment trends:** High full repayment rates can be used to refine models for identifying creditworthy applicants while excluding risky profiles.



Debt Consolidation Dominates Loan Purposes



Analysis

1. **Debt consolidation leads:** A large chunk of applications (~47%) pertain to consolidation of their debts makes it a risky affair.
2. **Credit card refinancing follows:** Loan taken for the purpose of clearing credit card debts further adds fuel to the fire.
3. **Other purposes are diverse but less frequent:** Categories like home improvement (3,982), major purchases (2,973), and small businesses (2,182) contribute smaller volumes, with niche categories (e.g., renewable energy, house, and vacation) being minimal.

Business Implications

1. **Focus on high-risk categories:** Since debt consolidation and credit card loans dominate, these categories should be analyzed closely for default risk trends to develop targeted credit risk models.
2. **Opportunity in niche loans:** Smaller categories like renewable energy and house loans may present opportunities to expand business in lower-risk areas if repayment trends are favorable.

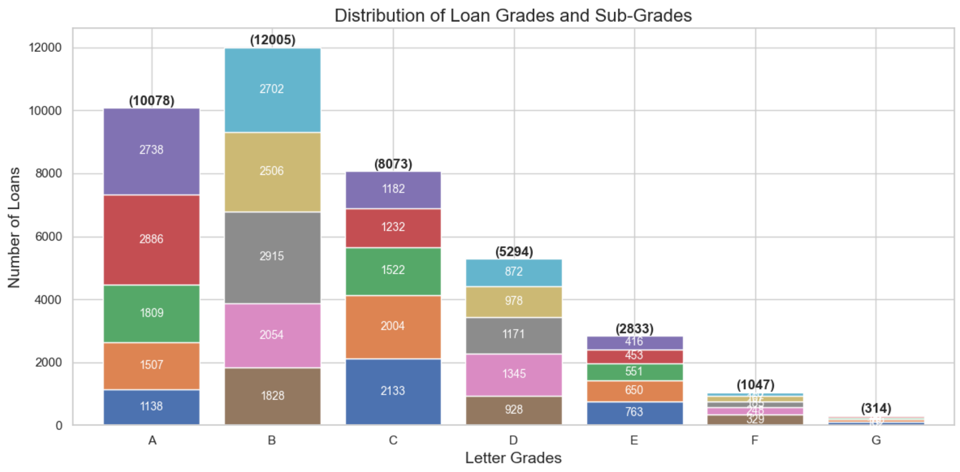
Loan Grades Distribution Indicates Varying Risk Profiles

Analysis

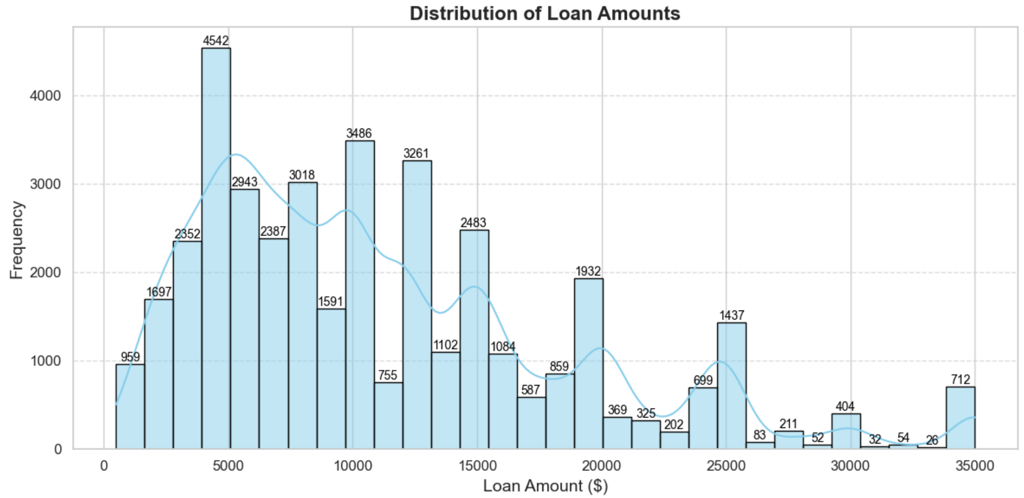
1. **Grade A and B dominate:** Grade A (10,078 loans) and B (12,005 loans) represent the majority of loans, indicating a significant portion of the portfolio is allocated to low-risk borrowers.
2. **Declining volume with higher grades:** Loan volumes decrease significantly from Grade C (8,073 loans) to Grade G (314 loans), reflecting a lower preference for high-risk borrowers.
3. **Sub-grade variability:** Sub-grades within each grade show varying volumes, suggesting a nuanced distribution of risk within the same grade.

Business Implications

1. **Portfolio stability in lower grades:** Grades A and B loans, being the majority, contribute to portfolio stability; a focus on maintaining their high repayment rates is essential.
2. **Risk assessment in higher grades:** Grades D-G represent fewer loans but higher risk; further analysis of default rates in these grades can help optimize lending strategies for riskier borrowers.



Loan Amounts Cluster Around $5,000 to $15,000



Analysis

1. **No consistency in loan amount distribution:** The loan amounts do not resemble any distribution at all
2. **Peak at $5,000:** The most common loan amount is around $5,000, with a frequency of 4,542 loans.
3. **Mid-range dominance:** A significant number of loans fall between $5,000 and $15,000, indicating a preference for moderate loan sizes.
4. **Higher amounts are rare:** Loan amounts above $25,000 are much less frequent, representing a smaller proportion of the portfolio.

Business Implications

1. **Focus on mid-range borrowers:** The concentration of loans in the $5,000–$15,000 range suggests these borrowers are the primary customer base and require careful credit risk assessment.
2. **Opportunity in larger loans:** While less frequent, larger loan amounts could provide higher profitability but may require stricter evaluation criteria to mitigate risk.

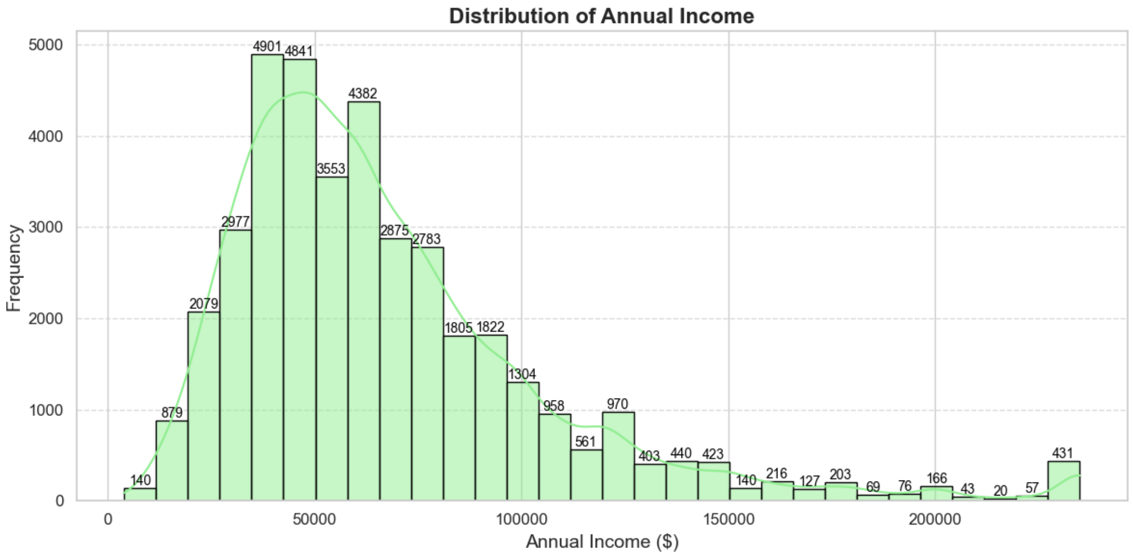
Annual Income Distribution Peaks Around $50,000

