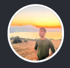
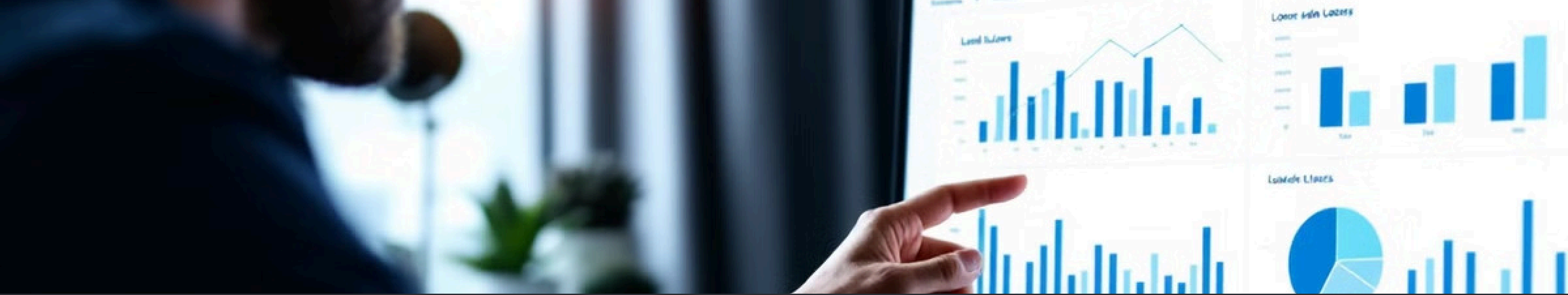
Lending Club Case Study



by Sankalp Sharma



Introduction

This case study analyses Lending Club's loan dataset to help a financial company identify potential defaulters and minimise lending risks. Using historical loan data, we aim to understand patterns that lead to loan defaults.

1. Analyze borrower characteristics and loan features
2. Identify patterns leading to defaults
3. Help lenders make better loan approval decisions
4. Reduce financial risk through data-driven insights

The analysis includes borrower income, credit history, loan purpose, and other relevant factors. By understanding these patterns, the company can improve its lending decisions, reduce default rates, and enhance profitability.

Problem Statement

Understanding the Loan Default Problem

This project helps a consumer finance company identify borrowers likely to default on their loans, a costly issue.

The Problem: Loan Risks

1. Missed Good Borrowers: Denying loans to those who would repay results in lost opportunities.
2. Bad Borrowers: Approving loans for people who won't repay leads to financial losses.

The Goal: Identifying Risky Borrowers

Using past loan data, we aim to find patterns that predict which borrowers are more likely to default.

Types of Loan Outcomes

- Fully Paid
- Current
- Charged Off
- Loan Rejected

Data Understanding

The "loan_status" column is vital for understanding loan outcomes, with three main categories:

1 Fully Paid

Borrowers who repaid the loan in full.

2 Charged Off

Borrowers who defaulted on the loan.

3 Current

Borrowers still repaying (excluded from this analysis).

Key Consumer Finance Attributes for Loan Risk Assessment

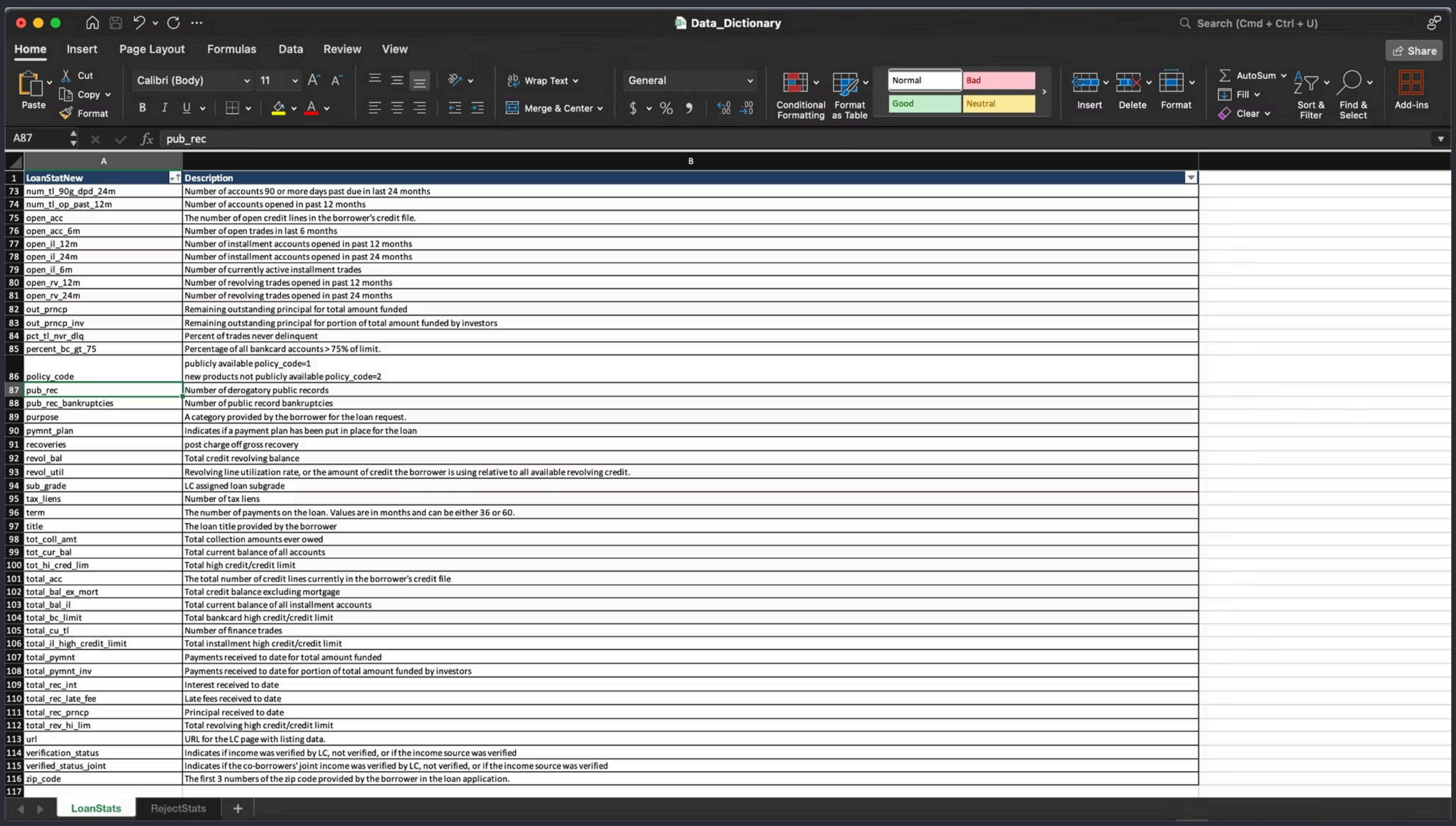
1. Earliest Credit Line (earliest_cr_line): Longer credit history may indicate better financial management.
2. Public Record Bankruptcies (pub_rec_bankruptcies): More bankruptcies suggest poorer creditworthiness.
3. Income Verification (verification_status): Verified income reduces risk.
4. Revolving Balance (revol_bal) & Utilization (revol_util): High balances and utilization may indicate financial strain.
5. Credit Accounts (total_acc): A balanced number of credit accounts suggests stable financial behavior.
6. Annual Income (annual_inc): Higher incomes increase repayment capacity.
7. Home Ownership (home_ownership): Owning a home can serve as collateral.
8. Employment Length (emp_length): Longer employment indicates stability.
9. Debt-to-Income Ratio (dti): Lower DTI suggests better debt management.

Loan Characteristics:

1. Loan Amount (loan_amnt): Larger loans may carry higher risk.
2. Funded Amount (funded_amnt): Indicates lender confidence.
3. Monthly Installment (installment): Larger installments can strain finances.
4. Loan Grade (grade): Higher grades indicate lower risk.
5. Loan Term (term): Longer terms can increase default risk.
6. Interest Rate (int_rate): Higher rates often indicate higher-risk borrowers.
7. Loan Purpose (purpose): Some loan purposes, like debt consolidation, may be riskier.

Exclusions from Analysis

- Customer Behavior Data: Post-loan behaviors are not relevant for initial loan approval.
- Granular Data: Detailed columns like sub_grade are excluded in favor of broader metrics like grade.



	LoanStatNew	Description
73	num_tl_90g_dpd_24m	Number of accounts 90 or more days past due in last 24 months
74	num_tl_op_past_12m	Number of accounts opened in past 12 months
75	open_acc	The number of open credit lines in the borrower's credit file.
76	open_acc_6m	Number of open trades in last 6 months
77	open_il_12m	Number of installment accounts opened in past 12 months
78	open_il_24m	Number of installment accounts opened in past 24 months
79	open_il_6m	Number of currently active installment trades
80	open_rv_12m	Number of revolving trades opened in past 12 months
81	open_rv_24m	Number of revolving trades opened in past 24 months
82	out_prncp	Remaining outstanding principal for total amount funded
83	out_prncp_inv	Remaining outstanding principal for portion of total amount funded by investors
84	pct_tl_nvr_dlq	Percent of trades never delinquent
85	percent_bc_gt_75	Percentage of all bankcard accounts > 75% of limit.
86	policy_code	publicly available policy_code=1 new products not publicly available policy_code=2
87	pub_rec	Number of derogatory public records
88	pub_rec_bankruptcies	Number of public record bankruptcies
89	purpose	A category provided by the borrower for the loan request.
90	pymnt_plan	Indicates if a payment plan has been put in place for the loan
91	recoveries	post charge off gross recovery
92	revol_bal	Total credit revolving balance
93	revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
94	sub_grade	LC assigned loan subgrade
95	tax_liens	Number of tax liens
96	term	The number of payments on the loan. Values are in months and can be either 36 or 60.
97	title	The loan title provided by the borrower
98	tot_coll_amt	Total collection amounts ever owed
99	tot_cur_bal	Total current balance of all accounts
100	tot_hi_cred_lim	Total high credit/credit limit
101	total_acc	The total number of credit lines currently in the borrower's credit file
102	total_bal_ex_mort	Total credit balance excluding mortgage
103	total_bal_il	Total current balance of all installment accounts
104	total_bc_limit	Total bankcard high credit/credit limit
105	total_cu_tl	Number of finance trades
106	total_il_high_credit_limit	Total installment high credit/credit limit
107	total_pymnt	Payments received to date for total amount funded
108	total_pymnt_inv	Payments received to date for portion of total amount funded by investors
109	total_rec_int	Interest received to date
110	total_rec_late_fee	Late fees received to date
111	total_rec_prncp	Principal received to date
112	total_rev_hi_lim	Total revolving high credit/credit limit
113	url	URL for the LC page with listing data.
114	verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified
115	verified_status_joint	Indicates if the co-borrowers' joint income was verified by LC, not verified, or if the income source was verified
116	zip_code	The first 3 numbers of the zip code provided by the borrower in the loan application.

Dataset Analysis: Detailed Preprocessing and Feature Engineering

Row and Column Analysis

- Rows:
  - "Current" loan status rows removed (ongoing loans).
  - Duplicate rows removed.
- Columns Removed:
  - 54 columns with only NA values.
  - 9 columns with zero variance, such as pymnt_plan, initial_list_status, acc_now_delinq.
  - Columns with >30% missing data (e.g., mths_since_last_delinq, desc).
  - Unique identifiers (id, member_id) and irrelevant textual data (e.g., emp_title, url) removed.
  - Post-loan behavior columns removed (e.g., total_pymnt_inv, recoveries).

Data Type Conversions

- Numeric conversions: loan_amnt, funded_amnt, int_rate, installment, dti.
- String cleaning: Removed "%" in int_rate and revol_util, converted to float.
- Dates converted to datetime format: issue_d, earliest_cr_line.
- emp_length converted to numeric values.

Value Standardization

- Standardized values:
  - home_ownership: "NONE", "OTHER" → "OTHER".
  - verification_status: "Source Verified", "Verified" → "Verified".

New Columns Added

- From issue_d: issue_y (year), issue_m (month), issue_q (quarter).
- From earliest_cr_line: early_cr_line_month, early_cr_line_year.
- From loan_status: loan_paid (Yes/No).

Bucketed Columns

Created buckets for:

- loan_amount_bucket, annual_income_bucket, dti_bucket, revol_util_bucket, int_rate_bucket, and more.

Derived Columns

1. Combined Debt Burden Score: Summarizes financial stress, highlighting borrowers at risk of default.
2. Loan DTI Interaction: Captures the relationship between loan amount and debt-to-income ratio, identifying high-risk borrowers.
3. Credit History Years: Measures the length of the borrower's credit history to assess creditworthiness.
4. Payment to Income: Evaluates the borrower's financial burden by comparing monthly payments to income.

Handling Missing Data

- Columns with >30% missing values removed.
- Imputation for remaining columns with lower missing values.

Outlier Handling

Outliers removed from:

- loan_amnt, funded_amnt, int_rate, annual_inc, dti, revol_util.

Analyzing Data

Analyzing Data

1. Loading the CSV File: The dataset was imported using pandas.read_csv() to begin the analysis.
2. Inspecting Rows and Columns: We checked the dataset's shape to understand the number of rows and columns.
3. Reviewing Dataset Information: Using .info(), we examined column data types, missing values, and memory usage.
4. Previewing the Data: The head() function was used to view the first five rows, checking for errors, format inconsistencies, and data structure.
5. Identifying Issues: We identified any missing data, outliers, and format inconsistencies for further cleaning.

Data Cleaning

Handling Missing Data

I identified and handled missing data by using the isna() function, converting it to percentages, and dropping columns with excessive missing values. The names of these columns were printed, and the dataset's shape was updated accordingly.

Removing Unnecessary and Redundant Columns

I removed irrelevant columns (e.g., loan sanctions, unique identifiers) and redundant ones (duplicates or variations). This streamlined the dataset, focusing only on essential columns for analysis, improving clarity and efficiency.

Standardizing Columns and Removing Unnecessary Rows

I standardized column names, removed duplicate rows, and dropped rows with over 30% missing values. I also excluded records with a "current" loan status, focusing on completed loans relevant to the default analysis.

Data Conversion

I cleaned and converted several columns for analysis:

- Removed '%' symbols in int_rate and revol_util, converting them to float for numerical analysis.
- Converted the term column (loan duration) from string format to an integer (term_months).
- Transformed date columns (issue_d, earliest_cr_line) into datetime format for easier manipulation.

