
LENDING CLUB CASE STUDY

By
Vishwanath Madhu
Rahul

Content

- Problem Statement
- Data Description
- Data Understanding
- Data Cleaning & Pre-processing
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Derived Columns
- Correlation Analysis
- Suggestions
- References & Useful Links

Problem Statement

- Lending Club, a Consumer Finance marketplace specialising in offering a variety of loans to urban customers, faces a critical challenge in managing its loan approval process. When evaluating loan applications, the company must make sound decisions to minimise financial losses, primarily stemming from loans extended to applicants who are considered “Risky”.
- These financial losses, referred to as Credit Losses, occur when borrowers fail to repay their loans or default. In simpler terms, borrowers labeled as “Charged-Off” are the ones responsible for the most significant losses to the company.
- The primary objective of this exercise is to assist Lending Club in mitigating credit losses. This challenge arises from two potential scenarios:
 1. Identifying applicants likely to repay their loans is crucial, as they can generate profits for the company through interest payments.
Rejecting such applicants would result in a loss of potential business.
 2. On the other hand, approving loans for applicants not likely to repay and at risk of default can lead to substantial financial losses for the company.
- The objective is to pinpoint applicants at risk of defaulting on loans, enabling a reduction in credit losses. This case study aims to achieve this goal through Exploratory Data Analysis (EDA) using the provided dataset.
- In essence, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

DATA DESCRIPTION

Lending Club provided us with customers' historical data.

This dataset contained information pertaining to the borrower's past credit history and Lending Club loan information. The total dataset consisted of over 39717 records and 111 columns, which was sufficient for our team to conduct analysis.

Variables present within the dataset provided an ample amount of information, which we could use to identify relationships and gauge their effect upon the success or failure of a borrower fulfilling the terms of their loan agreement.

LoanStatNew	Description
acc_now_delinq	The number of accounts on which the borrower is now delinquent.
acc_open_past_24mths	Number of trades opened in past 24 months.
addr_state	The state provided by the borrower in the loan application
all_util	Balance to credit limit on all trades
annual_inc	The self-reported annual income provided by the borrower during registration.
annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers
avg_cur_bal	Average current balance of all accounts
bc_open_to_buy	Total open to buy on revolving bankcards.
bc_util	Ratio of total current balance to high credit/credit limit for all bankcard accounts.
chargeoff_within_12_mths	Number of charge-offs within 12 months
collection_recovery_fee	post charge off collection fee
collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections
delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent.
desc	Loan description provided by the borrower
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
dti_joint	A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding mortgages and the requested LC loan, divided by the co-borrowers' combined self-reported monthly income
earliest_cr_line	The month the borrower's earliest reported credit line was opened
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
emp_title	The job title supplied by the Borrower when applying for the loan.*
fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to.
fico_range_low	The lower boundary range the borrower's FICO at loan origination belongs to.
funded_amnt	The total amount committed to that loan at that point in time.
funded_amnt_inv	The total amount committed by investors for that loan at that point in time.

We require only the variables that have a direct or indirect response to a borrower's potential to default. To achieve this, we prepare the data by choosing select variables that will best fit for this criteria.

DATA UNDERSTANDING

Dataset Attributes:

Primary Attribute:

Loan Status: The Principal Attribute of Interest (loan_status). This column consists of three distinct values:

- ✓ **Fully-Paid:** Signifies customers who have successfully repaid their loans.
- ✓ **Charged-Off:** Indicates customers who have been labeled as "Charged-Off" or have defaulted on their loans.
- ✓ **Current:** Represents customers whose loans are presently in progress and, thus, cannot provide conclusive evidence regarding future defaults.

For the purposes of this case study, rows with a "Current" status will be excluded from the analysis.

Decision Matrix:

Loan Acceptance Outcome- There are three potential scenarios:

Fully Paid- This category represents applicants who have successfully repaid both the principal and the interest rate of the loan.

Current- Applicants in this group are actively in the process of making loan instalments; hence, the loan tenure has not yet concluded. These individuals are not categorised as 'defaulted.'

Charged-off- This classification pertains to applicants who have failed to make timely instalments for an extended period, resulting in a 'default' on the loan.

Loan Rejection- In cases where the company has declined the loan application (usually due to the candidate not meeting their requirements), there is no transactional history available for these applicants. Consequently, this data is unavailable to the company and is not included in this dataset.

DATA UNDERSTANDING

Key Columns of Significance:

The provided columns serve as pivotal attributes, often referred to as predictors.

These attributes, available during the loan application process, significantly contribute to predicting whether a loan will be approved or rejected. It's important to note that some of these columns may be excluded due to missing data in the dataset.

Customer Demographics:

- ✓ **Annual Income (annual_inc):** Reflects the customer's annual income. Typically, a higher income enhances the likelihood of loan approval.
- ✓ **Home Ownership (home_ownership):** Indicates whether the customer owns a home or rents. Home ownership provides collateral, thereby increasing the probability of loan approval.
- ✓ **Employment Length (emp_length):** Represents the customer's overall employment tenure. Longer tenures signify greater financial stability, leading to higher chances of loan approval.
- ✓ **Debt to Income (dti):** Measures how much of a person's monthly income is already being used to pay off their debts. A lower DTI translates to a higher chance of loan approval.
- ✓ **State (addr_state):** Denotes the customer's location and can be utilised.

Loan Characteristics:

- ✓ **Loan Amount (loan_amt):** Represents the amount of money requested by the borrower as a loan.
- ✓ **Grade (grade):** Represents a rating assigned to the borrower based on their creditworthiness, indicating the level of risk associated with the loan.
- ✓ **Term (term):** Duration of the loan, typically expressed in months.
- ✓ **Loan Date (issue_d):** Date when the loan was issued or approved by the lender.
- ✓ **Purpose of Loan (purpose):** Indicates the reason for which the borrower is seeking the loan, such as debt consolidation, home improvement, or other purposes.
- ✓ **Verification Status (verification_status):** Represents whether the borrower's income and other information have been verified by the lender.
- ✓ **Interest Rate (int_rate):** Represents the annual rate at which the borrower will be charged interest on the loan amount.
- ✓ **Instalment (instalment):** Represents the regular monthly payment the borrower needs to make to repay the loan, including both principal and interest.
- ✓ **Public Records (public_rec):** Refers to derogatory public records, which contribute to loan risk. A higher value in this column reduces the likelihood of loan approval.
- ✓ **Public Records Bankruptcy (public_rec_bankruptcy):** Indicates the number of locally available bankruptcy records for the customer. A higher value in this column is associated with a lower success rate for loan approval.

DATA UNDERSTANDING

Excluded Columns:

In our analysis, we will not consider certain types of columns. It's important to note that this is a general categorisation of the columns we will exclude from our approach, and it does not represent an exhaustive list.

- **Customer Behaviour Columns-** Columns that describe customer behaviour will not be factored into our analysis. The current analysis focuses on the loan application stage, while customer behaviour variables pertain to post-approval actions. Consequently, these attributes will not influence the loan approval/rejection process.
- **Granular Data** - Columns providing highly detailed information that may not be necessary for our analysis will be omitted. For example, while the "grade" column may have relevance in creating business outcomes and visualisations, the "sub grade" column is excessively granular and will not be utilized in our analysis.
- 54 columns contain 'NA' values only, and these columns will be removed namely `acc_open_past_24mths`, `all_util`, `annual_inc_joint`, `avg_cur_bal`, `bc_open_to_buy`, `bc_util`, `dti_joint`, `il_util`, `inq_fi`, `inq_last_12m`, `max_bal_bc`, `mo_sin_old_il_acct`, `mo_sin_old_rev_tl_op`, `mo_sin_rcnt_rev_tl_op`, `mo_sin_rcnt_tl`, `mort_acc`, `mths_since_last_major_derog`, `mths_since_rcnt_il`, `mths_since_recent_bc`, `mths_since_recent_bc_dlq`, `mths_since_recent_inq`, `mths_since_recent_revol_delinq`, `num_accts_ever_120_pd`, `num_actv_bc_tl`, `num_actv_rev_tl`, `num_bc_sats`, `num_bc_tl`, `num_il_tl`, `num_op_rev_tl`, `num_rev_accts`, `num_rev_tl_bal_gt_0`, `num_sats`, `num_tl_120dpd_2m`, `num_tl_30dpd`, `num_tl_90g_dpd_24m`, `num_tl_op_past_12m`, `open_acc_6m`, `open_il_12m`, `open_il_24m`, `open_il_6m`, `open_rv_12m`, `open_rv_24m`, `pct_tl_nvr_dlq`, `percent_bc_gt_75`, `tot_coll_amt`, `tot_cur_bal`, `tot_hi_cred_lim`, `total_bal_ex_mort`, `total_bal_il`, `total_bc_limit`, `total_cu_tl`, `total_il_high_credit_limit`, `total_rev_hi_lim`, `verification_status_joint`.
- Certain columns contain only 0 values, and these columns will also be dropped.
- 9 Columns with single value that do not contribute to the analysis will be removed.
- Columns with values that are single value but have other values as NA will be treated as constant and dropped.
- Columns with more than 65% of data being empty (`mths_since_last_delinq`, `mths_since_last_record`) will be dropped.
- Columns (`id`, `member_id`) will be dropped as they are index variables with unique values and do not contribute to the analysis.
- Columns (`emp_title`, `desc`, `title`) will be dropped as they contain descriptive text (nouns) and do not contribute to the analysis.
- The redundant column (`url`) will be dropped. Further analysis reveals that the URL is a static path with the loan ID appended as a query, making it redundant compared to the (`id`) column.
- 660 records for `pub_rec_bankruptcies` are dropped due to missing values
- These columns capture customer behavior recorded after loan approval and are not available at the time of loan approval. Thus, these variables will not be included in the analysis.
- Columns to be dropped: `delinq_2yrs`, `earliest_cr_line`, `inq_last_6mths`, `open_acc`, `pub_rec`, `revol_bal`, `revol_util`, `total_acc`, `out_prncp`, `out_prncp_inv`, `total_pymnt`, `total_pymnt_inv`, `total_rec_prncp`, `total_rec_int`, `total_rec_late_fee`, `recoveries`, `collection_recovery_fee`, `last_pymnt_d`, `last_pymnt_amnt`, `last_credit_pull_d`, `application_type`