Analysis

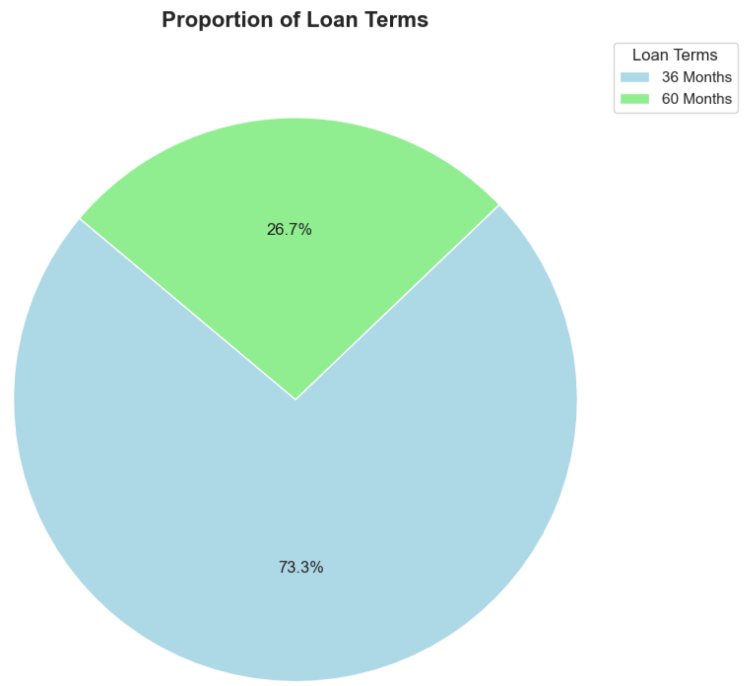
1. **Concentration around $50,000:** Many applicants have an annual income in the $40,000–$60,000 range, with peaks at approximately $50,000 (4,901 individuals).
2. **Long tail at higher incomes:** Fewer individuals have incomes exceeding $100,000, indicating that high-income borrowers are a smaller segment of the applicant pool.
3. **Low-income representation:** Many borrowers have incomes below $30,000, potentially indicating higher credit risk for them.

Business Implications

1. **Income-based risk profiling:** Borrowers with annual incomes below $30,000 may need stricter risk assessments due to their potential difficulty in repayment.
2. **Focus on middle-income segment:** The $40,000–$60,000 range represents the primary customer base and provides a balance of volume and repayment potential, necessitating tailored loan products and evaluation models for this group.



Loan Terms Are Predominantly Short Duration



Analysis

1. **Shorter loan terms dominate:** 73.3% of loans have a 36-month term, indicating that most borrowers prefer or qualify for shorter repayment periods.
2. **Smaller proportion for 60-month terms:** Only 26.7% of loans have a 60-month term, reflecting either higher risk or lower demand for longer-term loans.

Business Implications

1. **Focus on shorter-term loans:** The high preference for 36-month loans suggests tailoring credit products to this segment, which may have lower default risks due to shorter commitments.
2. **Evaluate risk in longer terms:** The 60-month loans may need closer monitoring for default trends, as longer durations typically entail higher credit risk.

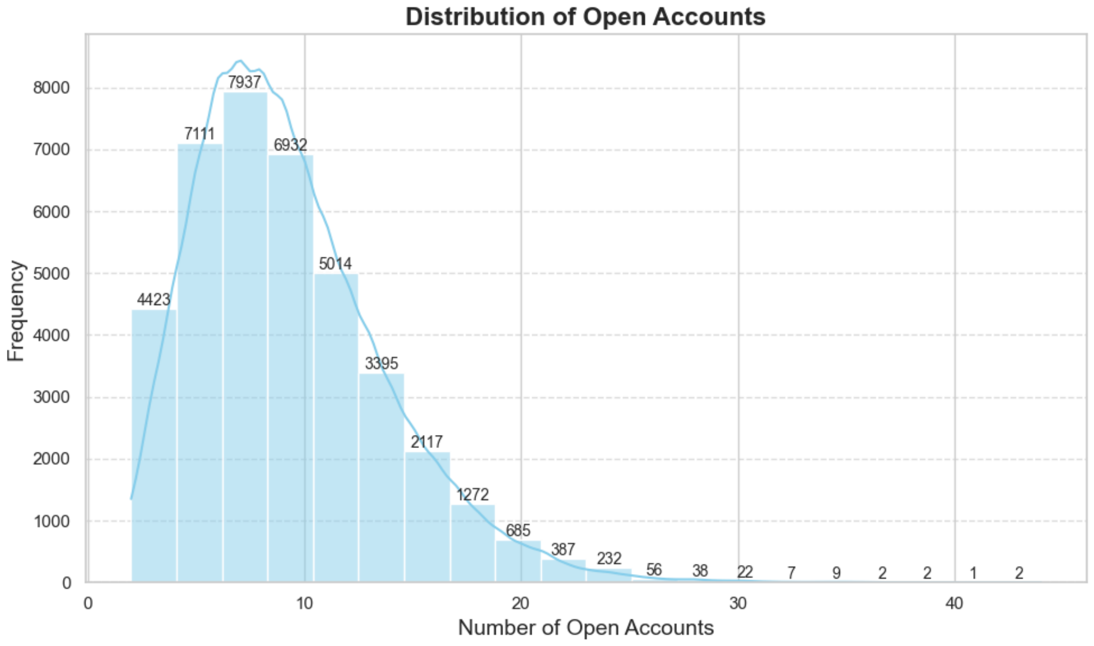
Majority of Borrowers Have 5-15 Open Accounts

Analysis

1. **Peak at 8 open accounts:** The majority of borrowers (7,937) have around 8 open accounts, which appears to be the most common number.
2. **Typical range of 5-15 accounts:** Most borrowers have between 5 and 15 open accounts, with a sharp decline in frequency beyond this range.
3. **Outliers with high open accounts:** Very few borrowers have more than 20 open accounts, indicating that such cases are rare.

Business Implications

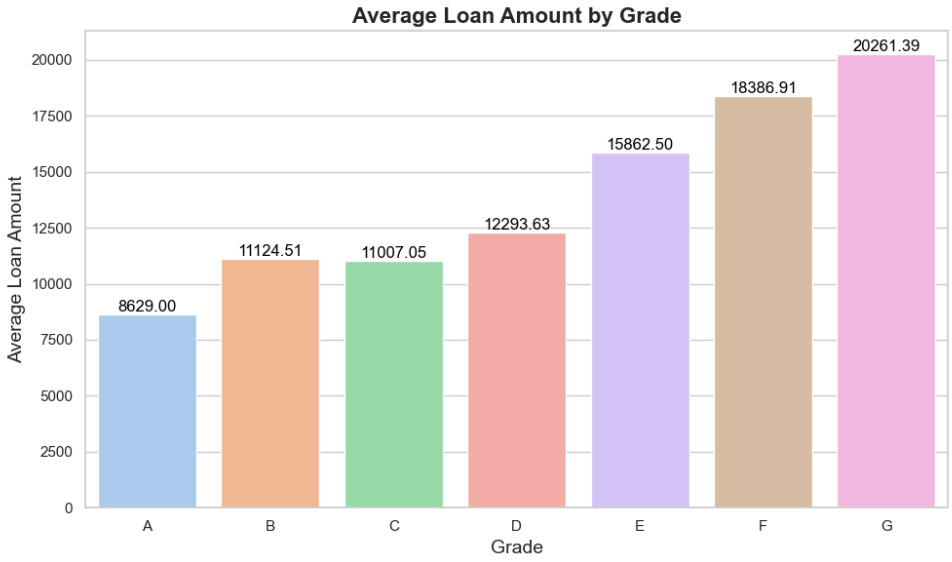
1. **Optimal credit load assessment:** Borrowers with 5-15 open accounts may represent a stable credit segment, requiring balanced risk evaluation.
2. **Monitor high-account borrowers:** Applicants with more than 20 open accounts may indicate potential over-leverage, necessitating stricter scrutiny to mitigate risk.