Imputing Values

I imputed missing values for emp_length with the mode value and converted it to integers. For other columns like home_ownership and verification_status, imputed values were filled, while rows with missing pub_rec_bankruptcies were removed due to non-imputability.

Common Functions

To automate analysis, I created functions like plot_bar_graph and plot_histogram for visualizations, and calculate_charged_off_percentage to compute key metrics. These functions streamlined repetitive tasks and improved efficiency in analyzing loan risk factors.

Handling Outliers

I removed outliers for key variables based on typical ranges:

- Loan amounts: $5,000 - $15,000
- Interest rates: 9% - 14%
- Annual income: $40,000 - $80,000
- Debt-to-income ratio: 8% - 19%
- Funded amount: $5,000 - $14,000
- Revolving utilization: 25% - 75%

These insights help identify potential default risks, such as high loan amounts or DTIs, which may indicate financial strain.

Derived Columns

I created new derived columns, such as loan_dti_interaction (loan amount vs. DTI ratio) and payment_to_income (monthly payment vs. income), along with datetime-based features. I also bucketed numerical variables for easier analysis.

EDA Analysis

Univariate Analysis

Univariate analysis focuses on examining individual variables to understand their distribution and characteristics. This includes analyzing and visualizing the data using histograms, box plots, or bar charts.

Bivariate Analysis

Bivariate analysis explores the relationship between two variables. It helps understand how one variable impacts or correlates with another. This analysis is useful for identifying the influences loan approval or default risk.

Segmented Analysis

Segmented analysis divides the dataset into subgroups based on specific criteria of loan status. By comparing these segments, we can identify differences in behavior or risk factors across different groups.

Correlation Analysis

Correlation analysis identifies the strength and direction of relationships between variables. Using correlation matrices or heatmaps, we can identify which variables are strongly correlated, such as income and loan amount, or debt-to-income ratio and default risk.

Multivariate Analysis

Multivariate analysis examines the interaction between multiple variables simultaneously. Techniques like clustering allow us to understand how several factors jointly influence the target variable (e.g., loan default).

Univariate Analysis

Categorical Variables

Ordered Categorical

- grade (A, B, C, etc.)
- emp_length
- term_months (36/60 months)
- issue_year (2011, etc.)
- issue_month (1-12)
- issue_quarter (Q1, Q2, Q3, Q4)
- early_cr_line_month
- early_cr_line_year

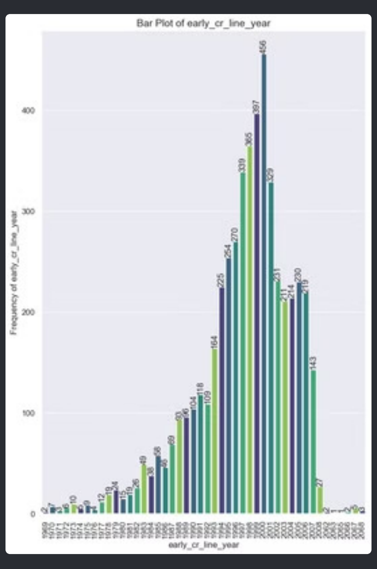
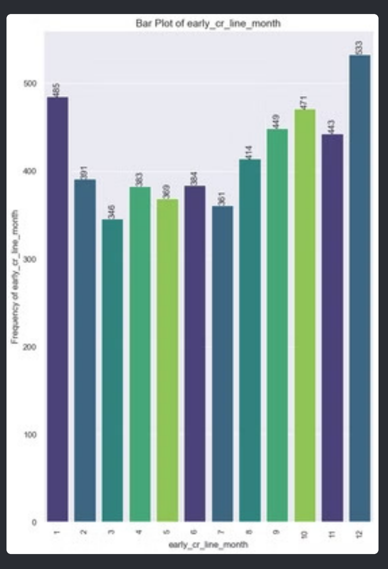
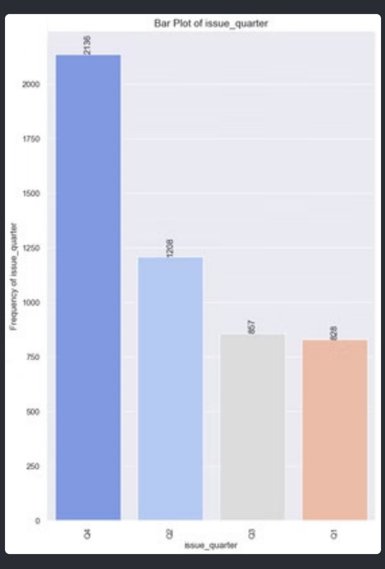
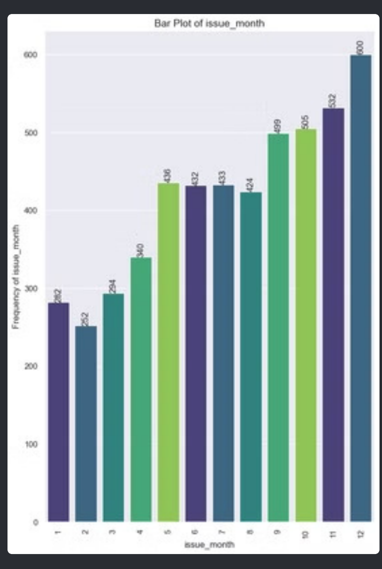
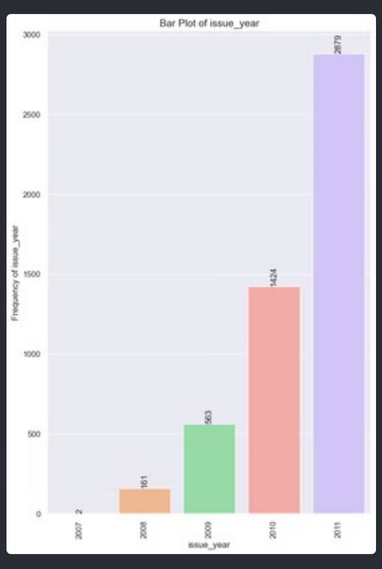
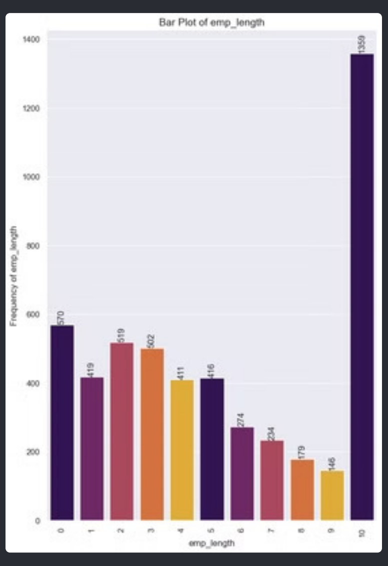
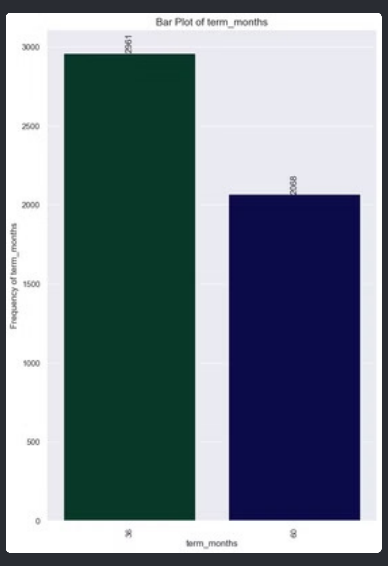
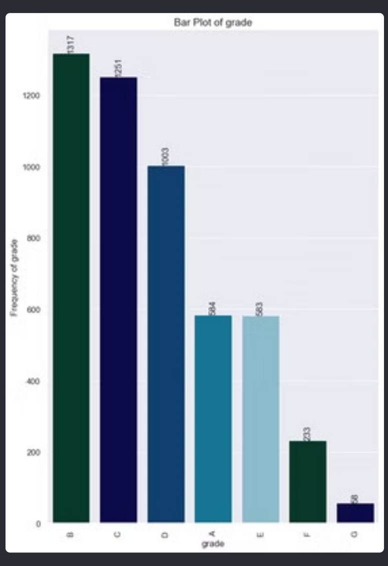
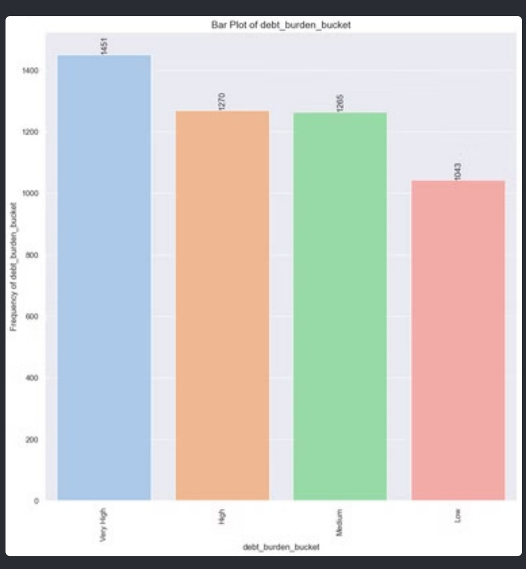
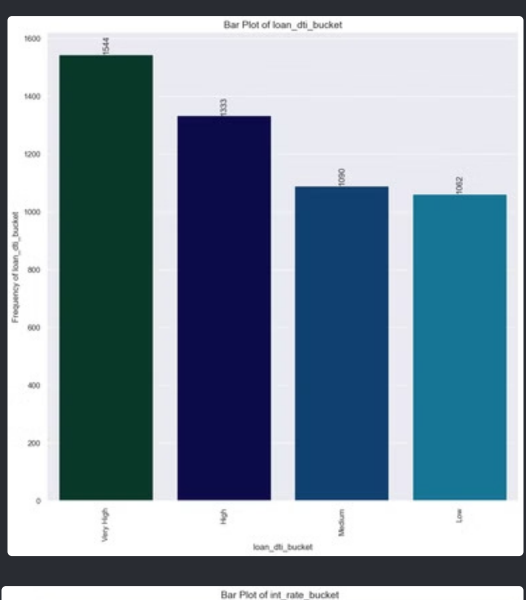
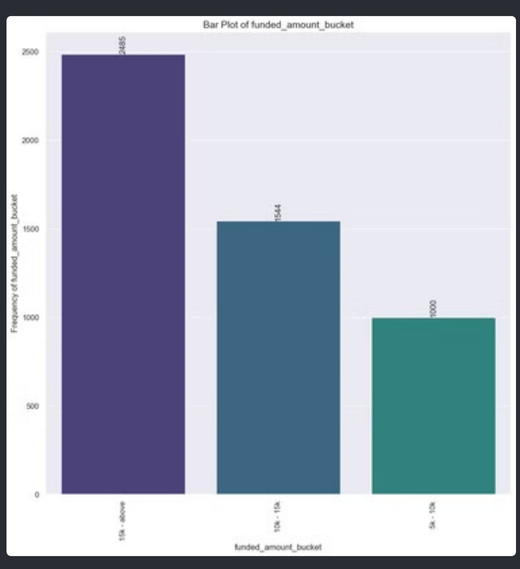
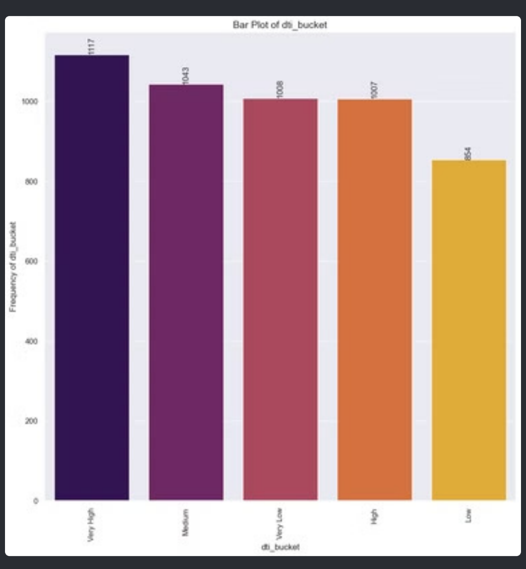
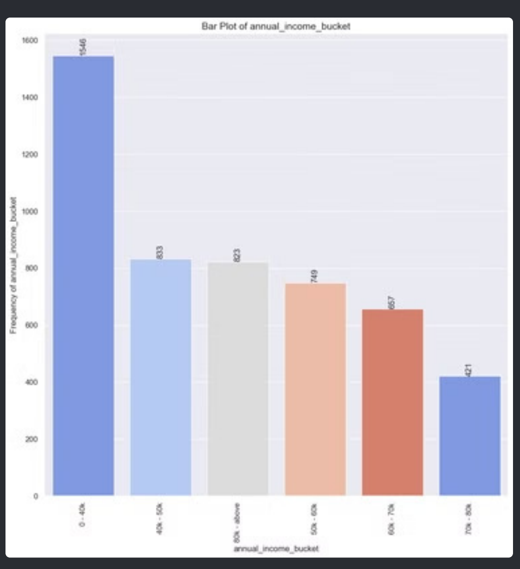
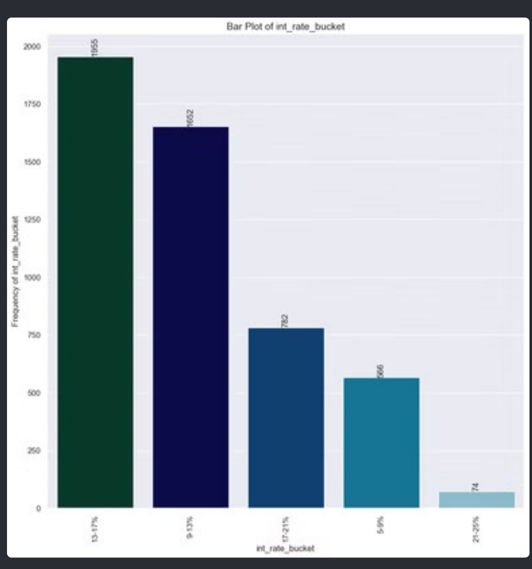
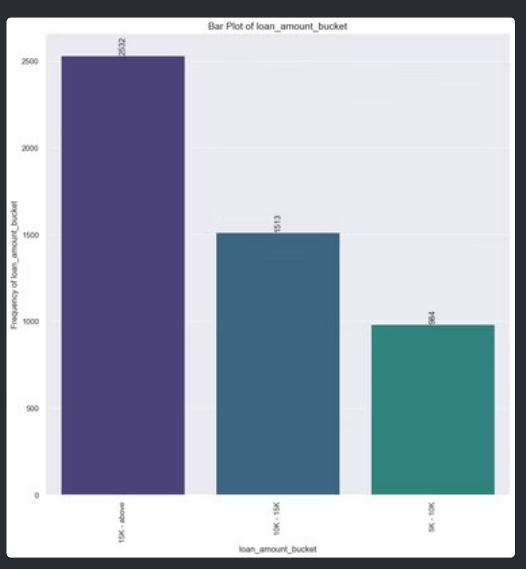
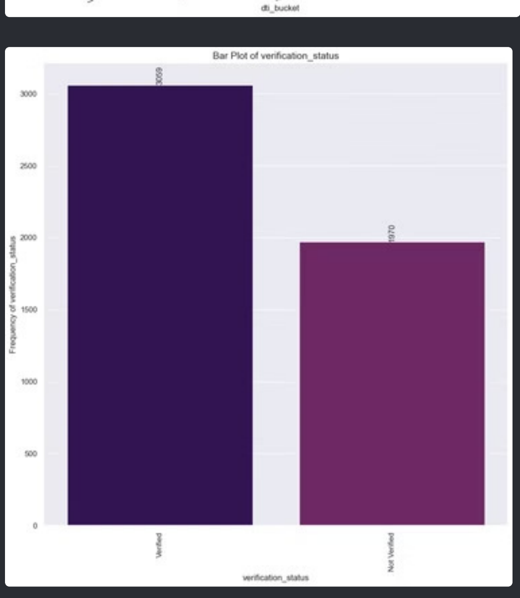
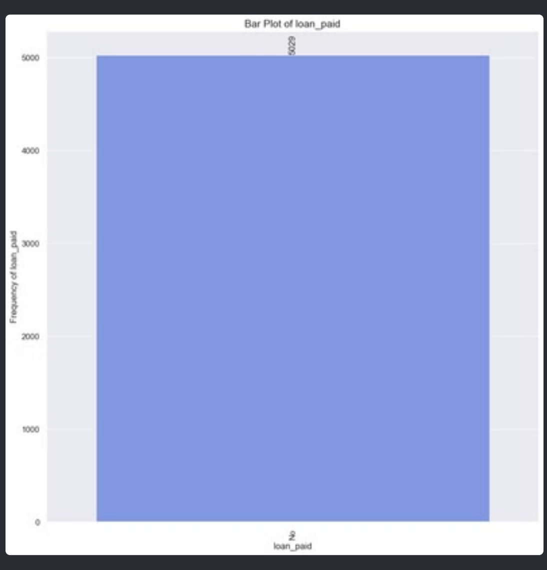
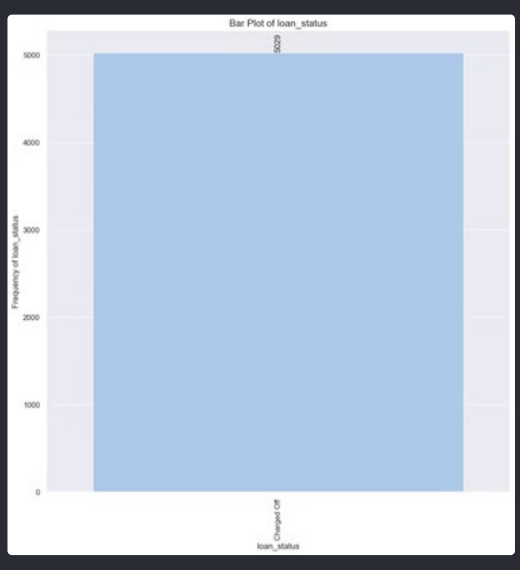
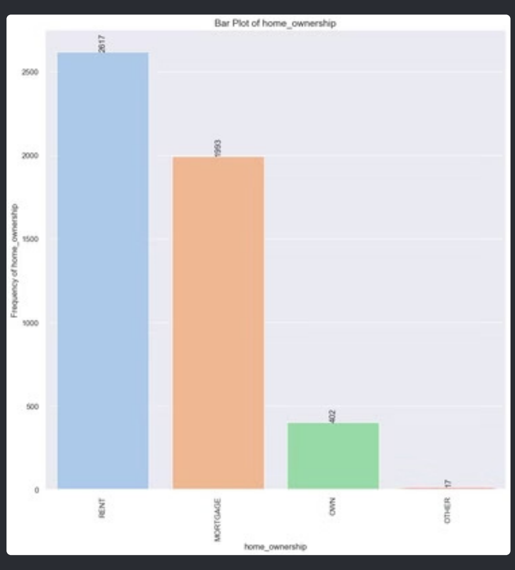
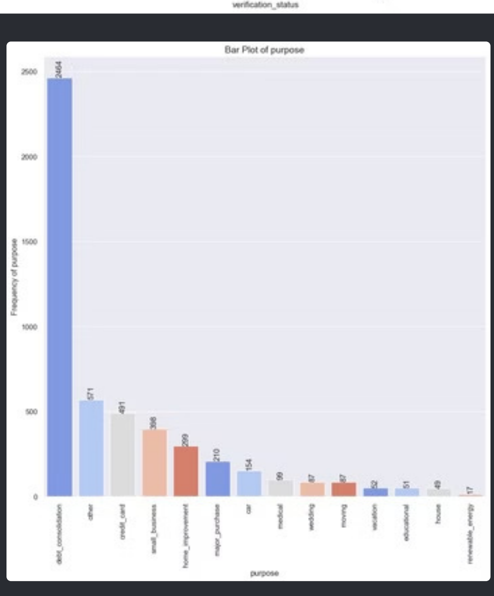
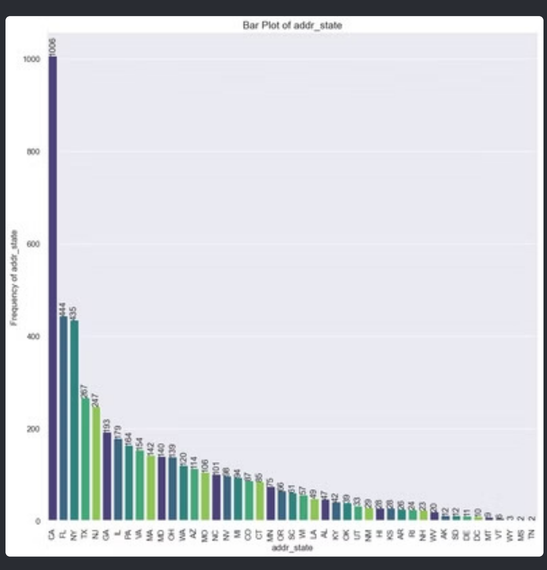
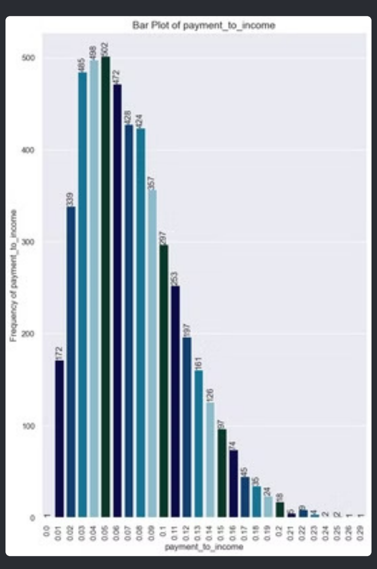
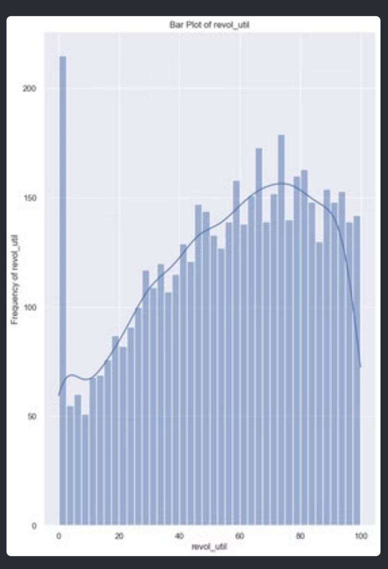
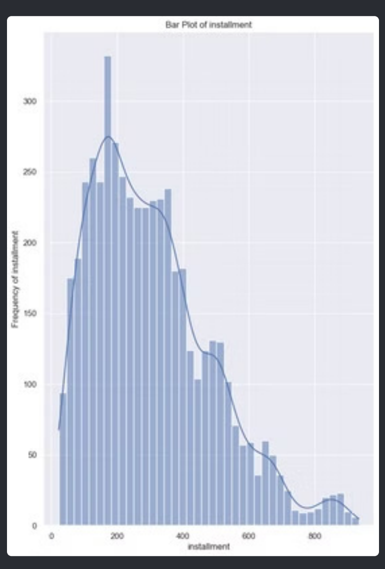
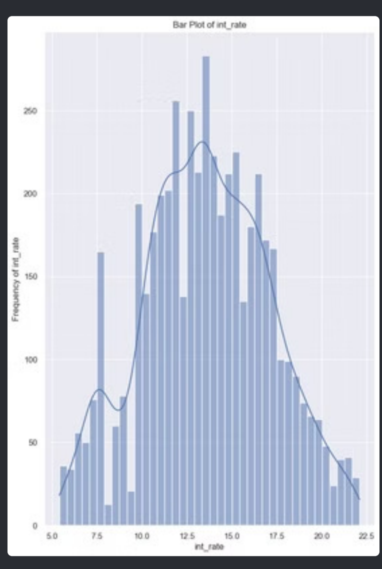
Unordered Categorical