Data Cleaning & Pre-processing

1. Loading data from loan CSV
 2. Checking for null values in the dataset
 3. Checking for unique values
 4. Checking for duplicated rows in data
 5. Dropping Records & Columns
 6. Common Functions
 7. Data Conversion
 8. Outlier Treatment
 9. Imputing values in Columns
-

Data Cleaning & Pre-processing

1. **Loading data from loan CSV:** While loading the dataset, some of the variables had mixed datatypes so they have to be converted accordingly as per analysis.
2. **Checking for null values in the dataset:** There're many columns with null values. So they had to be dropped as they won't play a role in the analysis of the dataset. Roughly 48% of the columns were dropped.
3. **Checking for unique values:** If the column has only a single unique value, it does not make any sense to include it as part of our data analysis. We need to find out those columns and drop them from the dataset. 9 columns had such unique values and they were removed.
4. **Checking for duplicated rows in data:** No duplicate rows were found.
5. **Dropping Records and Columns:**
 - Dropped records where **loan_status=“Current”** as the loan in progress cannot provide us insights as to whether the borrower is likely to default or not.
 - Dropping columns where missing data is **>=65%** as these columns will skew our data analysis and they need to be removed.
 - Dropping extra columns containing text like **collection_recovery_fee, delinq_2yrs, desc, earliest_cr_line, emp_title, id, inq_last_6mths, last_credit_pull_d, last_pymnt_amnt, last_pymnt_d, member_id, open_acc, out_prncp, out_prncp_inv, pub_rec, recoveries, revol_bal, revol_util, title, total_acc, total_pymnt, total_pymnt_inv, total_rec_int, total_rec_late_fee, total_rec_prncp, url, zip_code** as these will not contribute to loan pass or fail.

Data Cleaning & Pre-processing

6. **Common Functions:** Common functions were created for repeating common operations like plotting bar graphs, box plots, histograms, count plots, binning etc.
7. **Data Conversion:** Converted columns like debt to income (dti), funded amount (funded_amnt), funded amount investor (funded_amnt_inv) and loan amount (loan_amnt) to float to match the data. Also converted loan date (issue_d) to DateTime (format: yyyy-mm-dd).
8. **Outlier Treatment:** Calculated the Inter-Quartile Range (IQR) and filtering out the outliers outside of lower and upper bound. During Outlier analysis the following observations were made.

- ✓ The annual income of most of the loan applicants is between 40K - 75K USD
- ✓ The loan amount of most of the loan applicants is between 5K - 15K
- ✓ The funded amount of most of the loan applicants is between 5K - 14K USD
- ✓ The funded amount by investor for most of the loan applicants is between 5K - 14K USD
- ✓ The interest rate on the loan is between 9% - 14%
- ✓ The monthly instalment amount on the loan is between 160 - 440
- ✓ The debt to income ration is between 8 - 18

9. **Imputing values in Columns:**

- ✓ Replaced missing values of **annual_inc** with the corresponding mode value of **annual_inc** of the **emp_length annual_inc** field: They Employment length has **1015** missing values, which means either they are not employed or self-employed (business owners). Considering they have a decent average annual income, we have assumed that these are business owners and we have added their employment duration with the mode value of **emp_length** which is **10+** years.
- ✓ Mapped employment length with the respective number of years in int.
- ✓ Imputed **NONE** values as **OTHER** for **home_ownership**.
- ✓ Replaced the '**Source Verified**' values as '**Verified**' since both values mean the same thing i.e. the loan applicant has some source of income which is verified.
- ✓ There are **660 null values** for **pub_rec_bankruptcies**. Dropped those rows as they cannot be imputed.

*Post Data cleaning and Pre-processing of dataset, we were left with **38577 rows × 17 columns**.

Univariate Analysis

Univariate analysis is a statistical method used to analyse and summarise data sets consisting of **one variable**. It deals with the analysis of a single variable, rather than multiple variables, to understand its distribution, central tendency and dispersion.

- It was carried out for both **Categorical** and **Quantitative** Variables.