BIVARIATE ANALYSIS

SECTION 3

Higher Loan Grades Are Associated with Larger Loan Amounts



Analysis

1. **Loan amounts increase with grade risk:** The average loan amount rises steadily from Grade A ($8,629) to Grade G ($20,261), indicating that higher-risk borrowers tend to take larger loans.
2. **Significant jump in higher grades:** There is a noticeable increase in loan amounts from Grade D ($12,293) onward, with Grades F and G having substantially higher averages.
3. **Low amounts in Grade A and B:** Borrowers in the safest grades tend to borrow relatively smaller amounts, reflecting a conservative borrowing pattern.

Business Implications

1. **Higher risk, higher exposure:** Larger loan amounts in higher grades amplify financial exposure, making robust risk assessments critical for Grades E, F, and G.
2. **Focus on balancing profitability and risk:** Lending to higher grades might be more profitable due to larger amounts, but default risks must be mitigated through stricter underwriting and pricing strategies.

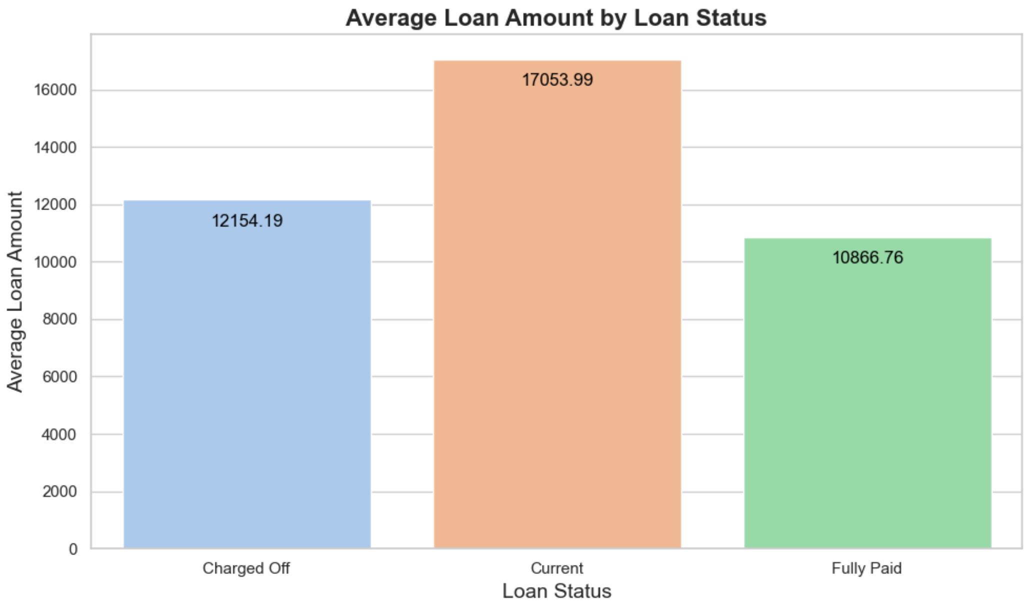
Higher Loan Amounts Linked to Defaults

Analysis

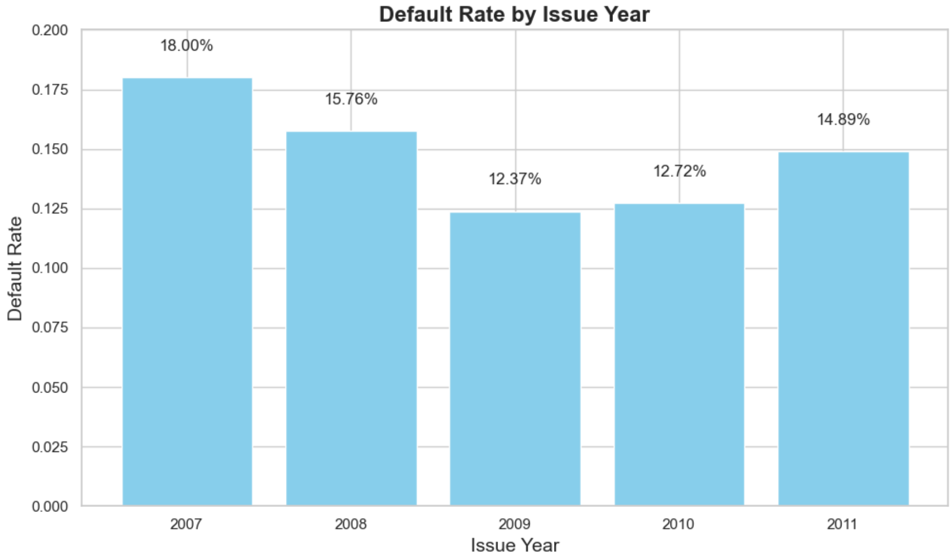
1. **Higher loan amounts for defaults:** The average loan amount for charged-off loans ($12,154) is notably higher than fully paid loans ($10,867), suggesting a possible correlation between higher loan amounts and increased default risk.
2. **Current loans have the highest average:** Loans in the "Current" status have the highest average amount ($17,054), indicating that ongoing loans often involve larger sums.
3. **Possibility of default of current loans:** Currently open loans defaulting will further increase the average default loan amount.

Business Implications

1. **Limit loan size for high-risk profiles:** The higher average for charged-off loans indicates a need to impose stricter loan limits on applicants flagged as high risk to reduce default exposure.
2. **Monitor larger ongoing loans:** The high average for current loans suggests close tracking of repayment behavior is necessary for this segment to preempt potential defaults.



Default Rates Reflect Economic Trends



Analysis

1. **High default rates in 2007 and 2008:** Default rates were 18.00% and 15.76%, respectively, likely influenced by the global financial crisis impacting borrowers' repayment abilities.
2. **Decline in 2009 and 2010:** Defaults dropped to 12.37% in 2009 and remained low at 12.72% in 2010, possibly due to economic recovery and improved lending practices.
3. **Rebound in 2011:** Default rates rose to 14.89%, suggesting economic challenges or shifts in borrower demographics or credit policies.

Out-of-Chart Insight

- The global financial crisis (2007-2008) likely played a significant role in the high default rates, as widespread economic uncertainty and job losses impacted borrowers' ability to repay loans.

Business Implications

1. **Economic resilience in credit policies:** Build lending models that account for macroeconomic risks to withstand economic downturns.
2. **Proactive default management:** Monitor early warning signs during economic recovery phases (like 2011) to prevent default spikes.

Wide Variation in Default Rates Across States

Analysis

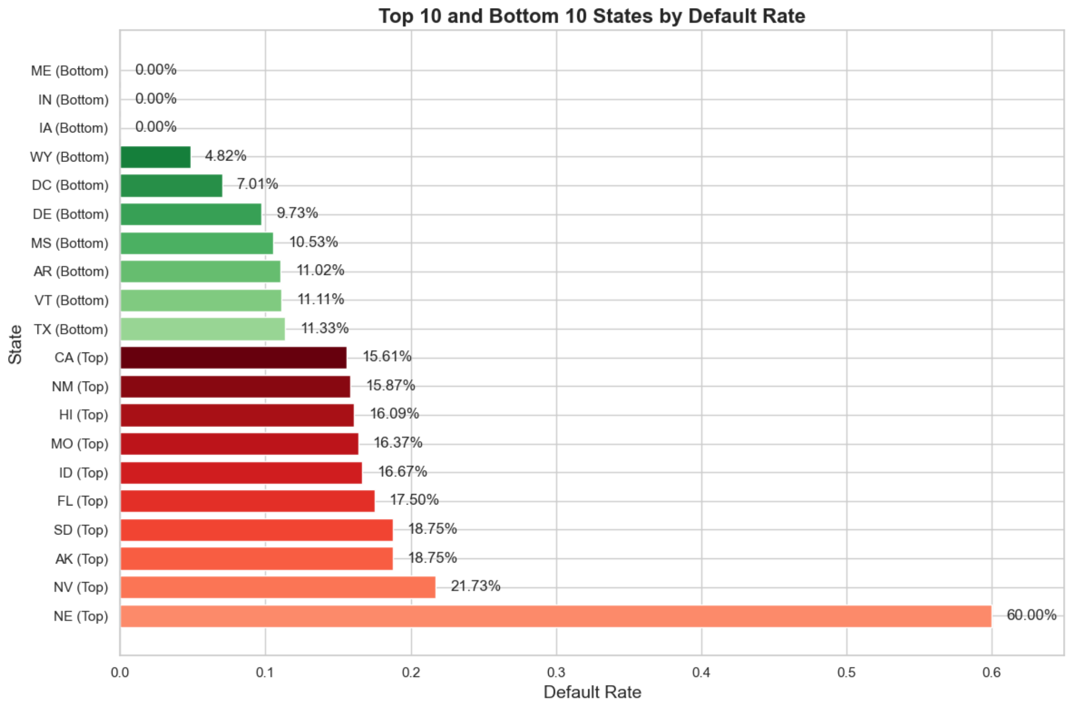
1. **Highest default rate in Nebraska (NE):** NE leads with an exceptionally high default rate of 60.00%, significantly outpacing other states.
2. **Other high-risk states:** States like Nevada (21.73%), South Dakota (18.75%), and Florida (17.50%) also exhibit high default rates, suggesting regional risk concentrations.
3. **Low-risk states dominate the bottom:** States like Maine (ME), Indiana (IN), and Iowa (IA) show 0% default rates, indicating negligible risk.

