

LENDING CLUB CASE STUDY

MANIKANTA GRANDHI
VISHAL LOKE

Problem Statement

- In the context of consumer finance, effectively managing credit risk is crucial for minimizing financial losses. The task is to analyze historical loan application data to identify patterns and factors associated with loan defaults.
- The goal of this analysis is to apply Exploratory Data Analysis (EDA) techniques to uncover key indicators of loan defaults. By identifying these risk factors, the company aims to enhance its decision-making process, reduce credit loss, and optimize its loan approval criteria

Objective

- Determine which attributes are significant predictors of loan default.
- Analyze consumer and loan attributes to uncover patterns and trends that indicate higher risk of default.
- Develop insights into the characteristics of high-risk loan applicants.
- Provide actionable recommendations to improve the company's risk assessment strategies and reduce the likelihood of default.

Analysis Approach

Data Understanding and Cleaning

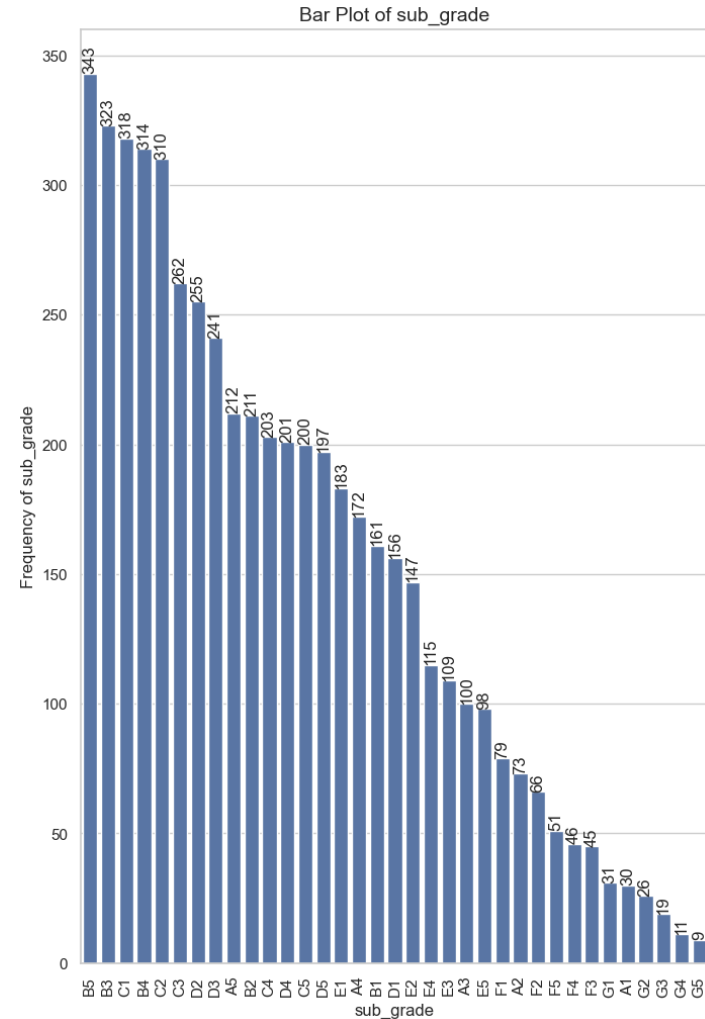
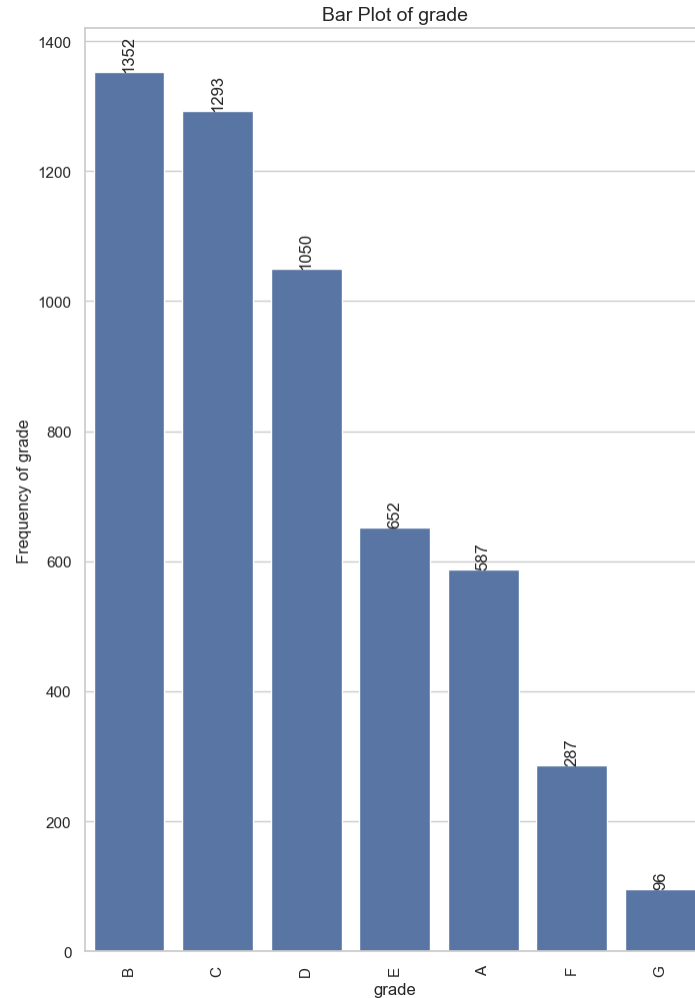
- Review Data Dictionary: Examine column names and their descriptions to comprehend the data structure.
- Data Quality Assessment: Handling Missing Values, Data Conversions, Outlier Treatment, Duplicate Records, Categorical Data Consistency and Imputing values in columns.
- Column Selection: Identify and select the most relevant columns for a focused analysis.

Exploratory Data Analysis (EDA)

- Univariate Analysis: Numerical and categorical variable analysis
- Bivariate Analysis: Explore the relationships and correlations between pairs of variables.
- Multivariate Analysis: Investigate interactions among multiple variables

Univariate and Segmented Univariate

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.

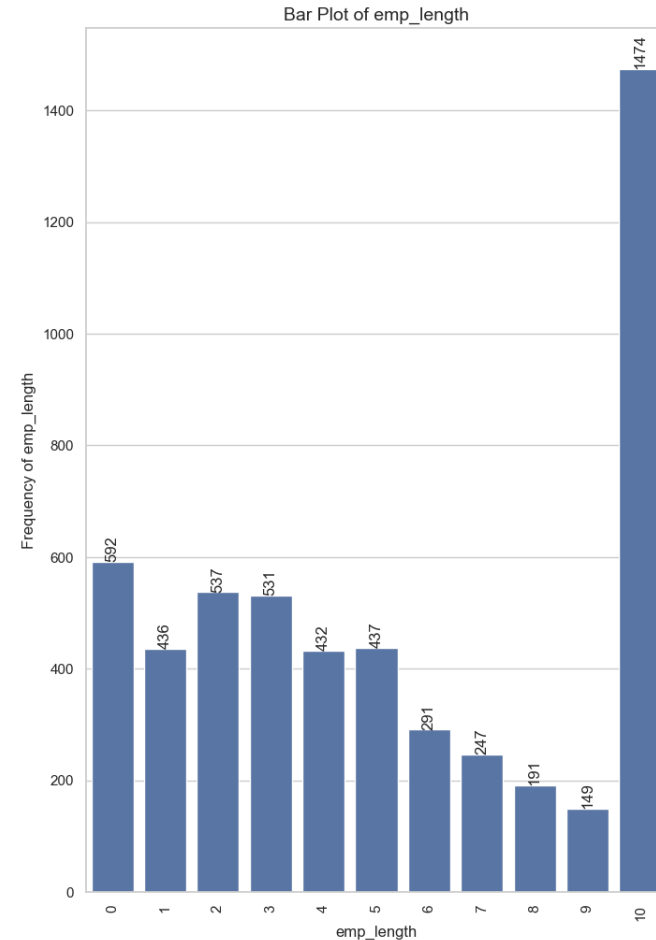
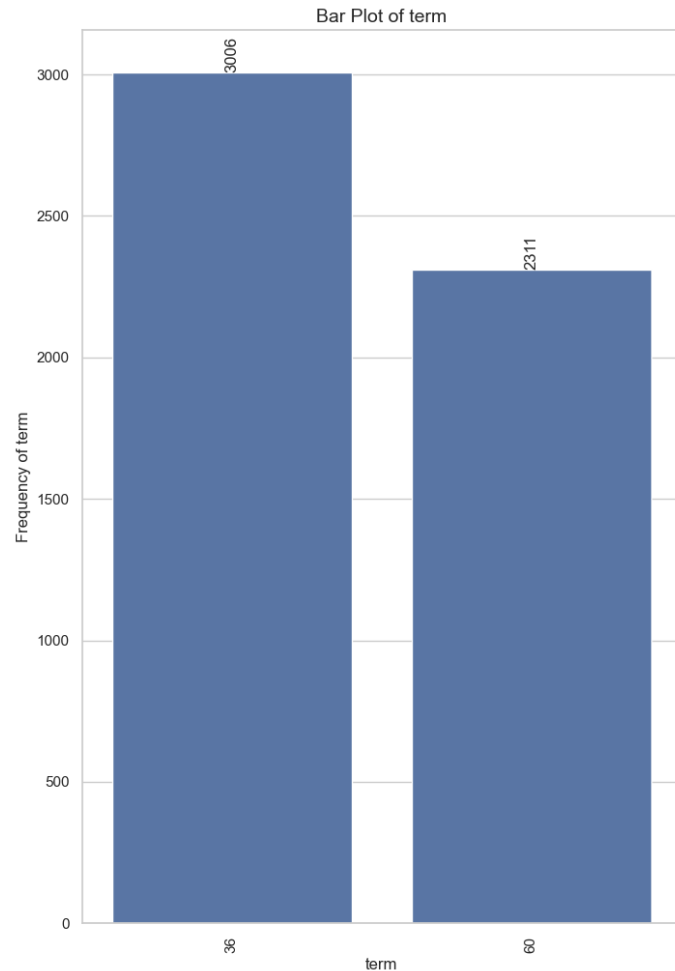


Grade and Sub grade

Univariate and Segmented Univariate

Term: Higher default rates observed for short term applicants, 3006.

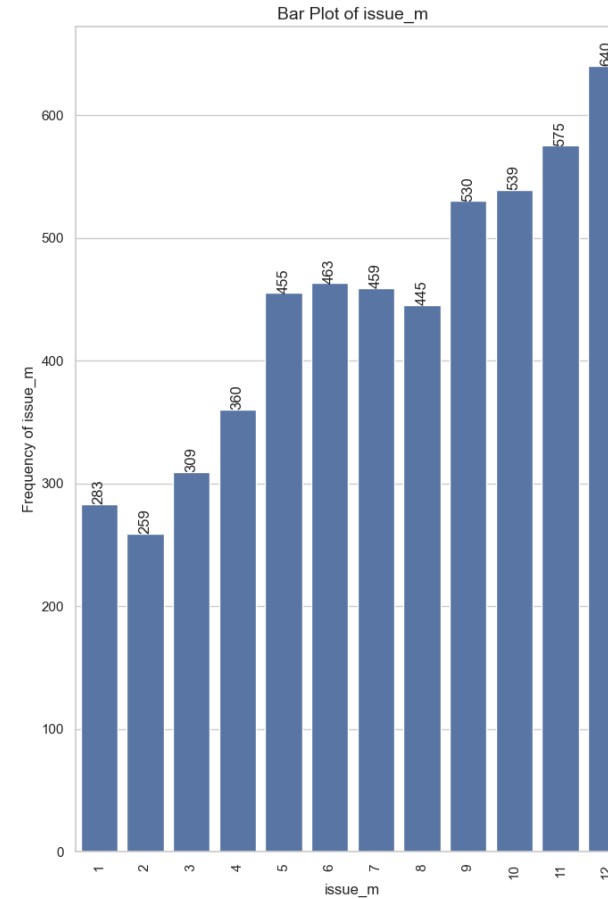
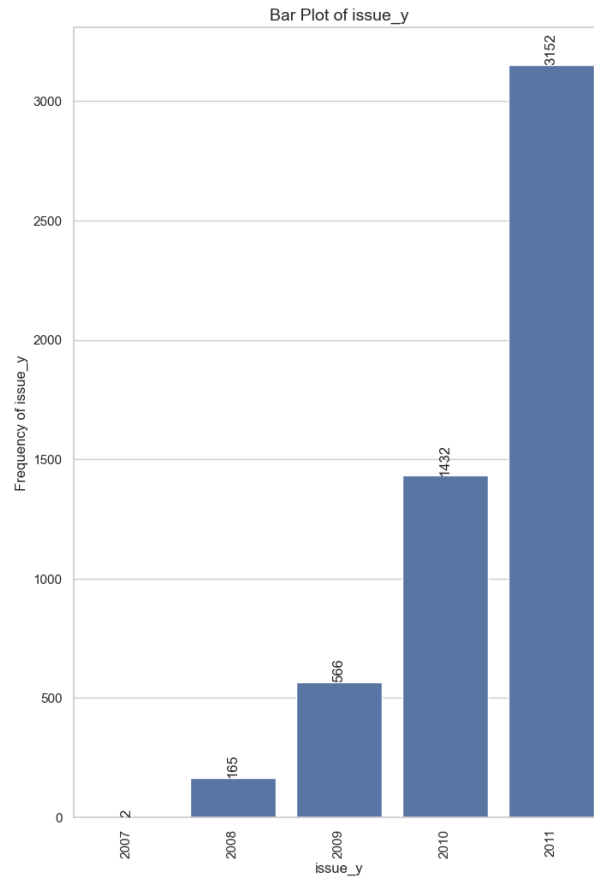
Emp_length: Higher default rates observed for employees having more than 10 years, 1,474.



Univariate and Segmented Univariate

Issue_y: Higher default rates are observed in 2011 year, total 3097.