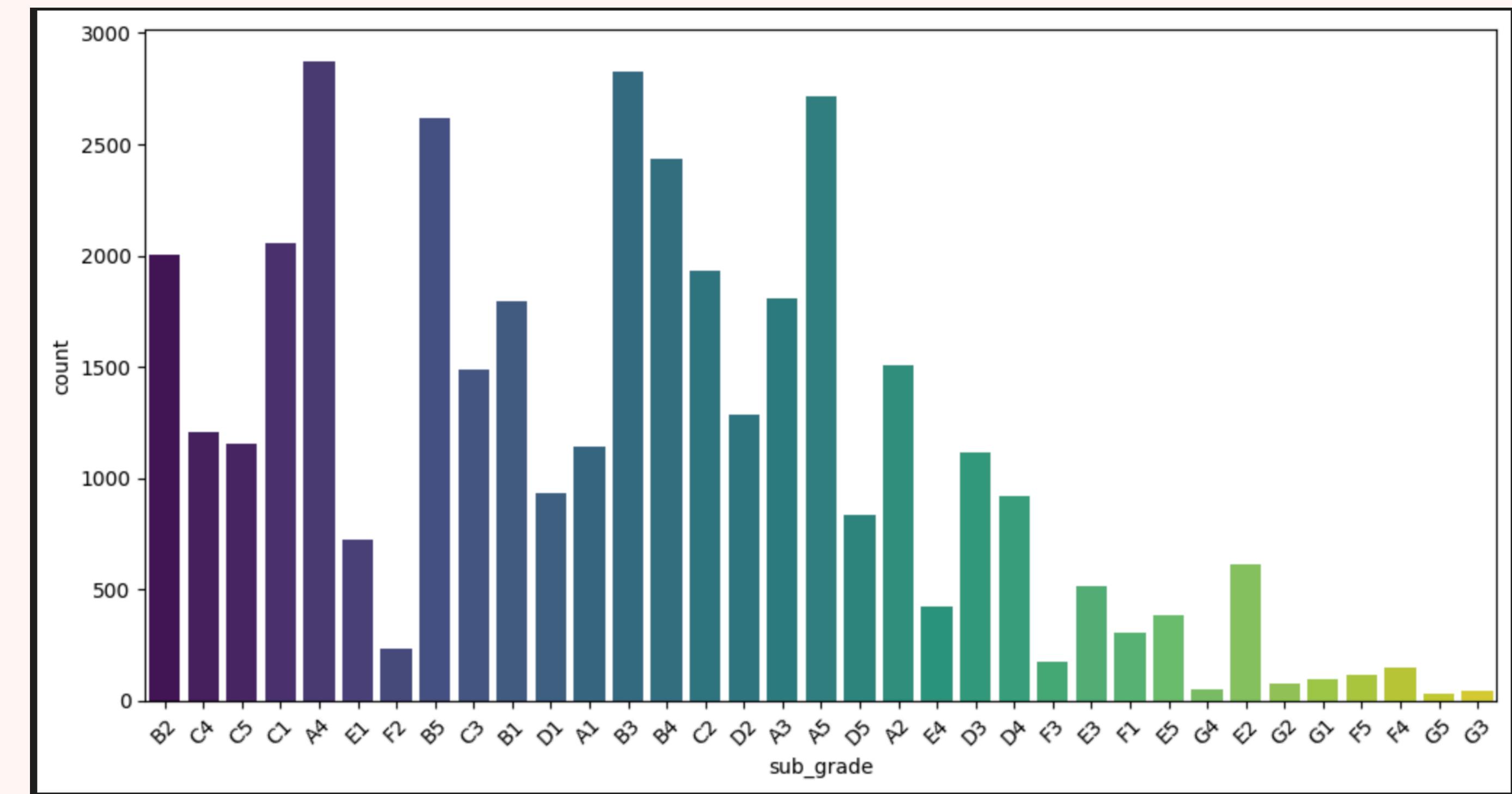
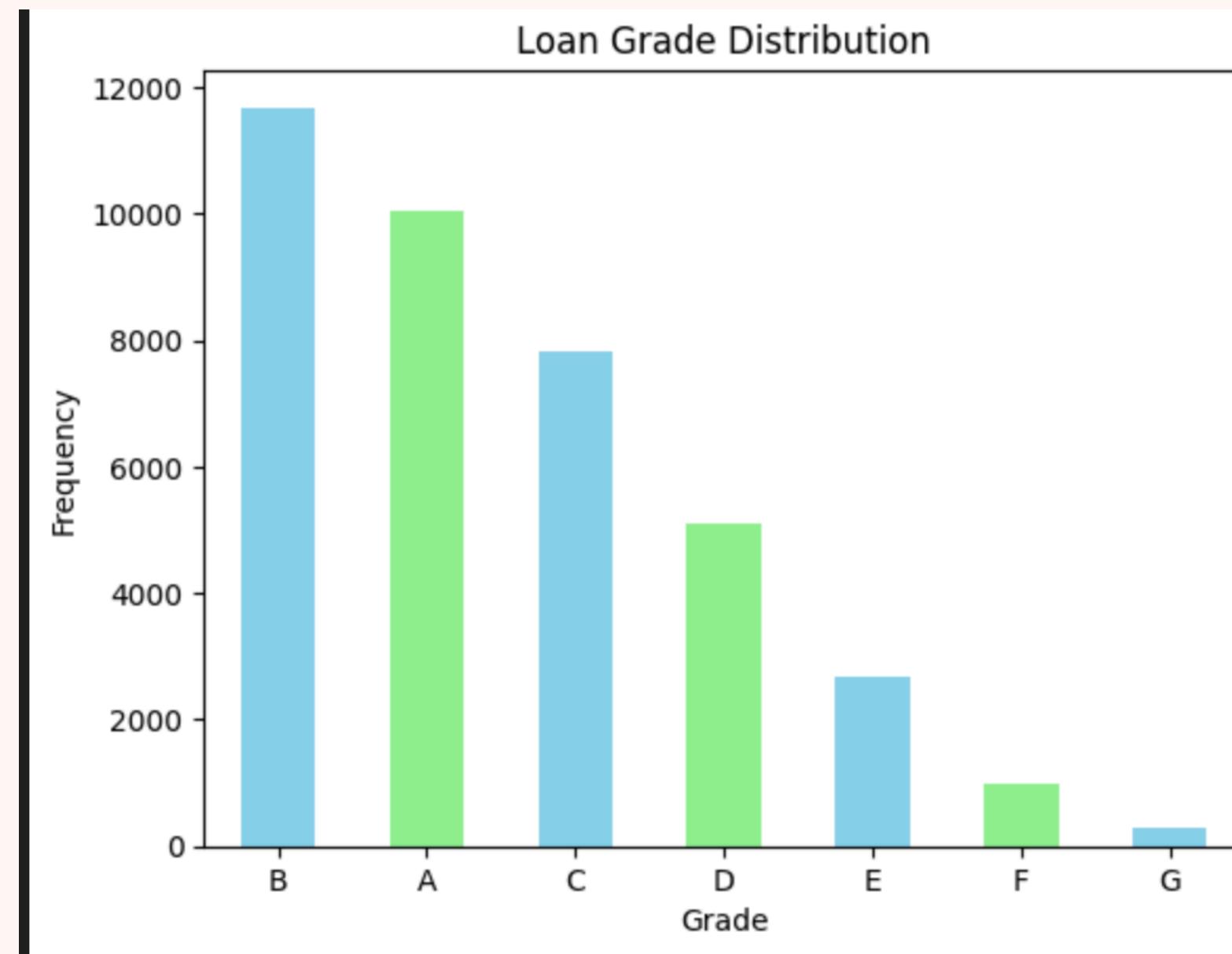
A. Categorical Variables:

Ordered	Unordered
<ul style="list-style-type: none">• Grade (grade)• Sub grade (sub_grade)• Term (36 / 60 months) (term)• Employment length (emp_length)• Issue year (issue_y)• Issue month (issue_m)• Issue quarter (issue_q)	<ul style="list-style-type: none">• Address State (addr_state)• Loan purpose (purpose)• Home Ownership (home_ownership)• Loan status (loan_status)• Loan paid (loan_paid)

B. Quantitative Variables:

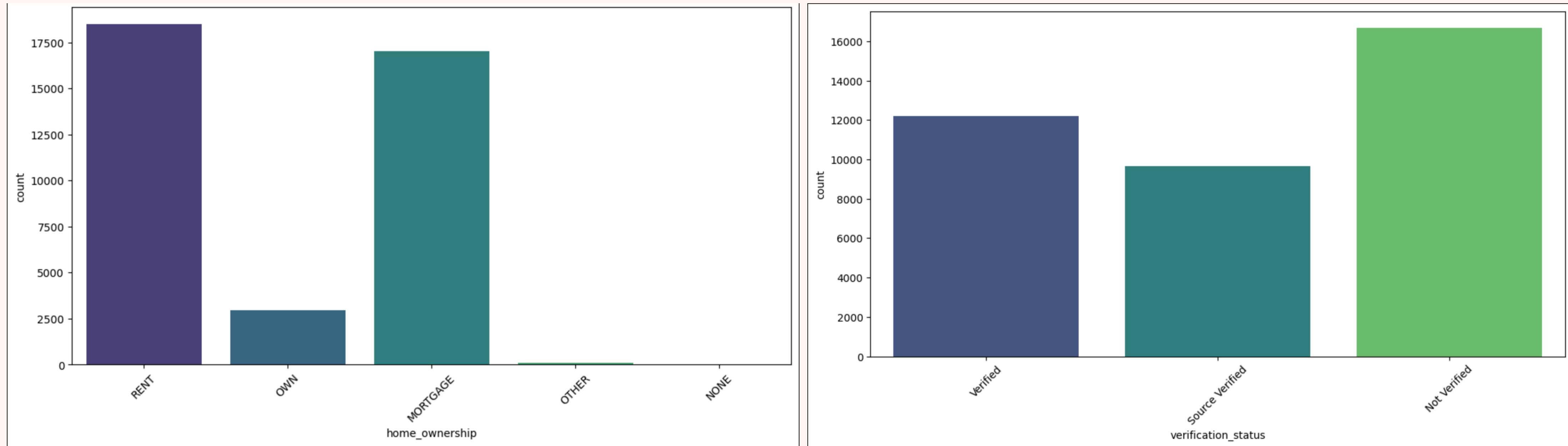
- ✓ Interest rate bucket (int_rate_bucket)
- ✓ Annual income bucket (annual_inc_bucket)
- ✓ Loan amount bucket (loan_amnt_bucket)
- ✓ Funded amount bucket (funded_amnt_bucket)
- ✓ Debt to Income Ratio (DTI) bucket (dti_bucket)
- ✓ Monthly Instalment (instalment)

Univariate Analysis - Unordered.