Business Implications

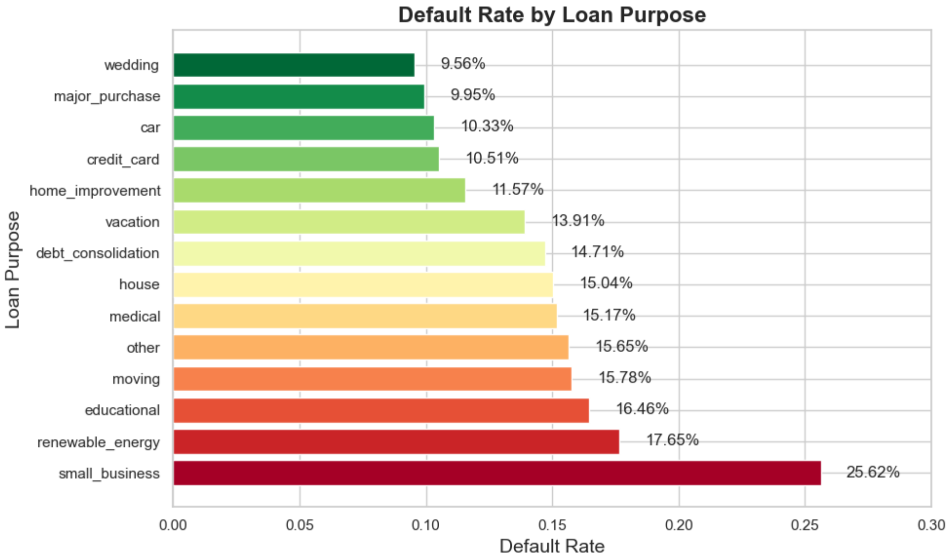
1. **Regional risk strategy:** High-risk states, especially NE, require stringent credit policies or higher interest rates to reduce potential losses.
2. **Opportunity in low-risk regions:** Lending in low-default states can be expanded to balance portfolio risk and enhance profitability.

Out-of-Chart Insight

- The high default rate in NE could indicate state-specific economic challenges, such as industry reliance or unemployment spikes, that necessitate further investigation.



Small Business Loans Have the Highest Default Rate



Analysis

1. **Highest default for small business loans:** Loans for small businesses have highest default rate at 25.62%, significantly above other purposes.
2. **Renewable energy and education loans also risky:** These categories show relatively high default rates of 17.65% and 16.46%, respectively.
3. **Debt consolidation default rate is relatively low:** This is the complete opposite to the belief that loans to pay off loans can worsen default rates.

Business Implications

1. **Tighter screening for small business loans:** Given their high default rate, small business loans require stricter credit checks and potentially higher interest rates to offset risk.
2. **Focus on low-risk purposes:** Expand offerings in lower-risk categories, like weddings and major purchases, to balance portfolio risk.

Out-of-Chart Insight

- The high default rate for small businesses could reflect economic volatility, lack of steady cash flows, geographical issues or startup failure rates, underscoring the need for tailored credit evaluation models for this segment.

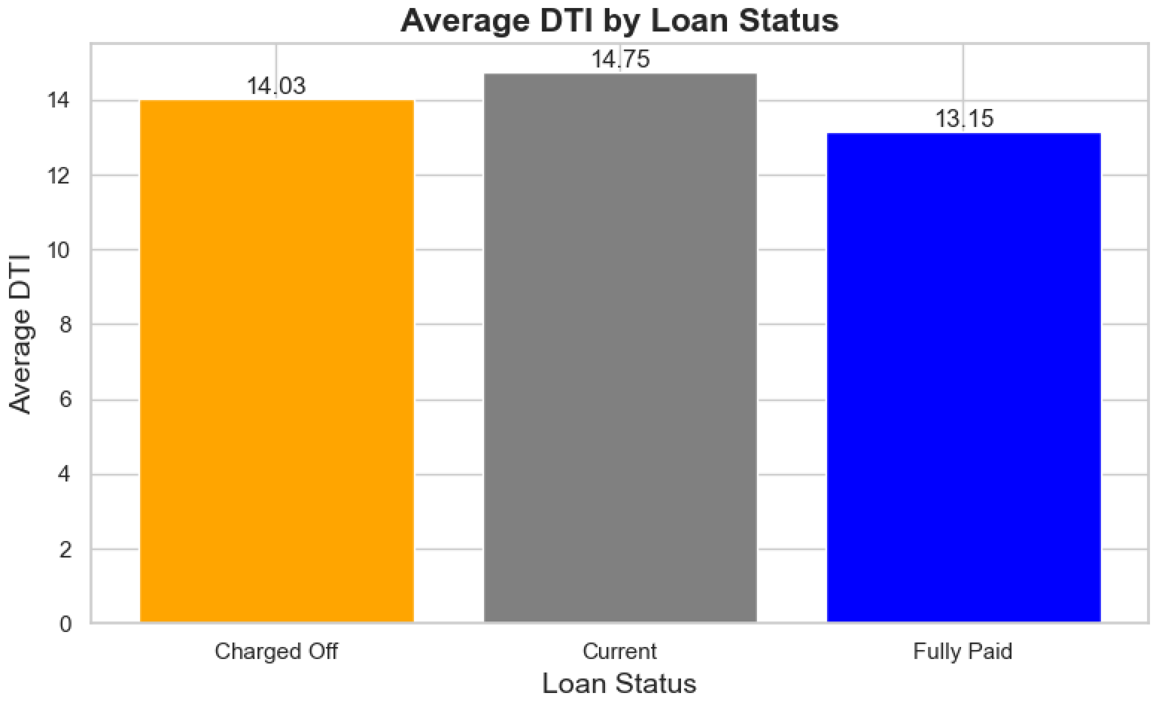
Higher DTI Linked to Charged-Off and Current Loans