- addr_state (AZ, GA, IL, etc.)
- home_ownership (RENT, OWN)
- loan_status (Fully Paid, Charged Off)
- purpose (credit_card, car, etc.)
- verification_status (Verified, Not Verified)
- loan_paid (Yes, No)
- loan_amount_bucket
- int_rate_bucket
- annual_income_bucket
- dti_bucket
- funded_amnt_bucket
- loan_dti_bucket
- debt_burden_bucket

Quantitative Variables

Continuous Variables

- installment
- payment_to_income
- revol_util

Key Insights

1. Categorical Variables

- Grade (LC Assigned Loan Grade):
  - Finding: Grade 'B' had the most defaults.
  - Reason: Mid-tier grades show higher default risk.
- Employment Length (emp_length):
  - Finding: 10 years was the most common length among defaulters.
  - Reason: Long employment doesn't guarantee repayment ability.
- Loan Term (term_months):
  - Finding: Most defaulters had 36-month loans.
  - Reason: Shorter terms can strain financial capacity due to higher payments.
- Issue Year (issue_year):
  - Finding: 2011 had the highest defaults.
  - Reason: Economic downturns led to increased defaults.
- Home Ownership (home_ownership):
  - Finding: Renters had the highest defaults.
  - Reason: Renters may face more financial instability.
- Loan Purpose (purpose):
  - Finding: Debt consolidation loans had the highest defaults.
  - Reason: Financial distress makes these borrowers higher risk.
- Verification Status (verification_status):
  - Finding: Most defaulters were verified.
  - Reason: Verification alone doesn't ensure repayment.

2. Quantitative Variables

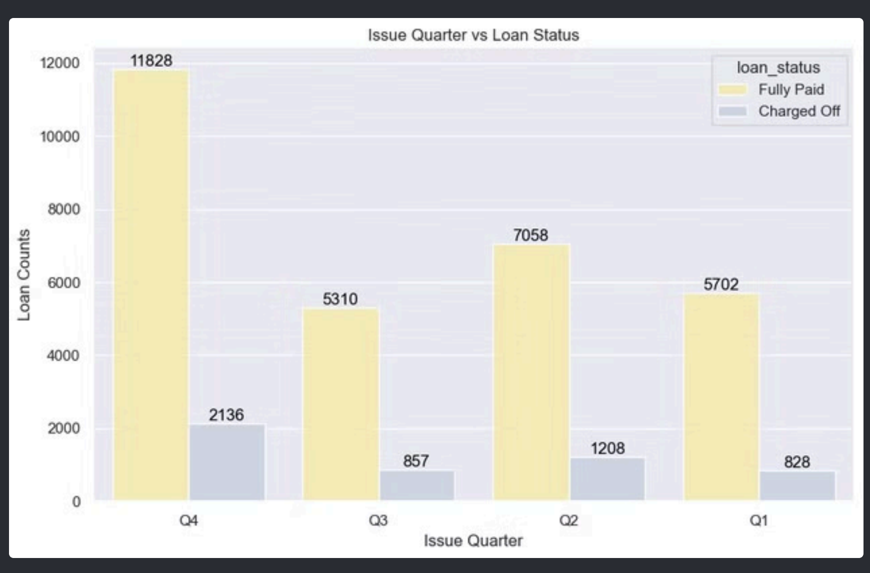
- Interest Rate (int_rate):
  - Finding: Defaults were highest for rates between 9-17%.
  - Reason: Higher rates increase borrowing costs and repayment challenges.
- Installment:
  - Finding: Most defaulters had monthly installments between 150-450 USD.
  - Reason: Higher payments strain borrowers, especially with low incomes.
- Payment-to-Income Ratio (payment_to_income):
  - Finding: Higher default rates for ratios above 5%.
  - Reason: A higher ratio indicates financial stress, increasing default risk.
- Revolving Utilization (revol_util):
  - Finding: Defaults were highest for utilization between 60-90%.