Issue_m: Issued month ranking as per the defaulter $10 < 11 < 12$

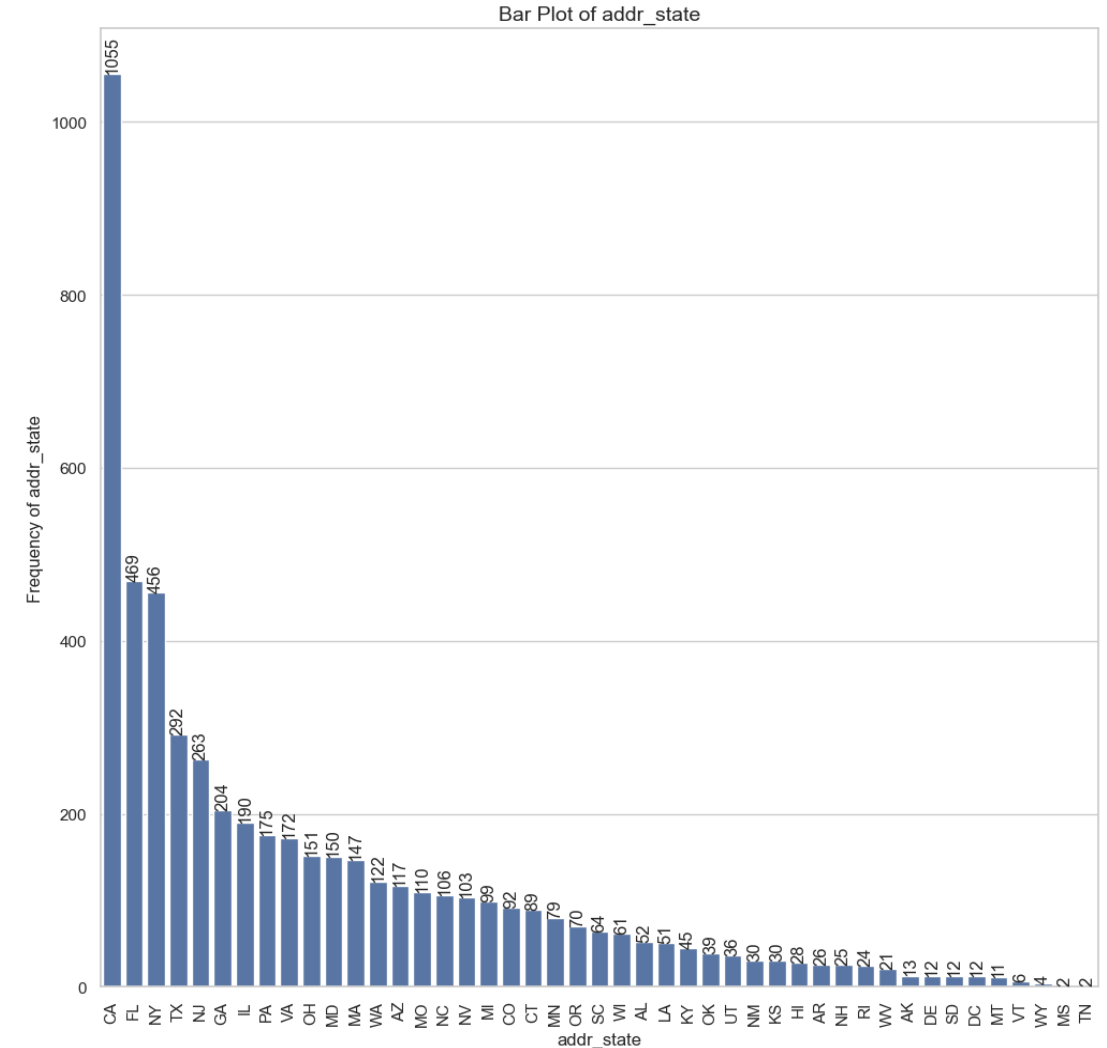
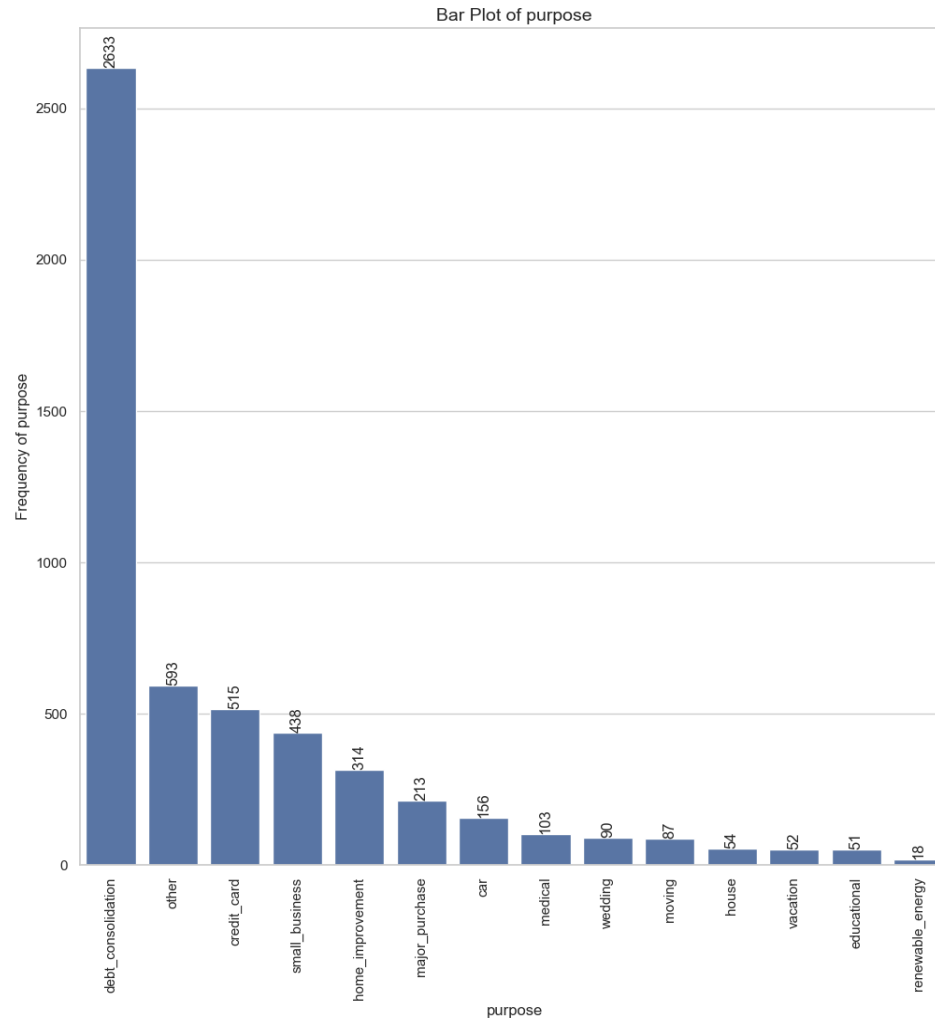


Issued year and Issued month

Univariate and Segmented Univariate

Address State: CA had highest defaulter with 1055 applicants, Ranking CA > FL > NY

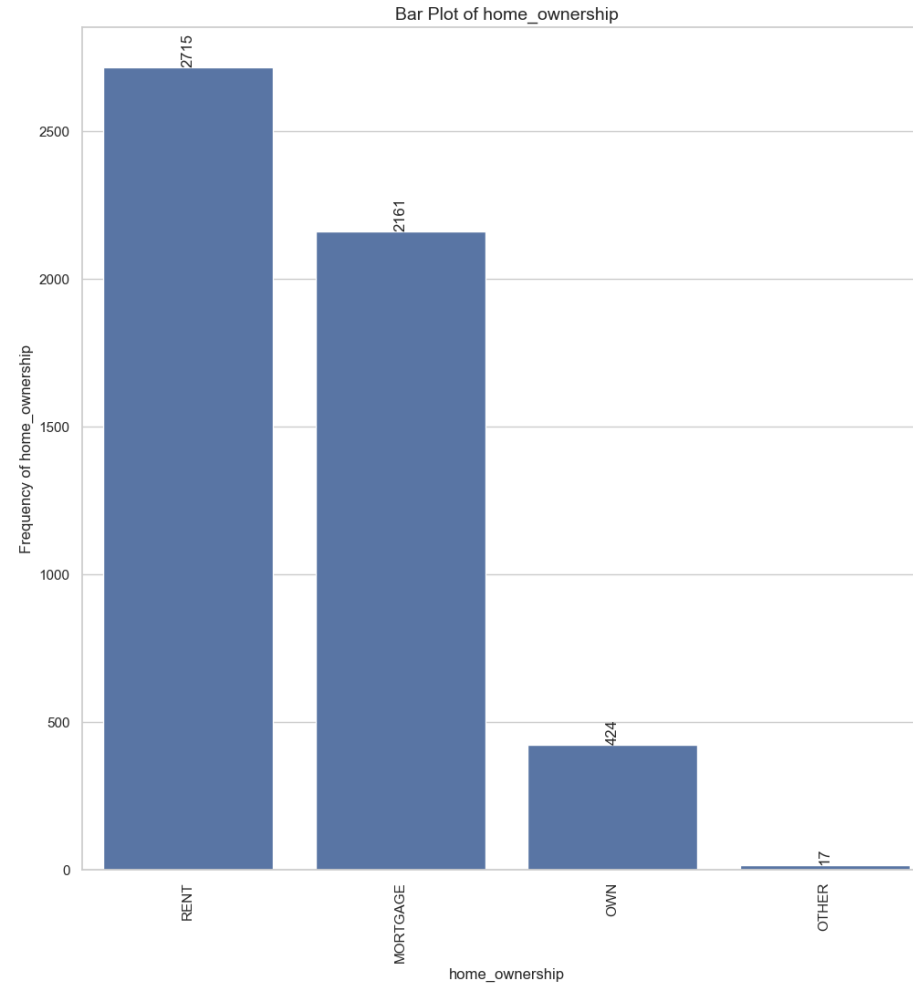
Purpose: Debt_consolidation was the highest defaulter's purpose



Purpose and Address state

Univariate and Segmented Univariate

Home_ownership: Majority of defaulters are lived in rented homes, Ranking Rent > Mortgage > Own

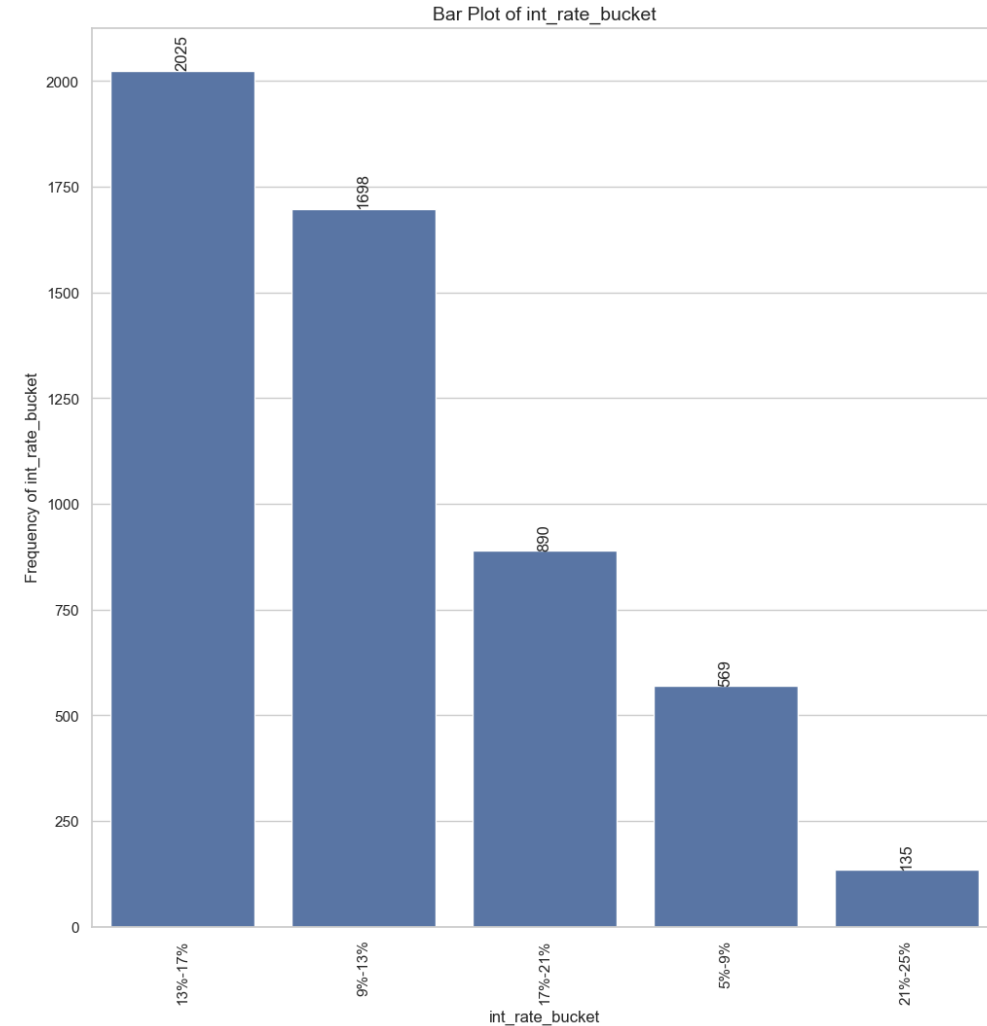
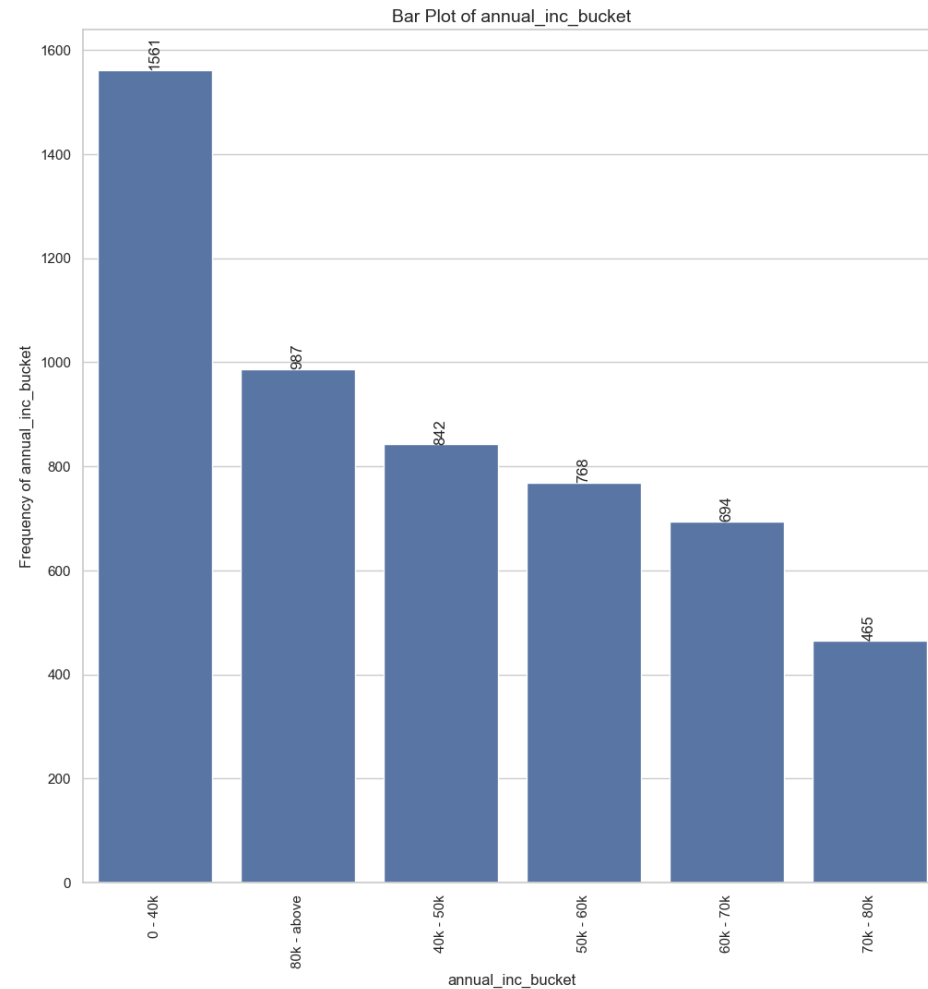


Various types of house ownership

Univariate and Segmented Univariate

Annual_inc_bkt: 1,561 loan applicants who charged off had annual salaries less than \$40,000.