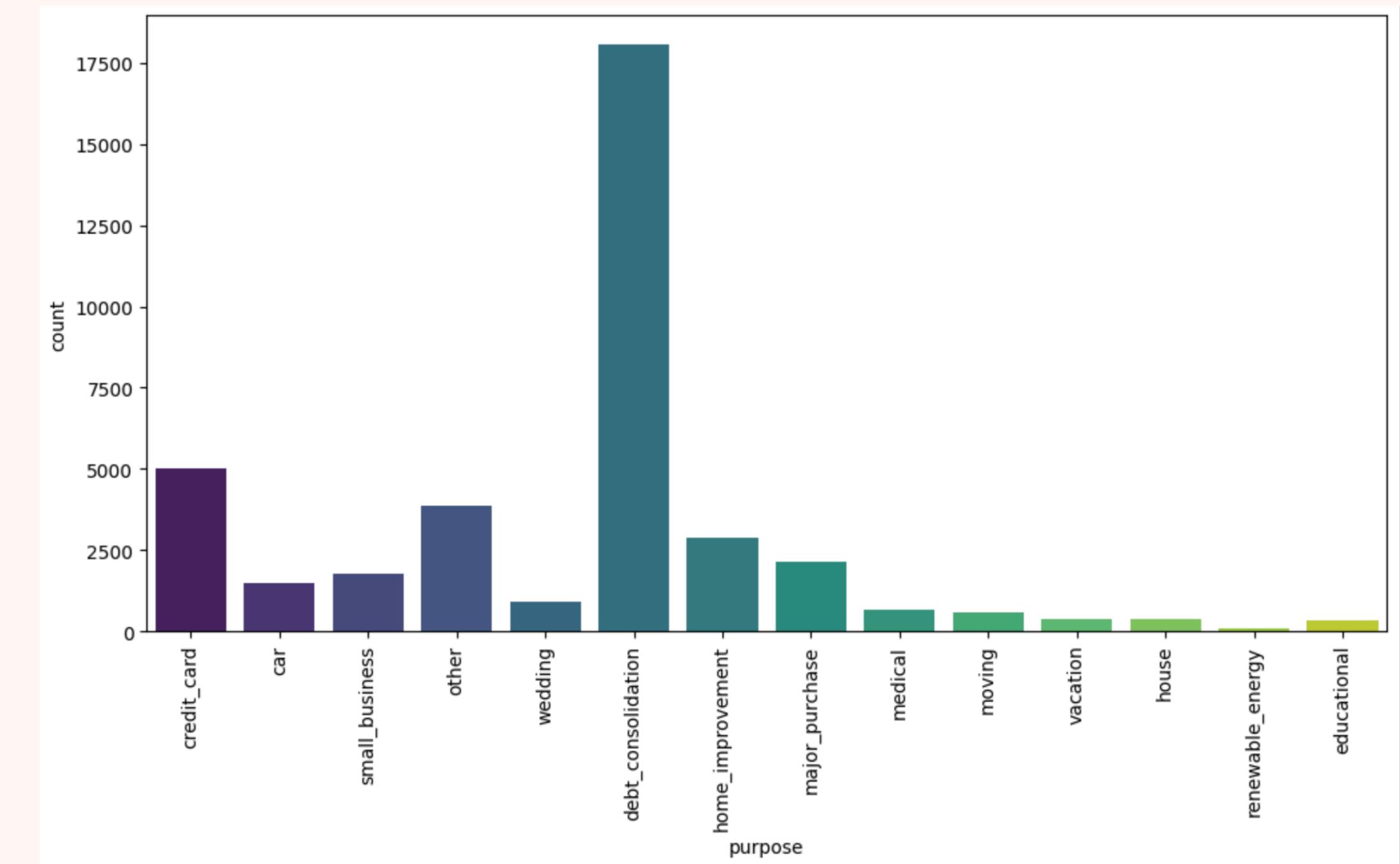
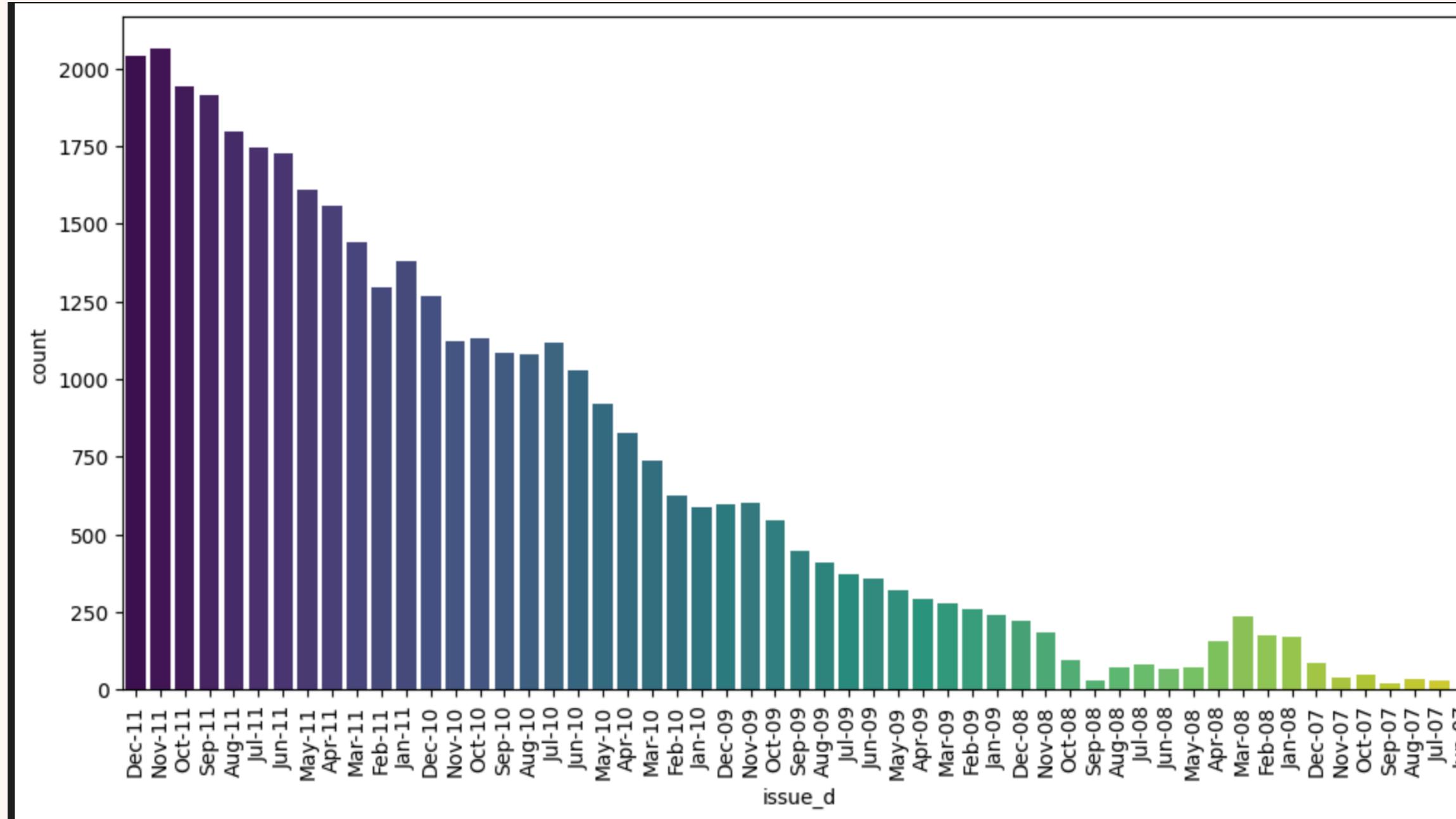
Grade and Subgrade

Univariate Analysis - Unordered.



Home_Ownership and Verification_Status

Univariate Analysis - Unordered.



Issue_d and Purpose

Univariate Analysis - ordered.

