Lending Club Case Study

Submitted By: Amit Kumar

Agenda

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

Lending Club Case Study

Managing Loan Approvals and Mitigating Credit Losses

Lending Club, a consumer finance marketplace specializing in offering a variety of loans to urban customers, faces a significant challenge in optimizing its loan approval process. The company must make sound decisions to minimize financial losses, particularly those arising from loans issued to applicants considered "risky."

The Problem of Credit Losses

- Credit losses occur when borrowers fail to repay their loans or default. Borrowers classified as "Charged-Off" contribute significantly to these losses, posing a critical financial challenge for Lending Club.

Objectives of the Analysis

The primary goal of this exercise is to help LendingClub reduce credit losses. The challenge stems from two key scenarios:

1. **Potential Profit Loss**: Identifying applicants who are likely to repay their loans is crucial, as these individuals generate profits for the company through interest payments. Rejecting such applicants leads to missed business opportunities.
2. **Risk of Financial Loss**: Approving loans for applicants who are unlikely to repay or are at risk of default results in substantial financial losses.

The objective is to identify applicants at risk of defaulting on loans, enabling LendingClub to minimize credit losses. This case study seeks to achieve this through **Exploratory Data Analysis (EDA)** using the provided dataset.

Key Goals

- Understand the driving factors (driver variables) behind loan defaults.
- Pinpoint variables that strongly indicate the likelihood of default.
- Use insights from the analysis to strengthen LendingClub's portfolio and risk assessment strategies.

By identifying these factors, LendingClub can make data-driven decisions to manage its loan approval process more effectively and ensure a balance between minimizing risks and maximizing profits.

Analysis of Lending Club Dataset

Lending Club provided historical customer data, which included details about borrowers past credit histories and information related to Lending Club loans. The dataset comprised over **39,717 records** and **111 columns**, offering a robust foundation for our analysis.

Data Exploration

The dataset contained a wealth of variables, providing extensive information to help identify relationships and understand their impact on a borrower's likelihood of successfully meeting loan repayment terms.

Data Preparation

To focus our analysis on factors influencing loan defaults, we selected only the variables directly or indirectly related to a borrower's potential to default. This involved carefully filtering and preparing the data to ensure the most relevant variables were used in the analysis, enabling a more precise assessment of default risks.

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:

FullyPaid-
Thiscategoryrepresentsapplicantswhohavesuccessfullyrepaidboththeprincipalandtheinterestrateoftheloan.

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

Charged-off-This classification pertains to applicants who have failed to make timely installments 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 aspredictors. 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 thesecolumns 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 utilized for creating a generalized demographic analysis. It may reveal demographic trends related to delinquency or default rates.

Data Understanding

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.
- ✓ **Installment (installment):**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 categorization of the columns we will exclude from our approach, and it does not represent an exhaustive list.

- ✓ **Customer Behavior Columns**-Columns that describe customer behavior will not be factored into our analysis. The current analysis focuses on the loan application stage, while customer behavior 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 visualizations, the "sub grade" columnisexcessively 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**

Data Understanding

- ✓ 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- processings

Data Cleaning & Pre- processing

- **Loading data from loan CSV**
- **Checking for null values in the dataset**
- **Checking for unique values**
- **Checking for duplicated rows in data**
- **Dropping Records & Columns**
- **Common Functions**
- **Data Conversion**
- **Outlier Treatment**
- **Imputing values in Columns**

Data Cleaning & Pre- processing

- **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.
- **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.
- **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.
- **Checking for duplicated rows in data: No duplicate rows were found.**
- **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

- **Common Functions:** Common functions were created for repeating common operations like plotting bar graphs, box plots, histograms, countplots, binning etc.
- **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).**
- **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 madeThe 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 installment amount on the loan is between 160 -440
  - ✓ The debt to income ration is between 8 -18

Data Cleaning & Pre- processing

- **Imputing values in Columns:**
  - ✓ **Replaced missing values of annual_inc with the corresponding mode value of annual_inc of the emp_lengthannual_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 **660null 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 **36094 rows × 18 columns.**

Univariant Analysis

Univariant Analysis

- **Univariate analysis** is a statistical method used to analyze and summarize 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

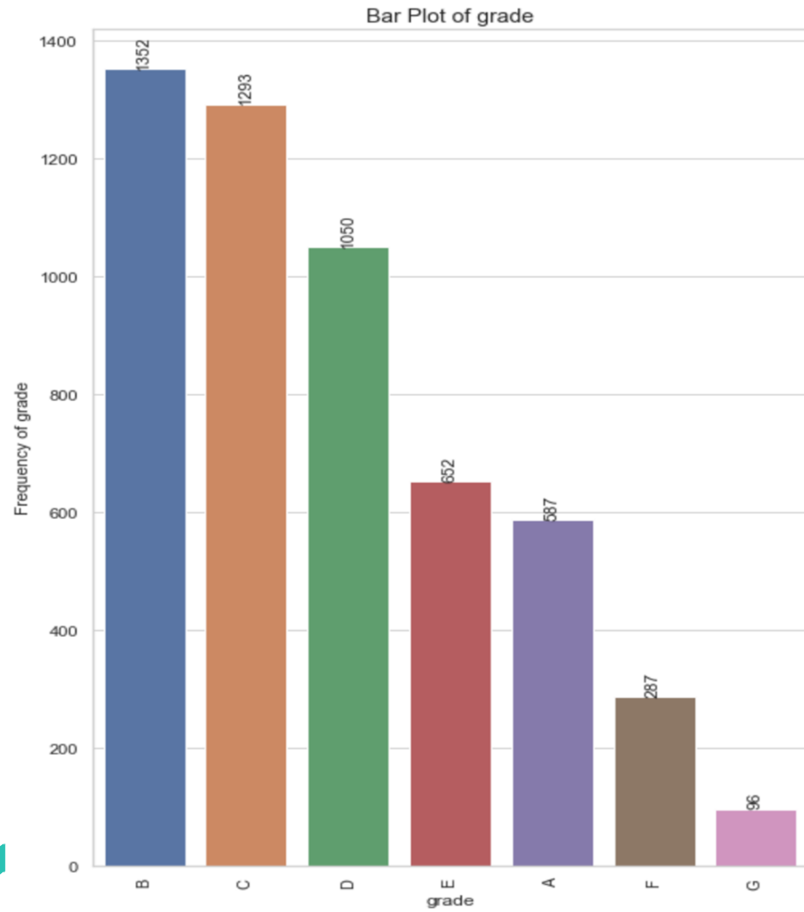
A. Categorical Variables:

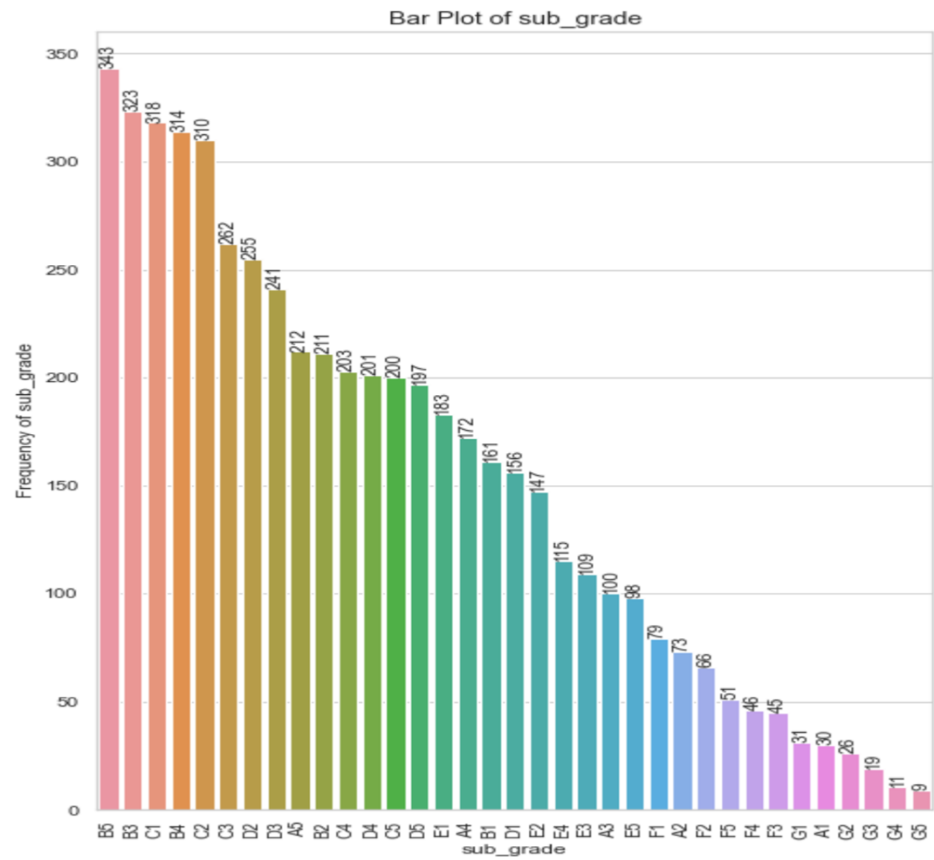
Ordered	Unordered
✓ Grade (grade)	✓ Address State (addr_state)
✓ Sub grade (sub_grade)	✓ Loan purpose (purpose)
✓ Term (36 / 60 months) (term)	✓ Home Ownership (home_ownership)
✓ Employment length (emp_length)	✓ Loan status (loan_status)
✓ Issue year (issue_y)	✓ Loan paid (loan_paid)
✓ Issue month (issue_m)	
✓ Issue quarter (issue_q)	

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 Installment (installment)**

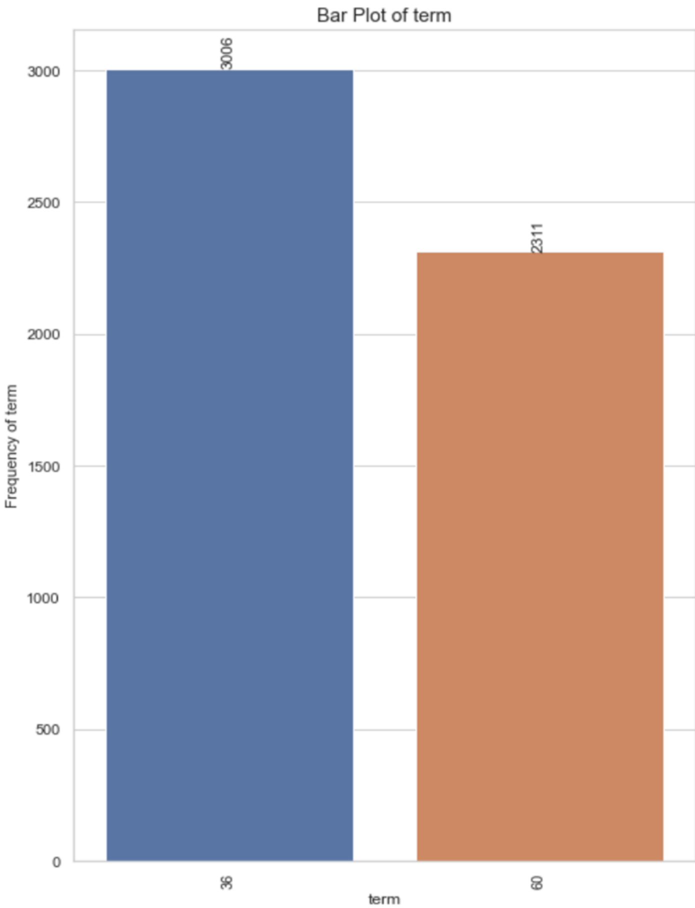
Univariant Analysis (Unordered Categorical)

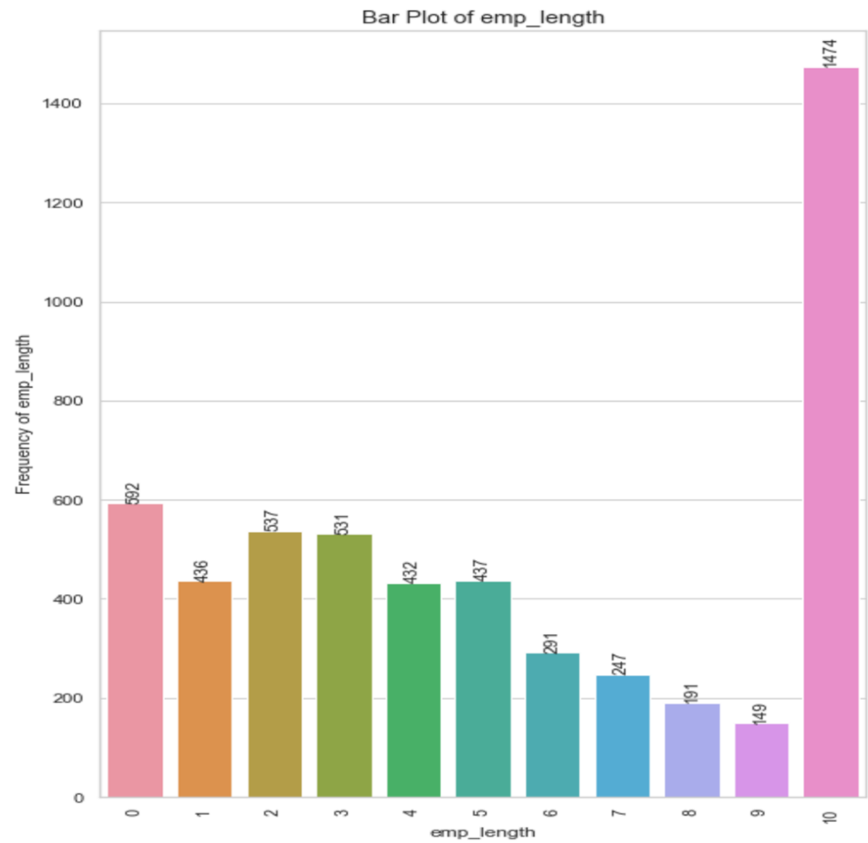




Grade and Sub-Grade

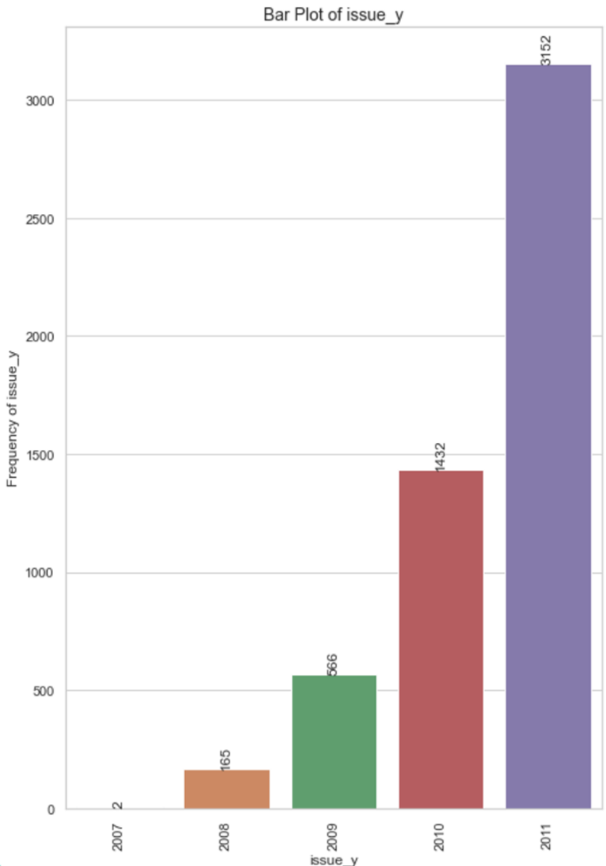
Univariant Analysis (Unordered Categorical)