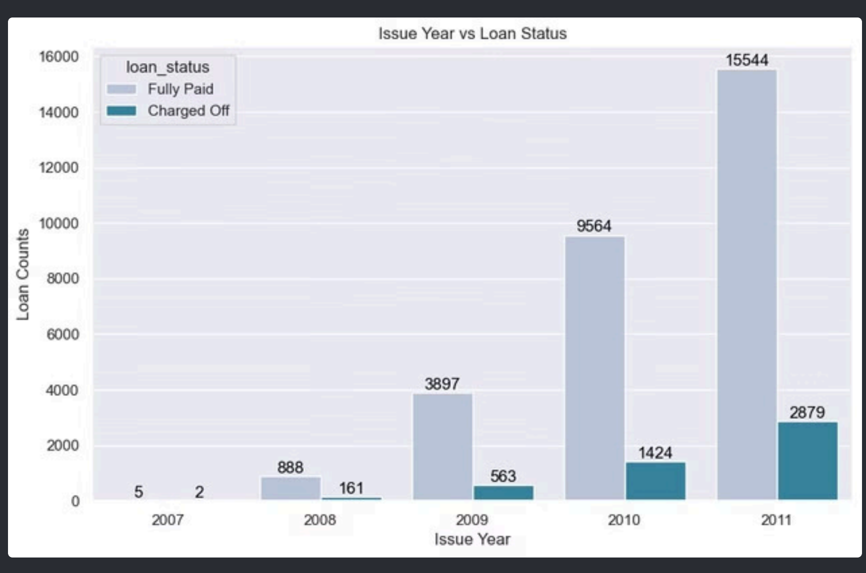
Key Insights & Recommendations

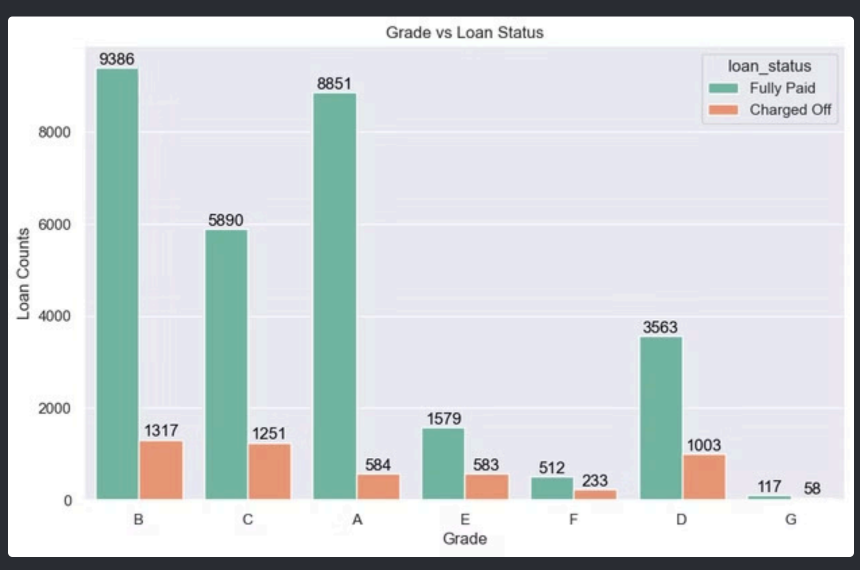
- Grade & Credit Risk: Lower grades = higher defaults. Stricter assessments for these applicants are needed.
- Loan Amount & Repayment: Larger loans and higher interest rates = higher default risk. Assess repayment capacity thoroughly.
- Income & Financial Stability: Lower-income borrowers are more likely to default. Stricter income assessments can help reduce risk.
- DTI & Credit Utilization: High DTI and revolving utilization predict defaults. Prioritize

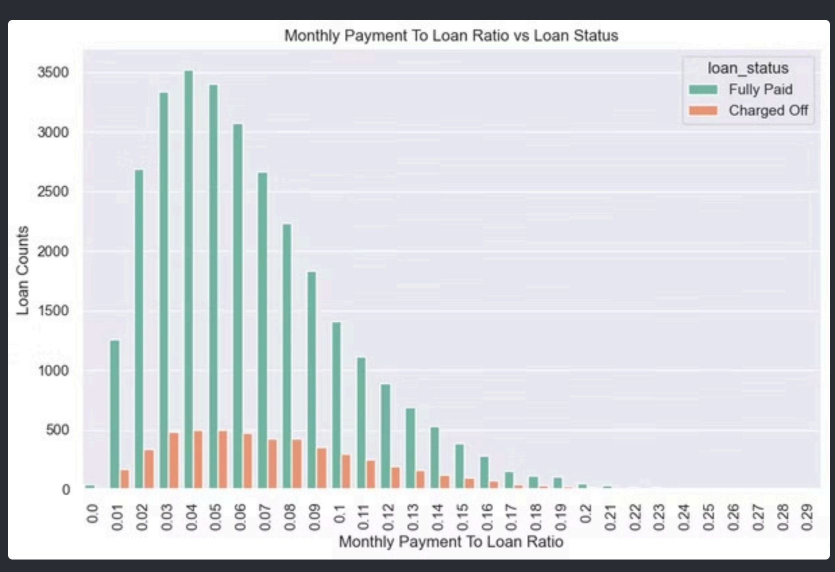
Bivariate Analysis

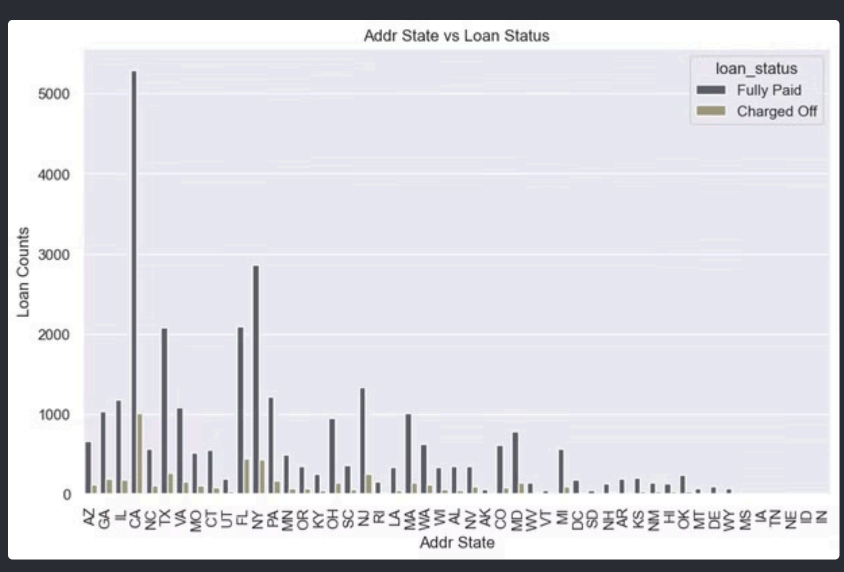
Variable Pair	Type
Ordered Variables	
grade vs. loan_status	Categorical
term_months vs. loan_status	Categorical
issue_year vs. loan_status	Categorical
issue_quarter vs. loan_status	Categorical
emp_length vs loan_status	Categorical
Unordered Variables	
home_ownership vs. loan_status	Categorical
purpose vs. loan_status	Categorical
addr_state vs. loan_status	Categorical
verification_status vs. loan_status	Categorical
Quantitative Variables	
payment_to_income vs. loan_status	Numerical

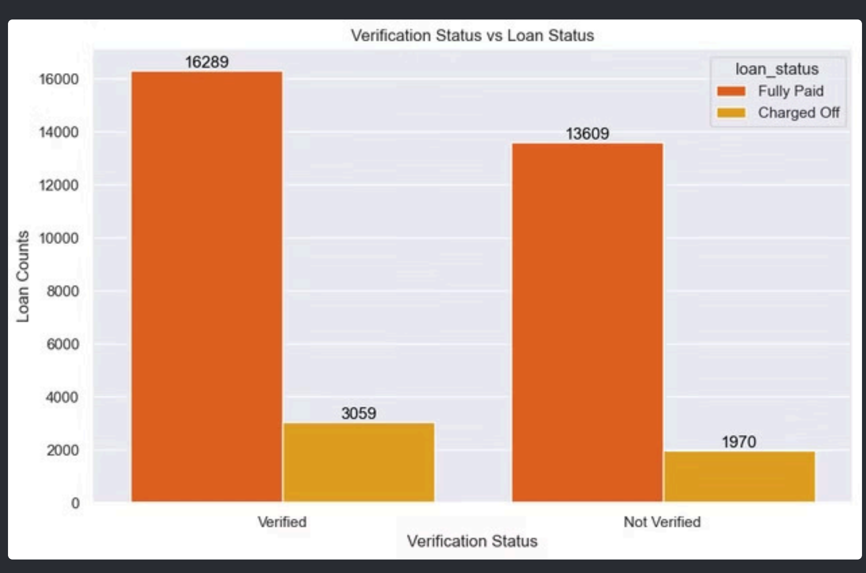


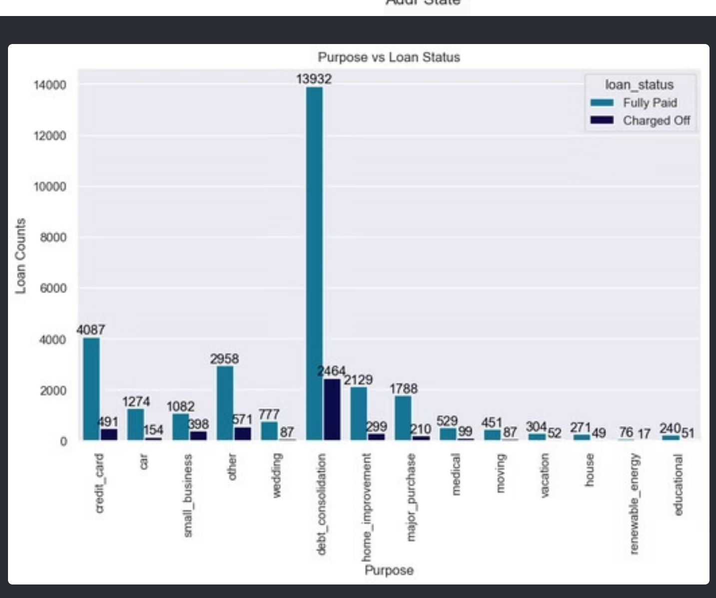


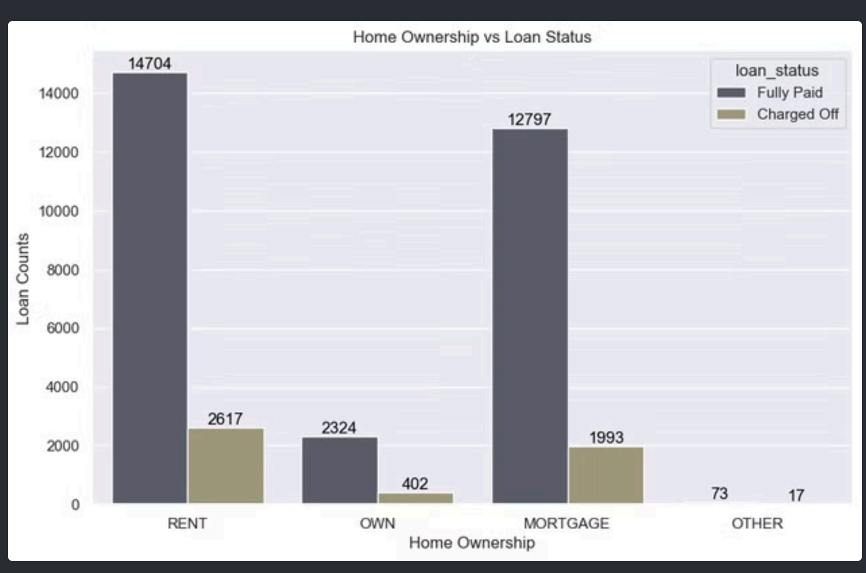












Key Findings & Inference

1. Grade vs. Loan Status
   - Grades B & C have the highest defaults (1,317 & 1,251). - Grade A shows fewer defaults.
   - Poor grades (B, C, D, E) are high-risk. - Stricter approval needed for these grades.
2. Term Month vs. Loan Status
   - 36-month loans have lower default rates. - 60-month loans have higher default rates.
   - Long-term loans (60 months) are riskier. - Focus on 36-month loans or stricter screening for 60-month loans.
3. Issue Year vs. Loan Status
   - Defaults peaked in 2011 (2,879). - Earlier years had lower defaults.
   - Recent loans (2010-2011) have higher defaults. - Update risk models with recent data.
4. Issue Quarter vs. Loan Status
   - Q4 has the highest defaults (2,136). - Q1 has the lowest defaults.
   - Seasonality impacts defaults in Q4. - Tighten lending policies during Q4.
5. Employment Length vs. Loan Status
   - 10 years employment had the highest defaults (1,359). - Shorter tenures showed moderate defaults.
   - Long-tenured employees (10 years) may over-leverage income. - Scrutiny for high-loan applicants with long tenure.

1. Home Ownership vs. Loan Status
   - Renters show the highest defaults (2,617). - Homeowners show lower defaults.
   - Renters are higher risk. - Offer smaller loans or stricter screening for renters.
2. Purpose vs. Loan Status
   - Debt consolidation loans had the highest defaults (2,464). - Small loans like wedding had lower defaults.
   - Debt consolidation loans are riskier. - Cap loan amounts for these purposes.
3. Verification Status vs. Loan Status
   - Verified loans show more defaults (3,059).
   - Verification alone is not a reliable predictor. - Strengthen screening for all loan types.
4. State Address vs. Loan Status
   - California, Florida, and New York had high defaults.
   - Region-based risk models needed. - Adjust interest rates based on state risks.
5. Monthly Payment to Loan Ratio vs. Loan Status
   - Defaulters mostly have smaller ratios (<8%). - Higher ratios (>10%) correlate with fewer defaults.
   - Borrowers with smaller payment ratios are at higher risk. - Enforce minimum repayment percentages.

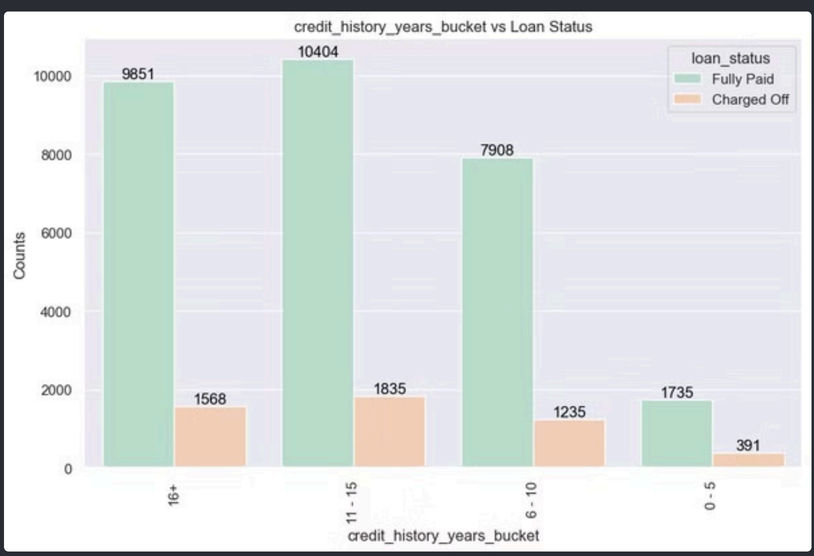
Key Takeaways for Risk Mitigation

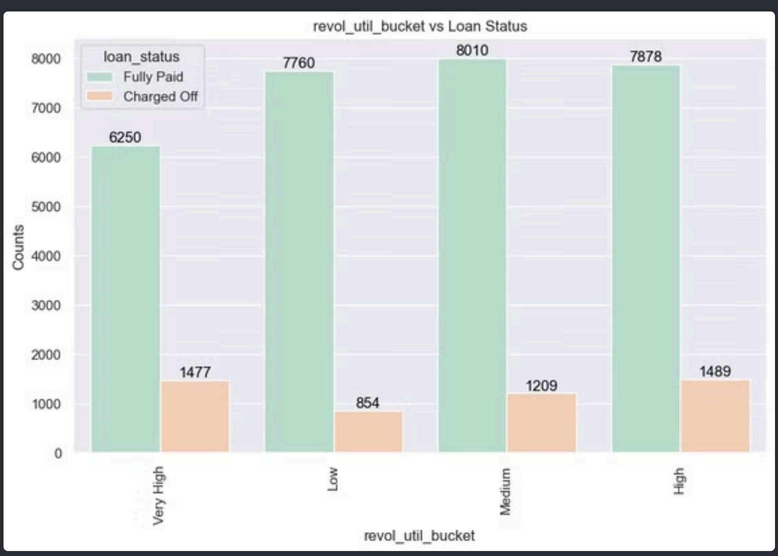
1. Loan Conditions:
   - Limit high-risk long-term loans (60 months) and promote shorter terms (36 months).
   - Cap loan amounts for debt consolidation and poor credit grades (C, D).
2. Risk-Based Pricing and Screening:
   - Apply higher interest rates for high-risk groups (renters, high DTI).
   - Implement stricter screening for high-risk states (e.g., California, Florida).