Int_rate_bkt: Highest defaulter observed in 13-17% bucket

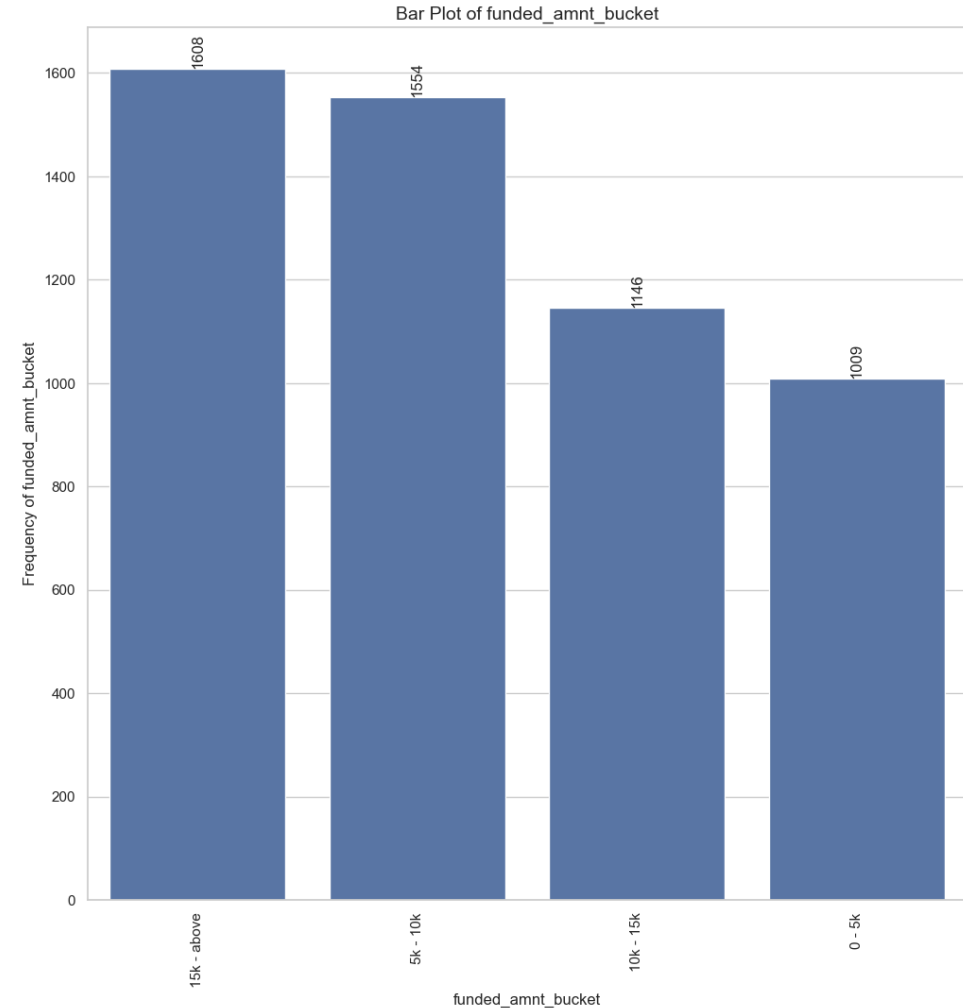
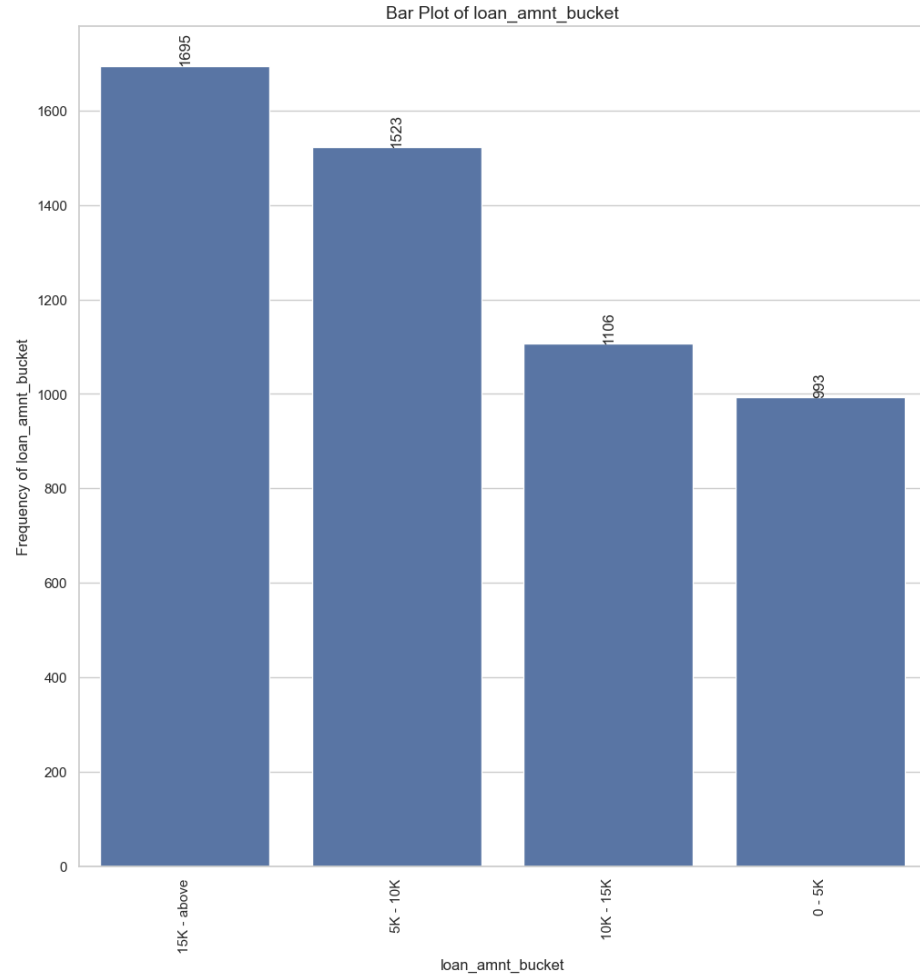


Annual_inc_bucket and Int_rate_bucket

Univariate and Segmented Univariate

Loan_amt_bkt: 1,695 loan participants are defaulted received loan amounts of 15,000 USD and above.

Funded_amt_bkt: 1,608 loan participants are defaulted received funded amounts of 15,000 USD and above.

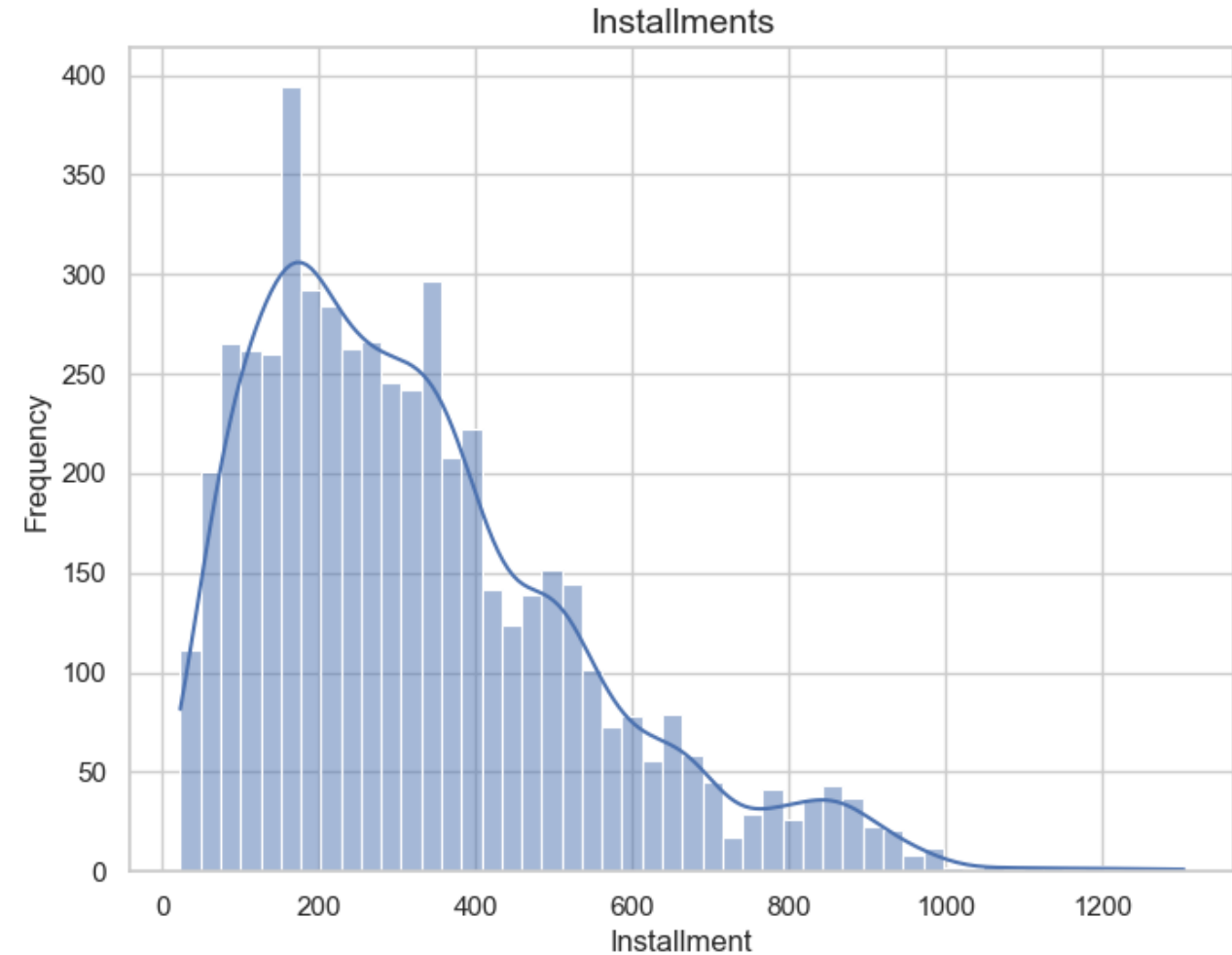
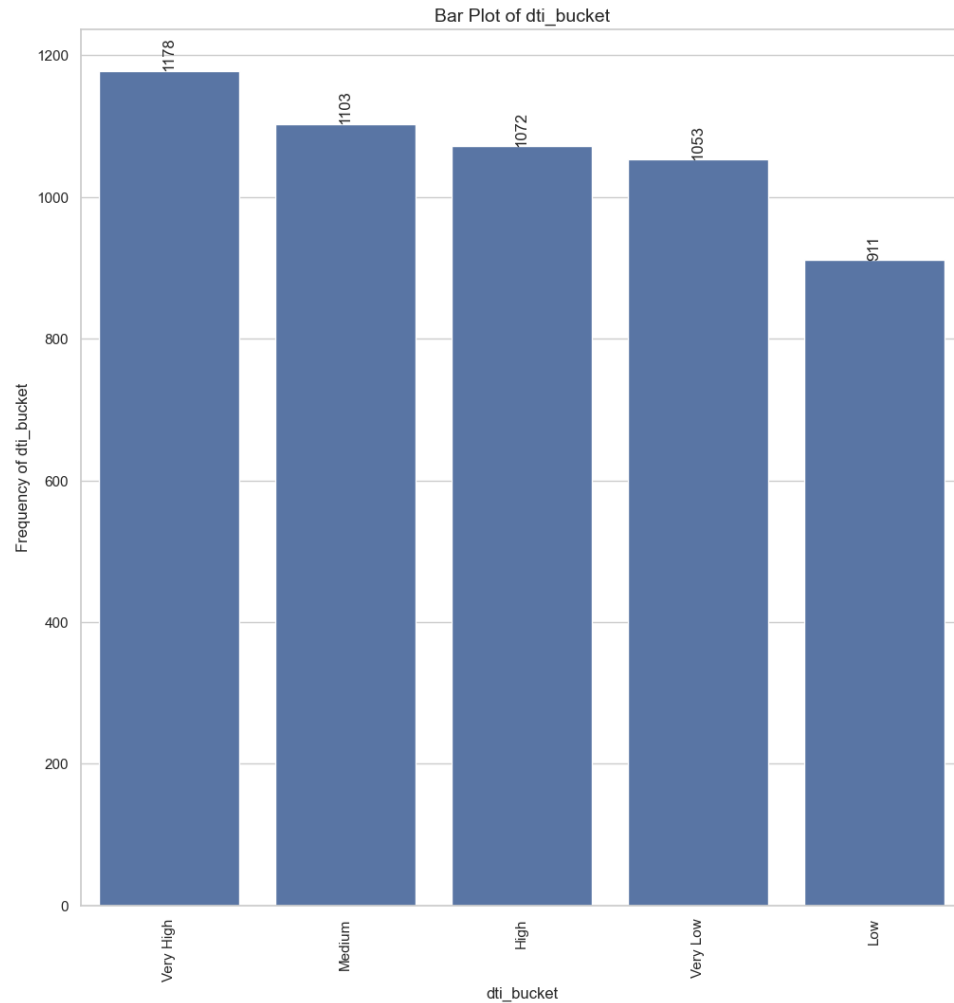


Loan_amt_bucket and funded_amount_bucket

Univariate and Segmented Univariate

Dti_bucket: 1,178 loan applicants had very high debt-to-income ratios are charged off.

Loan instalment: Majority of defaulters are from \$160 - \$440 instalment bracket.