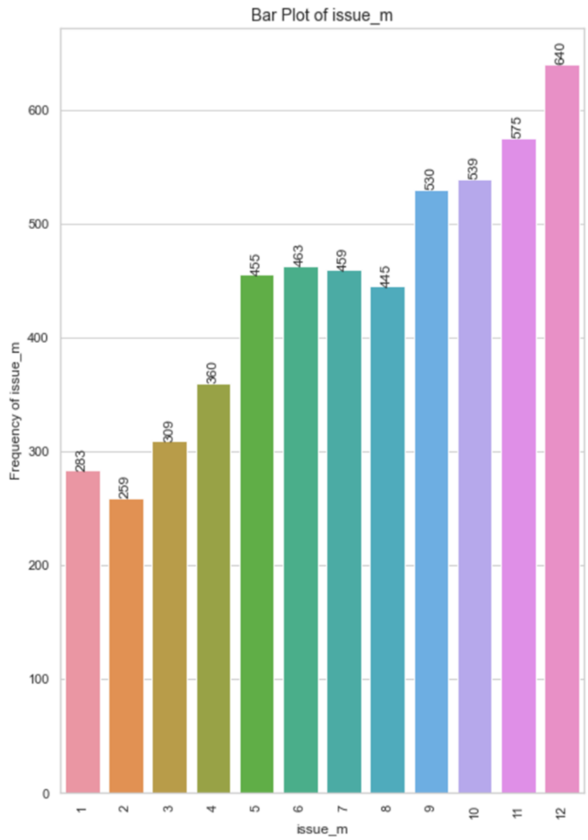




Term and Employment Length

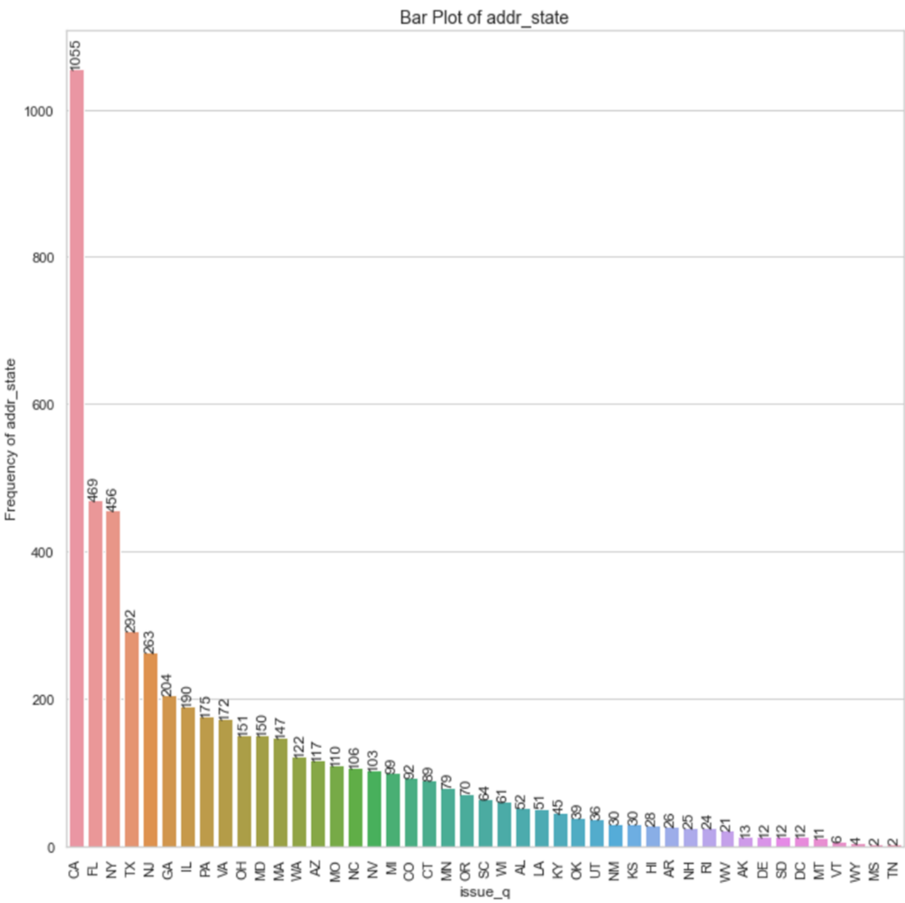
Univariant Analysis (Unordered Categorical)

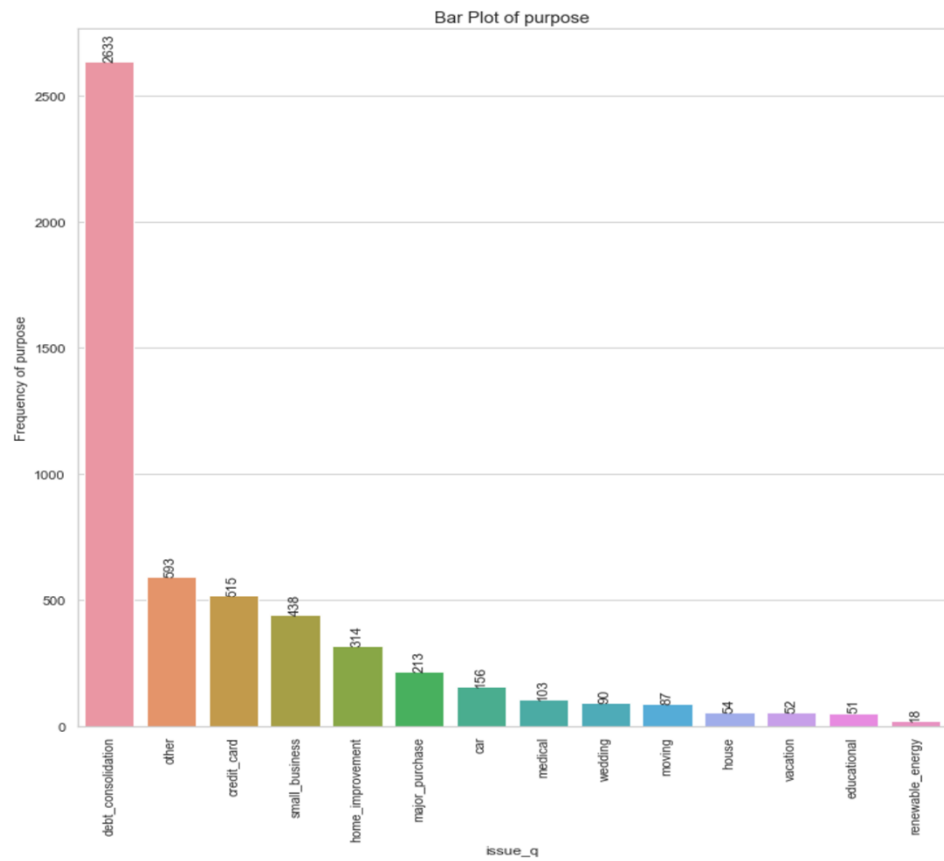




Term and Employment Length

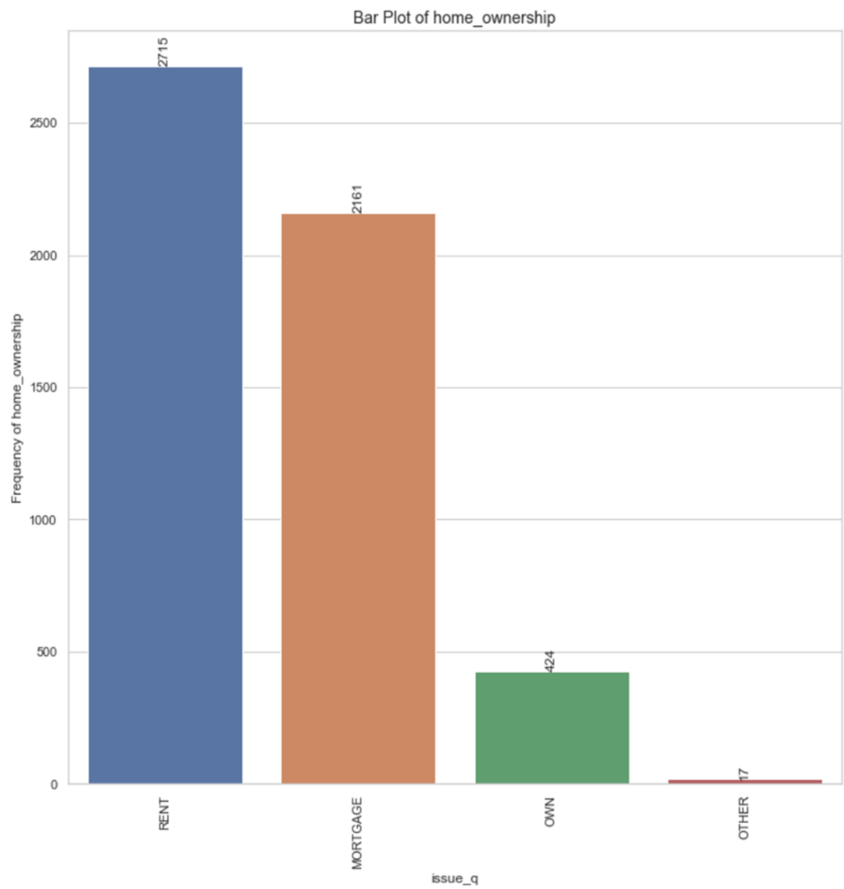
Univariant Analysis (Unordered Categorical)





Address State and Purpose of Loan

Univariant Analysis (Unordered Categorical)



Various types of Home Ownerships

Univariant Analysis (Categorical Variables)

Observations & Inferences:

A. Ordered Categorical Variables:

•**Grade B** had the highest number of "Charged Off" loan applicants, totaling **1,352**. This indicates that borrowers within this credit grade faced significant 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 defaulting borrowers chose shorter repayment terms.

•Applicants with **more than 10 years of employment** accounted for the highest number of "Charged Off" loans, totaling **1,474**. This highlights that a long-term employment history does not necessarily guarantee successful loan repayment.

•The year **2011** recorded the highest number of "Charged Off" loan applications, reaching **3,152**. This trend may reflect economic or financial challenges faced by borrowers during that period.

•"Charged Off" loans were predominantly taken during the **4th quarter**, with **2,284 applications**, primarily in **December**. This seasonal peak suggests that financial pressures during the holiday season may have contributed to loan defaults.

Univariant Analysis (Categorical Variables)

Observations & Inferences:

B. UnOrdered Categorical Variables:

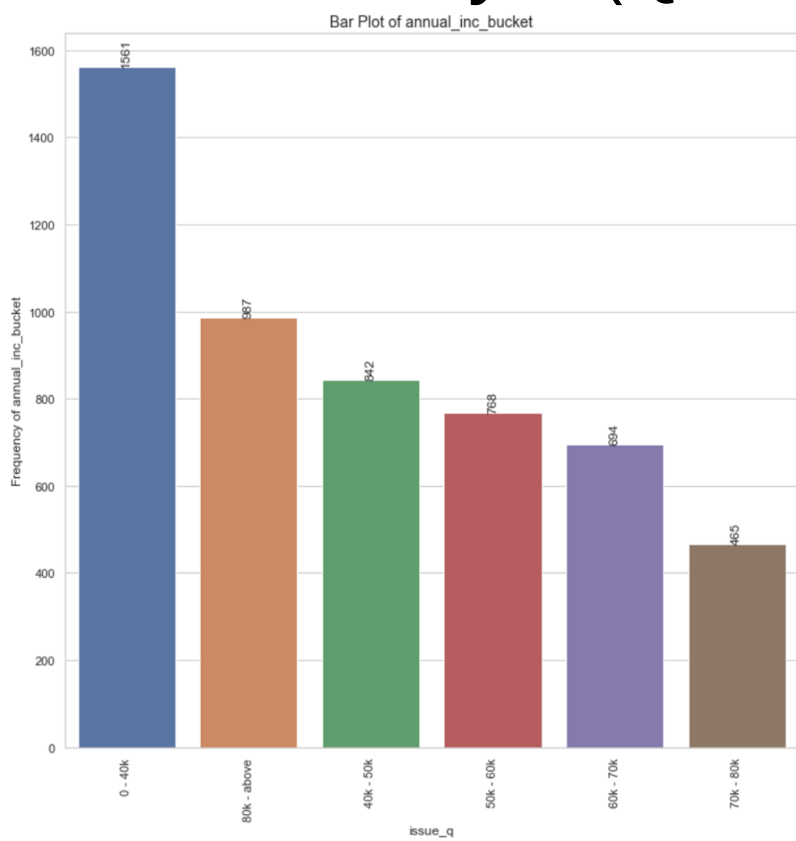
•**California** had the highest number of "Charged Off" loan applicants, with **1,055 applicants**. To address this, the lending company should consider implementing stricter eligibility criteria or enhanced credit assessments for applicants from this state due to the higher default rate.

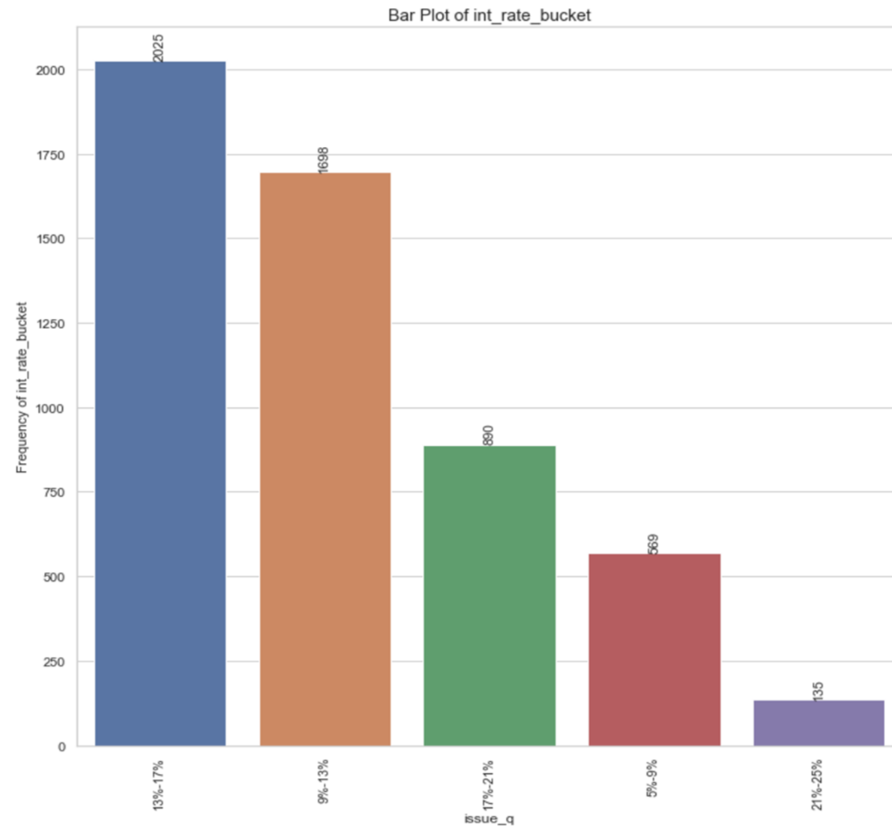
•**Debt consolidation** was the primary loan purpose for most "Charged Off" applicants, with **2,633 borrowers** selecting this option. The lending company should exercise caution when approving loans for debt consolidation purposes, as it represents a significant proportion of defaults.

•The majority of "Charged Off" loan participants, totaling **2,715 individuals**, lived in **rented houses**. The lending company should evaluate the financial stability of applicants living in rented accommodations, as they may be more vulnerable to economic fluctuations.

•A significant number of loan participants, specifically **5,317 individuals**, defaulted on their loans. The lending company should enhance its risk assessment practices, including stricter credit checks and lower loan-to-value ratios, for applicants with a history of loan defaults. Additionally, offering **financial education and support services** could help borrowers manage their finances better and improve repayment outcomes.

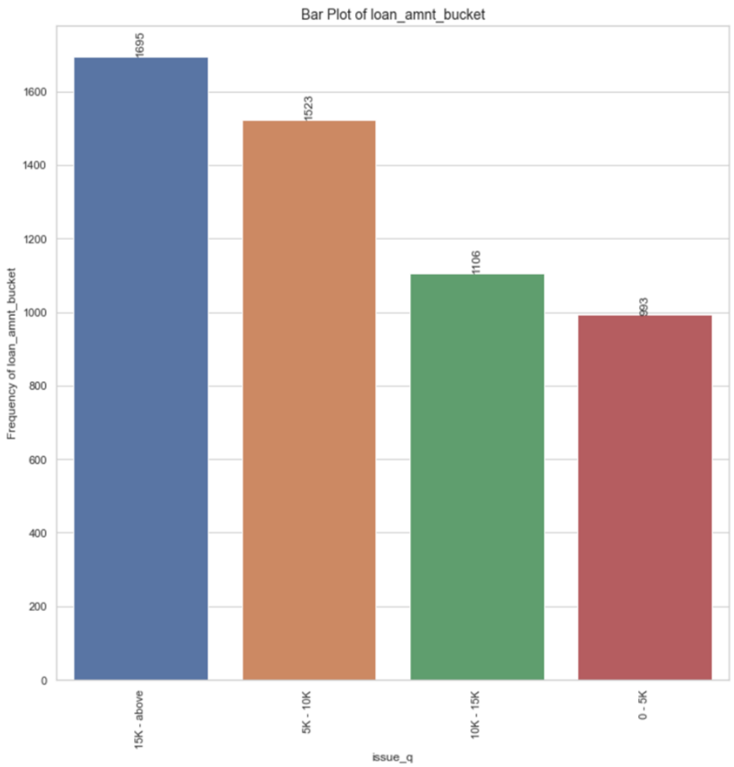
Univariant Analysis (Quotative Variable)