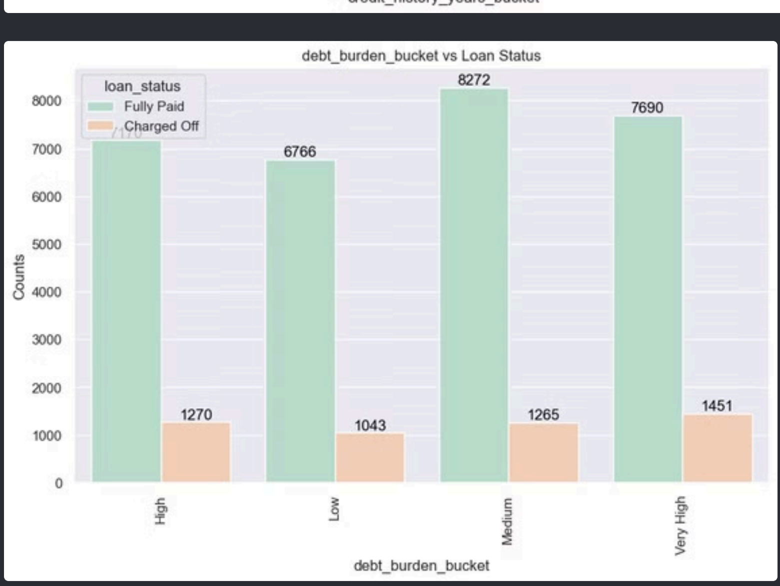
1. Targeted Support:
   - Offer counseling for high-risk borrowers (Grades C, D, E, G).
   - Propose smaller loans or flexible terms for renters and long-tenure employees (10 years).
2. Seasonal Adjustments:
   - Tighten approval criteria during Q4 (holiday season).
   - Assess year-end spending habits for default risks.

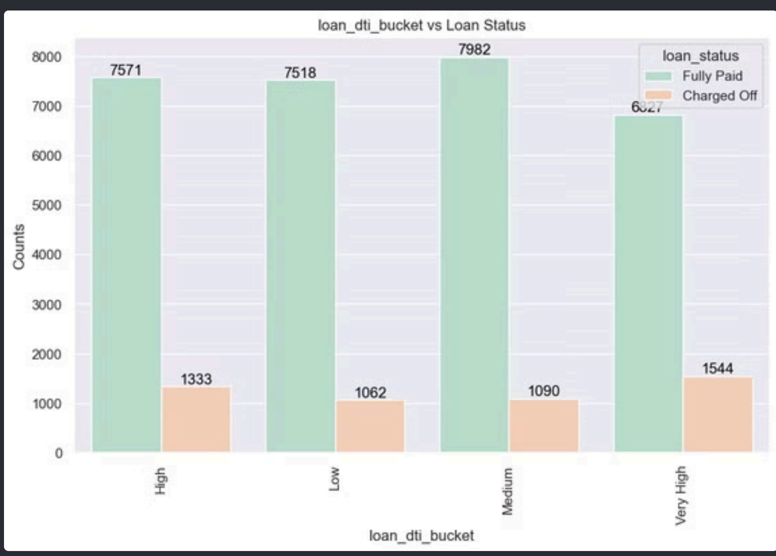
Segmentation Analysis

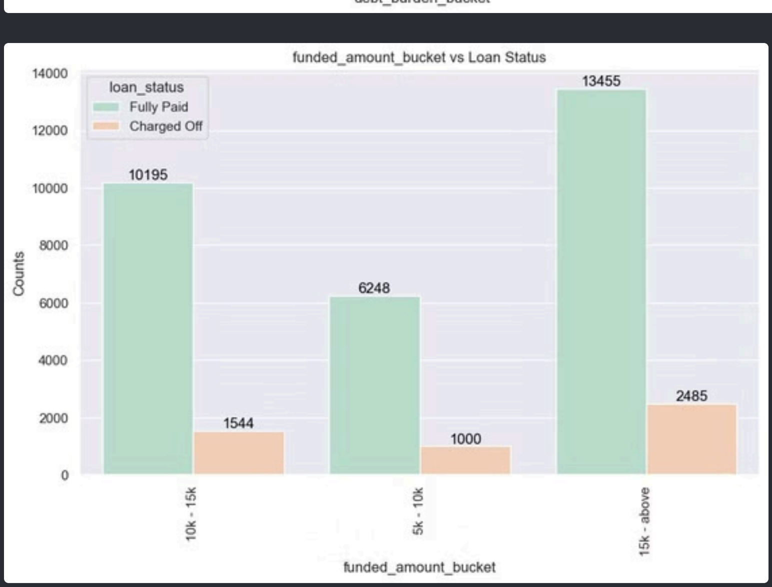
1. loan_amount_bucket vs. loan_status
2. int_rate_bucket vs. loan_status
3. annual_income_bucket vs. loan_status
4. total_account_bucket vs. loan_status
5. funded_amnt_bucket vs. loan_status
6. loan_dti_bucket vs. loan_status
7. debt_burden_bucket vs. loan_status
8. revol_util_bucket vs. loan_status
9. credit_history_years_bucket vs. loan_status

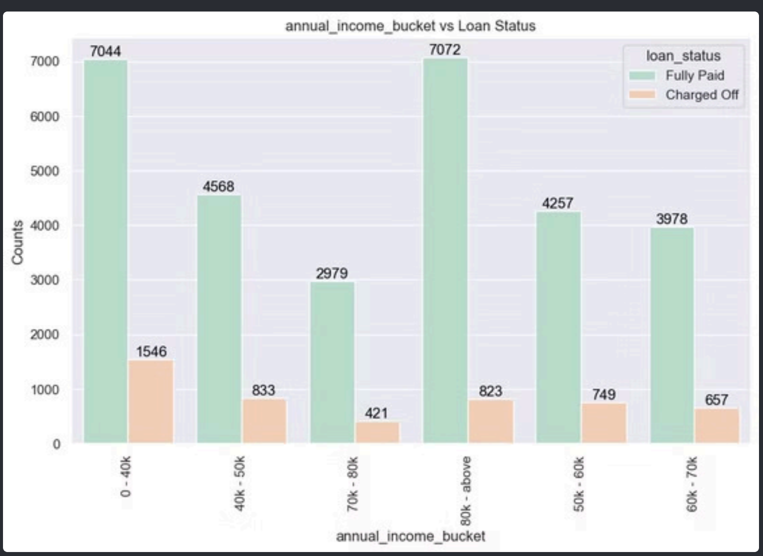


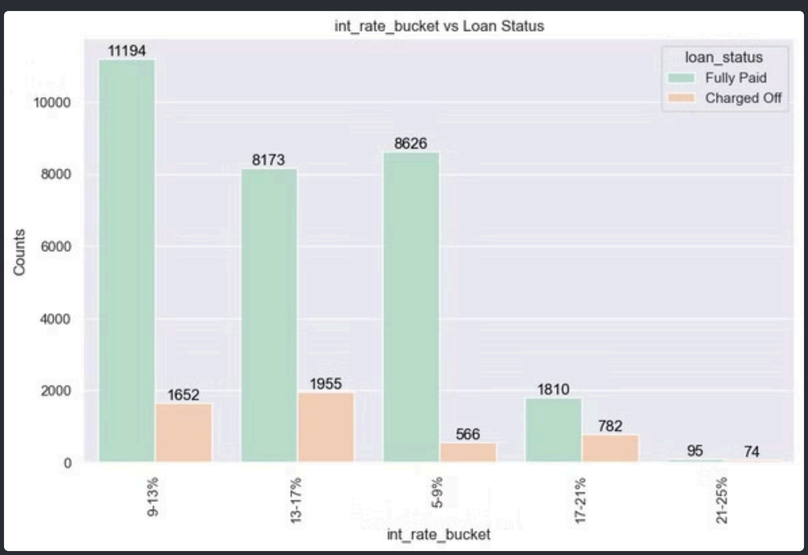


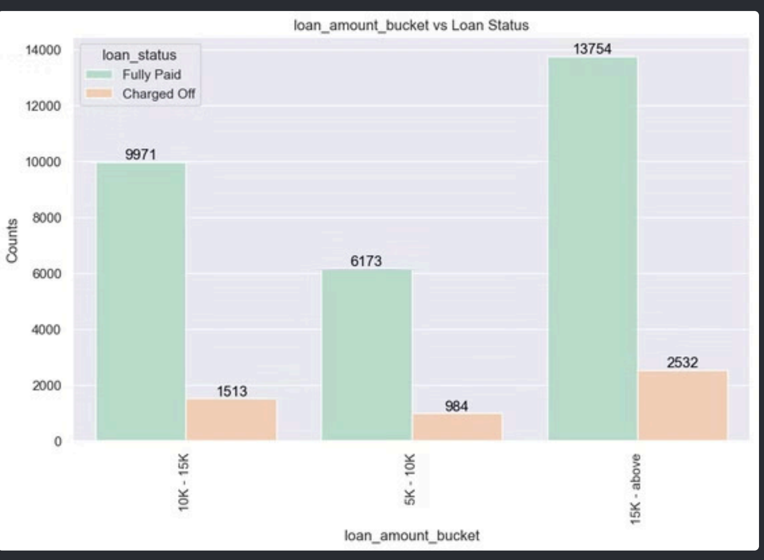


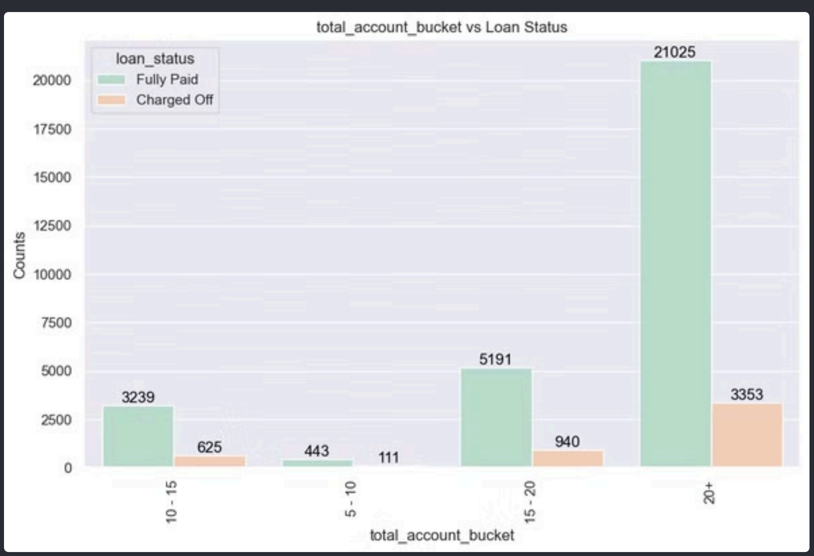












Key Insights

Feature	Insight	Action
Loan Amount Bucket	Larger loans (15K+) have higher default risks.	Tighten eligibility for larger loans.
Interest Rate Bucket	Higher interest rates (13%+) correlate with higher defaults.	Offer manageable interest rates or smaller loans for riskier applicants.
Annual Income Bucket	Lower-income borrowers (0-40K) show higher default rates.	Provide smaller loan amounts and flexible repayment plans.
Funded Amount Bucket	Larger funded amounts (15K+) correlate with more defaults.	Apply stricter underwriting for high-funded loans.
DTI (Debt-to-Income) Bucket	High DTI (>43%) borrowers show the highest defaults.	Implement cautious lending for high DTI groups.
Debt Burden Bucket	Higher debt burdens increase default risk.	Strengthen filters for high debt burden borrowers.
Revolving Utilization Bucket	High credit utilization increases default risk.	Enforce stricter repayment policies or limit further borrowing.
Credit History Years Bucket	Borrowers with 11-15 years of history show higher defaults.	Offer tailored support for borrowers with intermediate credit histories.