Analysis

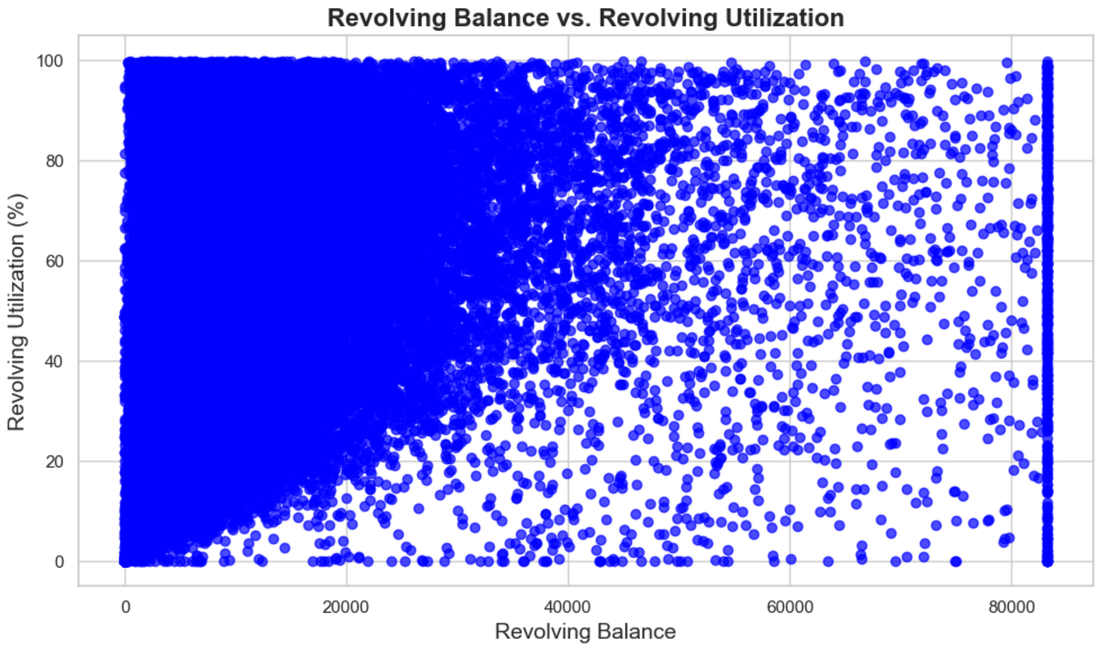
1. **Current loans have the highest DTI:** Borrowers with "Current" loans have the highest average Debt-to-Income (DTI) ratio at 14.75, indicating higher financial obligations relative to income.
2. **Charged-off loans have elevated DTI:** The average DTI for charged-off loans is 14.03, higher than fully paid loans, suggesting a correlation between higher DTI and default risk.
3. **Fully paid loans have the lowest DTI:** Borrowers who fully paid their loans have an average DTI of 13.15, indicating lower financial strain.

Business Implications

1. **Set DTI thresholds for approvals:** To reduce default risk, stricter DTI limits should be enforced during credit evaluations.
2. **Monitor borrowers with high DTI:** Active loans with elevated DTI levels (e.g., "Current" loans) should be closely monitored for early signs of repayment difficulties.



High Revolving Utilization Observed Across Revolving Balances



Analysis

1. **Broad range of utilization levels:** Borrowers exhibit varying revolving utilization rates (0–100%), regardless of their revolving balance amounts.
2. **High utilization clustering:** A significant proportion of borrowers have revolving utilization close to 100%, indicating heavy reliance on credit.
3. **Outliers at higher balances:** Borrowers with very high revolving balances (above $80,000) still show diverse utilization rates, suggesting unique financial behavior.

Business Implications

1. **High utilization as a risk indicator:** Borrowers with high revolving utilization rates may signal financial stress, warranting closer creditworthiness checks.
2. **Segment-based analysis:** Analyze borrowers with high balances and high utilization separately to design tailored risk management strategies.

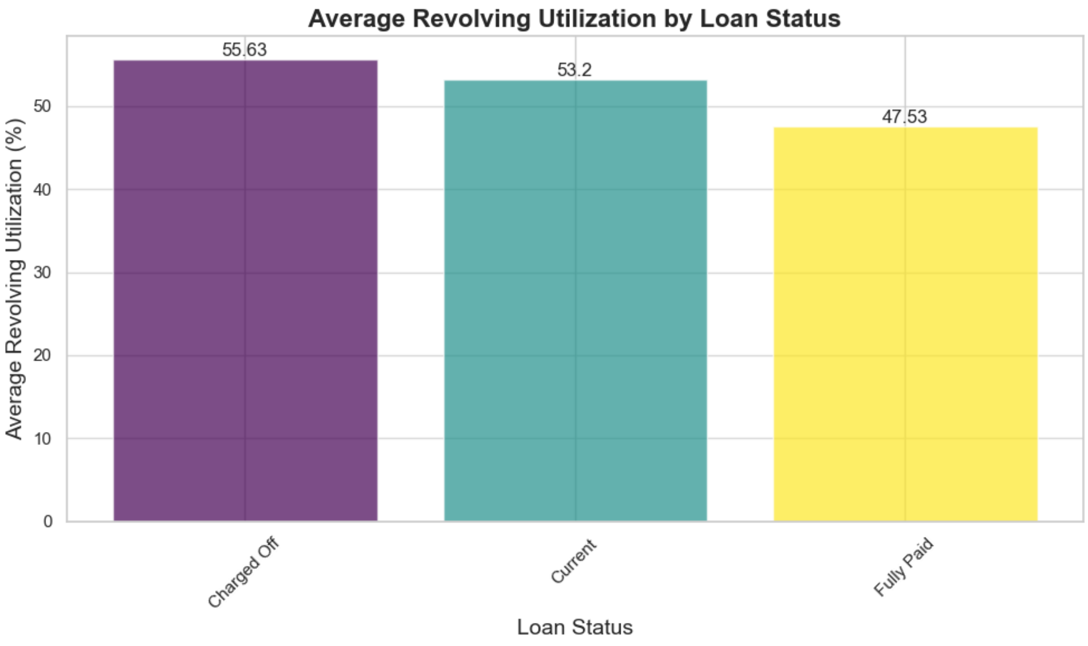
Charged-Off Loans Have the Highest Revolving Utilization

Analysis

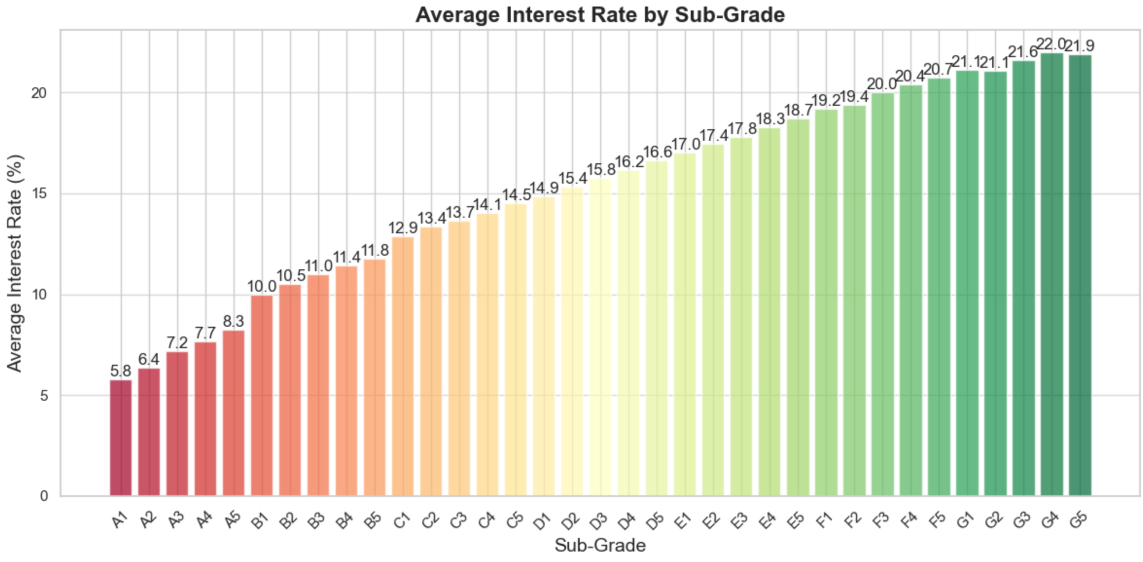
1. **Charged-off loans lead in utilization:** The average revolving utilization for charged-off loans is 55.63%, higher than both current loans (53.2%) and fully paid loans (47.53%).
2. **Fully paid loans show the lowest utilization:** Borrowers who fully paid their loans tend to have lower average revolving utilization, indicating healthier financial management.
3. **Current loans fall in between:** Current loans exhibit moderately high revolving utilization, suggesting a potential risk of default.