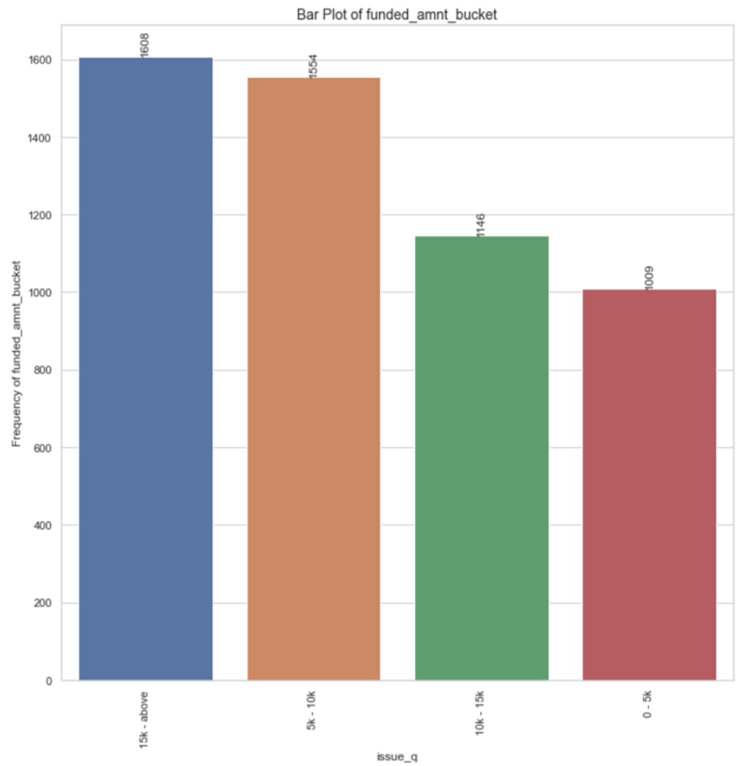




Buckets of Annual Income Status and Loan Interest Rates

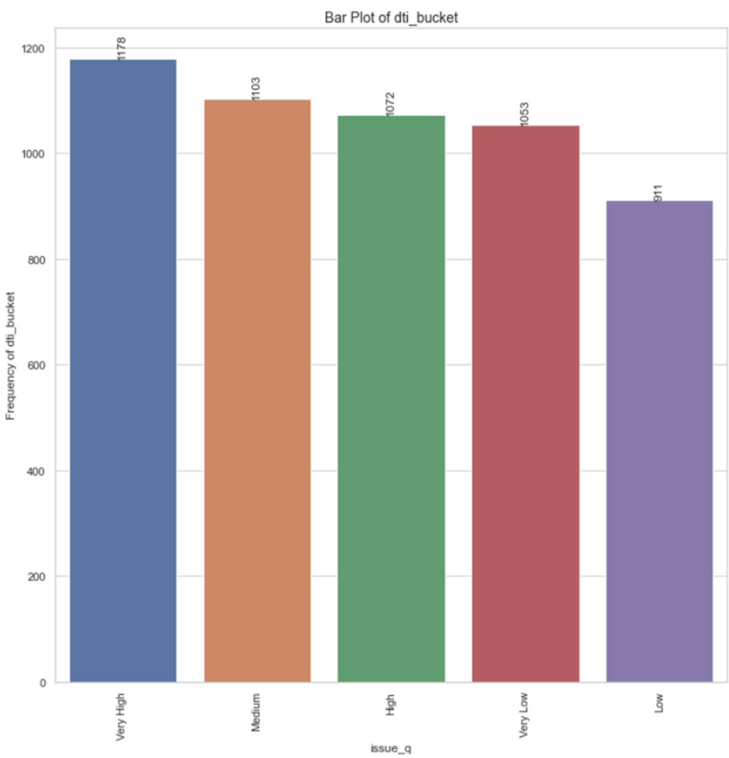
Univariant Analysis (Quotative Variable)



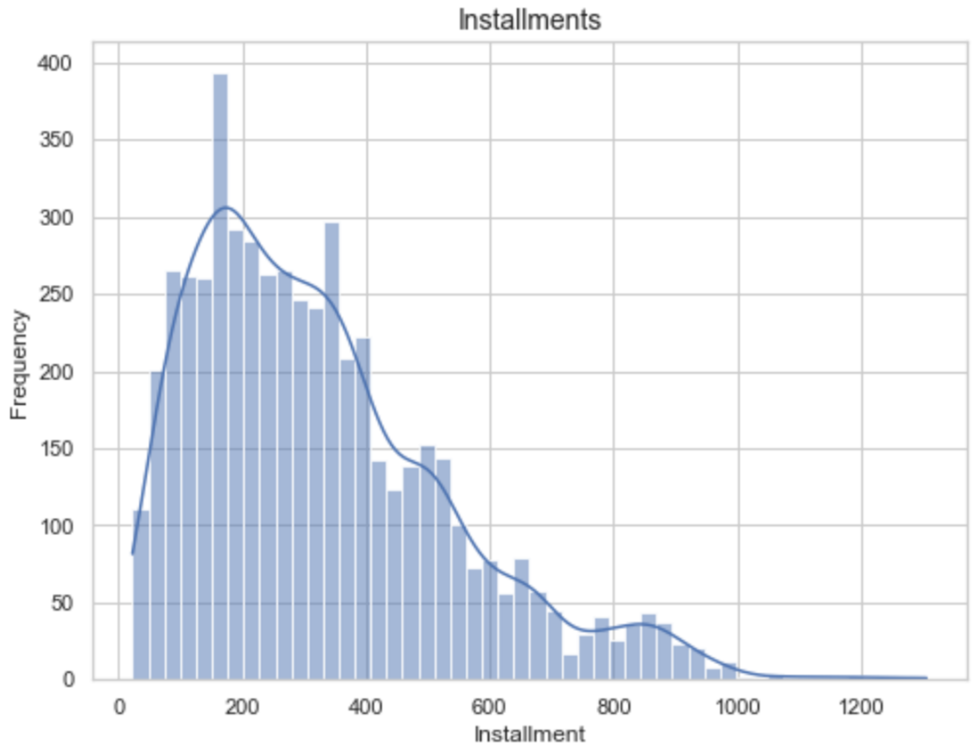


Buckets of Loan Amount and Funded Amount

Univariant Analysis (Quotative Variable)



Bucket of Debt to Income Ratio (DTI)



Histogram of Installment (For Defaulted Loans)

Univariant Analysis (Quotative Variable)

Observations & Inferences:

- ✓ 1,561 loan applicants who charged off had annual salaries less than 40,000 USD. The lending company should exercise caution when lending to individuals with low annual salaries. They should implement rigorous income verification and assess repayment capacity more thoroughly for applicants in this income bracket.
- ✓ Among loan participants who charged off (2,025), a considerable portion belonged to the interest rate bucket of 13%-17%.
- ✓ To reduce the risk of default, the lending company should consider offering loans at lower interest rates when possible.
- ✓ 1,695 loan participants who charged off received loan amounts of 15,000 USD and above. The lending company should evaluate applicants seeking higher loan amounts carefully. They should ensure the applicants must have a strong credit history and repayment capability to handle larger loans.
- ✓ 1,608 loan participants who charged off received funded amounts of 15,000 USD and above. The lending company should ensure that the funded amounts align with the borrower's financial capacity. They should conduct thorough credit assessments for larger loan requests.
- ✓ Among loan participants who charged off, 1,178 loan applicants had very high debt-to-income ratios. The lending company should implement strict debt-to-income ratio requirements to prevent lending to individuals with unsustainable levels of debt relative to their income.
- ✓ Among loan participants who charged off, it's observed that the majority of them had monthly installment amounts falling within the range of 160-440 USD. The lending company should closely monitor and assess applicants with similar installment amounts to mitigate the risk of loan defaults.

Bivariant Analysis

- ✓ Bivariate analysis is a statistical method that involves the simultaneous analysis of two variables (factors). It aims to determine the empirical relationship between them. The analysis can be used to test hypotheses, identify patterns, or explore relationships between the variables.
- ✓ It was carried out for both Categorical and Quantitative Variables

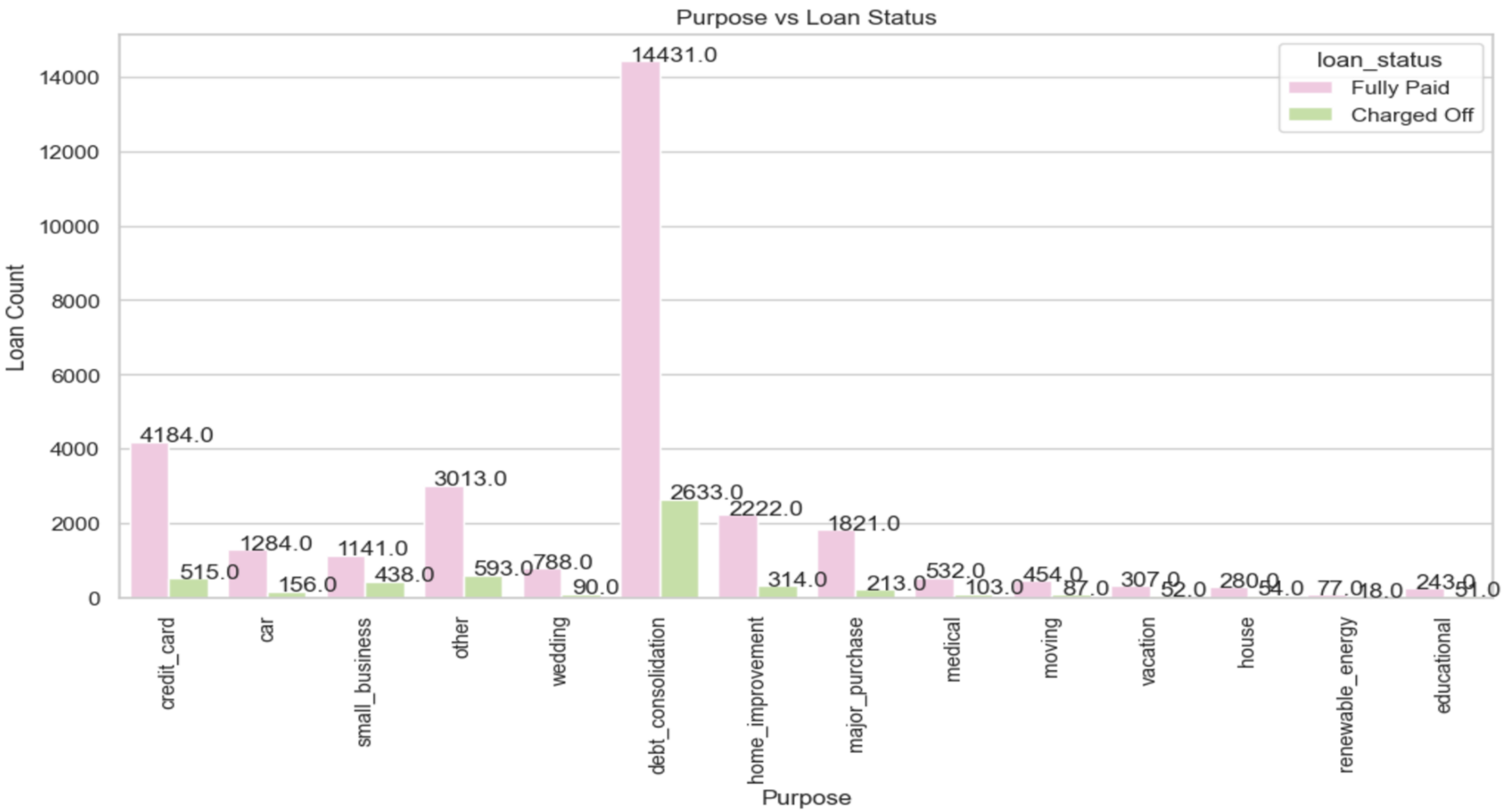
A. Categorical Variables:

Ordered	Unordered
✓ 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)	✓ Loan purpose (purpose) ✓ Home Ownership (home_ownership) ✓ Verification Status (verification_status) ✓ Address State (addr_state)

B. Quantitative Variables:

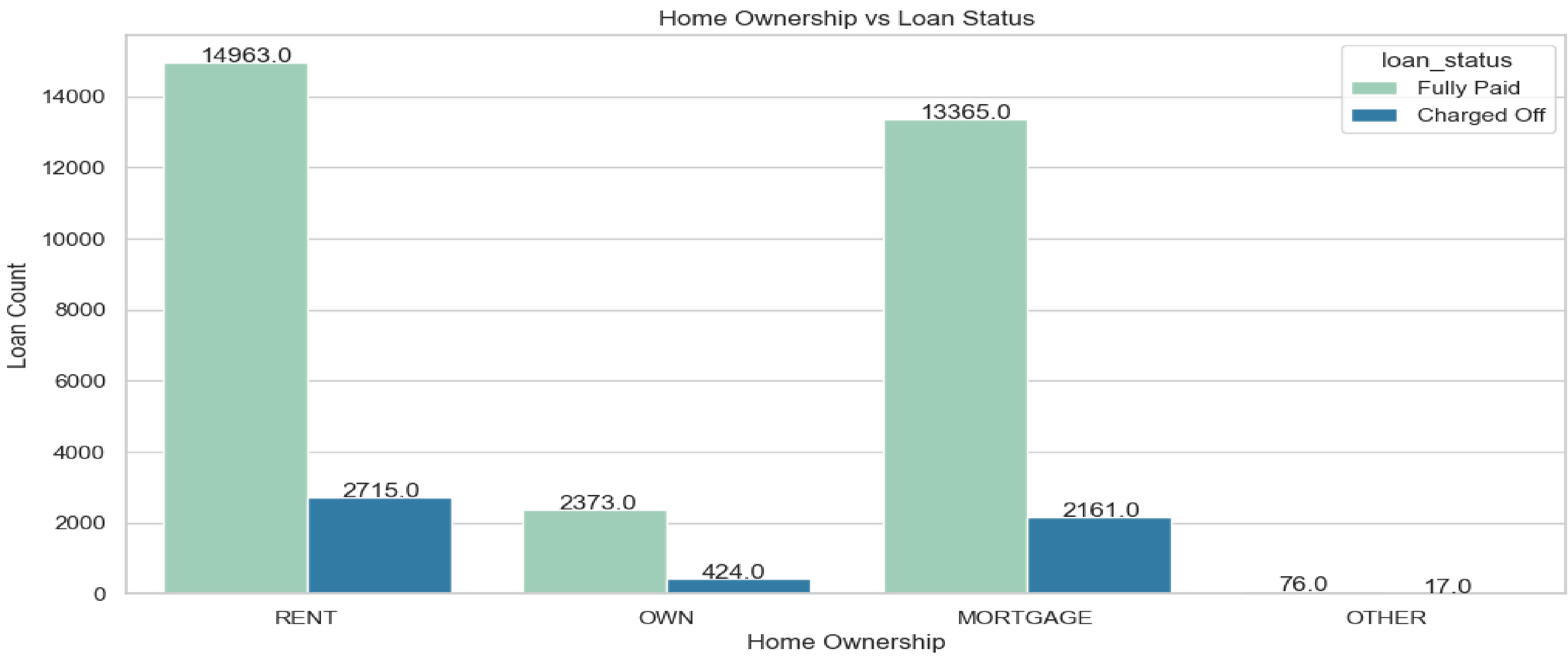
- ✓ **Int Rate Bucket (int_rate_bucket)**
- ✓ **Debt to Income Bucket (dti_bucket)**
- ✓ **Annual Income Bucket (annual_inc_bucket)**
- ✓ **Funded Amount Bucket (funded_amnt_bucket)**
- ✓ **Loan Amount Bucket (loan_amnt_bucket)**

Bivariant Analysis (Unordered Categorical)



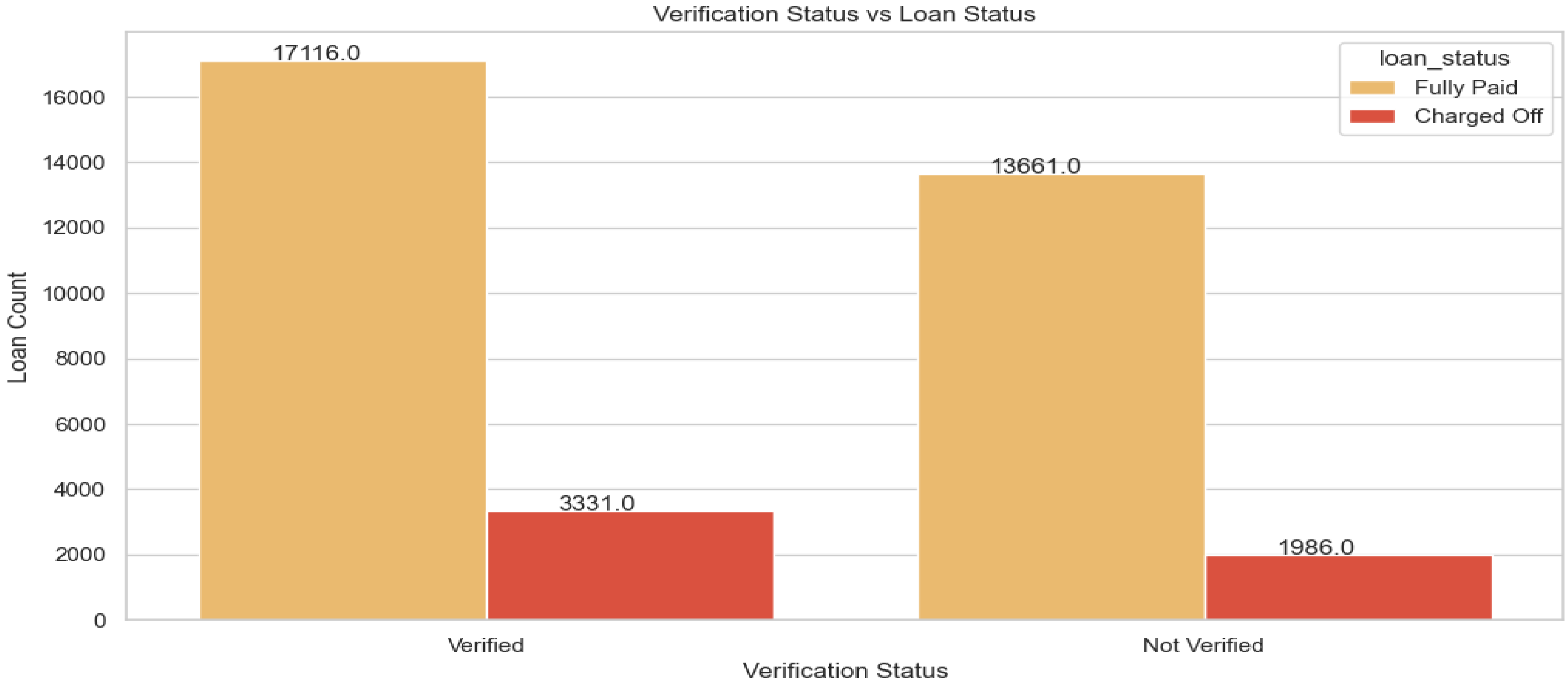
Purpose of Loan v/s Status of Loan

Bivariant Analysis (Unordered Categorical)