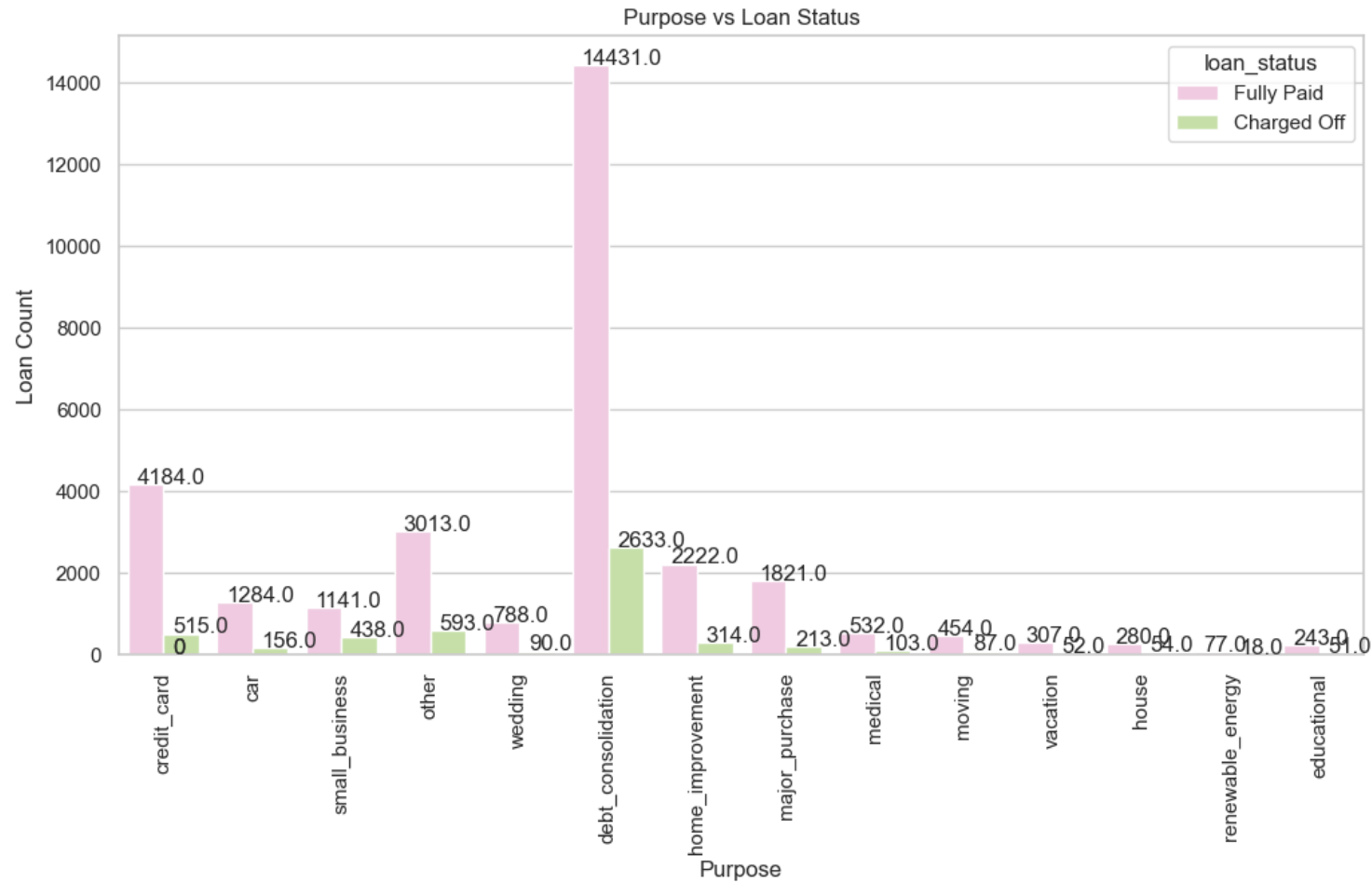


Dti_bucket

Loan defaulters histogram on loan instalment

Bivariate Analysis

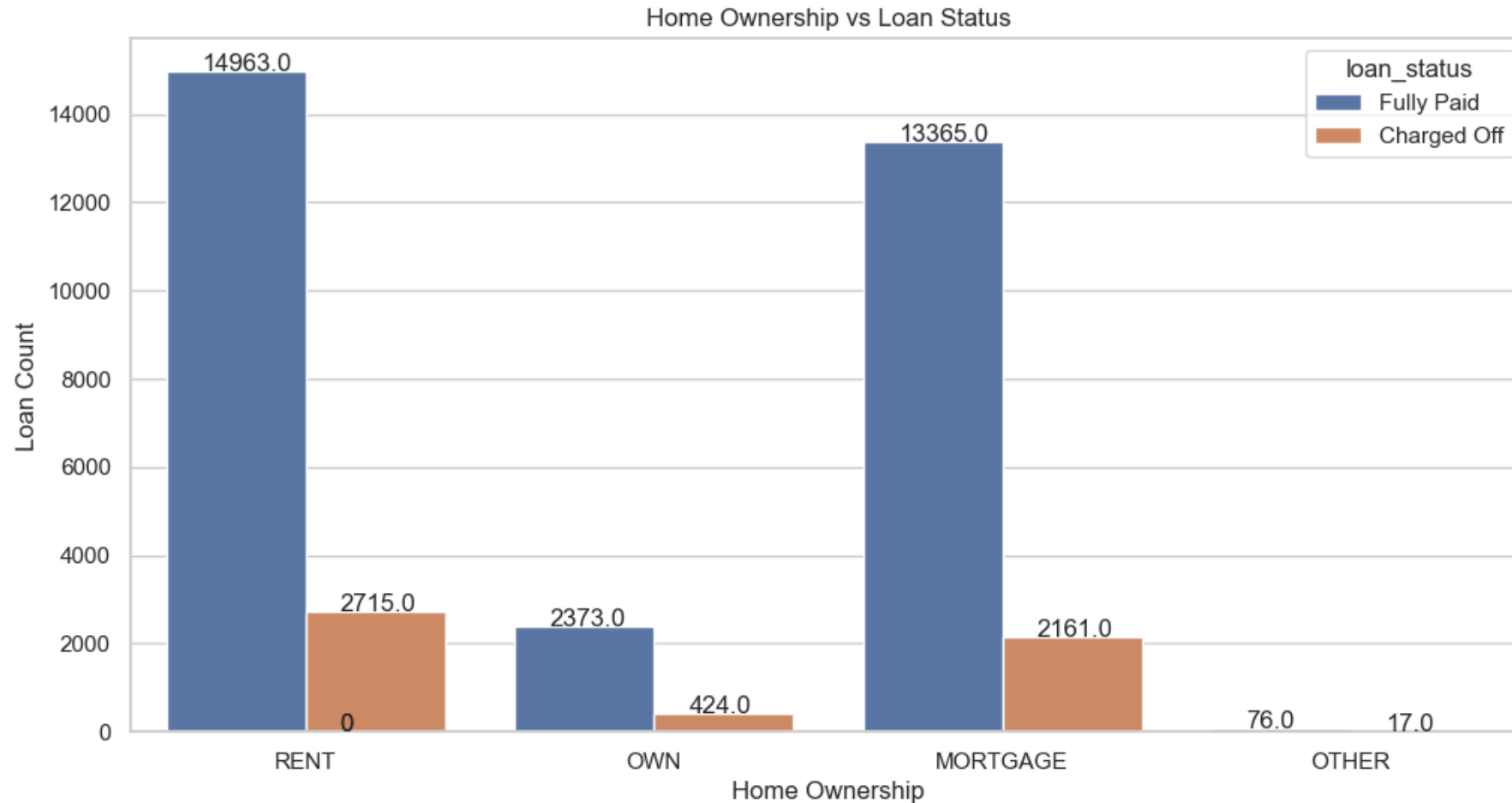
Debt consolidation is the category where maximum loans are issued and people have defaulted the most in the same category



Purpose of loan vs Status of loan

Bivariate Analysis

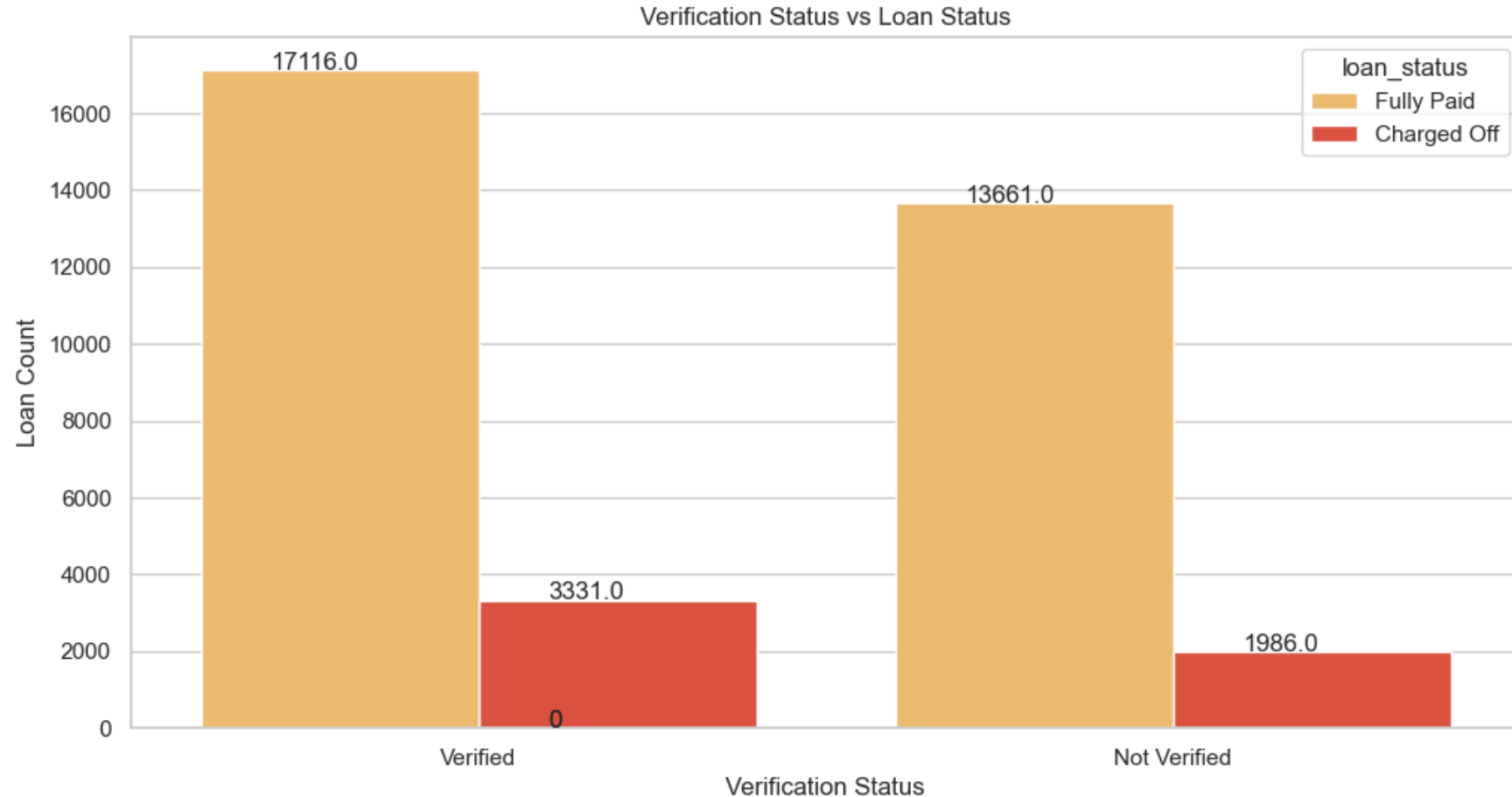
The loan applicants who live in a rented or mortgaged house are more likely to default



Home ownership vs Status of loan

Bivariate Analysis

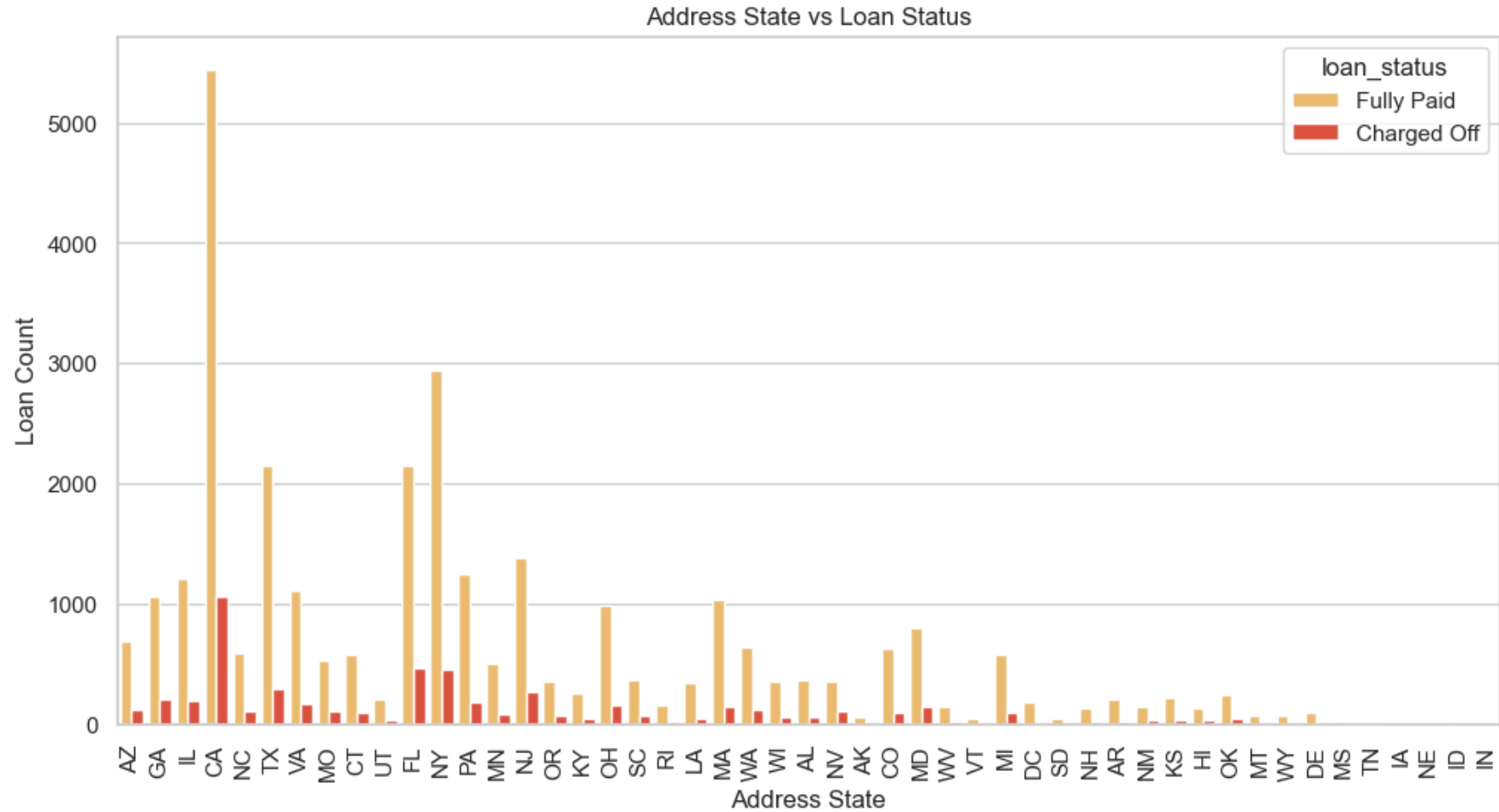
The loan applicants who have been verified are defaulting more than the applicants who are not verified