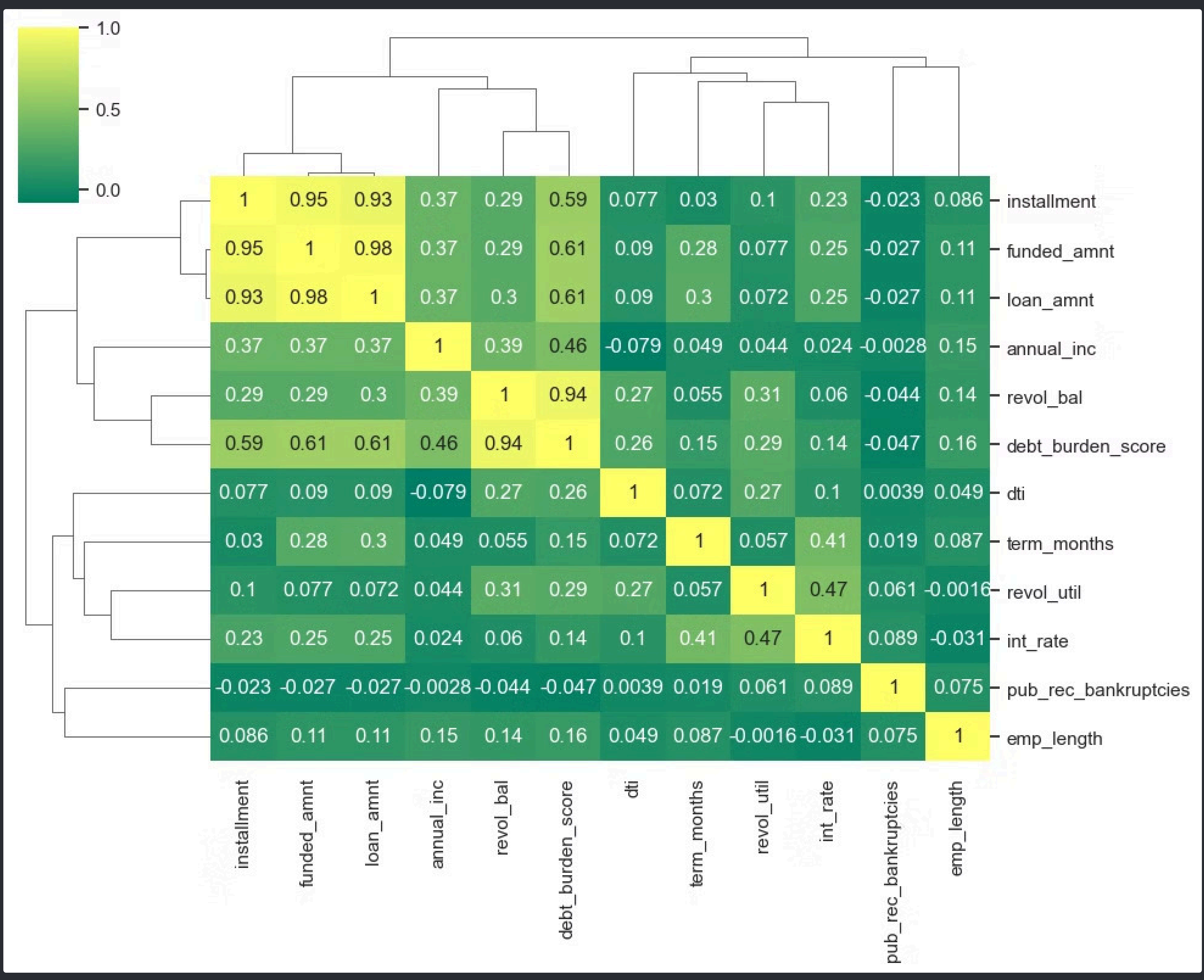
Key Recommendations:

- Focus on reducing defaults by targeting low-income, high-DTI, and high-utilization borrowers with smaller loan amounts and stricter policies.
- Offer more favorable terms to low-DTI, low-utilization borrowers.
- Tailor support for borrowers with intermediate credit histories (11-15 years).

Correlation Analysis

Examine relationships between numerical variables:

1. funded_amnt
2. loan_amnt
3. pub_rec_bankruptcies
4. revol_bal
5. revol_util
6. debt_burden_score
7. emp_length
8. int_rate
9. dti
10. annual_inc
11. term_months
12. installment



Key Insights from Correlation Analysis

Strong Positive Correlations (> 0.7):

1. installment, funded_amnt, loan_amnt: Larger loans typically result in higher installments and funded amounts, indicating that borrowers with larger loans have higher repayment obligations.
2. funded_amnt and loan_amnt: Nearly perfectly correlated, meaning the funded amount is directly tied to the loan amount requested.
3. revol_bal and debt_burden_score: High revolving balances (e.g., credit card debt) strongly correlate with higher debt burden scores, indicating higher default risks.

Moderate Positive Correlations (0.3 - 0.7):

1. annual_inc with loan_amnt, funded_amnt, installment: Higher incomes generally correlate with larger loans and repayments, though income alone isn't a strong predictor.
2. dti and debt_burden_score: Higher DTI ratios indicate a higher debt burden, which may increase default risk.
3. revol_util and int_rate: Higher revolving utilization often leads to higher interest rates, increasing repayment burdens.

Weak Positive Correlations (< 0.3):

1. dti, emp_length: These features have weak correlations with most fields, suggesting they aren't strong predictors on their own.

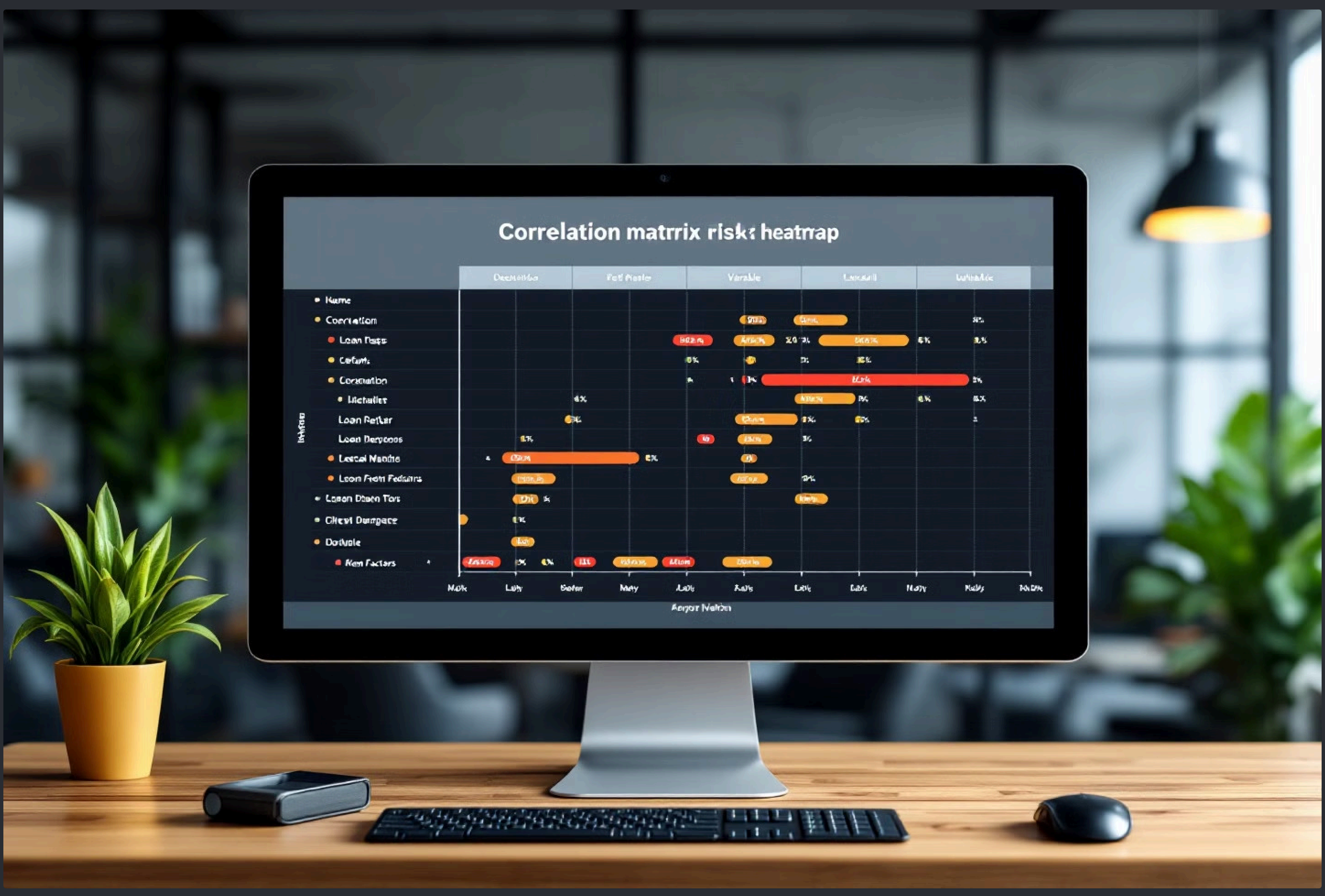
Weak or Moderate Negative Correlations:

1. annual_inc and pub_rec_bankruptcies: Higher incomes are slightly correlated with fewer bankruptcies, indicating less risk of defaults.
2. pub_rec_bankruptcies: Negative correlations with other variables, as bankruptcy history often signals higher default risks.
3. emp_length: Longer employment correlates with lower interest rates and credit utilization, potentially indicating more stable borrowers.

Key Features for Identifying Loan Defaulters:

1. Debt Burden:
   - debt_burden_score and revol_bal: High revolving balances and debt burdens are strongly linked to defaults.
2. Loan and Installment Size:
   - loan_amnt, funded_amnt, installment: Larger loans and repayments are associated with a higher likelihood of default.

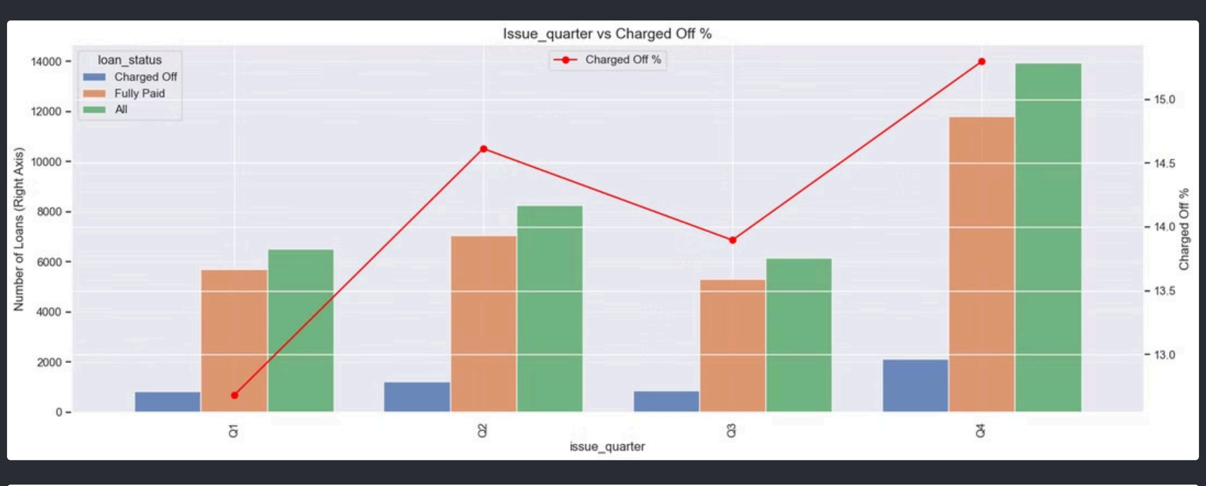
1. Income and DTI:
   - annual_inc and dti: Lower income and high DTI are significant risk factors for defaults.