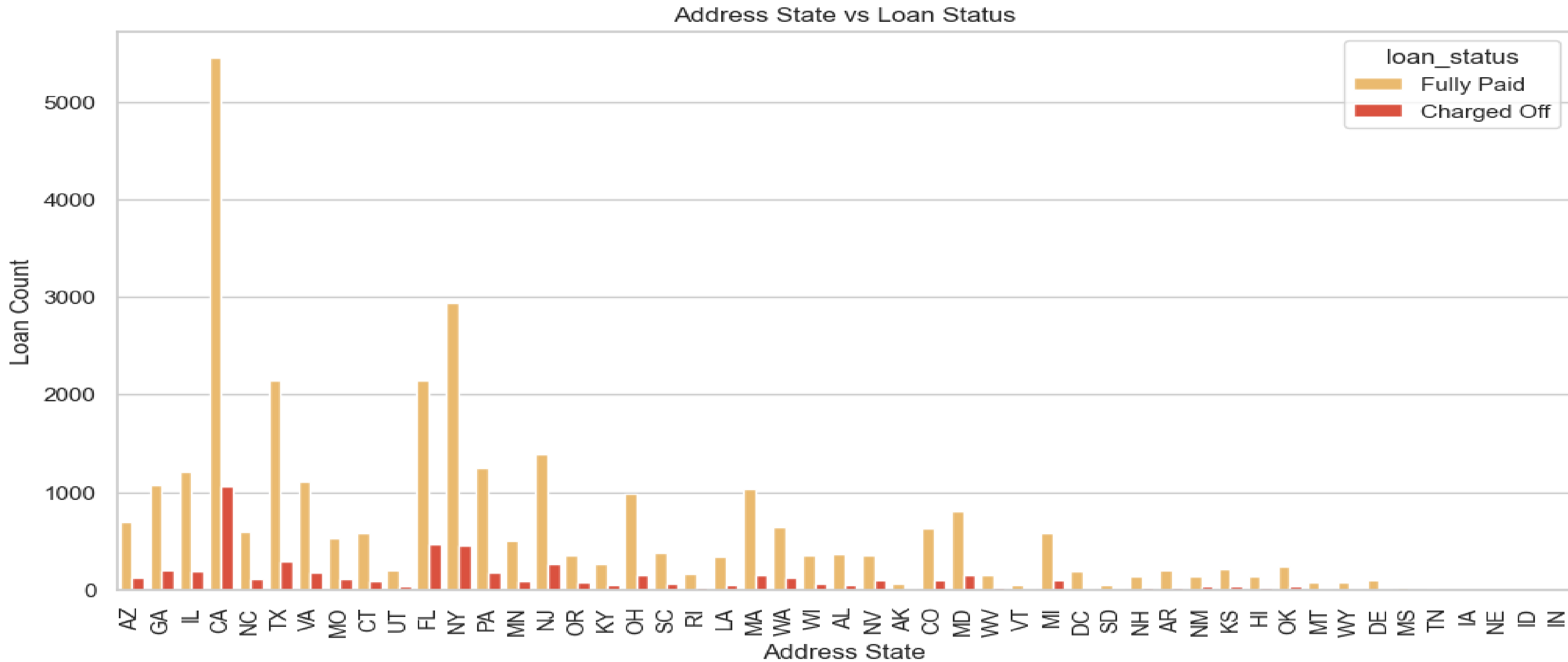
Home Ownership v/s Status of Loan

Bivariant Analysis (Unordered Categorical)



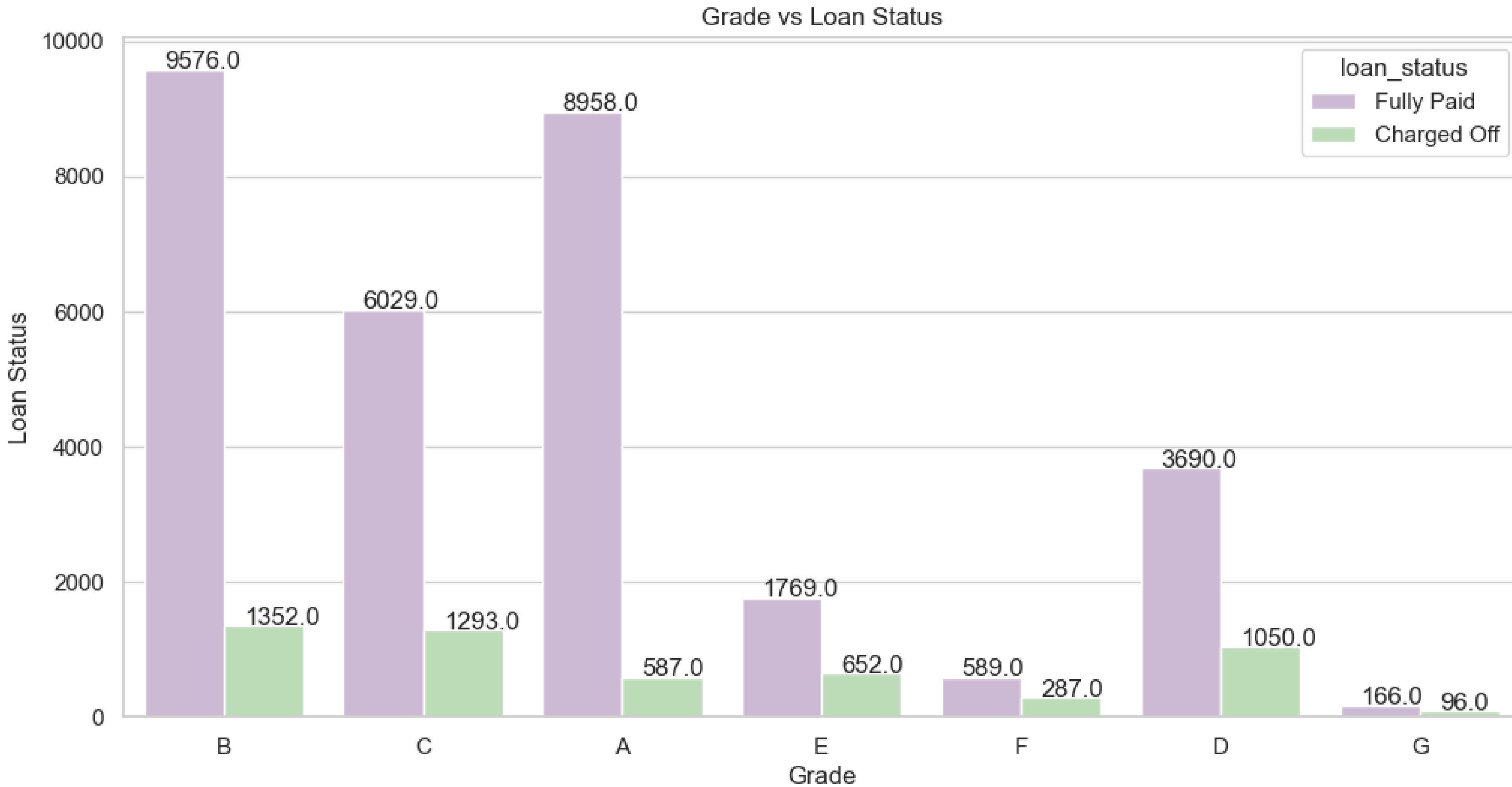
Verification Status of Loan v/s Status of Loan

Bivariant Analysis (Unordered Categorical)



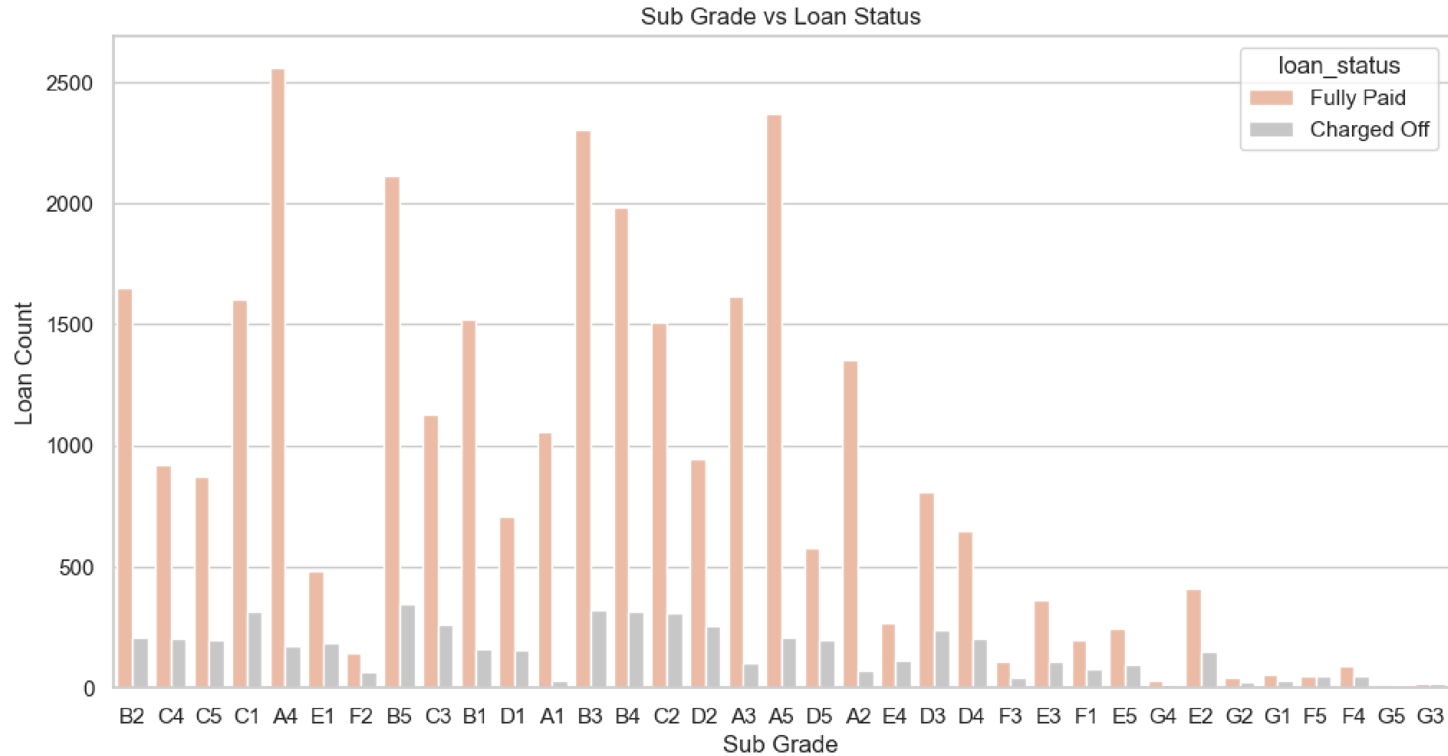
Address State v/s Status of Loan

Bivariant Analysis (Ordered Categorical)



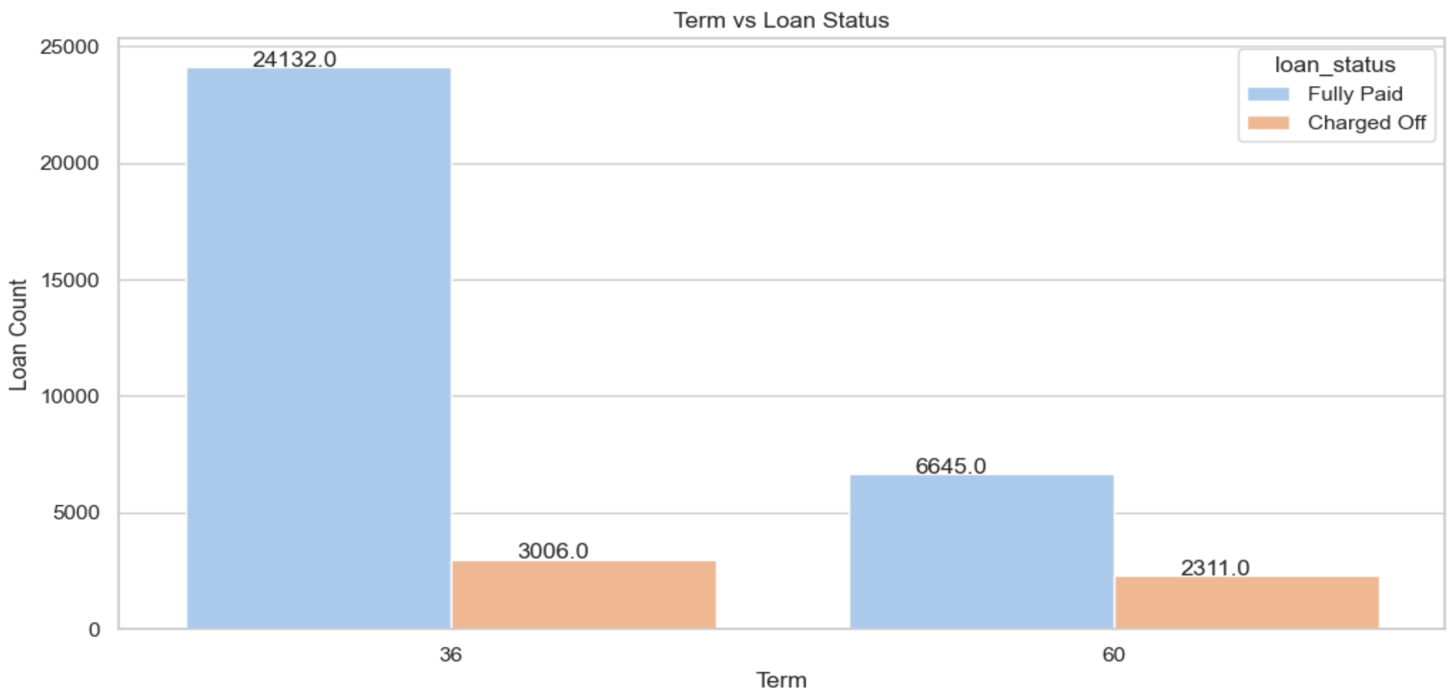
Loan Grade v/s Status of Loan

Bivariant Analysis (Ordered Categorical)



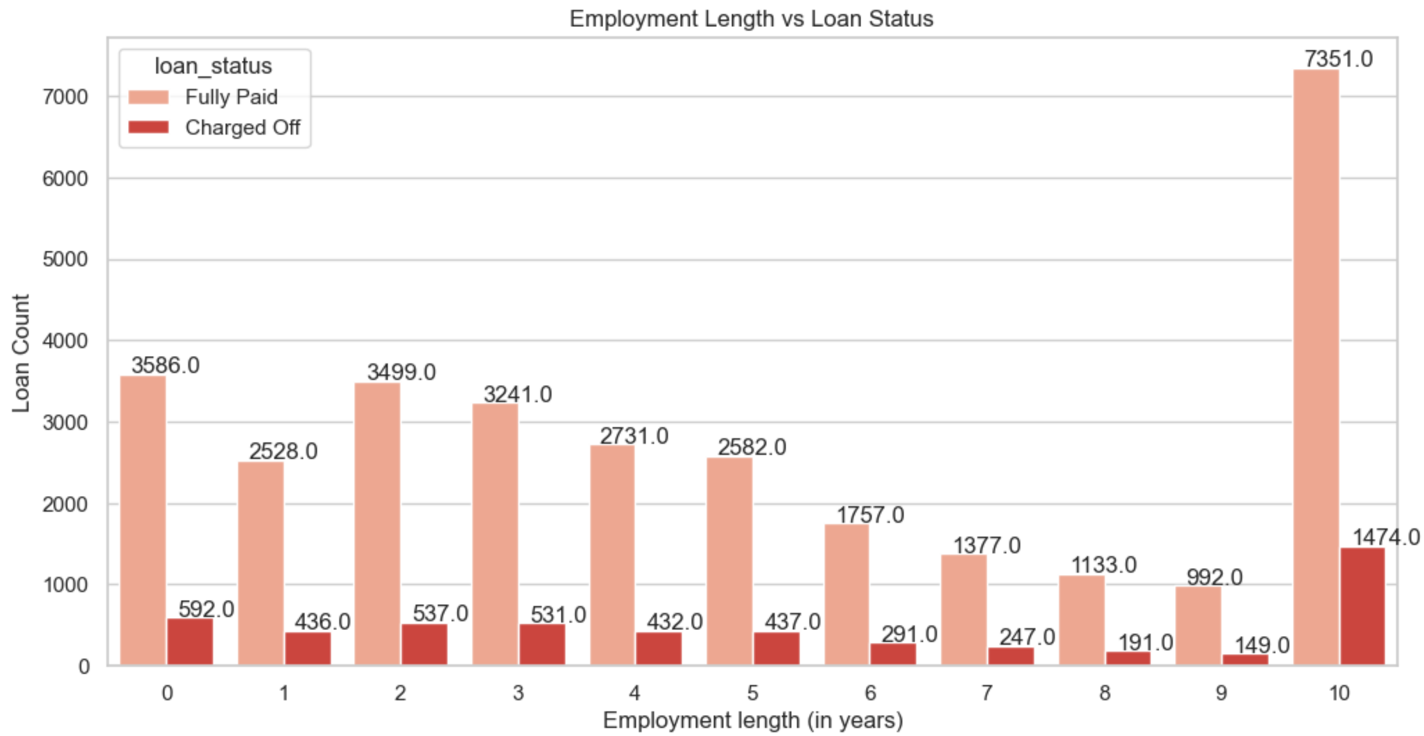
Loan Sub-Grade v/s Status of Loan

Bivariant Analysis (Ordered Categorical)