Verification Status of Loan vs Status of Loan

Bivariate Analysis

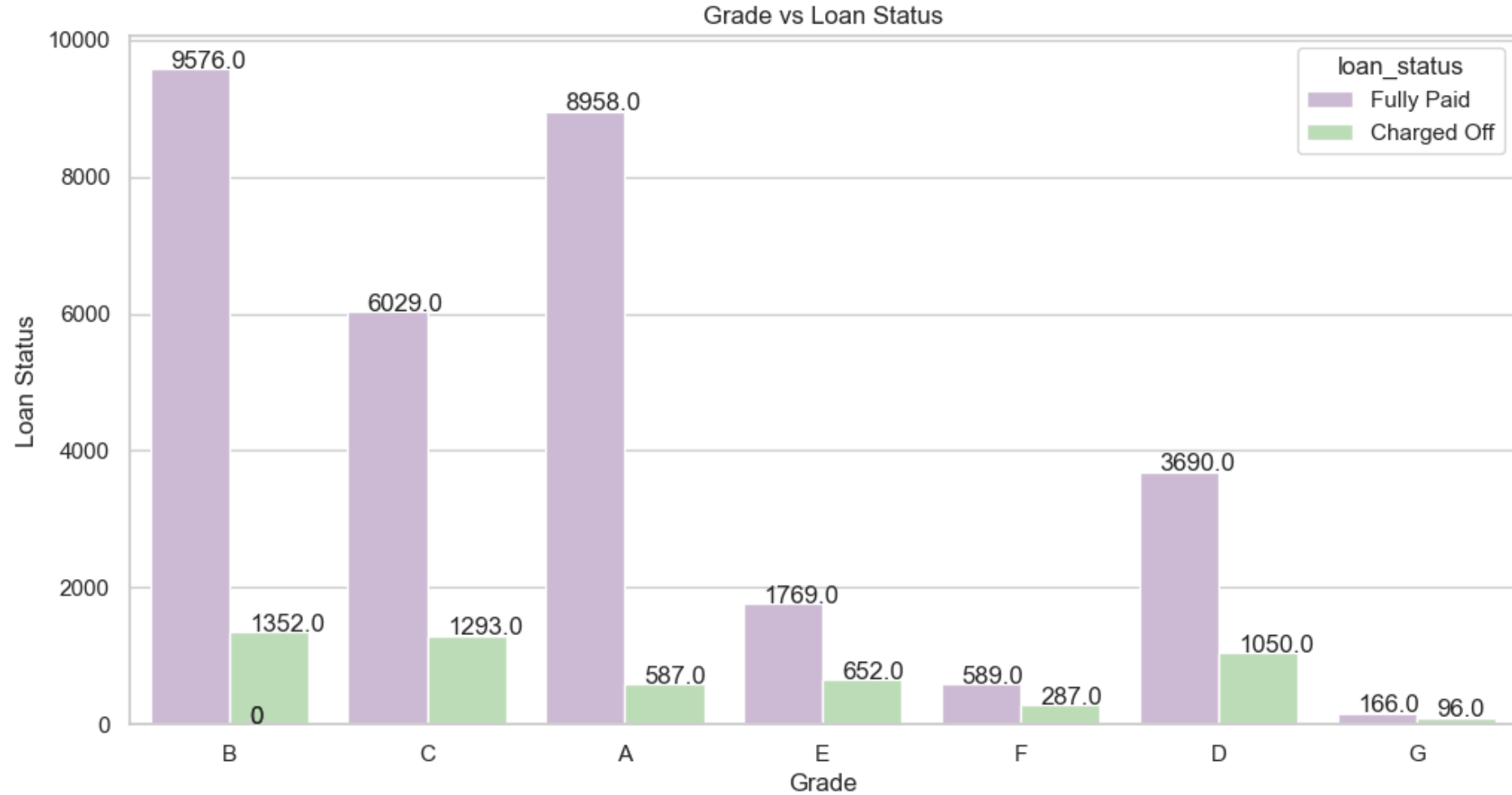
Loan applicants from the state of California (CA), Florida (FL), and New York (NY) are most likely to default



Address State vs Status of Loan

Bivariate Analysis

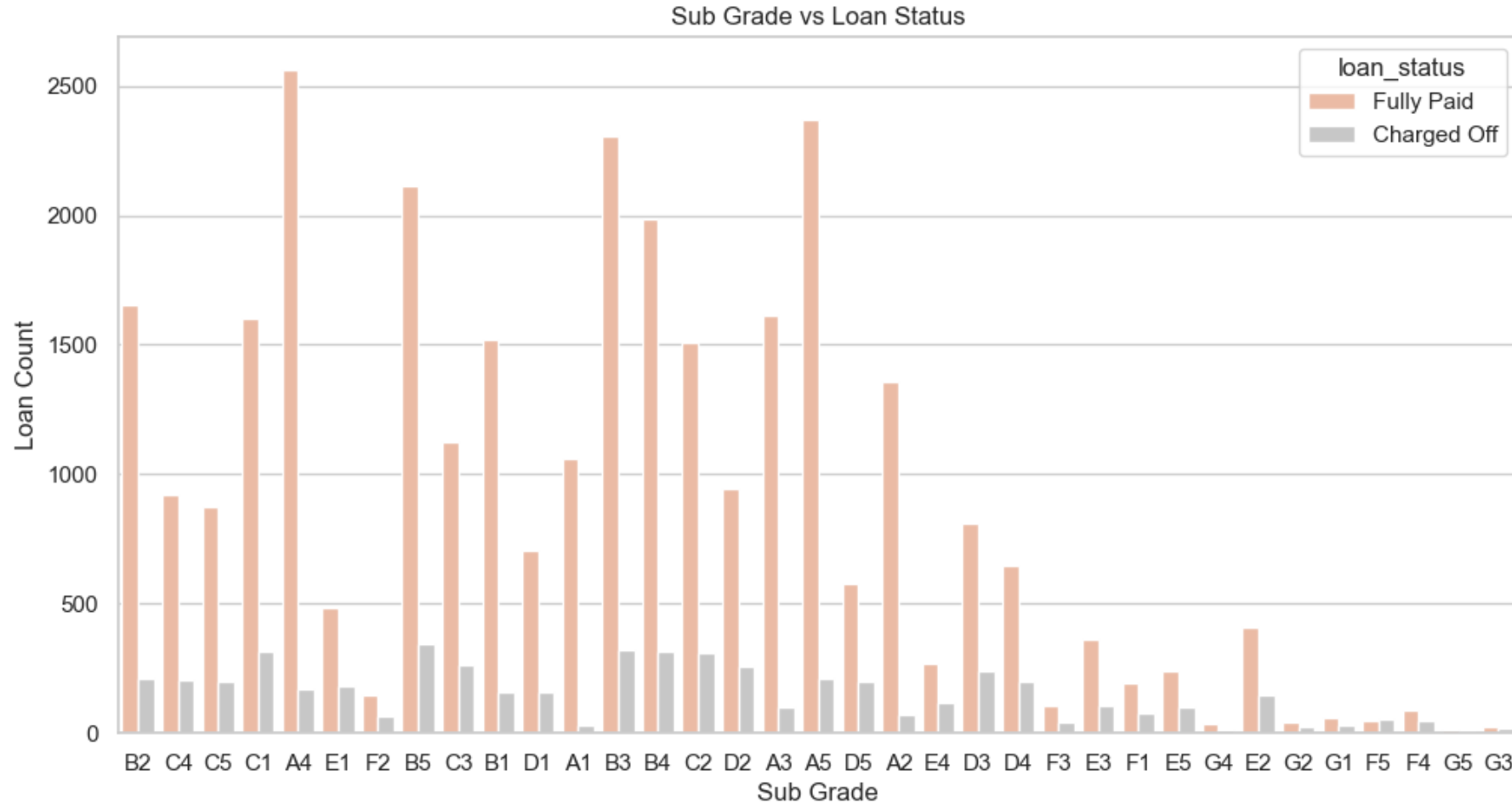
The loan applicants belonging to Grades B, C, and D contribute to most of the "Charged Off" loans.



Loan Grade vs Status of Loan

Bivariate Analysis

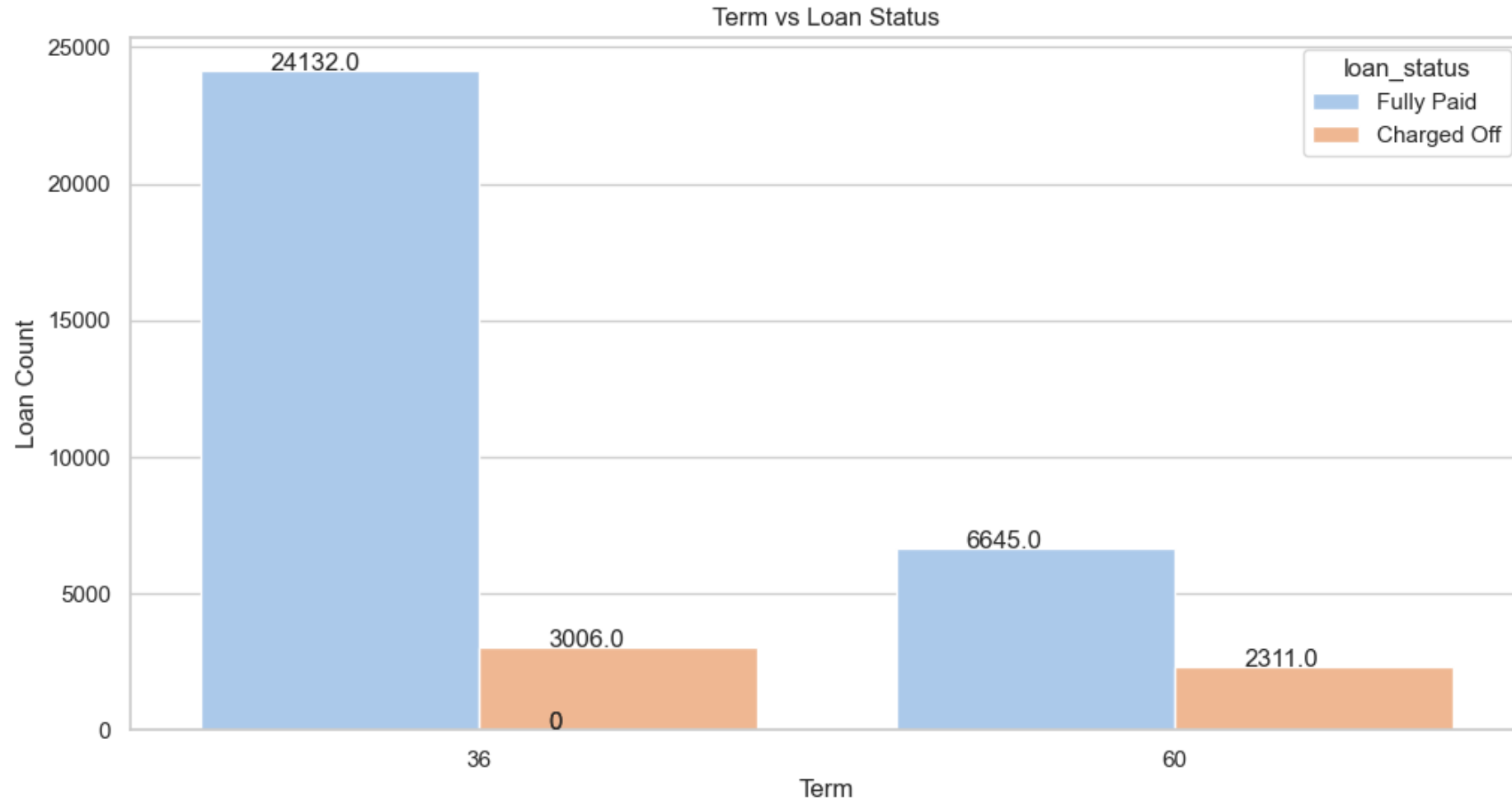
Loan applicants belonging to Sub Grades B3, B4, and B5 are more likely to charge off.



Loan Sub-Grade vs Status of Loan

Bivariate Analysis

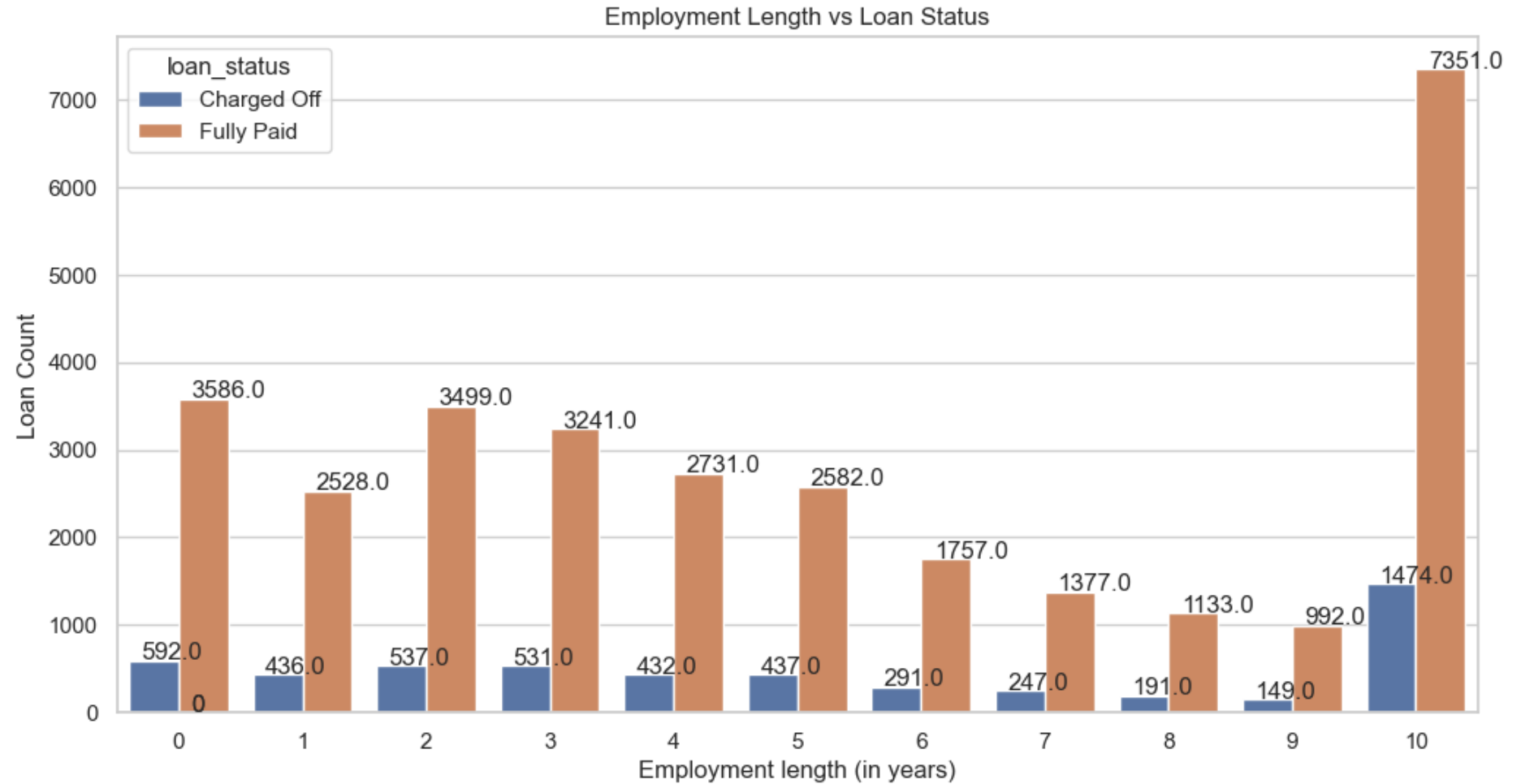
Loan applicants applying for loans with a 60-month term are more likely to default than those taking loans for 36 months.