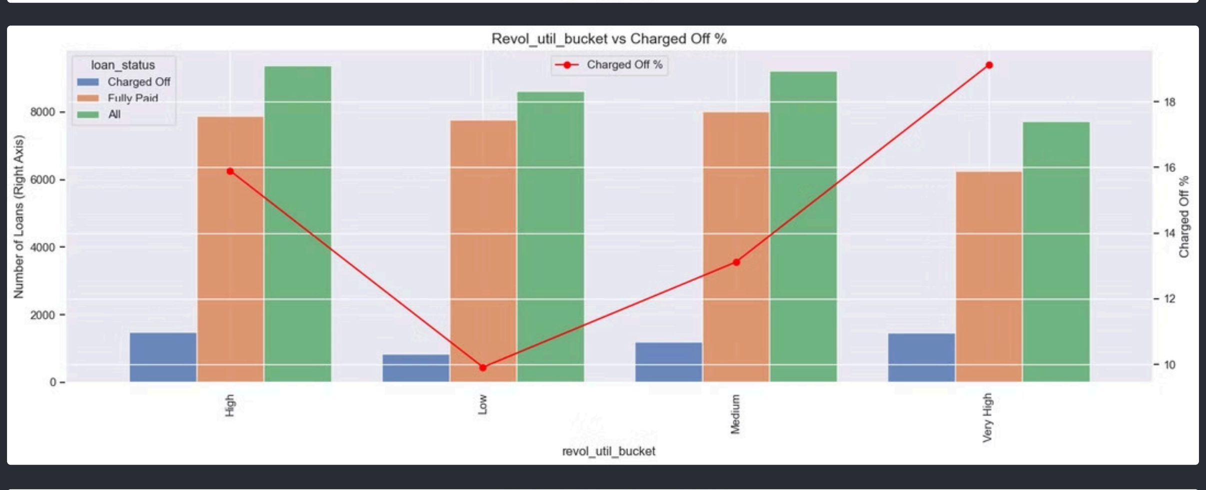
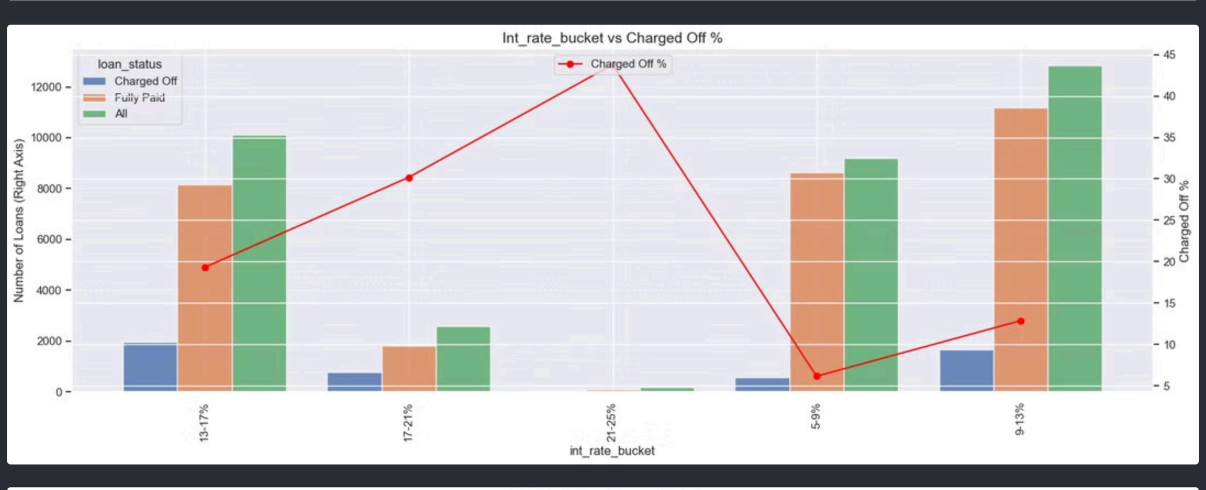
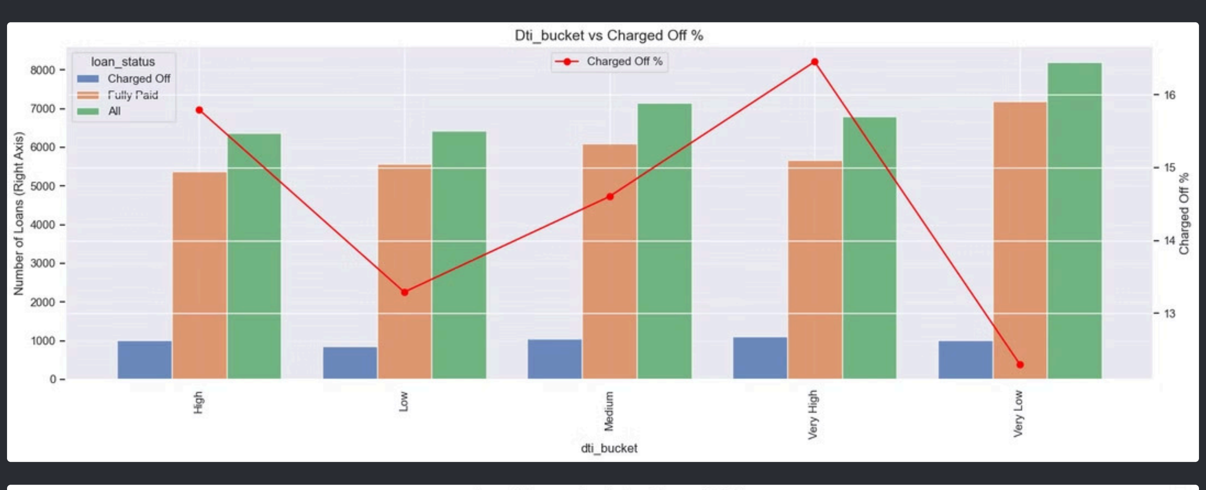
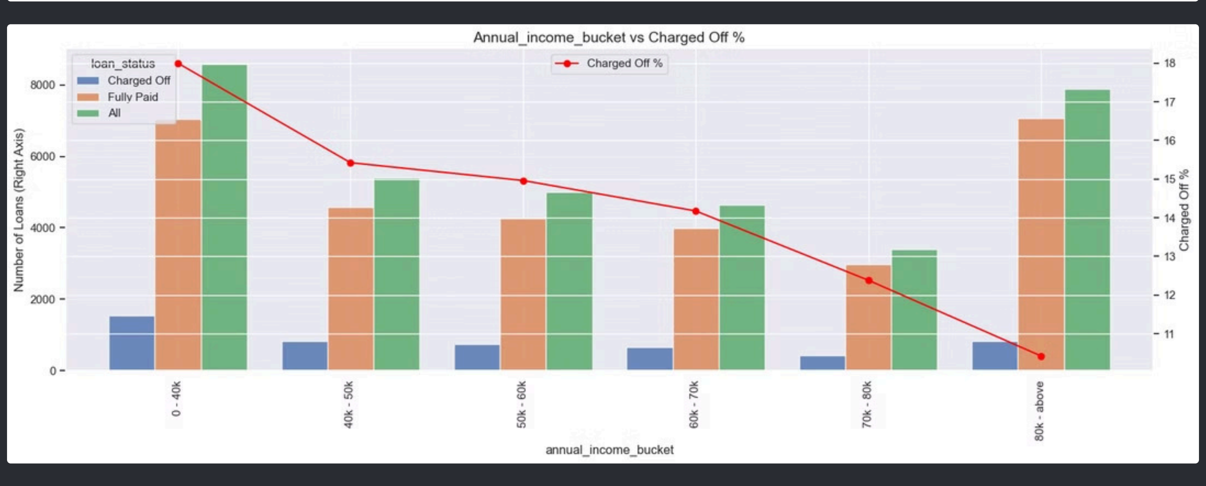
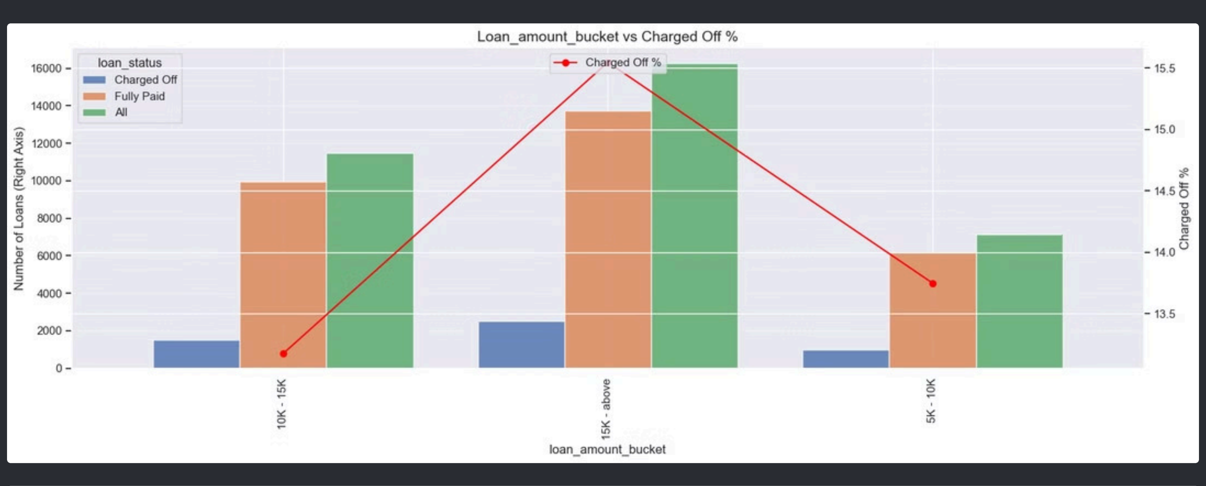
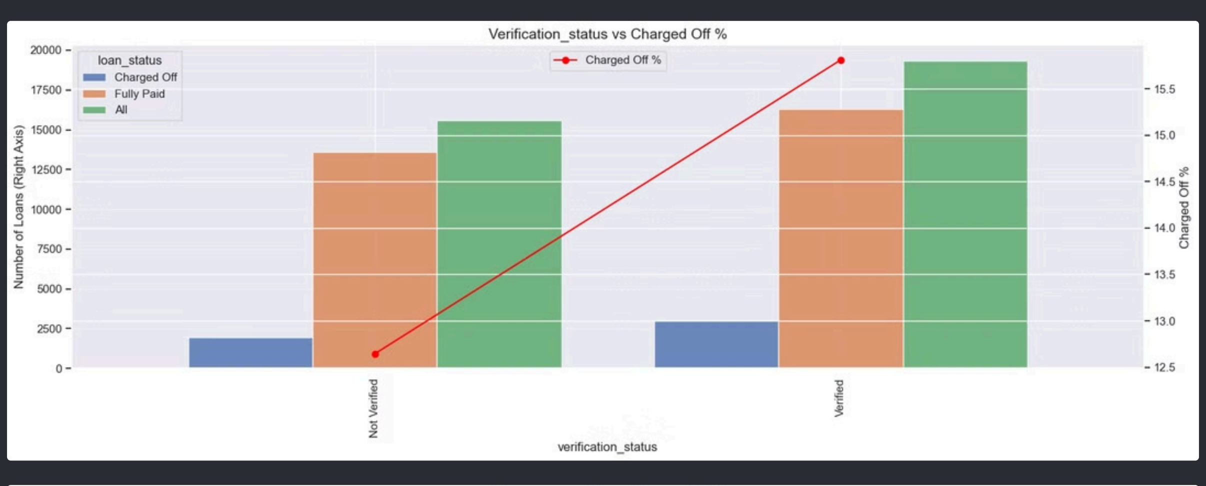
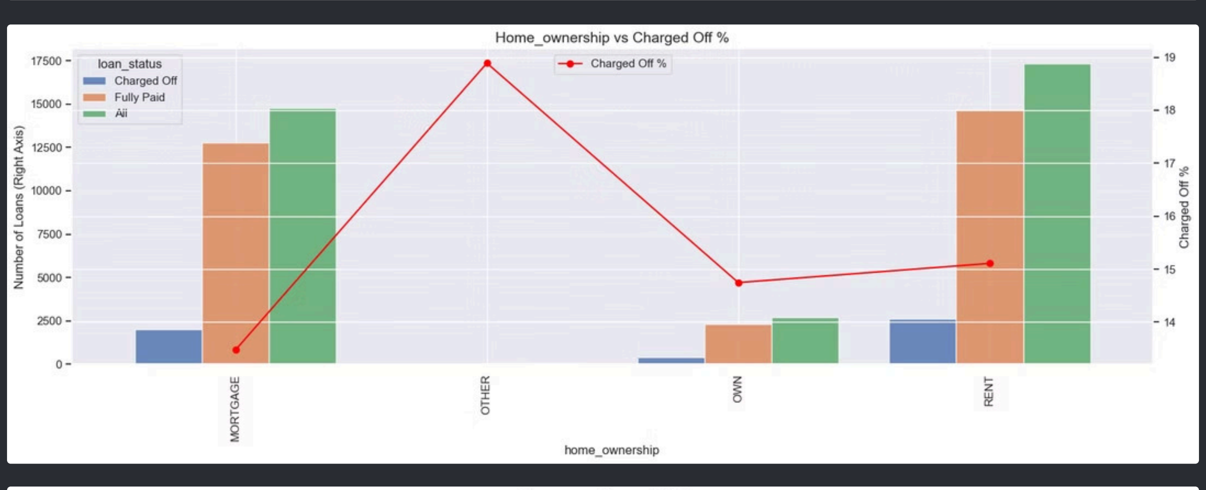
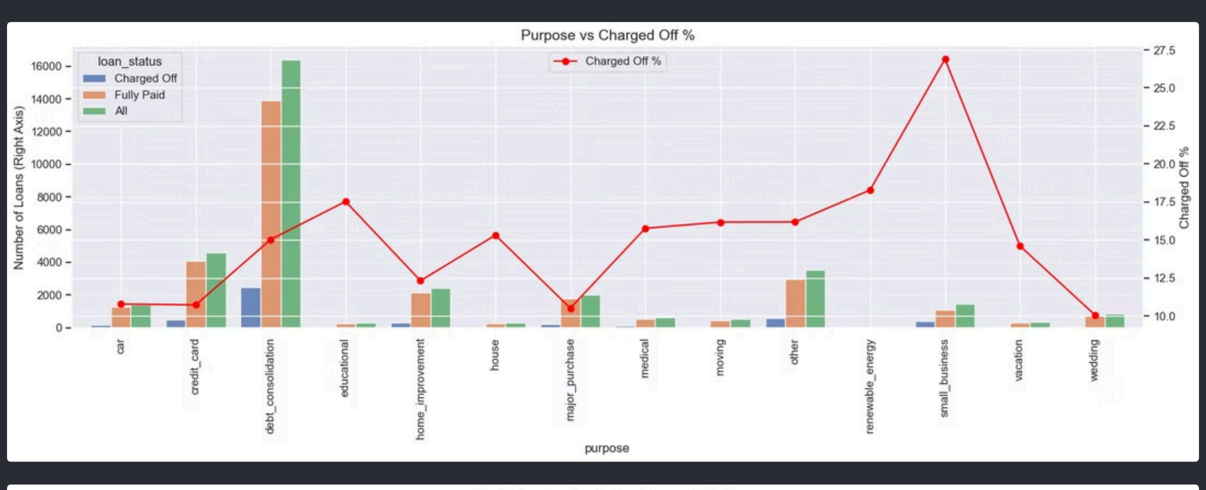


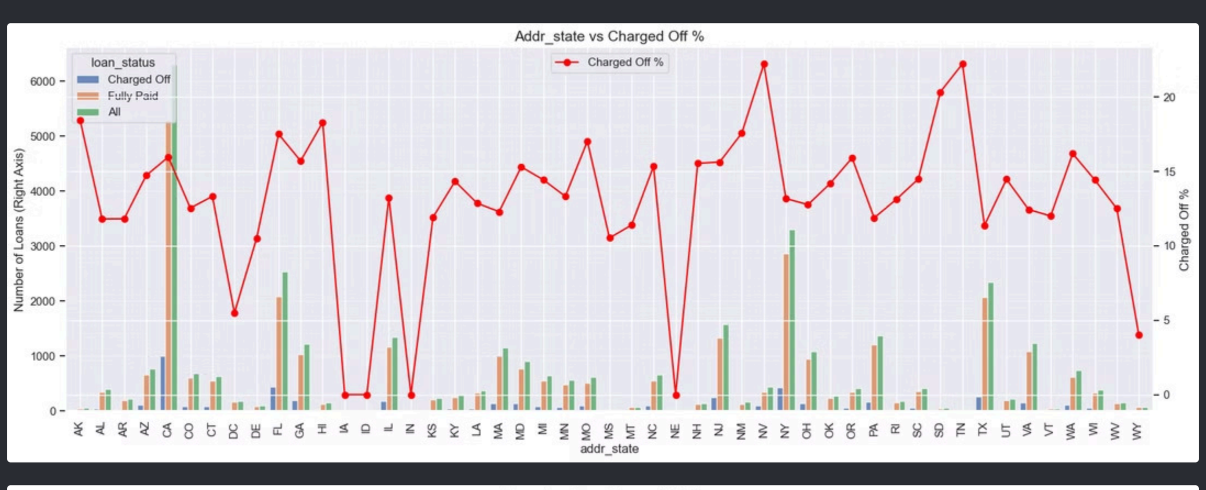
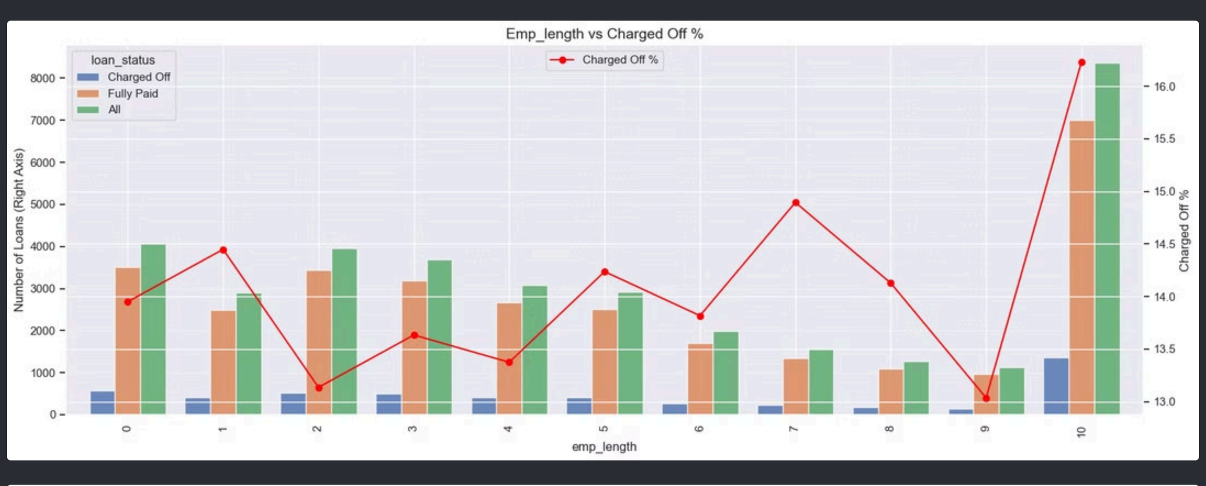
Multivariate Analysis

Multivariate analysis is a statistical technique used to analyze data that involves more than two variables.