Business Implications

1. **Revolving utilization as a risk metric:** Elevated revolving utilization is a strong indicator of repayment risk and should be incorporated into risk assessment models.
2. **Target utilization thresholds:** Encourage borrowers with high utilization to lower their revolving balances to reduce default risk.



Interest Rates Rise with Sub-Grade Risk



Analysis

1. **Low-risk sub-grades have lower interest rates:** Sub-grades like A1 (5.8%) and A2 (6.4%) offer the lowest interest rates, reflecting lower credit risk.
2. **High-risk sub-grades have elevated rates:** Sub-grades like G4 and G5 exhibit the highest average interest rates at 22.0% and 21.9%, respectively, to compensate for greater default risk.
3. **Gradual increase across sub-grades:** There is a consistent upward trend in interest rates as sub-grade risk increases, showing a direct correlation.

Business Implications

1. **Align rates with risk:** Continue using interest rates to price loans based on credit risk, ensuring adequate compensation for higher-risk lending.
2. **Focus on mid-range sub-grades:** Sub-grades in the B and C ranges, with moderate rates, could balance profitability and risk effectively.

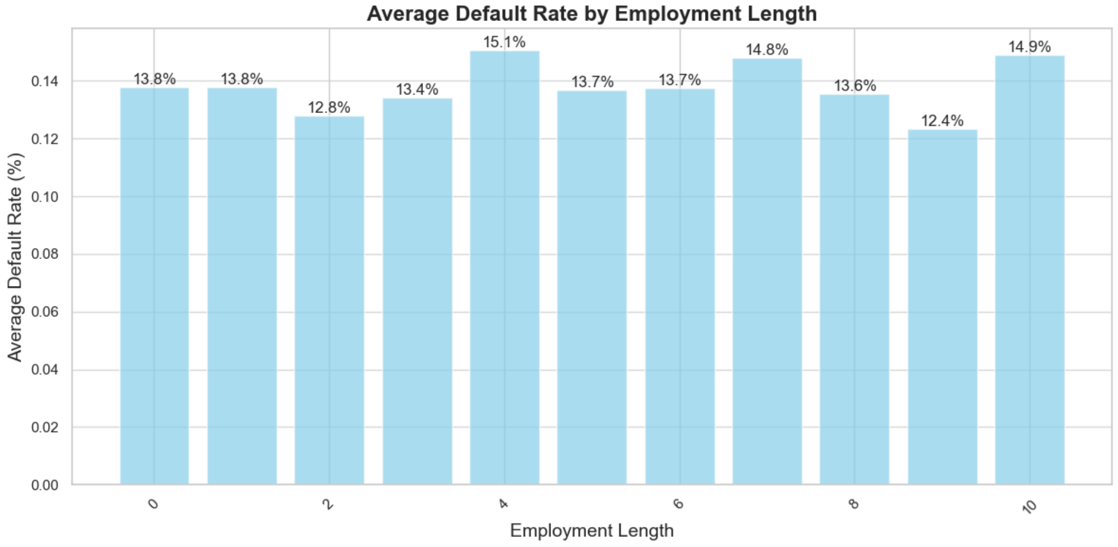
Default Rates Vary Marginally by Employment Length

Analysis

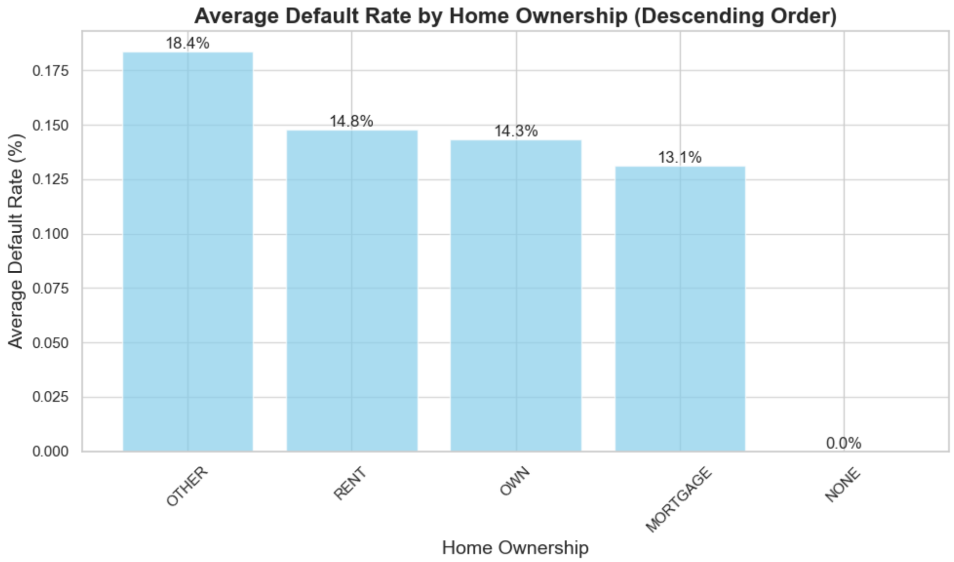
1. **Slight peak at 1 year:** Borrowers with 1 year of employment have the highest average default rate at 15.1%, suggesting early employment borrowers may carry higher risk.
2. **Lowest rate at 9 years:** Borrowers with 9 years of employment exhibit the lowest default rate at 12.4%, reflecting financial stability.
3. **Consistent rates elsewhere:** Default rates remain relatively stable across other employment lengths, averaging around 13-14%.

Business Implications

1. **Careful assessment for shorter employment:** Borrowers with 1 year or less of employment should undergo stricter evaluation due to higher risk.
2. **Leverage stable segments:** Focus on borrowers with 9+ years of employment as they represent lower default risk, offering safer lending opportunities.



Default Rates Vary by Home Ownership Type



Analysis

1. **Highest default rate in "Other" category:** Borrowers classified under "Other" have the highest default rate at 18.4%, indicating significant risk in this group.
2. **Renters show elevated default rates:** Renters exhibit a default rate of 14.8%, higher than homeowners or those with mortgages.
3. **Lowest rate for mortgages:** Borrowers with a mortgage show the lowest default rate among active categories at 13.1%, reflecting more financial stability.

Business Implications

1. **Target risk mitigation for renters and "Other":** Develop stricter credit policies or risk premiums for renters and borrowers in the "Other" category to address their higher default rates.
2. **Focus on mortgage-backed lending:** Borrowers with mortgages demonstrate the lowest risk, making them a reliable segment for safer loan products.

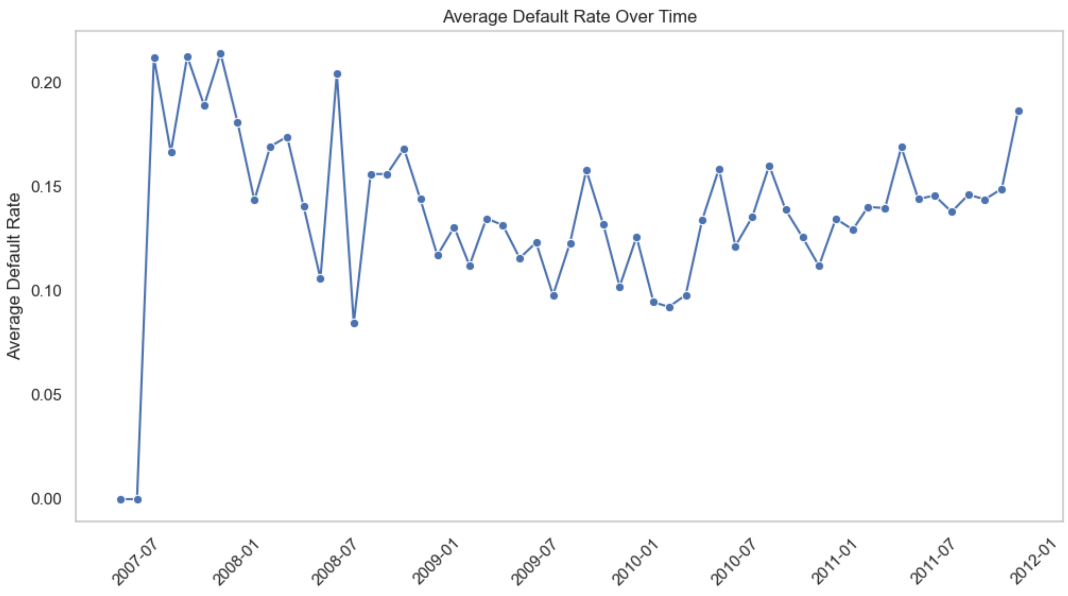
Default Rate Fluctuations Reflect Economic Trends