Observations & Inferences:

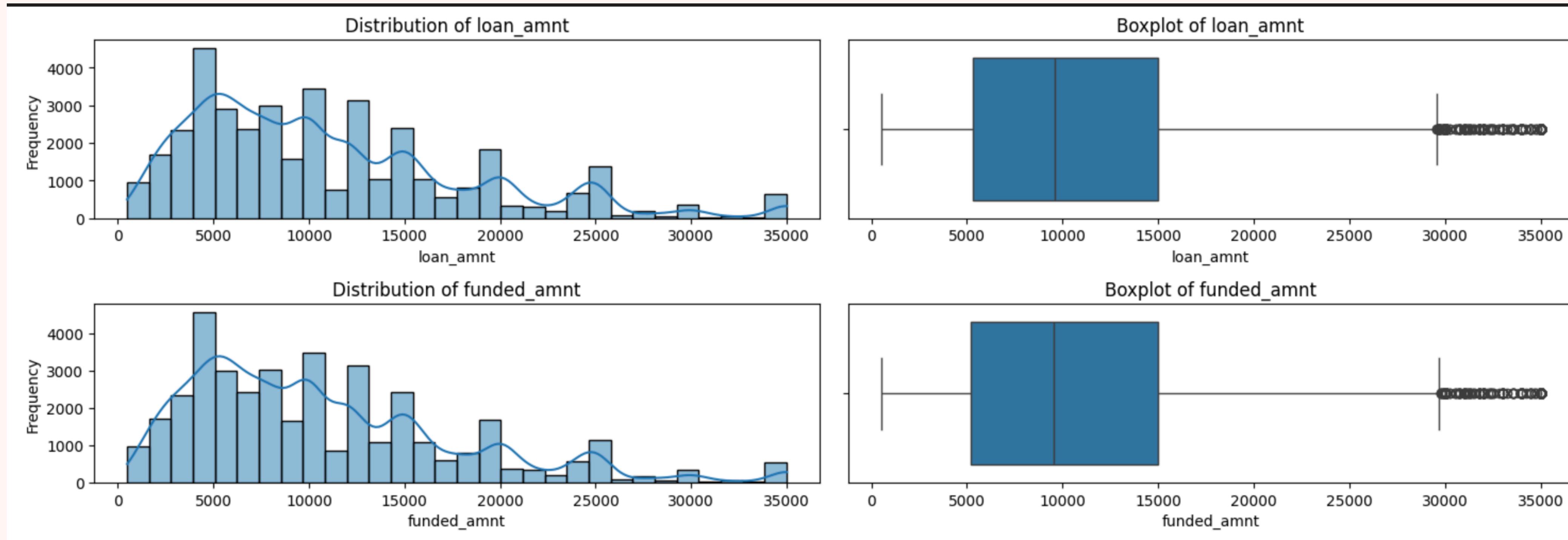
A. Ordered Categorical Variables:

- ✓ Grade B had the highest number of "Charged off" loan applicants, with a total of 1,352 applicants, indicating that applicants with this credit grade faced challenges in repaying their loans.
- ✓ Short-term loans with a duration of 36 months were the most popular among "Charged off" applicants, with 3,006 applications. This suggests that a significant portion of applicants who experienced loan default chose shorter repayment terms.
- ✓ Applicants who had been employed for more than 10 years accounted for the highest number of "Charged off" loans, totaling 1,474. This indicates that long-term employment history did not necessarily guarantee successful loan repayment.
- ✓ The year 2011 recorded the highest number of "Charged off" loan applications, totaling 3,152, signaling a positive trend in the number of applicants facing loan defaults over the years. This could be indicative of economic or financial challenges during that year.
- ✓ "Charged off" loans were predominantly taken during the 4th quarter, with 2,284 applications, primarily in December. This peak in loan applications during the holiday season might suggest that financial pressures during the holidays contributed to loan defaults

B. Unordered Categorical Variables:

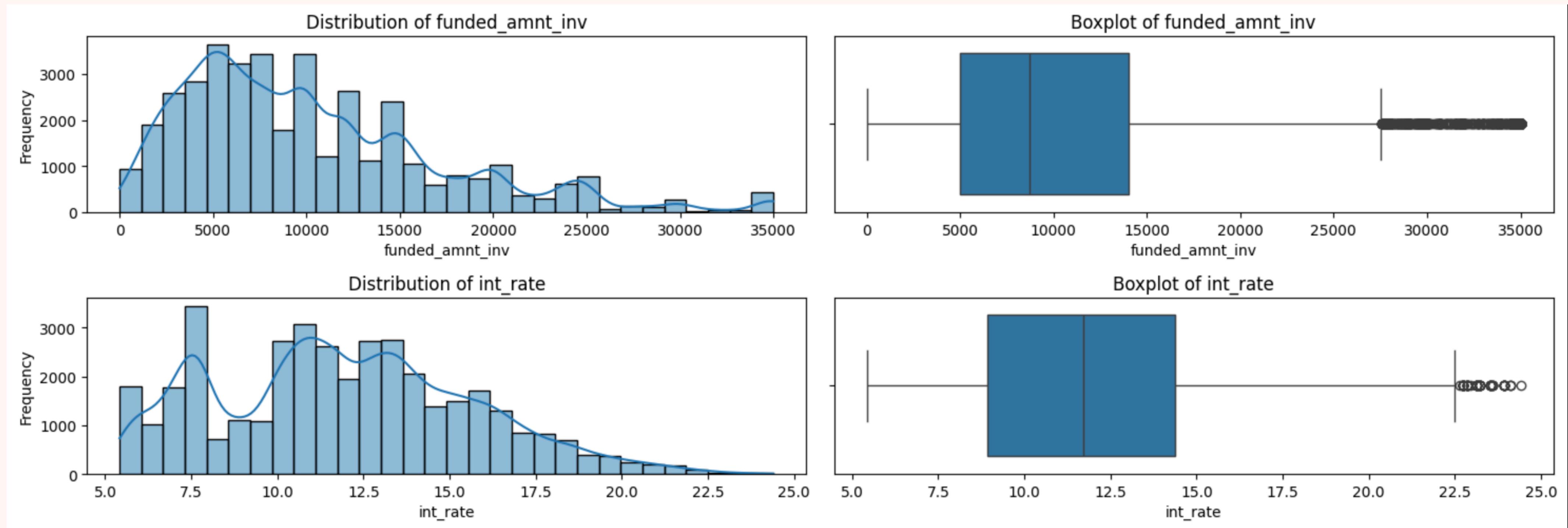
- ✓ California had the highest number of "Charged off" loan applicants, with 1,055 applicants. For such applicants, the lending company needs to implement stricter eligibility criteria or credit assessments due to a higher number of "Charged off" applicants from this state.
- ✓ Debt consolidation was the primary loan purpose for most "Charged off" loan applicants, with 2,633 applicants selecting this option. The lending company needs to exercise caution when approving loans for debt consolidation purposes, as it was the primary loan purpose for many "Charged off" applicants.
- ✓ The majority of "Charged off" loan participants, totaling 2,715 individuals, lived in rented houses. The lending company must assess the financial stability of applicants living in rented houses, as they may be more susceptible to economic fluctuations.
- ✓ A significant number of loan participants, specifically 5,317 individuals, were loan defaulters, unable to clear their loans. The lending company should enhance risk assessment practices, including stricter credit checks and lower loan-to-value ratios, for applicants with a history of loan defaults. They should offer financial education and support services to help borrowers manage their finances and improve loan repayment outcomes.

Univariate Analysis - Quantitative.



Loan_Amount and Funded_Amount

Univariate Analysis - Quantitative.



Funded_Amount by Investors and Interest on Loan

Univariate Analysis - Interpretation

Univariate Analysis

Loan Amount

Interpretation: The distribution of loan amounts shows that most loans are in the lower to mid-range, indicating that borrowers often seek moderate amounts. Outliers in higher loan amounts could financial needs or riskier borrowers.

Interest Rate:

Interpretation: The interest rate distribution skews higher, suggesting that many loans are offered at mid-to-high rates. Higher interest rates may imply greater credit risk or lower borrower creditworthiness.

Interpretation: The DTI ratio has a right-skewed distribution, with most borrowers having moderate ratios. High DTI values, seen as outliers, might indicate borrowers with significant financial strain compared to their income.

Loan Term Interpretation: Loans are typically either 36 or 60 months. The binary distribution here reflects borrower preferences and lender policies, where longer terms might indicate larger loans or lower monthly payments.

Bivariate Analysis

Bivariate analysis is a statistical method that involves the simultaneous analysis of **Two variables**. It aims to determine the empirical relationship between them. The analysis can be used to test Hypotheses, Identify patterns or explore relationship between the variables.

- ✓ It was carried out for both **Categorical** and **Quantitative** Variables.