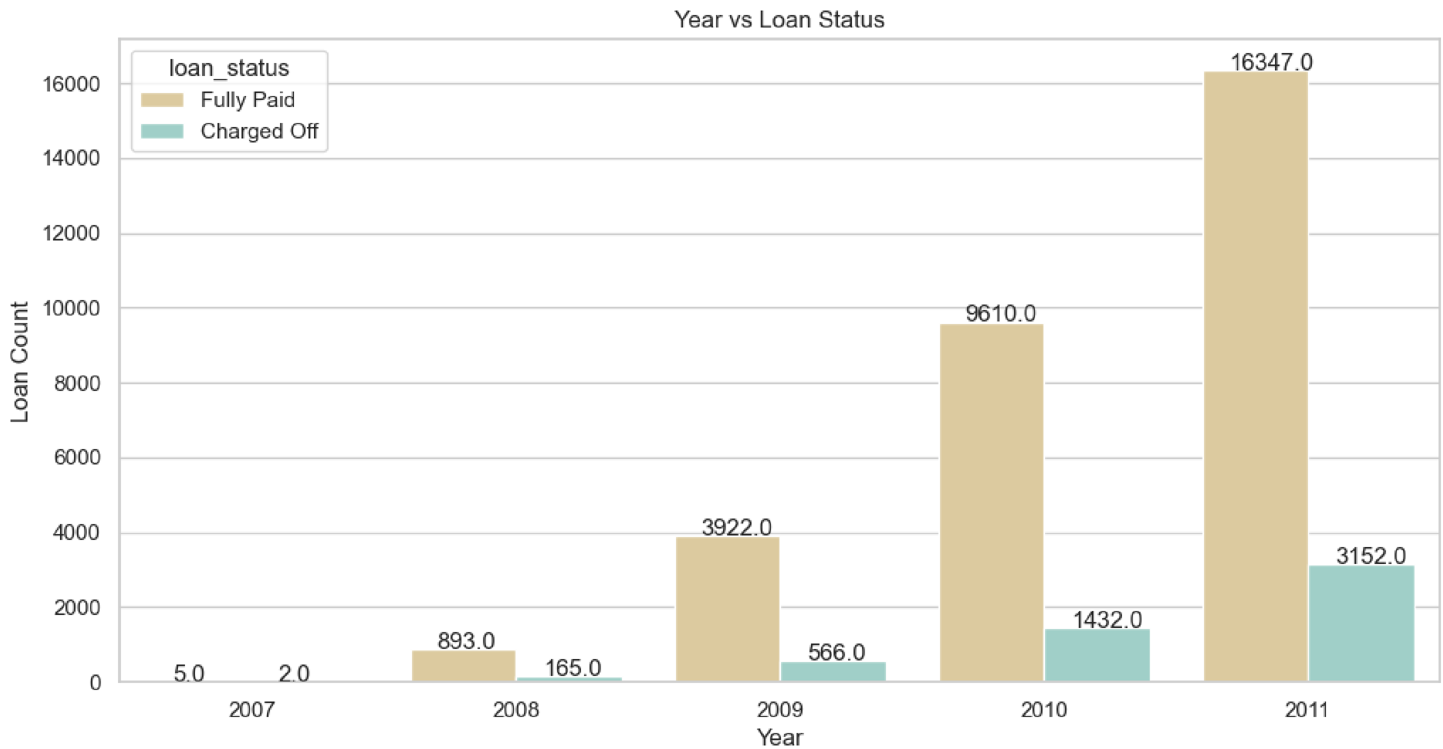
Term of Loan v/s Status of Loan

Bivariant Analysis (Ordered Categorical)



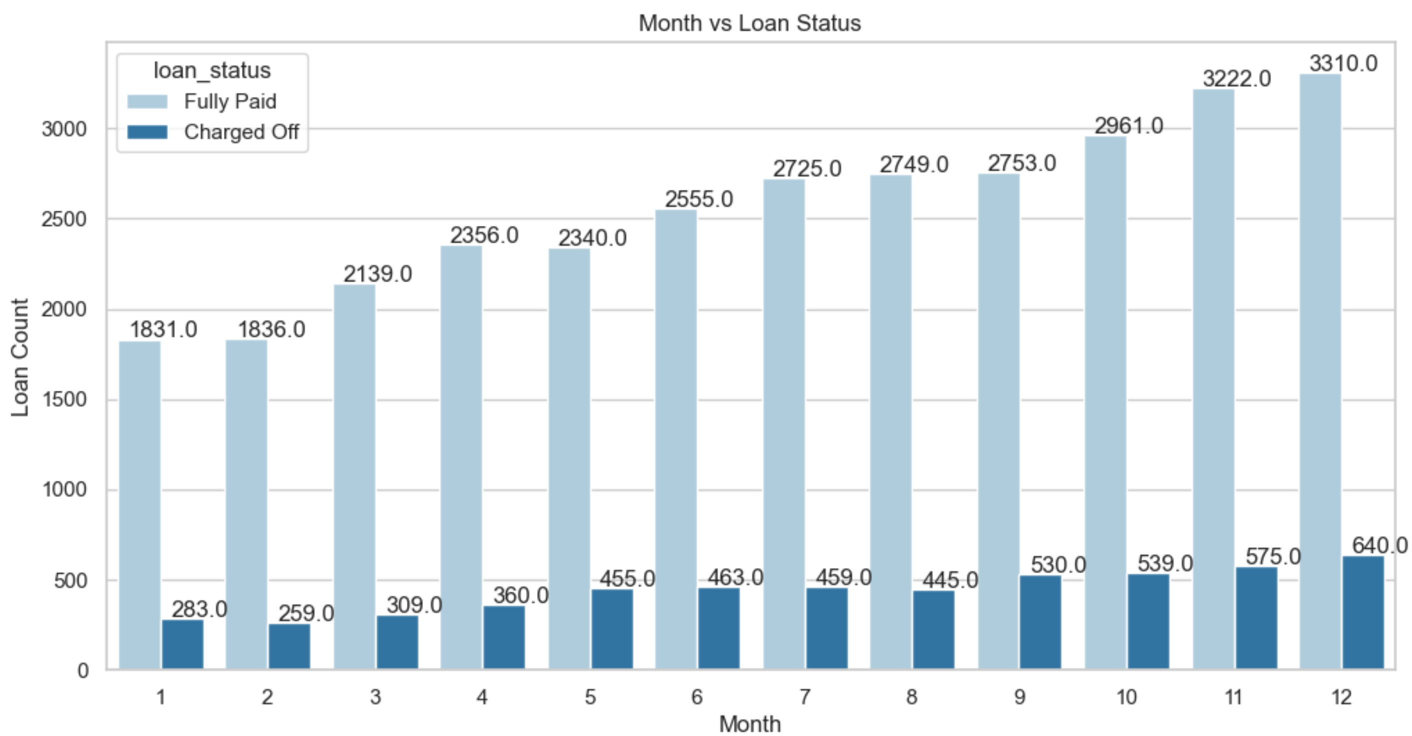
Employment Length of Customer
v/sStatus of Loan

Bivariant Analysis (Ordered Categorical)



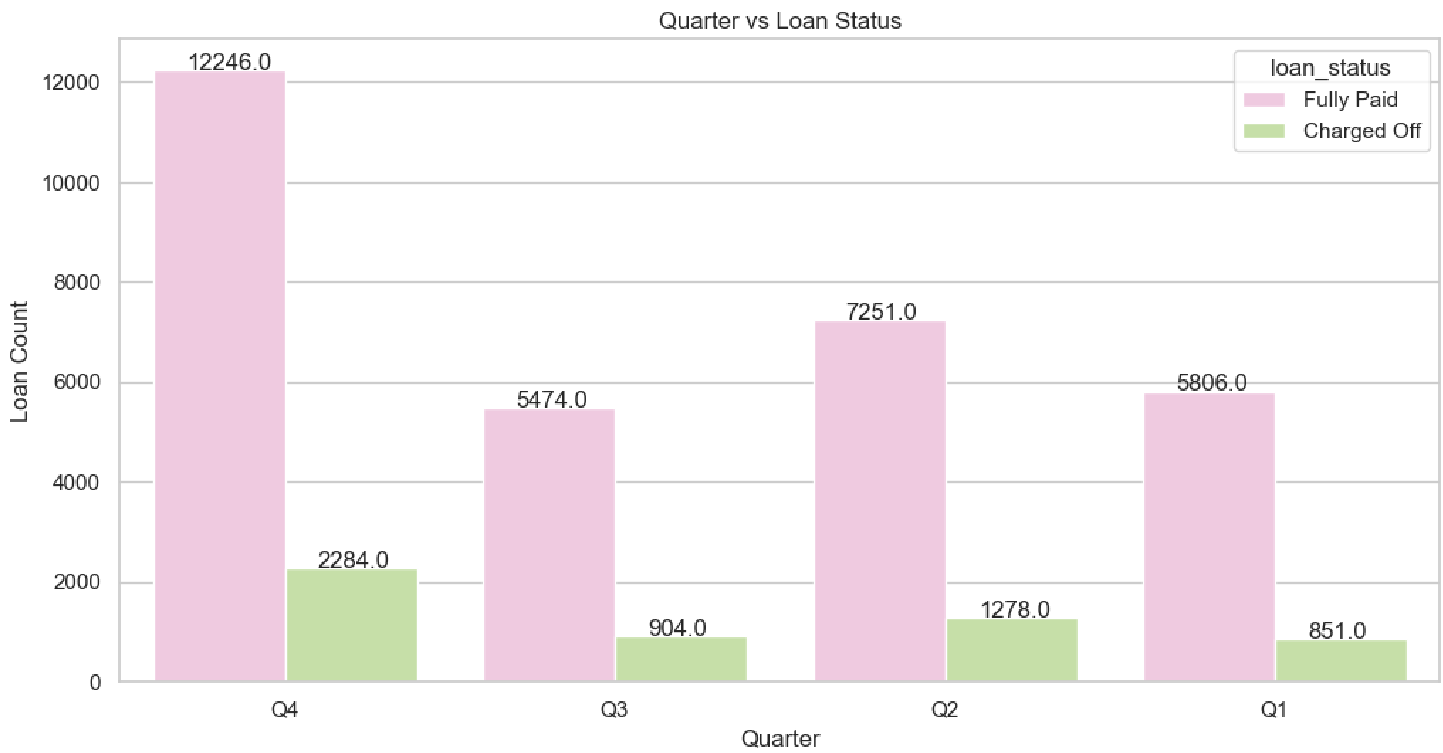
Year the Loan was given to Customer v/s Status of Loan

Bivariant Analysis (Ordered Categorical)



Month during which the Loan was given to Customer v/s Status of Loan

Bivariant Analysis (Ordered Categorical)



Quarter during which the Loan was given to Customer v/s Status of Loan

Bivariant Analysis (Categorical Variable)

Observations:

A. Ordered Categorical Variables:

- ✓ The loan applicants belonging to Grades B, C, and D contribute to most of the "Charged Off" loans.
- ✓ Loan applicants belonging to Sub Grades B3, B4, and B5 are more likely to charge off.
- ✓ Loan applicants applying for loans with a 60-month term are more likely to default than those taking loans for 36 months.
- ✓ Most loan applicants have ten or more years of experience, and they are also the most likely to default.
- ✓ The number of loan applicants has steadily increased from 2007 to 2011, indicating a positive trend in the upcomingyears.
- ✓ December is the most preferred month for taking loans, possibly due to the holiday season.
- ✓ The fourth quarter (Q4) is the most preferred quarter for taking loans, primarily because of the upcoming holiday season.

B Unordered Categorical Variables:

- ✓ Debt consolidation is the category where the maximum number of loans are issued, and people have defaulted the most in the same category.
- ✓ Loan applicants who live in rented or mortgaged houses are more likely to default.
- ✓ Verified loan applicants are defaulting more than those who are not verified.
- ✓ Loan applicants from the states of California (CA), Florida (FL), and New York (NY) are most likely to default.

Bivariant Analysis (Categorical Variable)

Inferences:

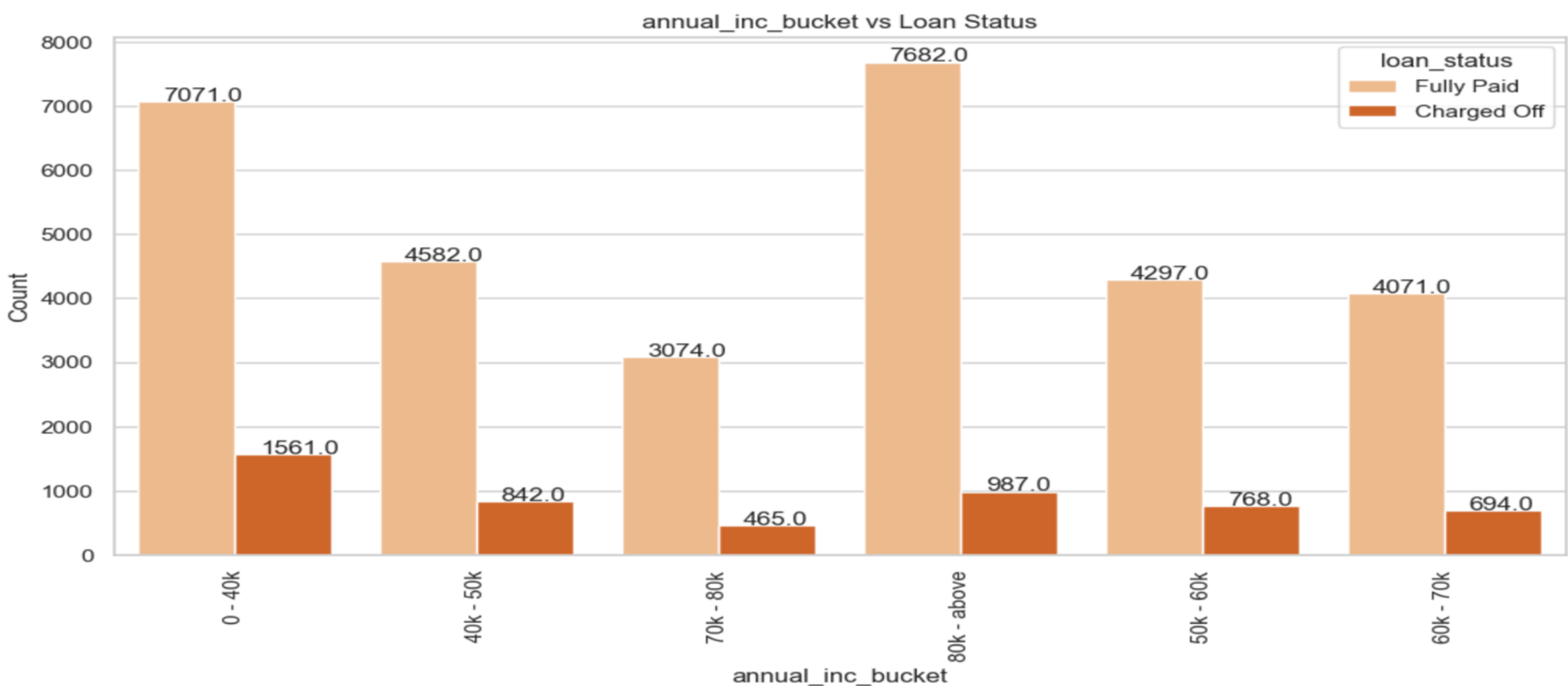
A. Ordered Categorical Variables:

- ✓ **Risk Assessment for Grades B, C, and D:** Since loan applicants from Grades B, C, and D contribute to most of the "Charged Off" loans, the company should consider implementing stricter risk assessment and underwriting criteria for applicants falling into these grades.
- ✓ **Subgrades B3, B4, and B5:** Pay special attention to applicants with Subgrades B3, B4, and B5, as they are more likely to charge off. Implementing additional risk mitigation measures or offering them lower loan amounts could be considered.
- ✓ **Term Length**: Given that applicants opting for 60-month loans are more likely to default, the company should consider evaluating the risk associated with longer-term loans and potentially limiting the maximum term or adjusting interest rates accordingly.
- ✓ **Experience and Default Probability:** Loan applicants with ten or more years of experience are more likely to default. This suggests that experience alone may not be a reliable indicator of creditworthiness. The company should use a more comprehensive credit scoring system that factors in other risk-related attributes.
- ✓ **Positive Growth Trend:** The steady increase in the number of loan applicants from 2007 to 2011 indicates growth in the market. The company can capitalize on this trend by maintaining a competitive edge in the industry while keeping risk management practices robust.
- ✓ **Seasonal Trends:** December and Q4 are peak periods for loan applications, likely due to the holiday season. The company should anticipate increased demand during these periods and ensure efficient processing to meet customer needs