Analysis

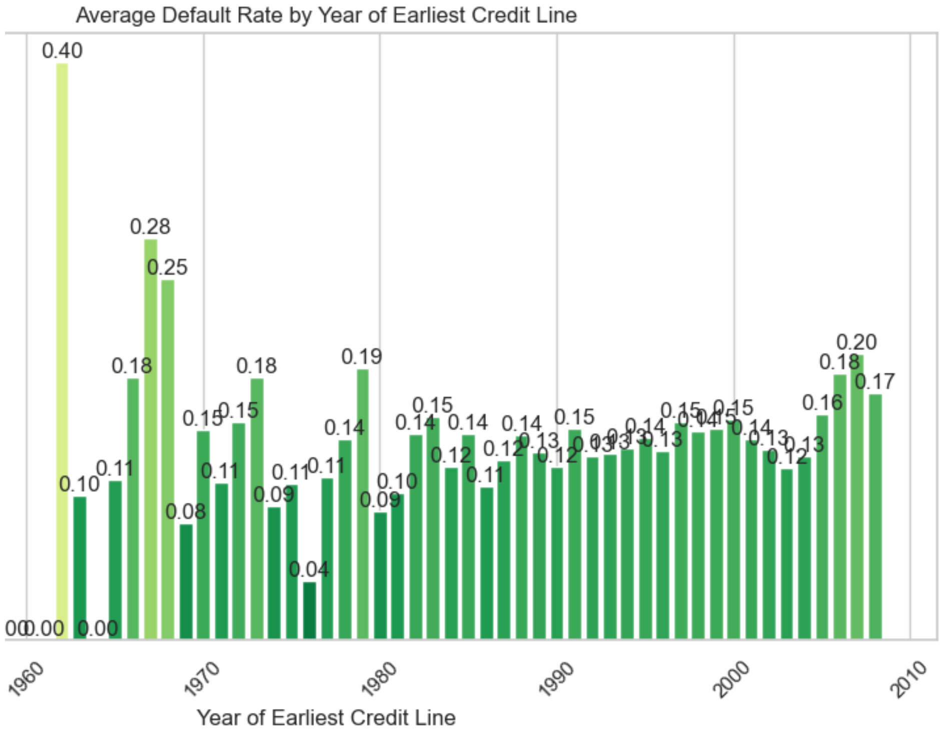
1. **Initial spike during 2007–2008:** Default rates peaked during the 2007–2008 period, likely influenced by the global financial crisis, with increased economic instability affecting borrower repayment abilities.
2. **Stabilization post-2009:** Default rates saw a steady decline after 2009, indicating improving economic conditions and better lending practices.
3. **Slight increase in 2012:** A rise in default rates toward 2012 suggests possible economic challenges or adjustments in borrower risk profiles.

Business Implications

1. **Economic context in risk models:** Incorporate macroeconomic factors into credit models to better predict and manage default risks during downturns.
2. **Prepare for cyclical risks:** Maintain proactive monitoring of external economic conditions to adjust lending strategies dynamically and mitigate rising default trends.



Default Rates Vary by Year of Earliest Credit Line



Analysis

1. **Highest default rate in early credit lines:** Borrowers with the earliest credit lines from the 1960s have significantly high average default rates (up to 40%), though these are likely outliers due to limited data.
2. **Steady default rates after 1970:** Post-1970, default rates stabilize, hovering between 10% and 20%, reflecting more consistent borrower profiles and credit behaviors.
3. **Slight increase in recent years:** Borrowers with credit lines initiated post-2000 show a gradual rise in default rates, peaking at 20% in the early 2010s.

Business Implications

1. **Investigate early credit line risks:** High default rates for older credit lines may require validation, likely due to historical inconsistencies or small data volumes.
2. **Monitor newer credit line trends:** Rising default rates in recent credit line cohorts suggest the need for more robust evaluation of younger borrowers or those with limited credit history.

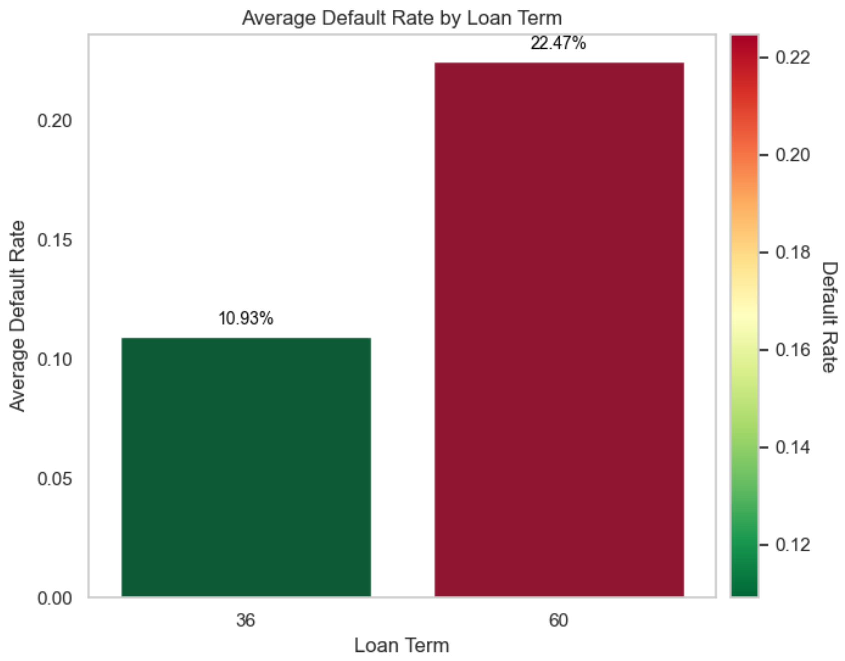
Longer Loan Terms Have Higher Default Rates

Analysis

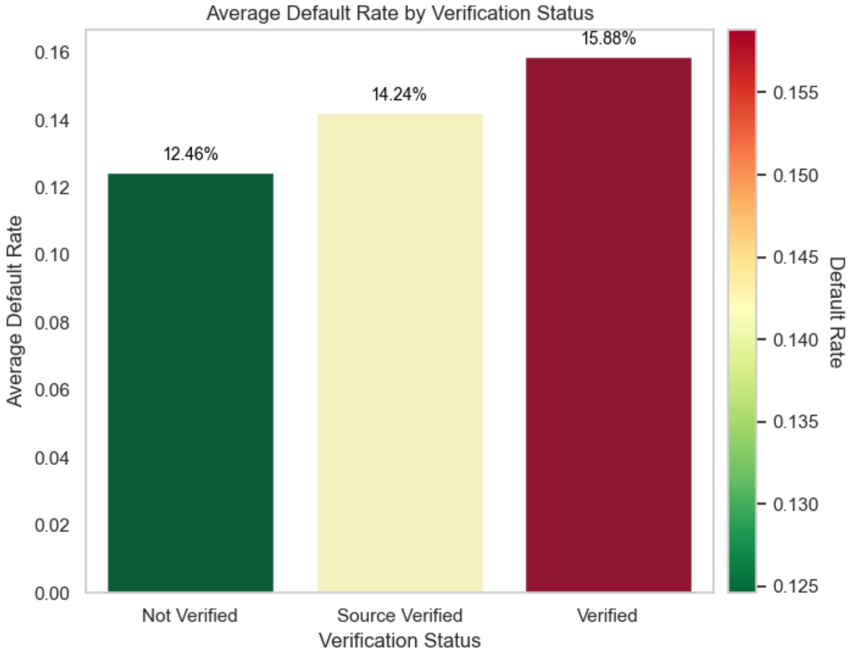
1. **60-month loans have a higher default rate:** Loans with a 60-month term show an average default rate of 22.47%, more than double the rate for 36-month loans (10.93%).
2. **Shorter terms show lower risk:** The significantly lower default rate for 36-month loans highlights their relative safety in lending portfolios.