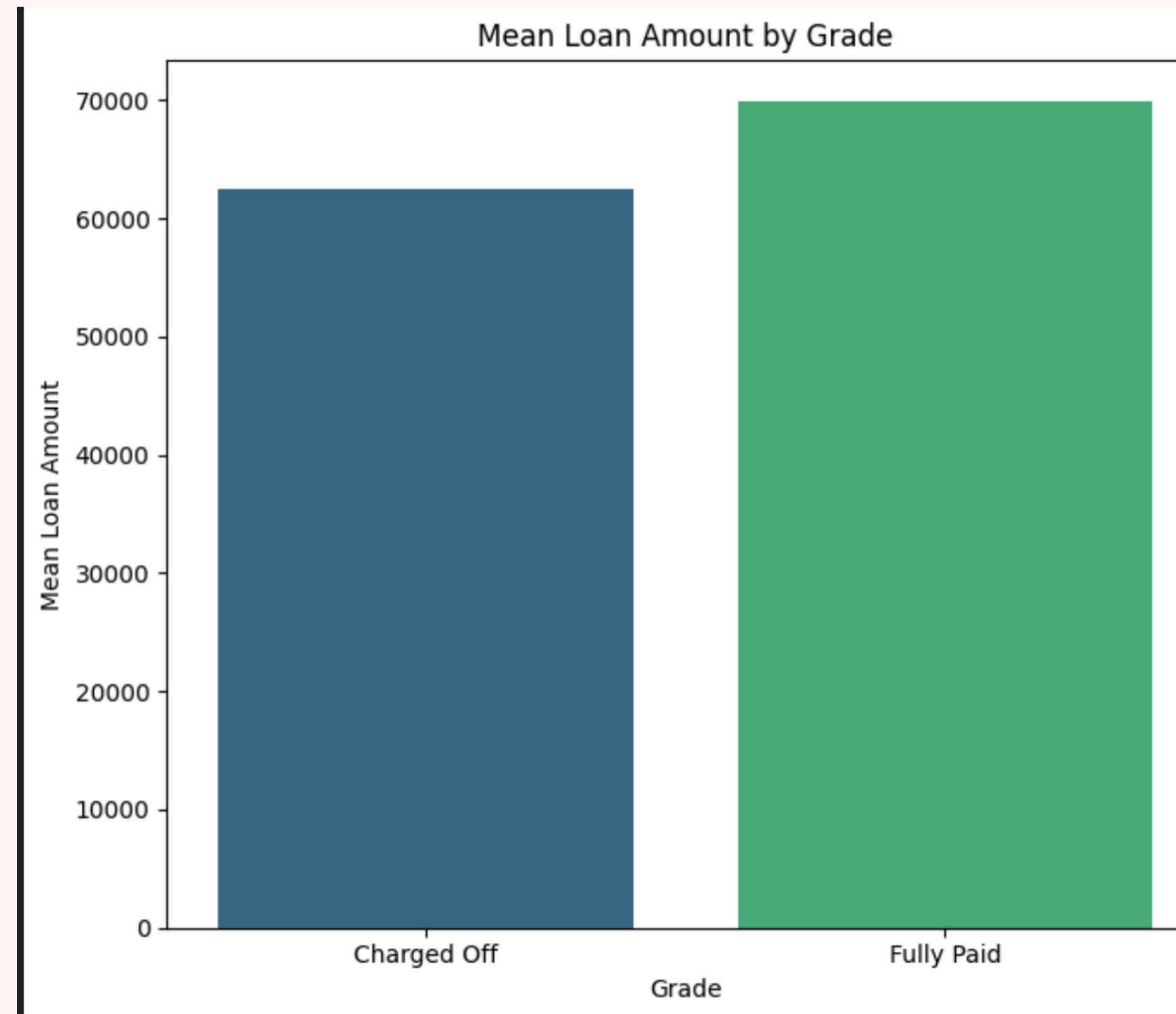
A. Categorical Variables:

Ordered	Unordered
<ul style="list-style-type: none">• Grade (grade)• Sub grade (sub_grade)• Term (36 / 60 months) (term)• Employment length (emp_length)• Issue year (issue_y)• Issue month (issue_m)• Issue quarter (issue_q)	<ul style="list-style-type: none">• Address State (addr_state)• Loan purpose (purpose)• Home Ownership (home_ownership)• Loan status (loan_status)• Verification status(verification_status)

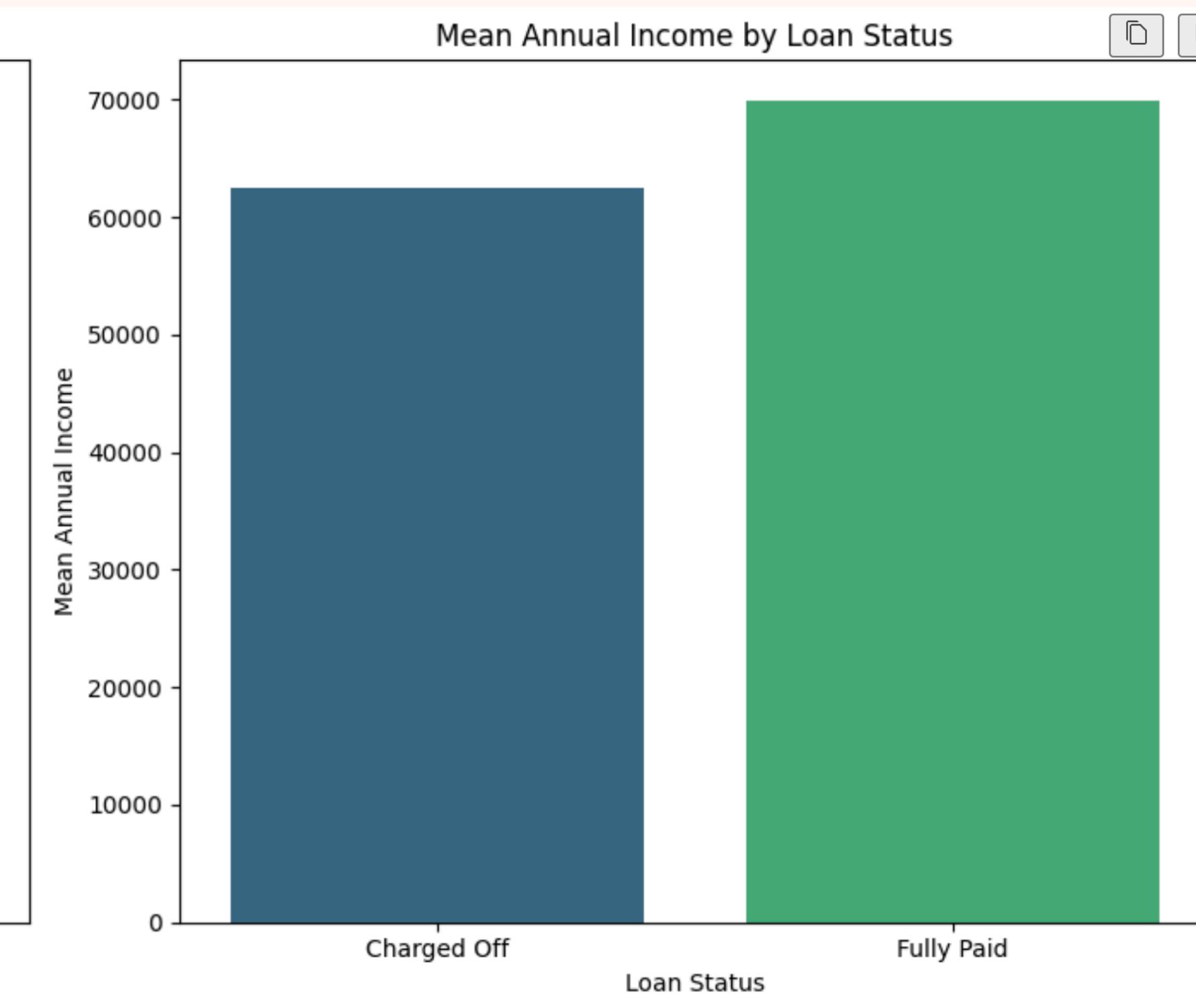
B. Quantitative Variables:

- ✓ Interest rate bucket (int_rate_bucket)
- ✓ Annual income bucket (annual_inc_bucket)
- ✓ Loan amount bucket (loan_amnt_bucket)
- ✓ Funded amount bucket (funded_amnt_bucket)
- ✓ Debt to Income Ratio (DTI) bucket (dti_bucket)