Term of Loan vs Status of Loan

Bivariate Analysis

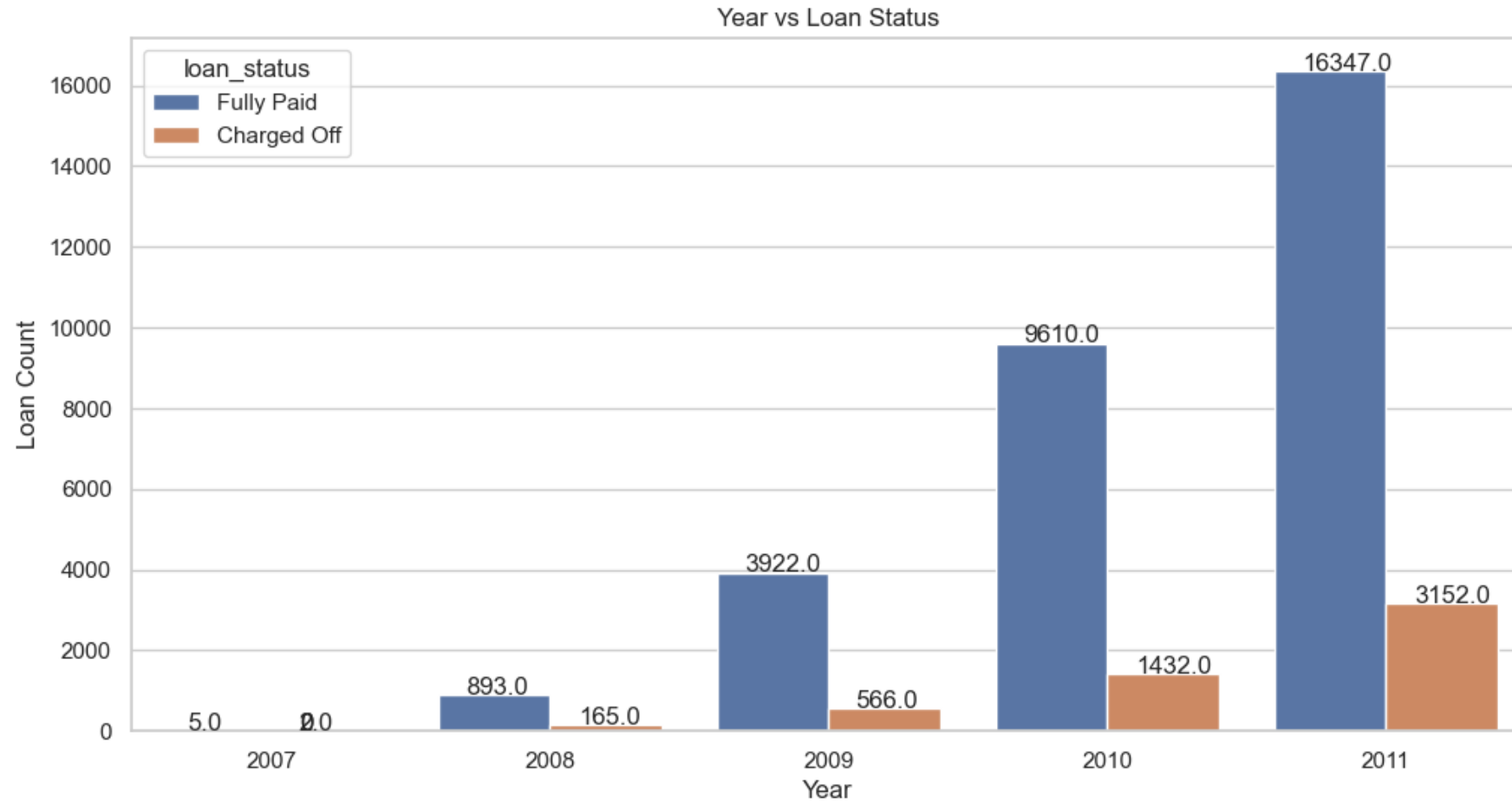
Most loan applicants have ten or more years of experience, and they are also the most likely to default.



Employment Length of Customer vs Status of Loan

Bivariate Analysis

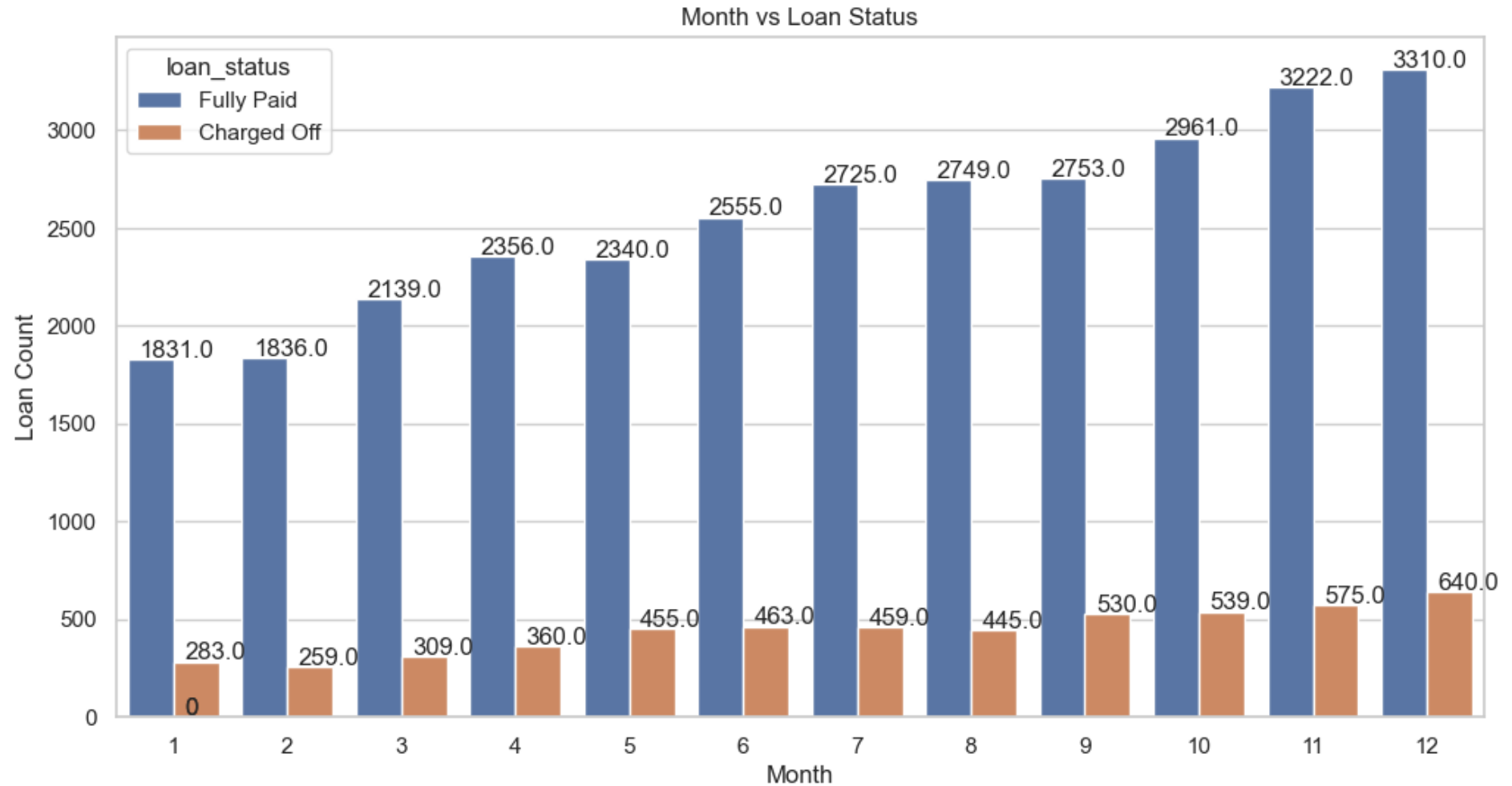
The number of loan applicants has steadily increased from 2007 to 2011, indicating a positive trend in the upcoming years.



Year the Loan was given to Customer vs Status of Loan

Bivariate Analysis

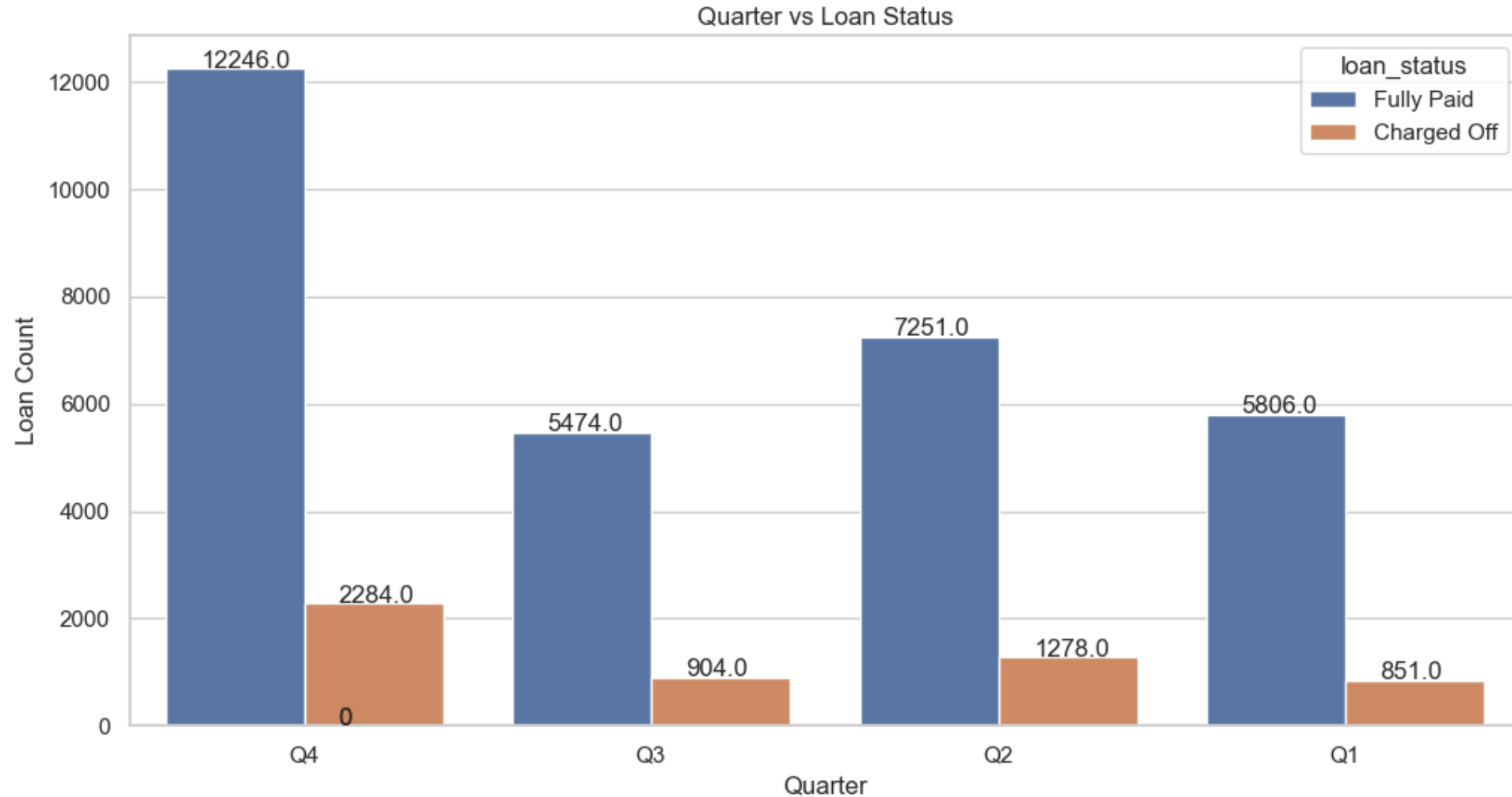
December is the most preferred month for taking loans, possibly due to the holiday season.



Month during which the Loan was given to Customer vs Status of Loan

Bivariate Analysis

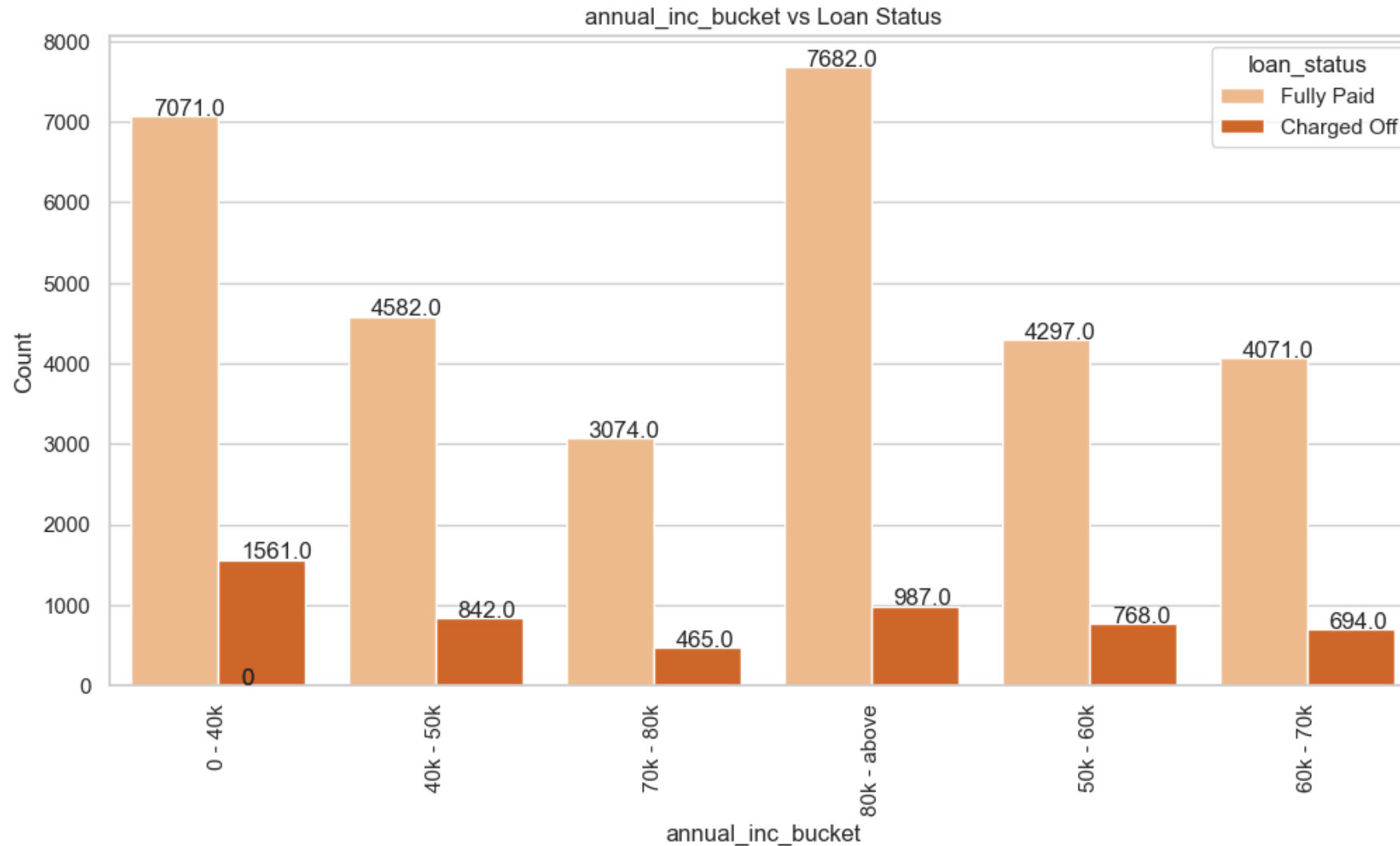
The fourth quarter (Q4) is the most preferred quarter for taking loans, primarily because of the upcoming holiday season.