Business Implications

1. **Focus on shorter-term loans:** Favoring 36-month loans can reduce overall portfolio risk while maintaining profitability.
2. **Stricter policies for longer terms:** Implement stricter credit checks, higher interest rates, or other risk mitigation measures for 60-month loans to offset their higher default risk.



Verified Loans Show the Highest Default Rate



Analysis

1. **Verified loans have the highest default rate:** Loans with verified borrower information exhibit a default rate of 15.88%, higher than source-verified (14.24%) and not verified loans (12.46%).
2. **Lower default rate for unverified loans:** Surprisingly, loans without verification show the lowest average default rate, which could reflect selection bias in the approval process.

Business Implications

1. **Reassess verification effectiveness:** Investigate why verified loans have higher default rates and refine the verification process to better predict repayment ability.
2. **Focus on source-verified loans:** Loans in the source-verified category strike a balance, offering moderate risk, potentially reflecting more accurate borrower evaluation.

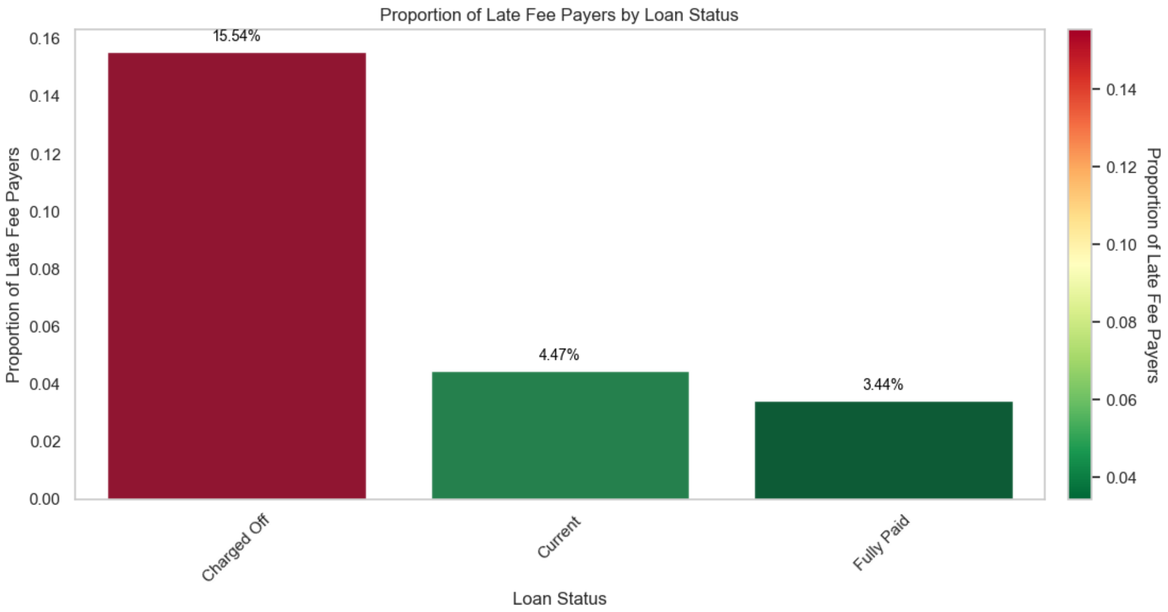
Late Fee Incidence Is Highest Among Charged-Off Loans

Analysis

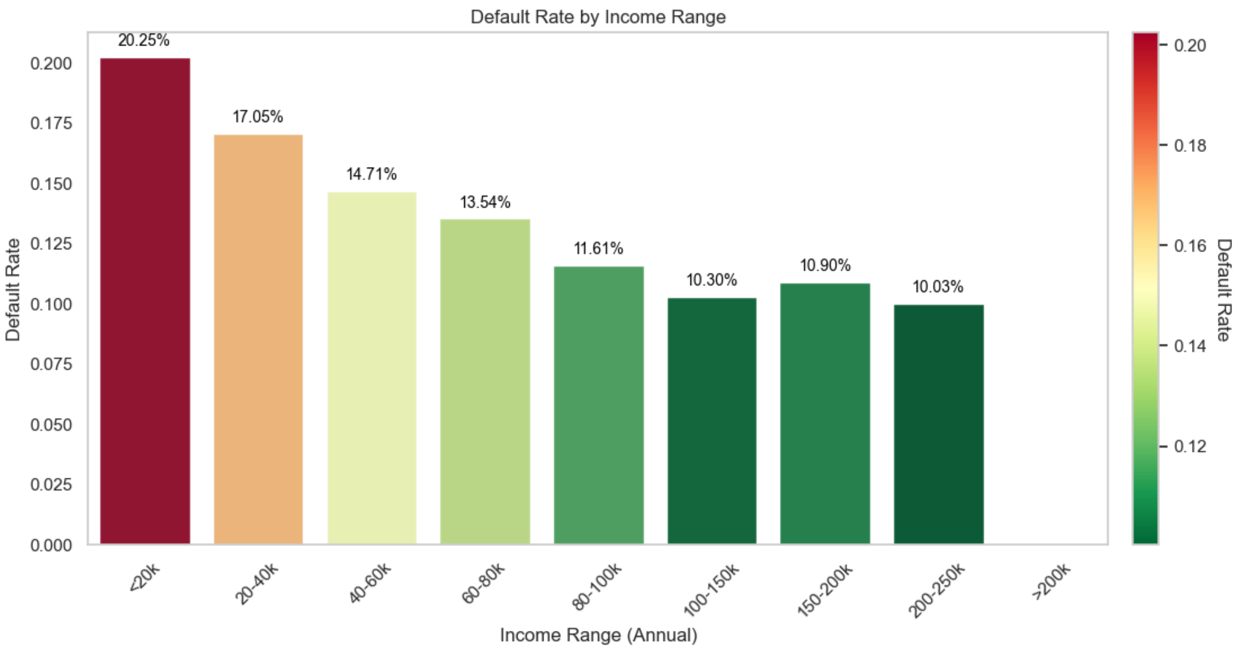
1. **Charged-off loans have the highest late fee proportion:** Borrowers with charged-off loans exhibit a 15.54% incidence of late fee payments, indicating chronic repayment challenges.
2. **Current loans show moderate late fees:** Current loans have a late fee proportion of 4.47%, reflecting occasional repayment delays.
3. **Fully paid loans have the lowest incidence:** Only 3.44% of fully paid loans involve late fees, suggesting better financial discipline among these borrowers.

Business Implications

1. **Late fees as a risk indicator:** High late fee proportions in charged-off loans suggest a strong correlation with default risk; consider incorporating this metric into risk models.
2. **Proactive intervention for current loans:** Address late payments early in the repayment process for current loans to prevent future defaults.



Lower-Income Borrowers Have Higher Default Rates



Analysis

1. **Highest default rate for incomes under $20k:** Borrowers earning less than $20k annually have the highest default rate at 20.25%, indicating significant financial strain in this group.
2. **Default rates decrease with income:** Borrowers in higher income brackets ($80k and above) show lower default rates, with the lowest at 10.03% for those earning over $200k.
3. **Mid-range income groups show moderate risk:** Borrowers earning $40k–$80k exhibit default rates between 13.54% and 14.71%, reflecting a balance of risk.

Business Implications

1. **Stricter evaluation for low-income borrowers:** Enforce stricter credit checks or higher interest rates for borrowers earning under $20k to mitigate elevated default risks.
2. **Focus on high-income borrowers:** Expand offerings for higher-income groups, who represent lower risk and greater repayment reliability.

THANK YOU