B. Unordered Categorical Variables:

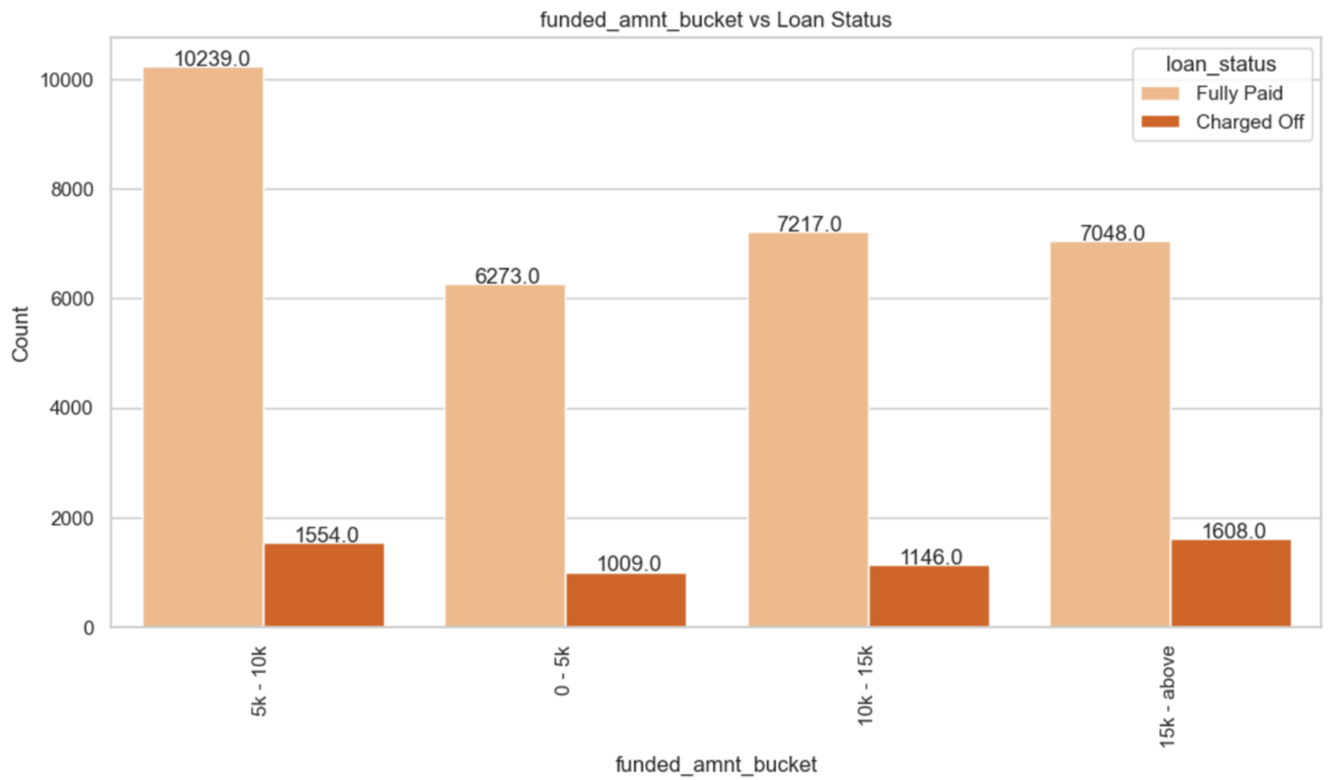
- ✓ **Debt Consolidation Risk**: Since debt consolidation is the category with the maximum number of loans and high default rates, the company should carefully evaluate applicants seeking debt consolidation loans and potentially adjust interest rates or offer financial counseling services.
- ✓ **Housing Status and Default Risk**: Applicants living in rented or mortgaged houses are more likely to default. This information can be considered in the underwriting process to assess housing stability and its impact on repayment ability.
- ✓ **Verification Process**: Verified loan applicants are defaulting more than those who are not verified. The company should review its verification process to ensure it effectively assesses applicant creditworthiness and consider improvements or adjustments.
- ✓ **Geographic Risk**: Loan applicants from states like California (CA), Florida (FL), and New York (NY) are more likely to default. The company should monitor regional risk trends and adjust lending strategies or rates accordingly in these areas.

Bivariant Analysis (Quotative Variables)

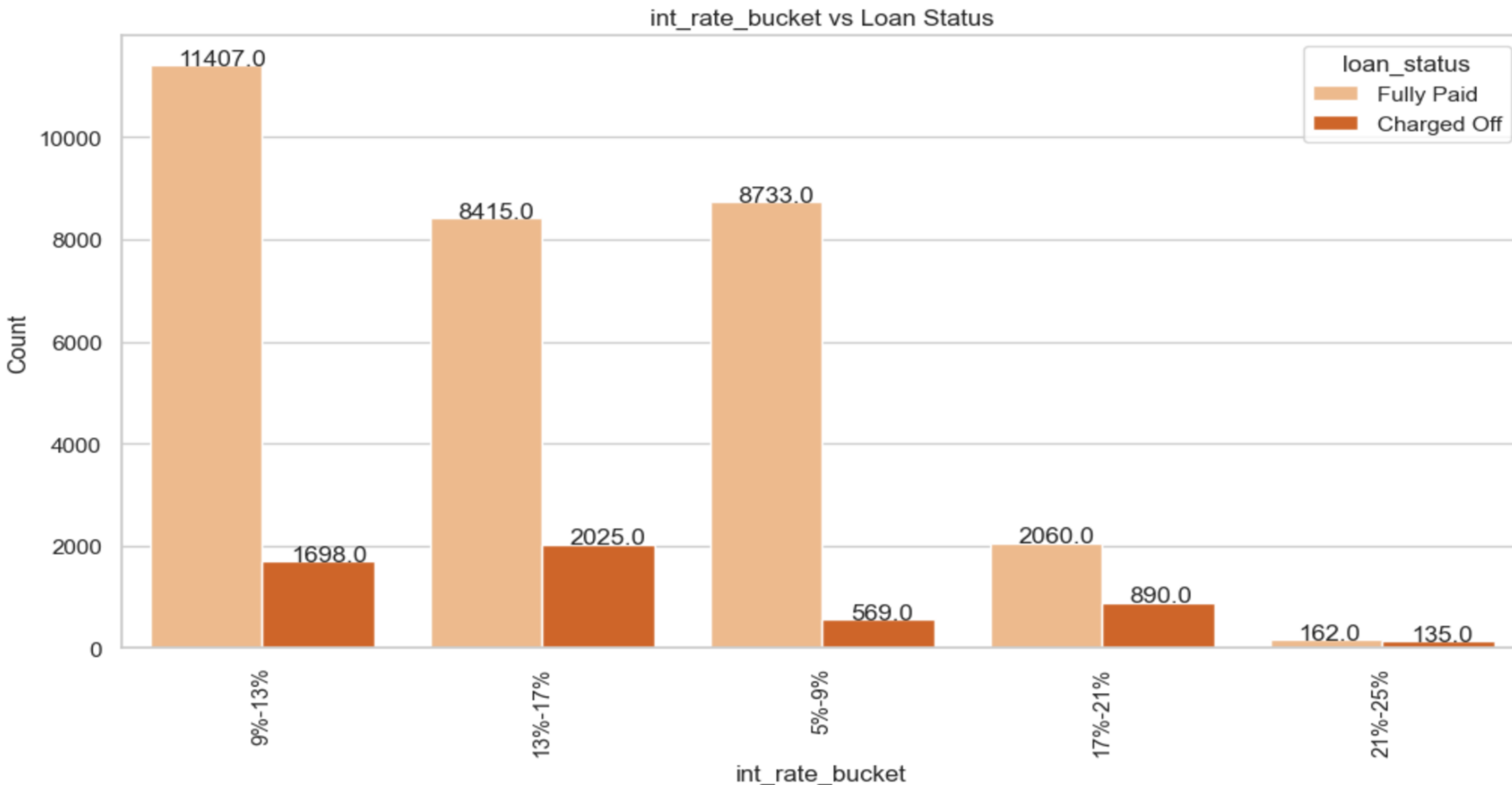


Bucket of Annual Income v/s Status of Loan

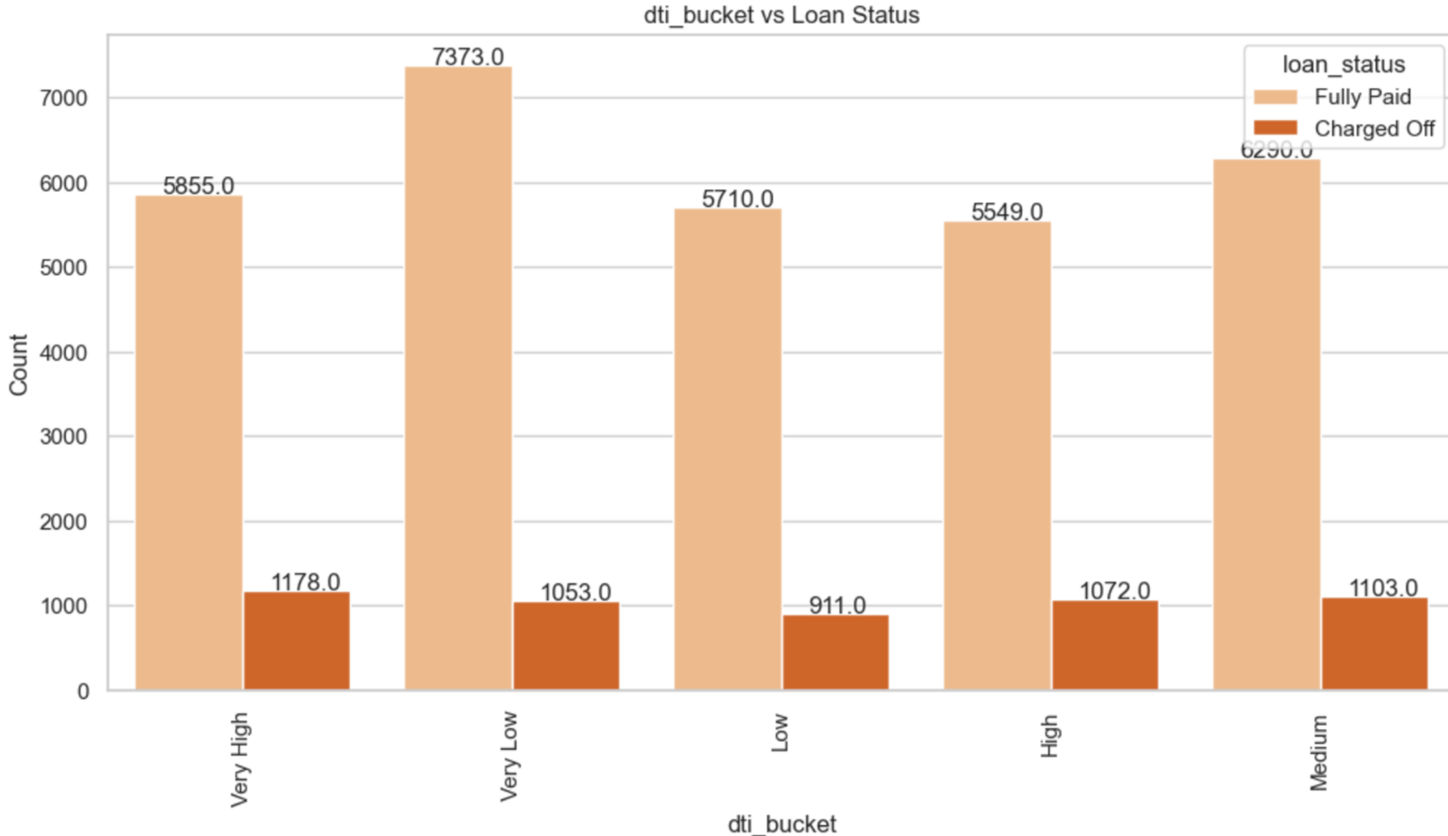
Bivariant Analysis (Quotative Variables)



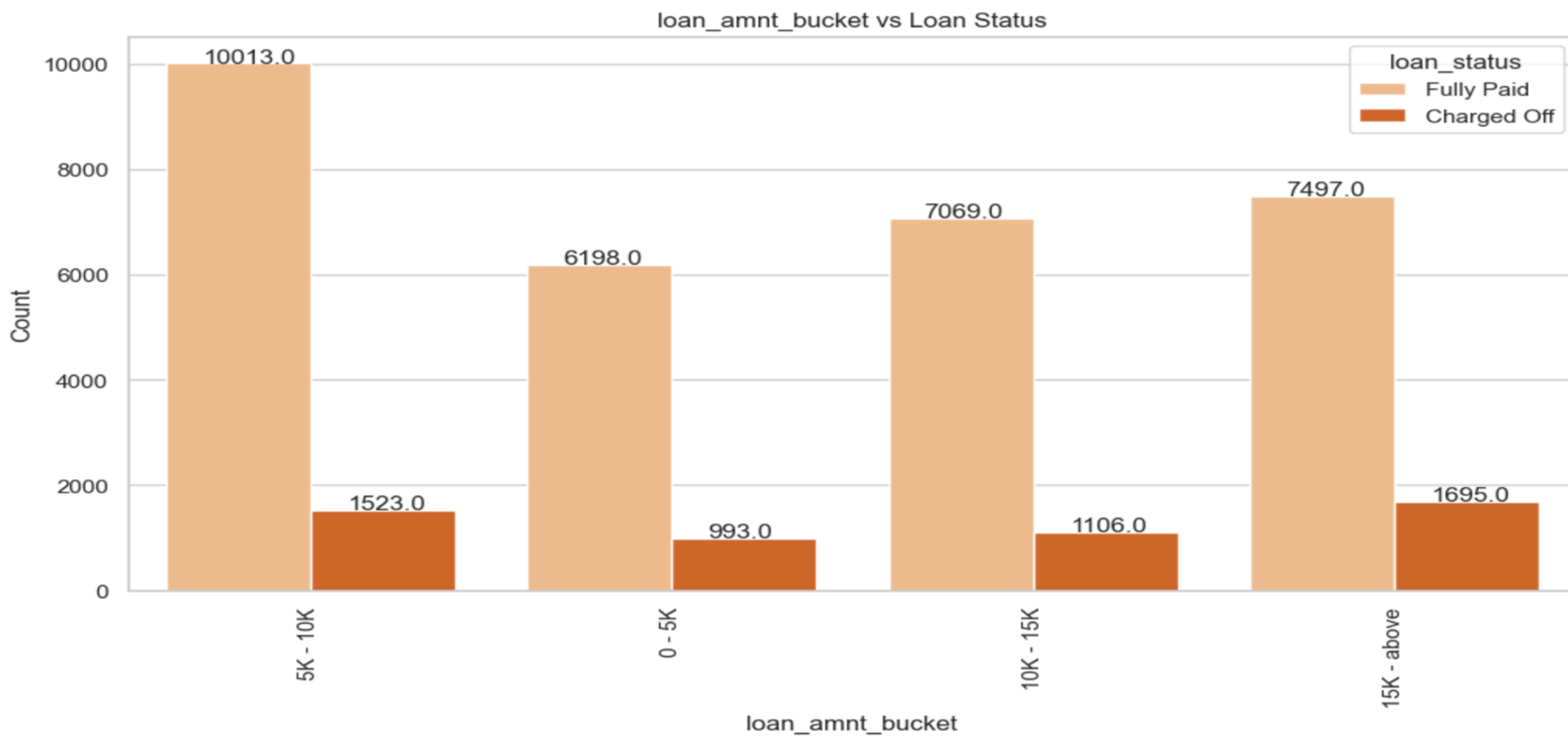
Bivariant Analysis (Quotative Variables)



Bivariant Analysis (Quotative Variables)



Bivariant Analysis (Quotative Variables)



Bivariant Analysis (Quantative Variable)

Observation:

- ✓ A majority of the loan applicants who defaulted received loan amounts of $15,000 or higher.
- ✓ The majority of loan applicants who charged off had significantly high Debt-to-Income (DTI) ratios.
- ✓ A significant portion of loan applicants who defaulted received loans with interest rates falling within the range of 13% to 17%.
- ✓ A majority of the loan applicants who charged off reported an annual income of less than $40,000.