Quarter during which the Loan was given to Customer vs Status of Loan

Bivariate Analysis

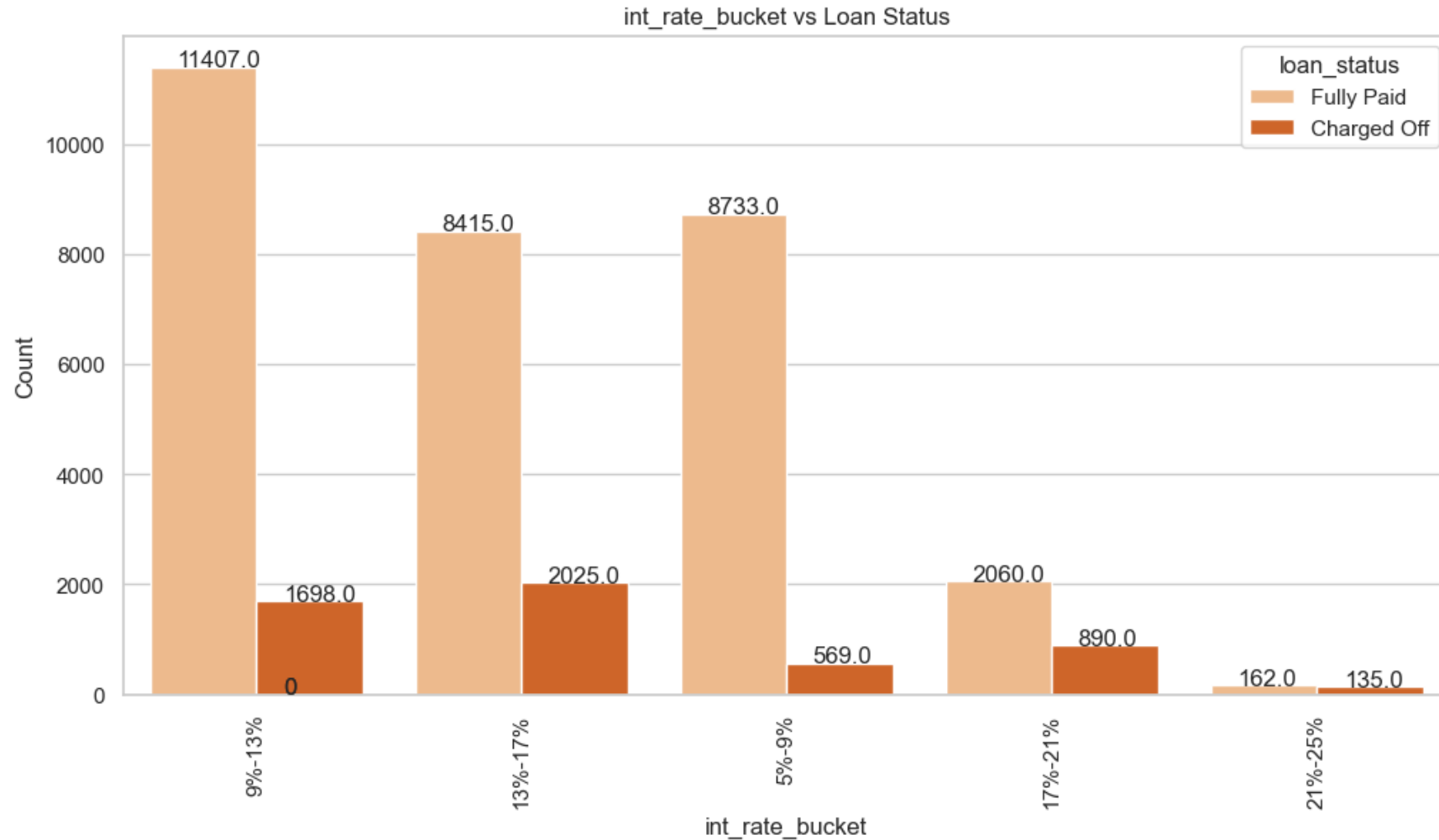
A majority of the loan applicants who charged off reported an annual income of less than \$40,000.



Bucket of Annual Income vs Status of Loan

Bivariate Analysis

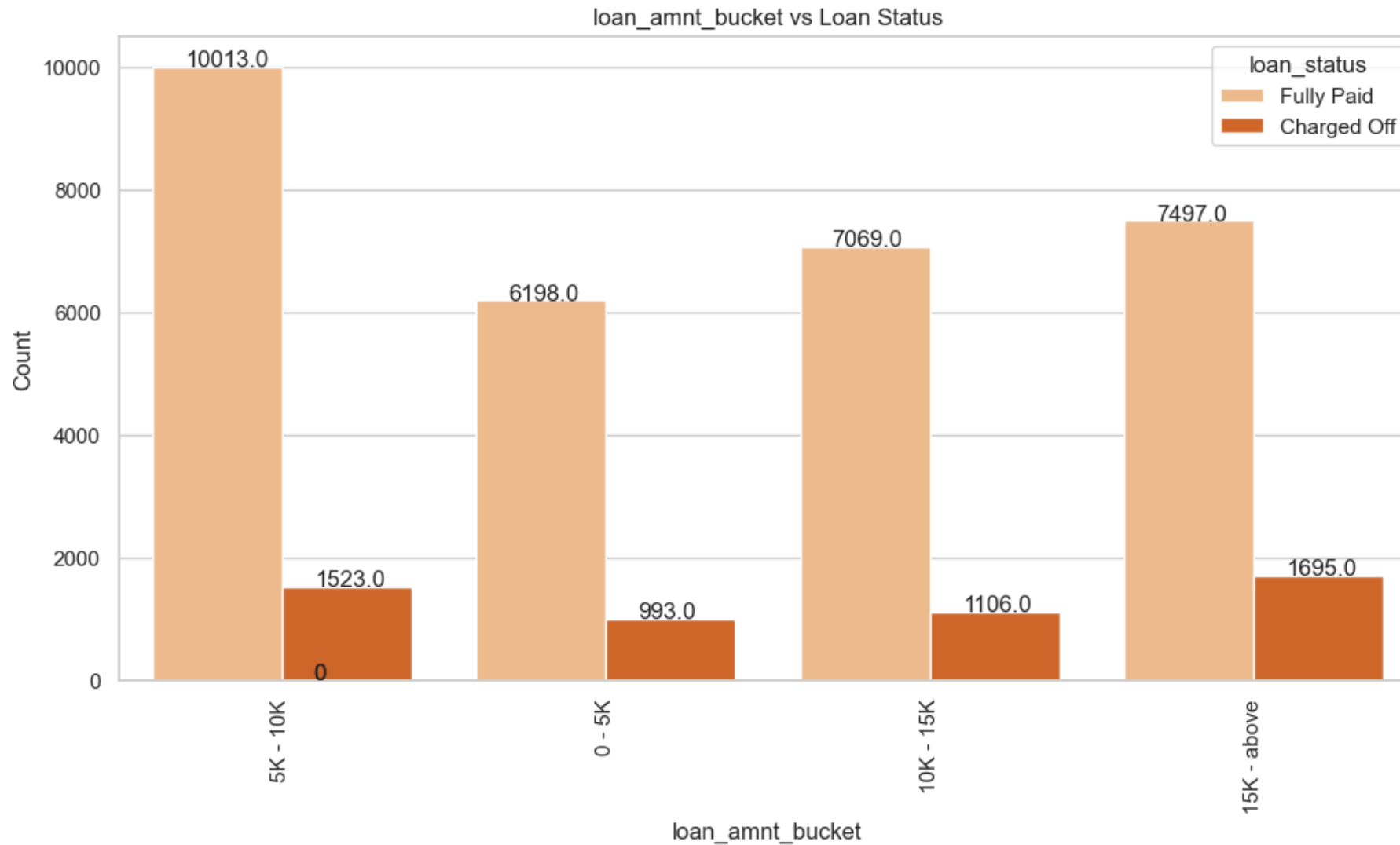
A significant portion of loan applicants who defaulted received loans with interest rates falling within the range of 13% to 17%.



Bucket of Interest Rate of the loan vs Status of Loan

Bivariate Analysis

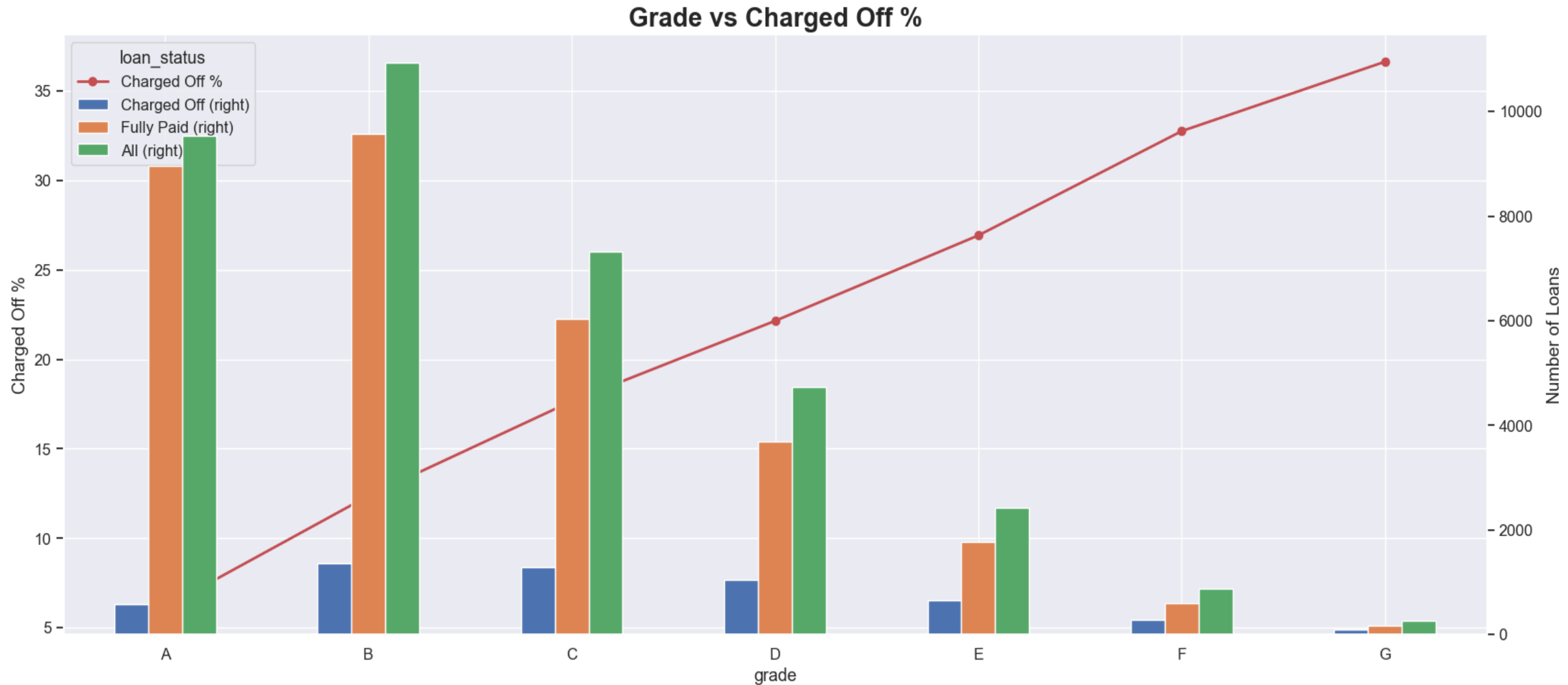
A majority of the loan applicants who defaulted received loan amounts of \$15,000 or higher.



Bucket of Loan Amount associated with the customer vs Status of Loan

Multivariate Analysis

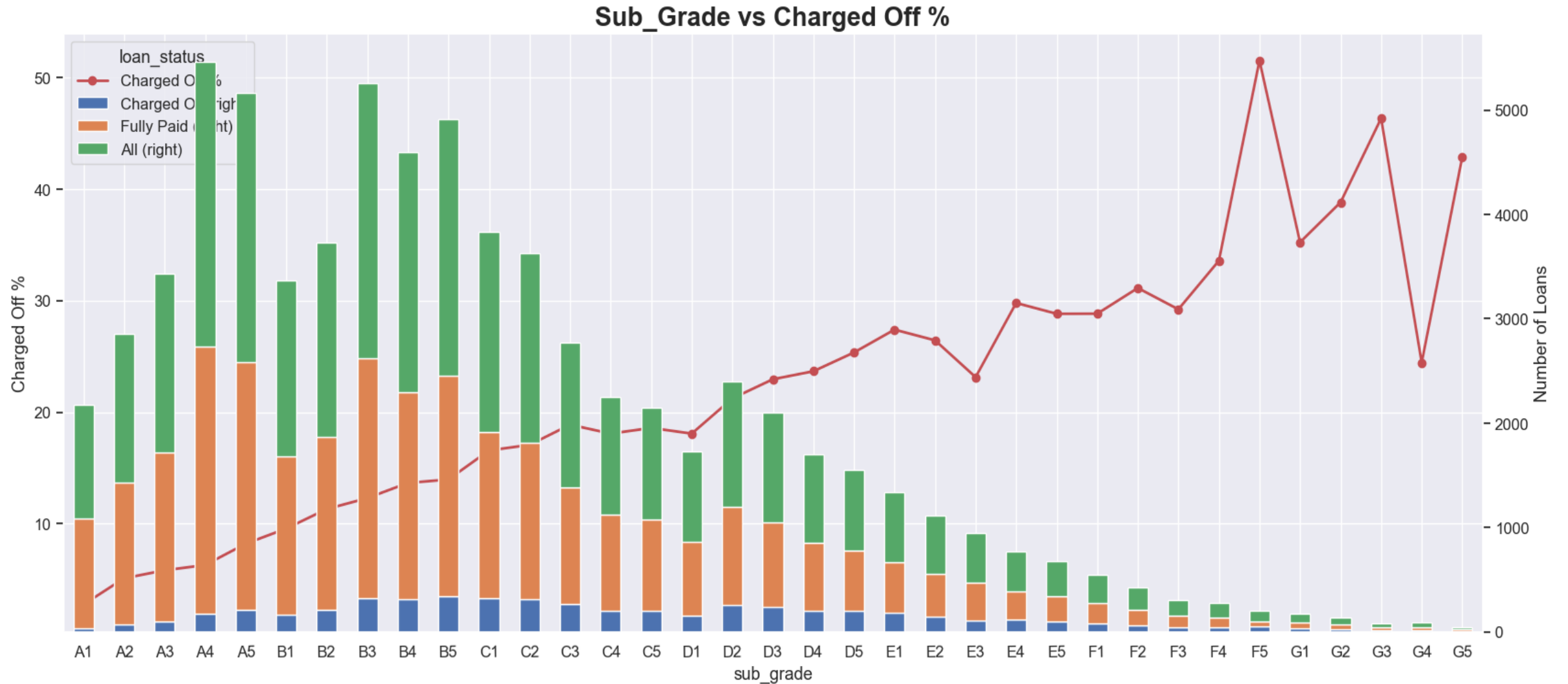
Tendency to default the loan is likely with loan applicants belonging to B, C, D grades.