Bivariate Analysis

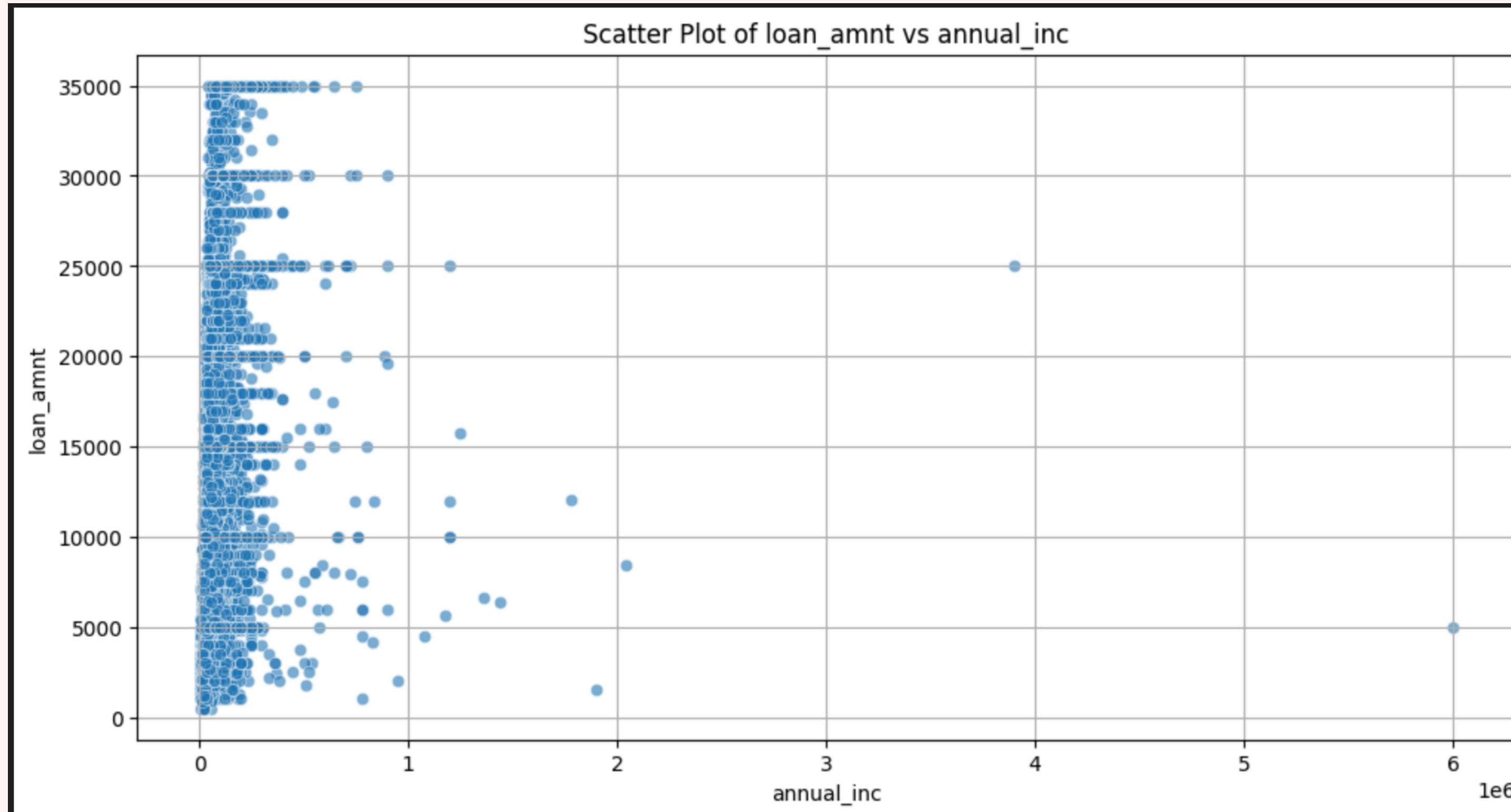


Grade and Mean Loan Amount



Loan status and Mean Annual Income

Bivariate Analysis



Annual Income and Loan Amount

Bivariate Analysis - Interpretation

Loan Status vs. Average Interest Rate

Interpretation: Loans with a "Charged Off" status show higher average interest rates, suggesting a possible correlation between higher interest rates and increased likelihood of default. This can help in assessing default risk based on interest rate tiers.

Funded Amount by Grade and Loan Status

Interpretation: Loans in higher-risk grades (e.g., D, E, F) have lower average funded amounts compared to lower-risk grades (A, B). However, higher-grade loans (A, B) are more commonly "Fully Paid," indicating that riskier loans are not only smaller but also more likely to default.

DTI Ratio by Loan Status

Interpretation: Loans marked as "Charged Off" tend to have higher DTI ratios, indicating that borrowers with more financial obligations relative to income are at greater risk of default. This insight can help lenders in adjusting DTI thresholds to mitigate risk.

Risk Grade vs. Funded Amount

Interpretation: Higher risk grades (mapped to higher numbers in risk_grade) have lower average funded amounts, reflecting a cautious approach by lenders in extending credit to riskier borrowers.

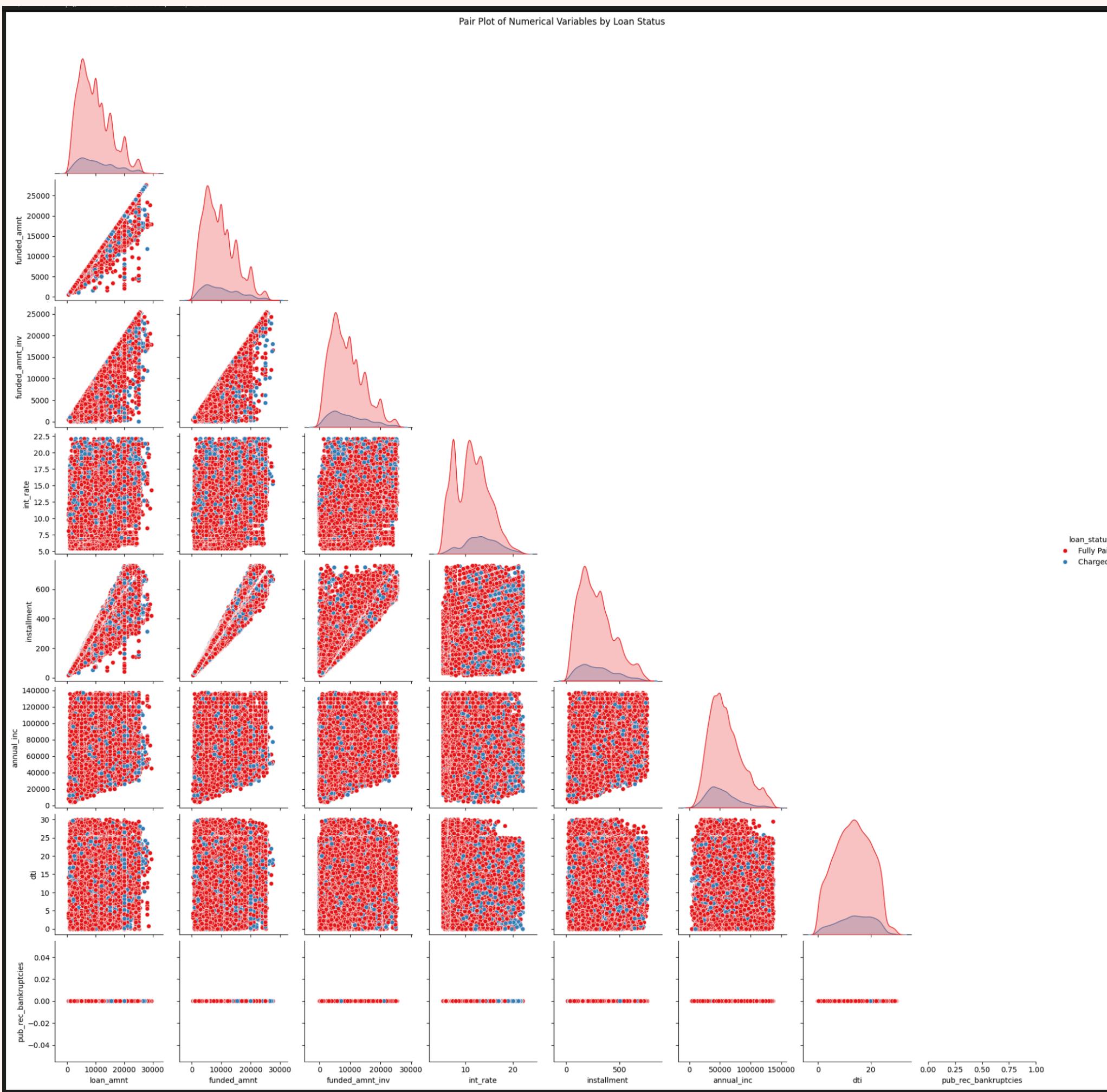
Income Verification vs. Loan Status

Interpretation: Verified income status correlates positively with "Fully Paid" loans, implying that income verification may play a crucial role in loan performance. Loans with unverified income status exhibit a higher incidence of defaults.