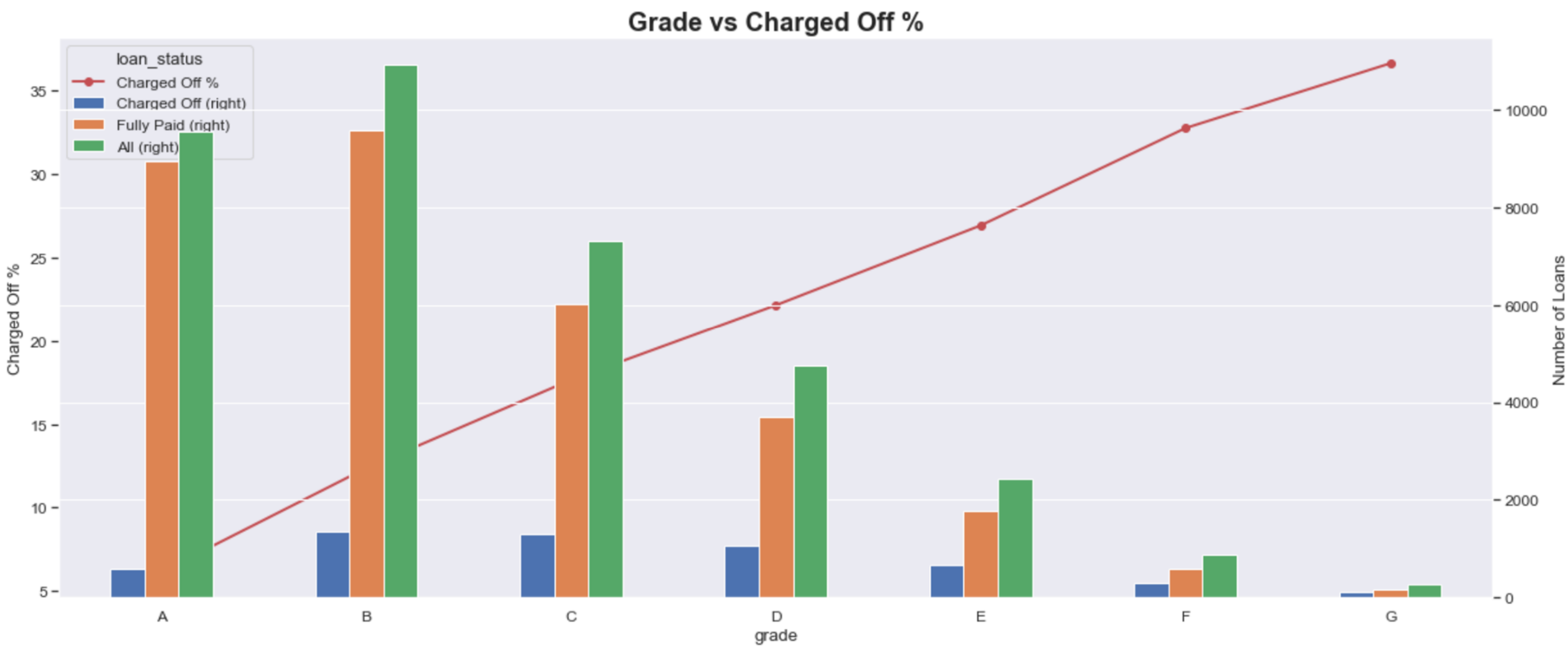
Inferences:

- ✓ **<u>High Loan Amounts</u>**: Applicants receiving loan amounts of $15,000 or higher are more likely to default. The company can mitigate this risk by conducting more thorough assessments for larger loan requests and potentially capping loan amounts for higher-risk applicants.
- ✓ **<u>DTI and Interest Rates</u>**: High Debt-to-Income (DTI) ratios and interest rates in the 13%-17% range are associated with defaults. The company should review its interest rate determination process and consider adjusting rates based on DTI ratios to better align with the borrower's ability to repay.
- ✓ **<u>Low Annual Income</u>**: Applicants with annual incomes less than $40,000 have a higher likelihood of defaulting. The company should consider offering financial education resources or setting maximum loan amounts based on income levels to ensure affordability for borrowers.

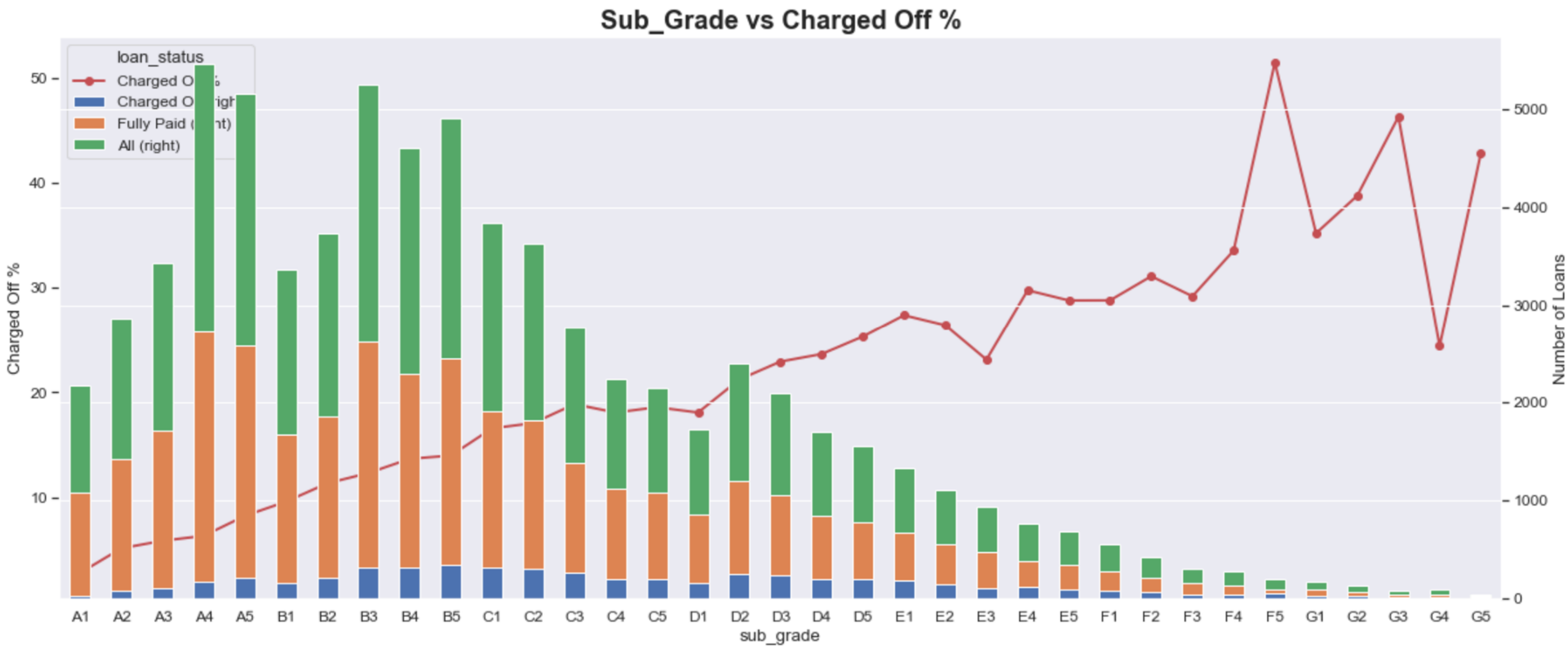
Multivariant Analysis

- ✓ **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.