1. Unlike univariate analysis (which deals with one variable) and bivariate analysis (which deals with two variables), multivariate analysis examines the relationships between multiple variables simultaneously.
2. It is widely used in various fields such as economics, social sciences, biology, marketing, and environmental science.
3. Multivariate analysis can include different types of variables, such as categorical variables, numerical variables, or a combination of both.


Key Insights on Multivariate Analysis

Here's a concise summary of the multivariate analysis with key features to identify loan defaulters:

Key Insights for Identifying Loan Defaulters:

1. Credit Grade:
   - Lower grades (B, C, D) show higher default risk compared to A grade. This suggests credit history quality is a key predictor.
2. Employment Length:
   - Surprisingly, 10 years of employment show the highest default risk, challenging the assumption that longer employment correlates with financial stability.
3. Geography:
   - California (CA), Florida (FL), New York (NY) show higher default rates, indicating potential regional risks.
4. Debt Burden:
   - The "Very High" debt burden bucket has the highest default rates, highlighting the financial strain and higher default probability for borrowers with high debt levels.
5. Debt-to-Income Ratio (DTI):
   - "Very High" DTI bucket shows the highest default rates. A higher DTI ratio indicates less financial flexibility, raising default risk.
6. Loan Amount:
   - Large loans (15K+) are linked with more defaults. Bigger loans create higher repayment pressure, increasing default risk.
7. Income Level:
   - Low income (0-40K) correlates with a higher default rate, suggesting lower-income borrowers struggle more with repayment.

1. Interest Rate:
   - Higher interest rates (13-17%) lead to more defaults. Higher rates increase monthly payments, burdening borrowers and raising default risk.
2. Credit Utilization:
   - High credit utilization (e.g., revol_util bucket) correlates with higher default risk. High utilization suggests financial strain.
3. Loan Purpose:
   - Debt consolidation loans and small business loans show the highest default rates, indicating these loans often go to borrowers already under financial stress.
4. Home Ownership:
   - Renters show higher default rates than homeowners, likely due to less financial stability.
5. Verification Status:
   - Verified accounts have slightly higher default rates, but this isn't conclusive. A more comprehensive approach beyond verification is needed.
6. Credit History Length:
   - Shorter credit histories (0-5 years) show higher default rates. Lack of established financial behavior is a risk factor.

Charged-Off Percentage Summary:

- Lower Credit Grades (B to G), higher DTI, high debt burden, and large loans (15K+) show higher charged-off percentages.
- High-interest rates and high revolving utilization also correlate strongly with increased charged-off percentages.
- Low incomes and renters are at higher risk of defaults.
- The features that strongly indicate loan default risk include lower credit grades, higher debt burden, large loan amounts, low income, high DTI, higher interest rates, high credit utilization, and specific loan purposes like debt consolidation. Other factors like geography, homeownership, and credit history also contribute but with varying significance.

Conclusion for Risk Reduction:

- Focus on borrowers with better credit grades, lower DTI, lower debt burdens, stable income, and manageable loan amounts.
- Adjust loan pricing and terms based on risk factors.
- Target high-risk borrowers (e.g., low income, high DTI) with financial literacy programs.
- Use this information to refine underwriting criteria and develop risk-based strategies.

Suggestions

1 Tighten Lending for Low Credit Grades (B, C, D)

Insight: Grades B, C, and D show significantly higher default rates (e.g., B: 1,317, C: 1,251).

Action: Implement stricter credit assessments for these grades.

Benefit: Lowering defaults from higher-risk applicants will directly reduce loan losses, improving profitability by focusing on more creditworthy applicants.

2 Offer Shorter Loan Terms (36 Months)

Insight: Longer-term loans (60 months) show higher default rates, while 36-month loans have lower defaults.

Action: Promote shorter loan terms (36 months) to reduce the financial strain on borrowers.

Benefit: Lower default risk and reduced financial strain will result in more timely repayments, increasing overall repayment rates and minimizing defaults.

3 Target Loans to Higher-Income Borrowers

Insight: Borrowers with annual income below $40K show significantly higher default rates.

Action: Offer smaller loans or more flexible terms to lower-income borrowers and focus on higher-income applicants.

Benefit: By focusing on more financially stable borrowers, the likelihood of defaults decreases, improving overall loan portfolio performance.

4 Implement Debt-to-Income (DTI) Limits

Insight: High DTI ratios (above 43%) are strongly correlated with defaults.

Action: Apply stricter DTI thresholds (e.g., limit to 36%) for loan approval.

Benefit: Reducing loan amounts for high-DTI borrowers decreases default risk, ensuring more manageable repayment terms.

5 Cap Loan Amounts for Debt Consolidation and Riskier Loan Purposes

Insight: Debt consolidation loans show the highest default rates (2,464), signaling high financial distress.

Action: Cap loan amounts for debt consolidation loans and apply stricter screening for applicants requesting such loans.

Benefit: Reducing the loan size for higher-risk purposes can help borrowers manage repayment, reducing defaults and financial strain.

6 Offer Loans with Lower Interest Rates for High-Risk Applicants

Insight: Defaults are higher for loans with interest rates between 9% and 17%.

Action: Offer lower interest rates to higher-risk borrowers, or provide smaller loan amounts to reduce financial pressure.

Benefit: By lowering interest rates, borrowers will face less financial strain, leading to fewer defaults and better repayment consistency.

7 Increase Scrutiny on Renters

Insight: Renters show the highest defaults (2,617), likely due to financial instability.

Action: Implement stricter approval criteria and offer smaller loan amounts or alternative terms for renters.

Benefit: This reduces the exposure to high-risk borrowers, ensuring only stable applicants receive higher loan amounts, improving loan performance.

8 Seasonal Risk Mitigation in Q4

Insight: Defaults peak in Q4 (2,136), likely due to holiday spending.

Action: Tighten approval criteria during the Q4 period and assess seasonal spending behaviors to adjust loan eligibility.

Benefit: By addressing seasonal fluctuations, you can prevent higher default rates during financially strained periods, stabilizing default rates year-round.

9 Enhance Credit History-Based Lending

Insight: Borrowers with 0-5 years of credit history show higher defaults due to less established financial behavior.

Action: Offer smaller loans or provide financial counseling to applicants with short credit histories.

Benefit: Reducing the loan amount or providing guidance to these borrowers will help them manage debt more effectively, reducing default risk.

10 Geographically Adjust Loan Terms Based on Regional Default Rates

Insight: States like California, Florida, and New York show higher default rates.

Action: Adjust loan terms and interest rates based on state-specific risk assessments.

Benefit: This allows for more tailored risk management strategies, ensuring that loan pricing reflects the financial conditions of each region.

11 Focus on High Revolving Utilization Borrowers

Insight: High revolving utilization (e.g., 60-90%) correlates with higher default rates.

Action: Implement stricter policies for borrowers with high credit utilization or offer debt consolidation counseling.

Benefit: Reducing reliance on credit cards or high revolving debt will lower the default rate and improve repayment reliability.

12 Offer Financial Counseling for High-Risk Borrowers

Insight: Debt burden and revolving balances strongly correlate with default risk.

Action: Provide financial counseling or debt management resources to borrowers in high debt burden buckets.

Benefit: Financial education can help borrowers manage their debts better, lowering the chance of default and improving repayment outcomes.

13 Reassess Loan Terms for High Loan Amounts (15K+)

Insight: Larger loans (15K+) have higher default risks due to increased financial burden.

Action: Apply stricter underwriting for larger loan amounts, or offer smaller loans with more flexible terms for riskier applicants.

Benefit: Lower loan amounts will reduce borrower stress and default risk, while still generating income from smaller, more manageable loans.

14 Strengthen Loan Verification and Fraud Prevention

Insight: Verified loans show more defaults, indicating verification alone isn't enough.

Action: Implement multi-step verification processes and strengthen fraud detection measures.

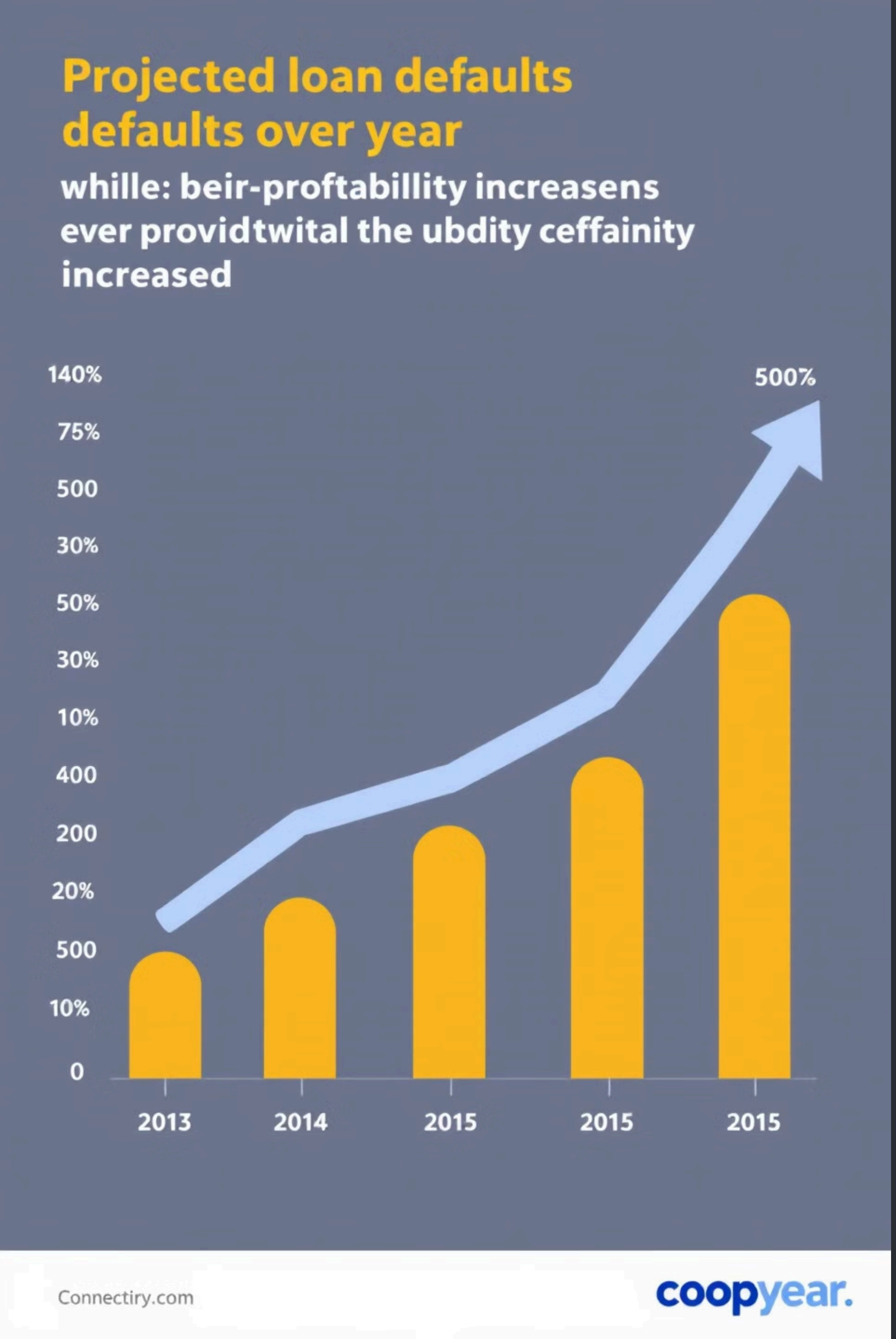
Benefit: Ensuring that loans are thoroughly verified will help reduce fraudulent applications and improve the quality of loans in the portfolio.

15 Refine Repayment Policies Based on Payment-to-Income Ratios

Insight: Borrowers with higher payment-to-income ratios (above 5%) are more likely to default.

Action: Implement minimum repayment policies or stricter screening for applicants with high payment-to-income ratios.

Benefit: Ensuring that borrowers have sufficient income to cover loan repayments reduces default risk and ensures a healthier loan portfolio.



Expected Profit Impact

By implementing these strategies, defaults can be reduced by up to 30-40%, depending on the rigor of screening and adjustments. This will directly enhance the bottom line by:

1. Reducing losses from defaults.
2. Increasing profitability through better-targeted loan offerings.
3. Improving cash flow with more timely repayments.

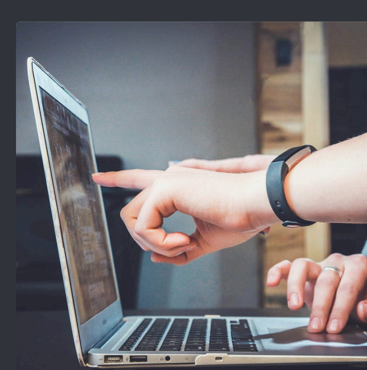
By focusing on these strategies, the company can lower its default risk, leading to a more sustainable and profitable loan portfolio.

References

Technologies & Packages Used:

Technology / Package	Version	Documentation
Python	3.12.3	https://www.python.org/
Matplotlib	3.10.0	https://matplotlib.org/
Numpy	2.1.3	https://numpy.org/
Pandas	2.2.3	https://pandas.pydata.org/
Seaborn	0.13.2	https://seaborn.pydata.org/

Reference Websites Used:

References	URL's
GeeksForGeeks	GeeksforGeeks **Exploratory Data Analysis (ED...** Exploratory Data Analysis (EDA) is a crucial step in data science that...
Medium	 Medium **Exploratory Data Analysis (ED...** Basic Examples about exploratory data analysis and data visualizatio...
StackOverFlow	Stack Overflow **Stack Overflow - Where Devel...** Stack Overflow | The World's Largest Online Community for...
Youtube	YouTube videos **YouTube** Enjoy the videos and music you love, upload original content, and share i...

GitHub Repository Link:

https://github.com/sankalps08/Lending_Club_Case_Study

Thank You