Multivariate Analysis

- ✓ **Multivariate analysis** is a statistical technique used to analyze data that involves more than two variables.
 - ✓ 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.
 - ✓ It is widely used in various fields such as economics, social sciences, biology, marketing, and environmental science.
 - ✓ Multivariate analysis can include different types of variables, such as categorical variables, numerical variables, or a combination of both.
-

Multivariate Analysis



Derived Columns

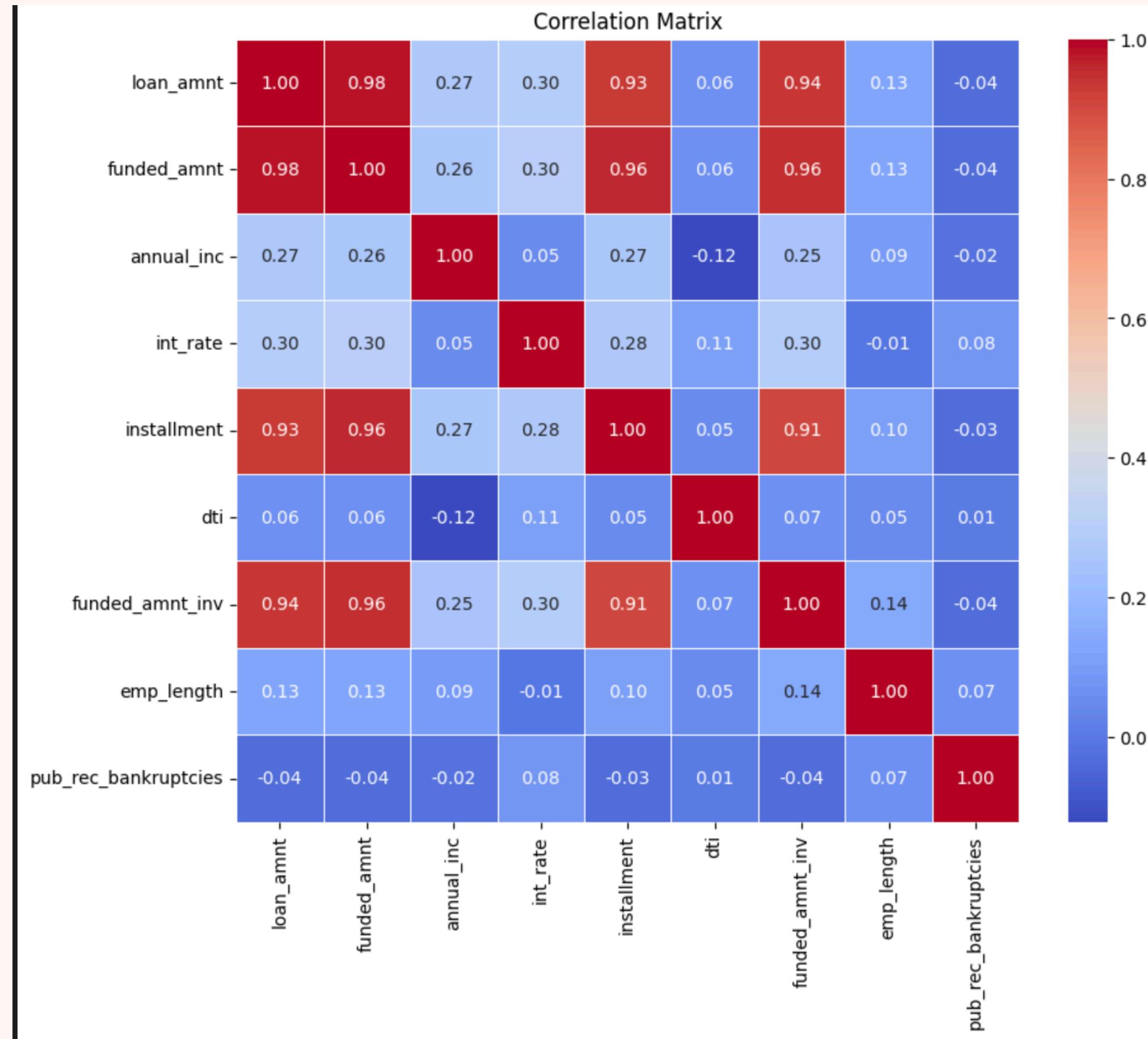
A derived column is a column in the data that is not actually a physical column, but is instead generated from other values.

- Formula: `loan_amnt / annual_inc`
Purpose: Indicates the portion of annual income committed to the loan, helping lenders assess risk. A high ratio may imply a higher default risk. Term in Months
- Formula: Extracts number of months from term (e.g., 36 or 60 months)
Purpose: Converts loan term into numeric format for easier comparison and analysis. Total Paid Amount
- Formula: `installment * term_in_months`
Purpose: Represents the total amount the borrower pays over the loan period. Useful for understanding borrower commitment and lender earnings. Interest Amount
- Formula: `total_paid - loan_amnt`
Purpose: Captures the total interest cost paid by the borrower, which is important for understanding loan profitability. Long-Term Loan Indicator

Correlation Analysis

- ✓ Correlation analysis is a statistical technique used to measure the strength and direction of the relationship between two or more variables.
- ✓ It quantifies the degree to which changes in one variable are associated with changes in another variable.
- ✓ Correlation analysis is widely used in various fields, including finance, economics, biology, psychology, and social sciences, to understand patterns and relationships in data.
- ✓ It ranges from **-1 to 1**.
 - **r=1**: indicates a perfect positive correlation
 - **r=-1**: indicates a perfect negative correlation
 - **r=0**: indicates no correlation between the variables

Correlation Matrix.



Correlation Analysis

Observations and Inferences.

❖ Strong Correlation:

- **installment** has a strong correlation with **funded_amnt**, **loan_amnt**, and **funded_amnt_inv**
- **term** has a strong correlation with **interest_rate**
- **annual_inc** has a strong correlation with **loan_amount**

❖ Weak Correlation:

- **dti** has weak correlation with most of the fields
- **emp_length** has weak correlation with most of the fields

❖ Negative Correlation:

- **pub_rec_bankruptcies** has a negative correlation with almost every field
- **annual_inc** has a negative correlation with **dti**

Suggestions.

- ✓ **Implement Stricter Criteria for Grades B, C, and D:** Consider implementing stricter risk assessment and underwriting criteria for applicants falling into Grades B, C, and D to minimise default risks.
- ✓ **Focus on Sub-grades B3, B4, and B5:** Pay special attention to applicants with Subgrades B3, B4, and B5. Consider additional risk mitigation measures or offering lower loan amounts for these sub-grades to reduce default rates.
- ✓ **Evaluate and Limit 60-Month Loans:** Evaluate the risk associated with 60-month loans. Consider limiting the maximum term or adjusting interest rates for longer-term loans to decrease the likelihood of defaults.
- ✓ **Comprehensive Credit Scoring System:** Develop a comprehensive credit scoring system that incorporates various risk-related attributes, as experience alone might not be sufficient to gauge creditworthiness.
- ✓ **Capitalising on Market Growth:** Capitalise on the market's growth trend observed from 2007 to 2011 by maintaining a competitive edge in the industry while ensuring robust risk management practices.
- ✓ **Anticipate Peak Periods:** Anticipate increased loan applications during peak periods such as December and Q4. Ensure efficient processing to meet customer demands during these busy seasons.

Suggestions.

- ✓ **Careful Evaluation for Debt Consolidation Loans:** Carefully evaluate applicants seeking debt consolidation loans, considering potential interest rate adjustments or offering financial counselling services to manage the associated risks.
- ✓ **Consider Housing Stability:** Take housing status into account during the underwriting process to assess housing stability and its impact on the applicant's ability to repay the loan.
- ✓ **Review Verification Process:** Review the verification process to ensure effective assessment of applicant creditworthiness. Consider improvements or adjustments based on the review findings.
- ✓ **Monitor & Adjust for Regional Risk Trends:** Monitor regional risk trends, especially in states like **California, Florida, and New York**. Adjust lending strategies or rates accordingly in high-risk regions.
- ✓ **Thorough Assessment for High Loan Amounts:** Conduct more thorough assessments for loan amounts of \$15,000 or higher. Consider capping loan amounts for higher-risk applicants to mitigate potential defaults.
- ✓ **Adjust Interest Rates Based on DTI Ratios:** Review the interest rate determination process and consider adjusting rates based on Debt-to-Income (DTI) ratios to align with the borrower's ability to repay.
- ✓ **Consider Income Levels for Affordability:** Consider offering financial education resources and set maximum loan amounts based on annual incomes below \$40,000 to ensure loan affordability for borrowers

References & Useful Links.

Technologies & Packages Used:

Technology/Package	Version	Documentation
Python	3.9.2	https://www.python.org/
Matplotlib	3.9.1	https://matplotlib.org/
Numpy	2.0.0	https://numpy.org/
Pandas	2.2.2	https://pandas.pydata.org/
Seaborn	0.13.2	https://seaborn.pydata.org/

Git Hub Repository link:

<https://github.com/vismadhu/LendingclubCaseStudy>

Thank you.