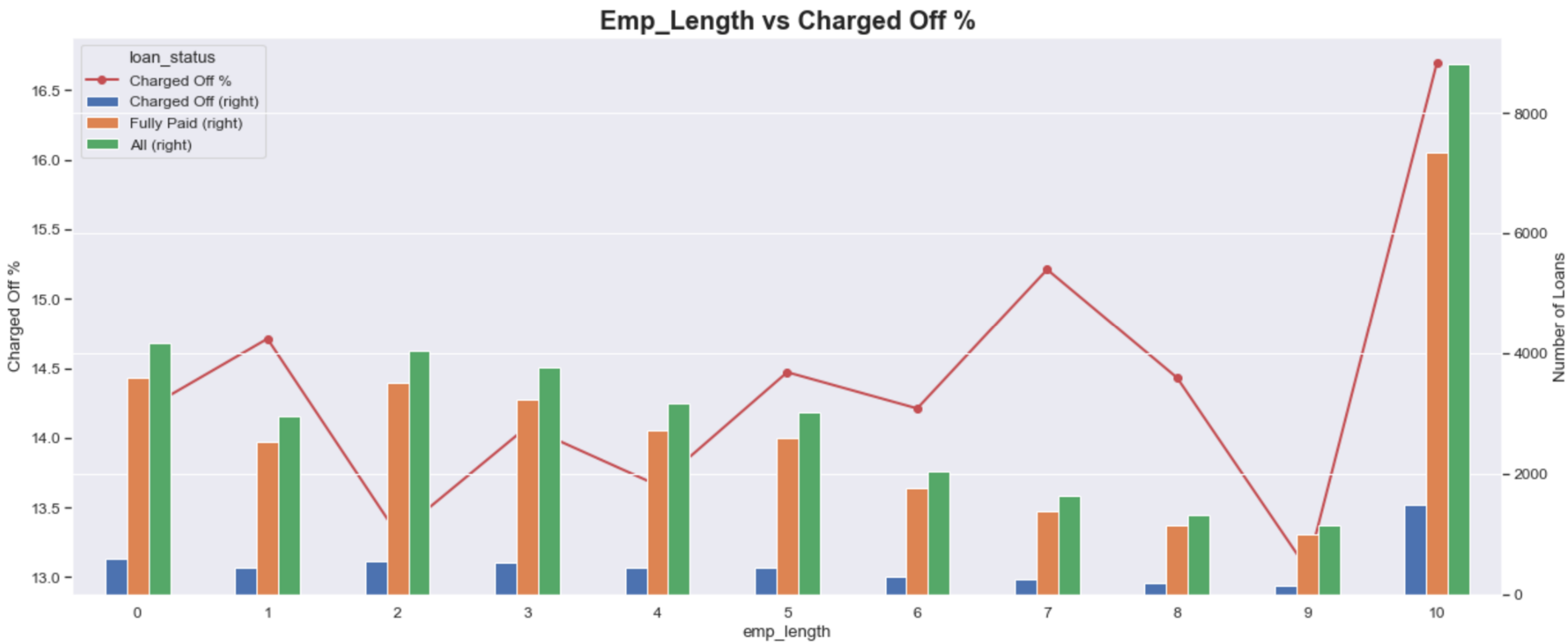
Multivariant Analysis



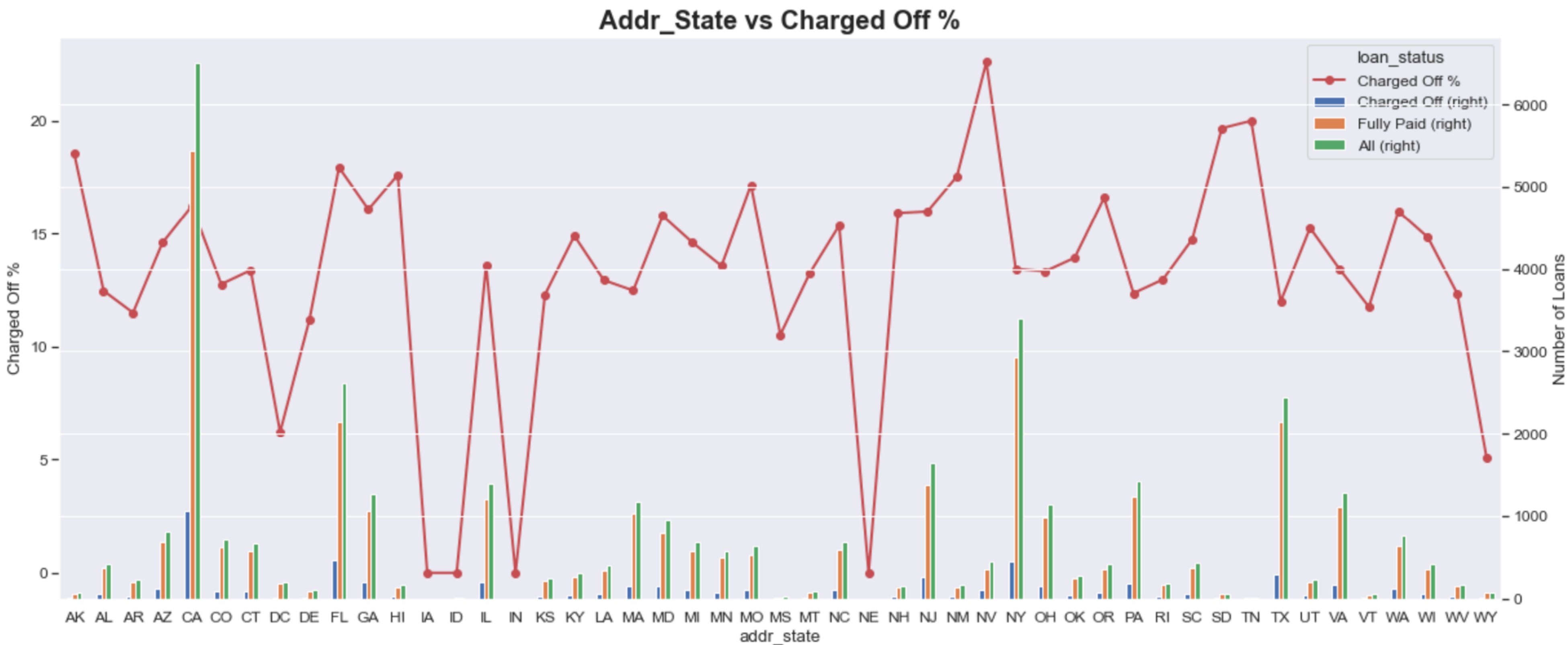
Multivariant Analysis



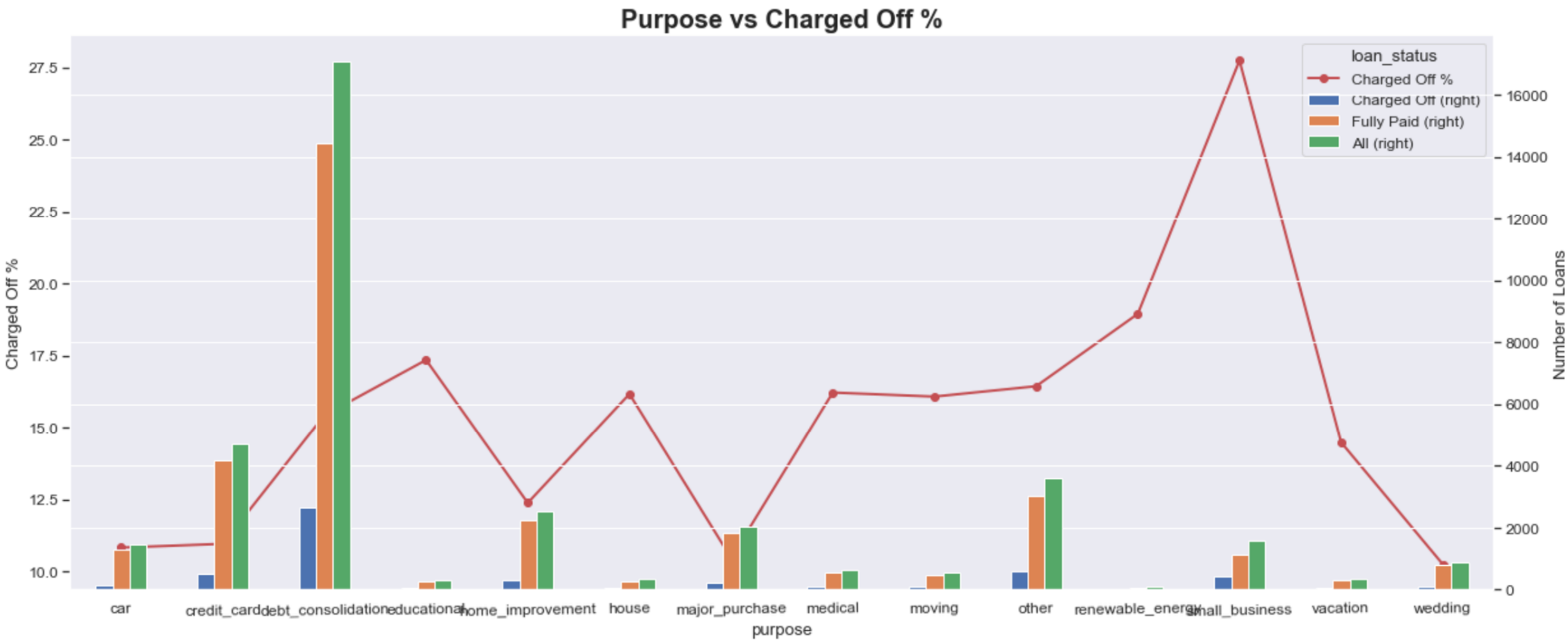
Multivariant Analysis



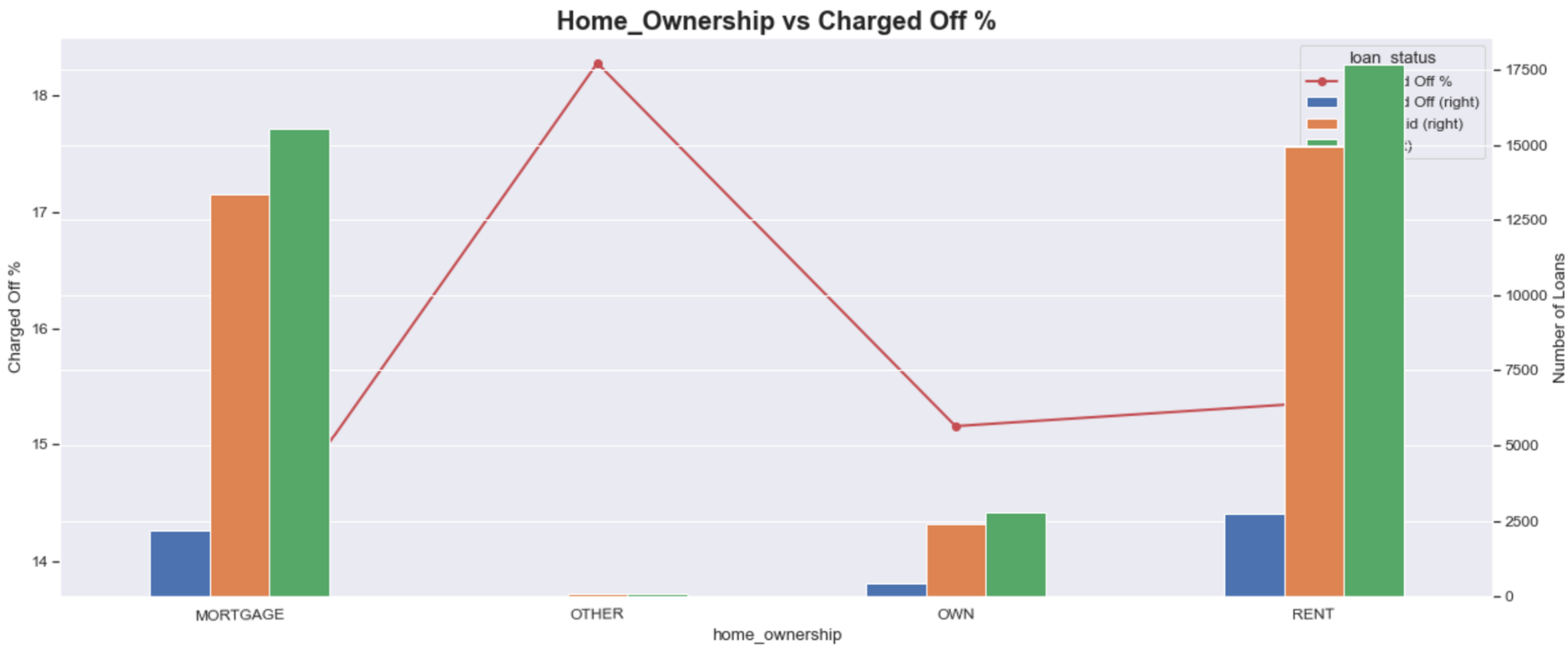
Multivariant Analysis



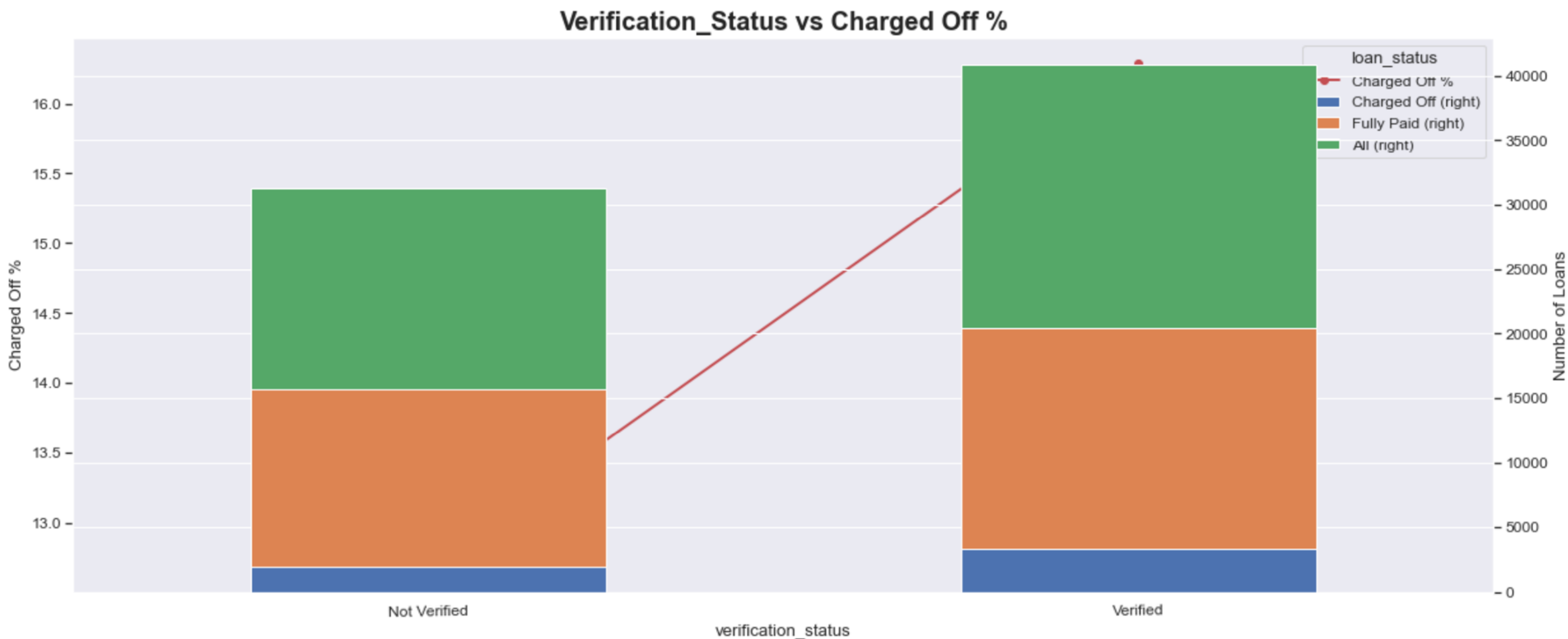
Multivariant Analysis



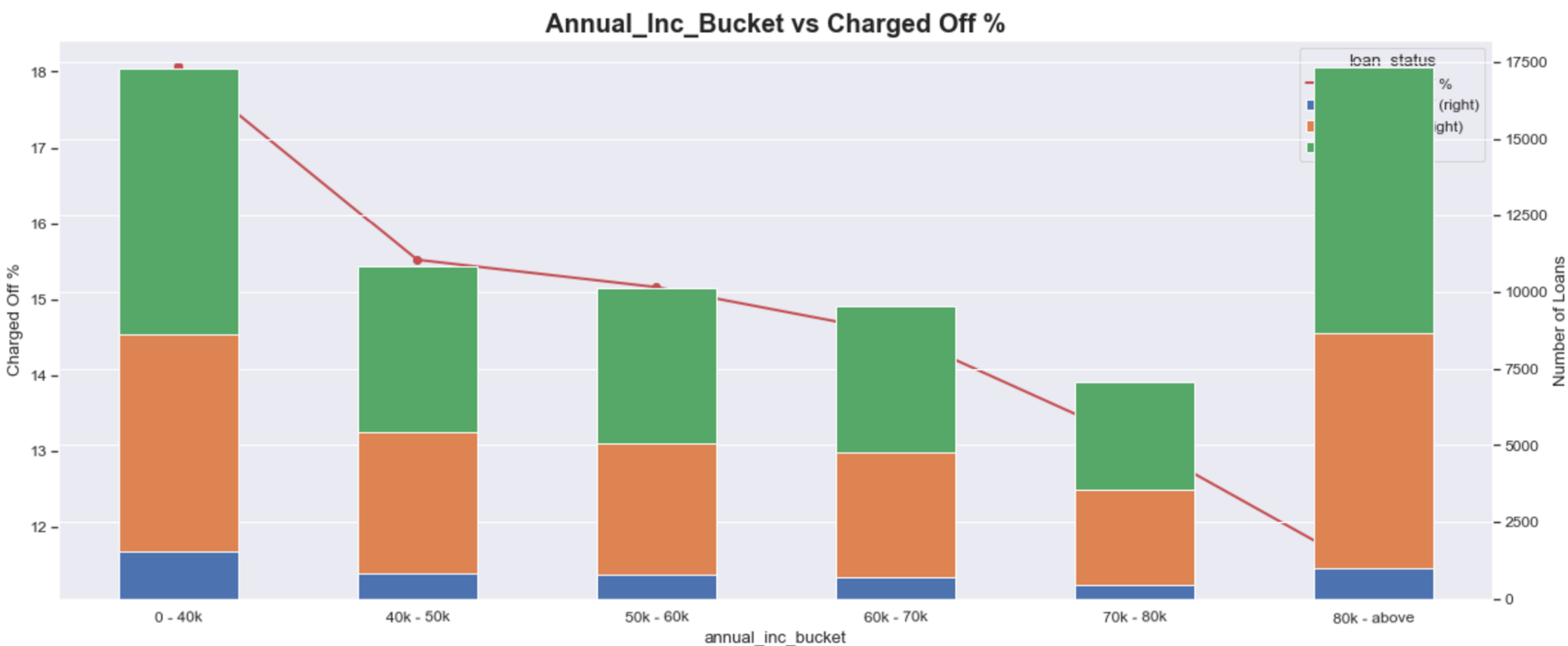
Multivariant Analysis



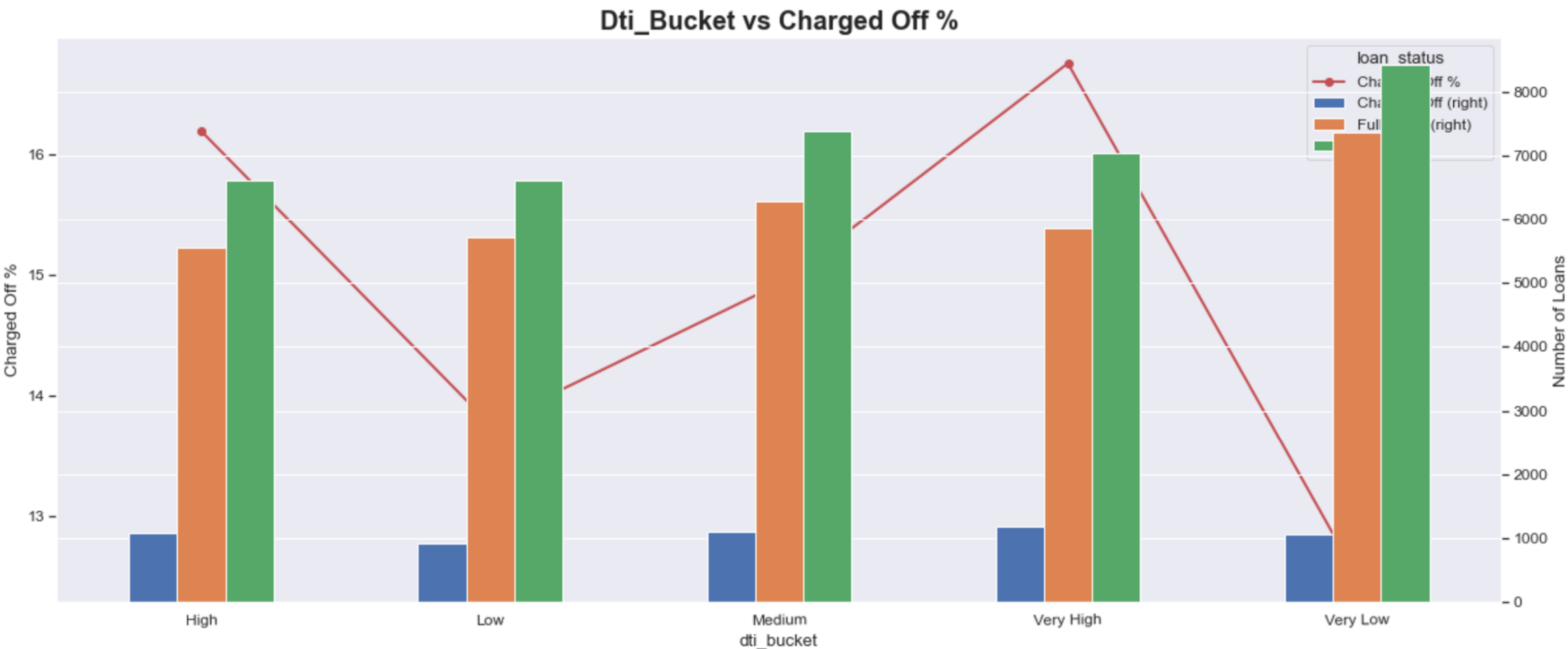
Multivariant Analysis



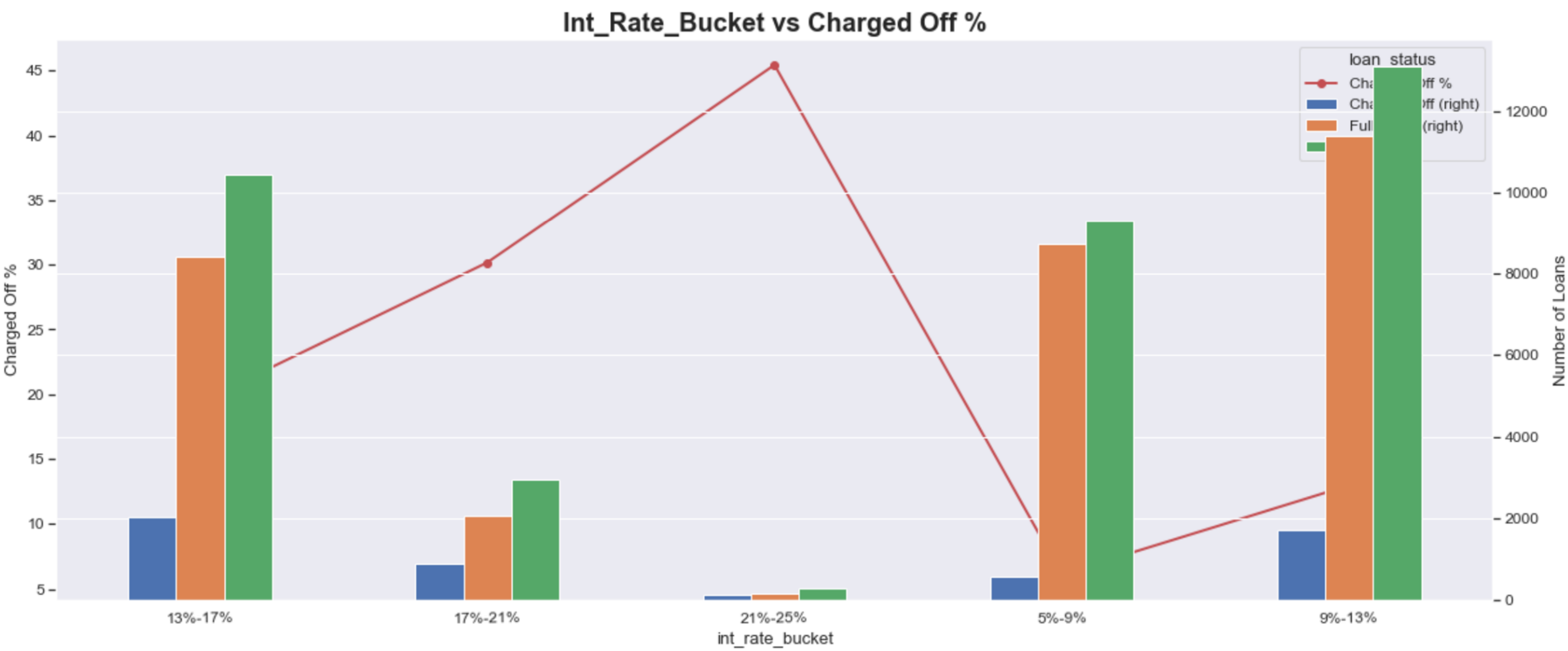
Multivariant Analysis



Multivariant Analysis



Multivariant Analysis



Multivariant Analysis

Observations & Inferences:

- ✓ Tendency to default the loan is likely with loan applicants belonging to B, C, D grades.
- ✓ Borrowers from sub grade B3, B4 and B5 have maximum tendency to default.
- ✓ Loan applicants with 10 years of experience has maximum tendency to default the loan.
- ✓ Borrowers from states CA, FL, NJ have maximum tendency to default the loan.
- ✓ Borrowers from Rented House Ownership have highest tendency to default the loan.
- ✓ The borrowers who are in lower income groups have maximum tendency to default the loan and it generally decreases with the increase in the annual income.
- ✓ The tendency to default the loan is increasing with increase in the interest rate.

Suggestion:

- ✓ **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 minimize default risks.
- ✓ **Focus on Subgrades 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 subgrades 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.
- ✓ **Capitalizing on Market Growth:** Capitalize 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.

Suggestion:

- ✓ **Careful Evaluation for Debt Consolidation Loans**: Carefully evaluate applicants seeking debt consolidation loans, considering potential interest rate adjustments or offering financial counseling services tomanage 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 & Links:

✓ **Technologies & Packages Used:**

Technology / Package	Version	Documentation
Python	3.11.4	https://www.python.org/
Matplotlib	3.7.1	https://matplotlib.org/
Numpy	1.24.3	https://numpy.org/
Pandas	1.5.3	https://pandas.pydata.org/
Seaborn	**0.12.2**	https://seaborn.pydata.org/

✓ **GitHub Repository Link:**

https://https://github.com/amit024/Lending-Club-Case-Study

Thank you

Amit Kumar