Grade vs Percentage of Charged-off Loans

Multivariate Analysis

Borrowers from sub grade B3, B4 and B5 have maximum tendency to default.

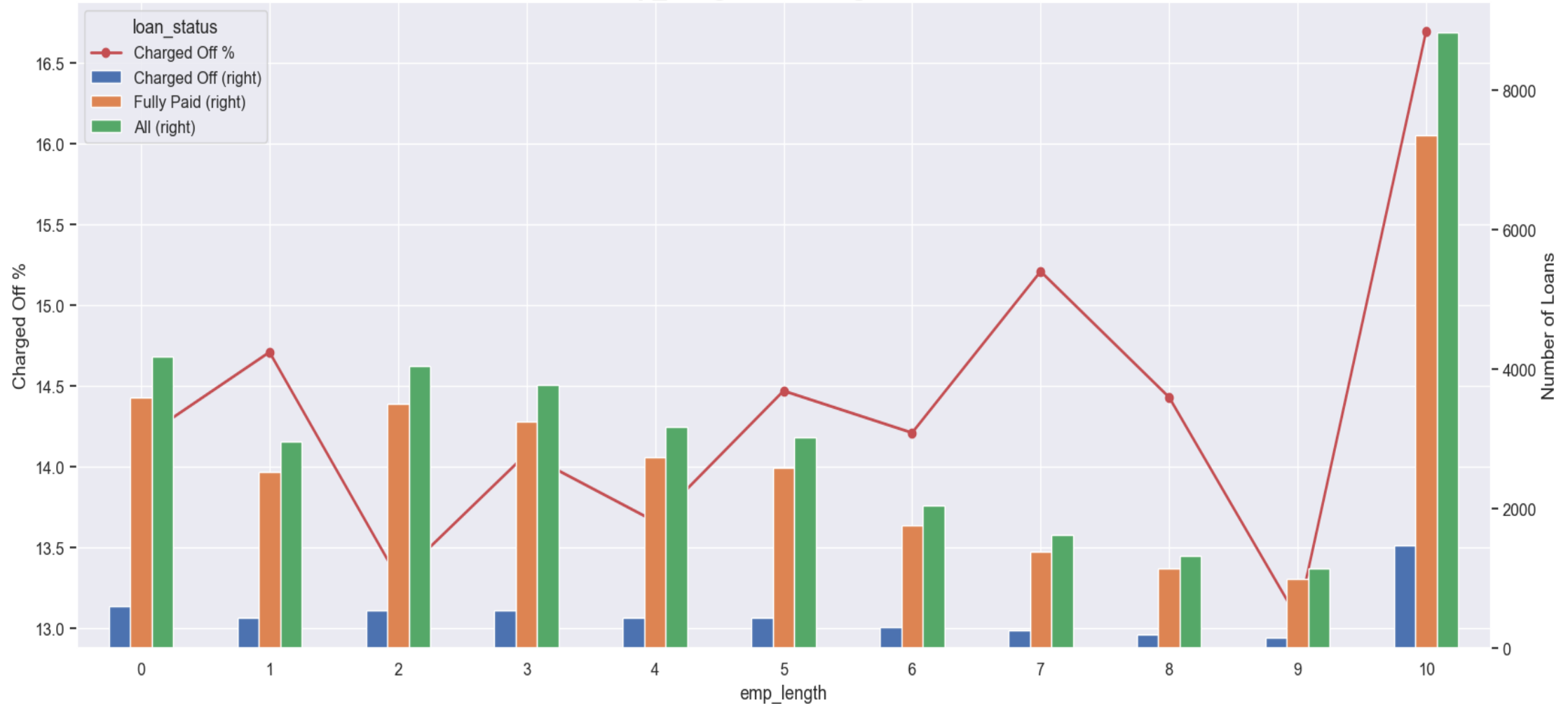


Sub-Grade vs Percentage of Charged-off Loans

Multivariate Analysis

Loan applicants with 10 years of experience has maximum tendency to default the loan.

Emp_Length vs Charged Off %

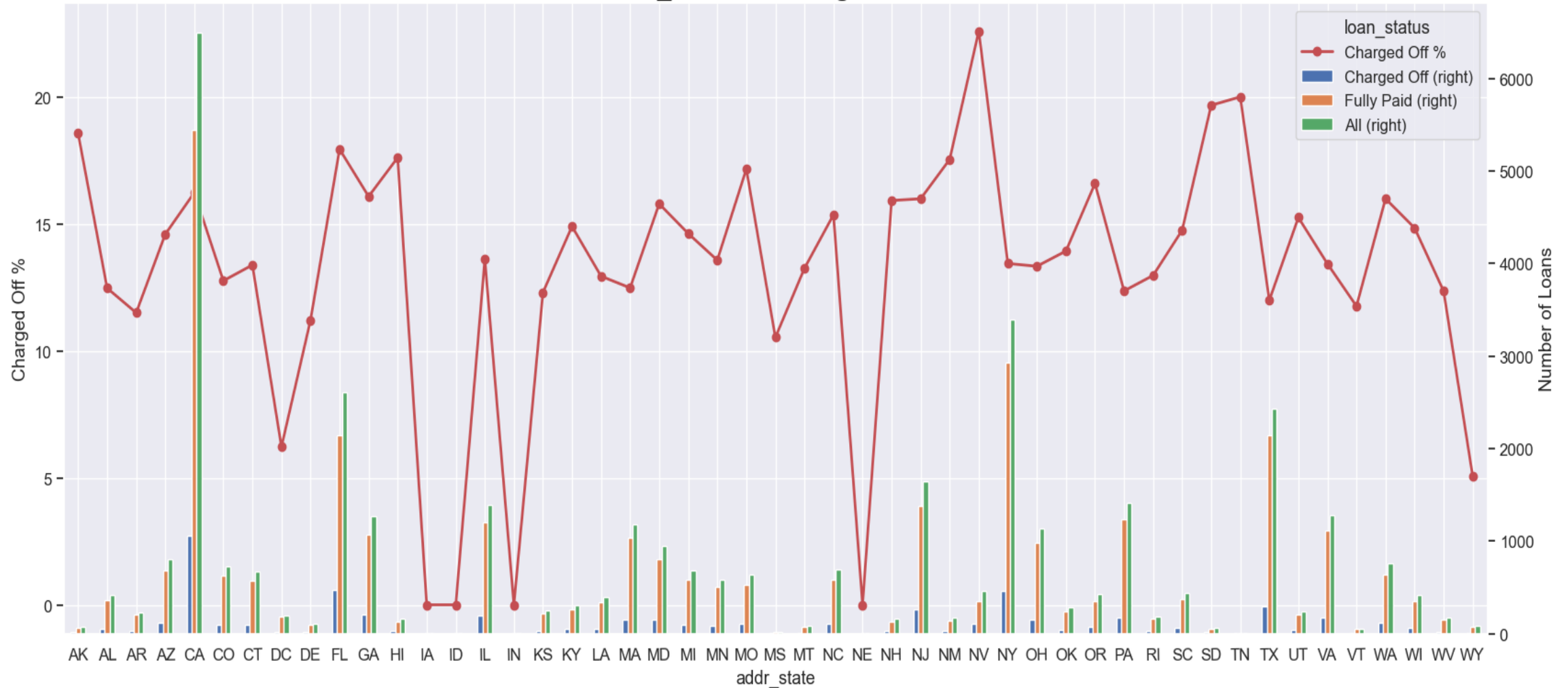


Employment Length (In Years) vs Percentage of Charged-off Loans

Multivariate Analysis

Borrowers from states CA, FL, NJ have maximum tendency to default the loan.

Addr_State vs Charged Off %

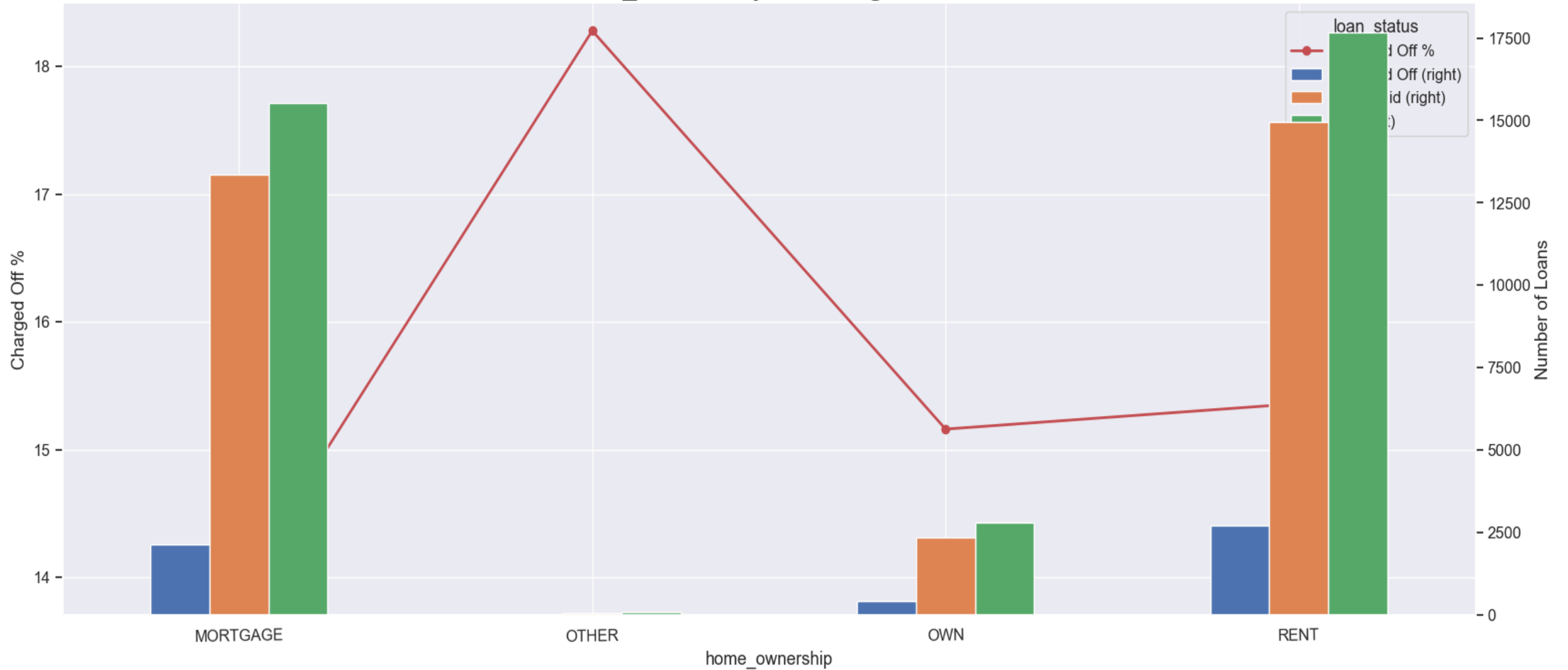


Address State vs Percentage of Charged-off Loans

Multivariate Analysis

Borrowers from Rented House Ownership have highest tendency to default the loan.

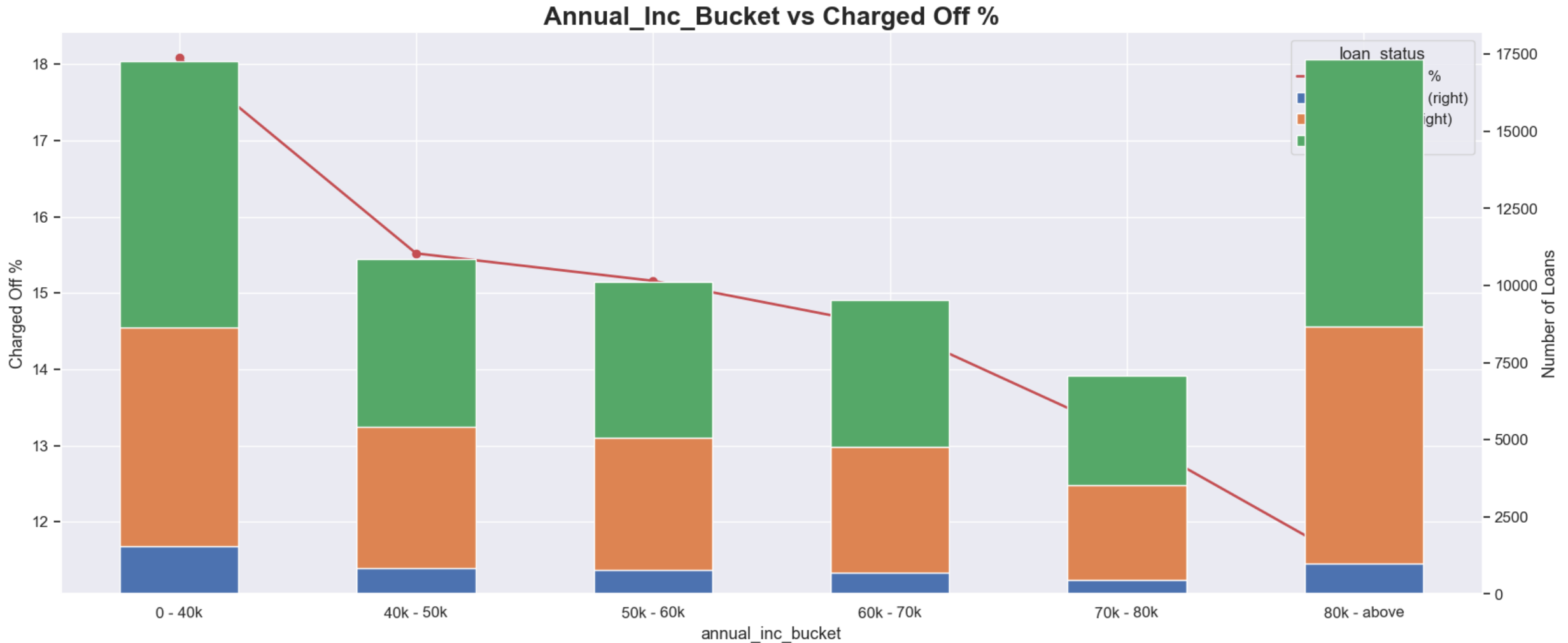
Home_Ownership vs Charged Off %



Home Ownership vs Percentage of Charged-off Loans

Multivariate Analysis

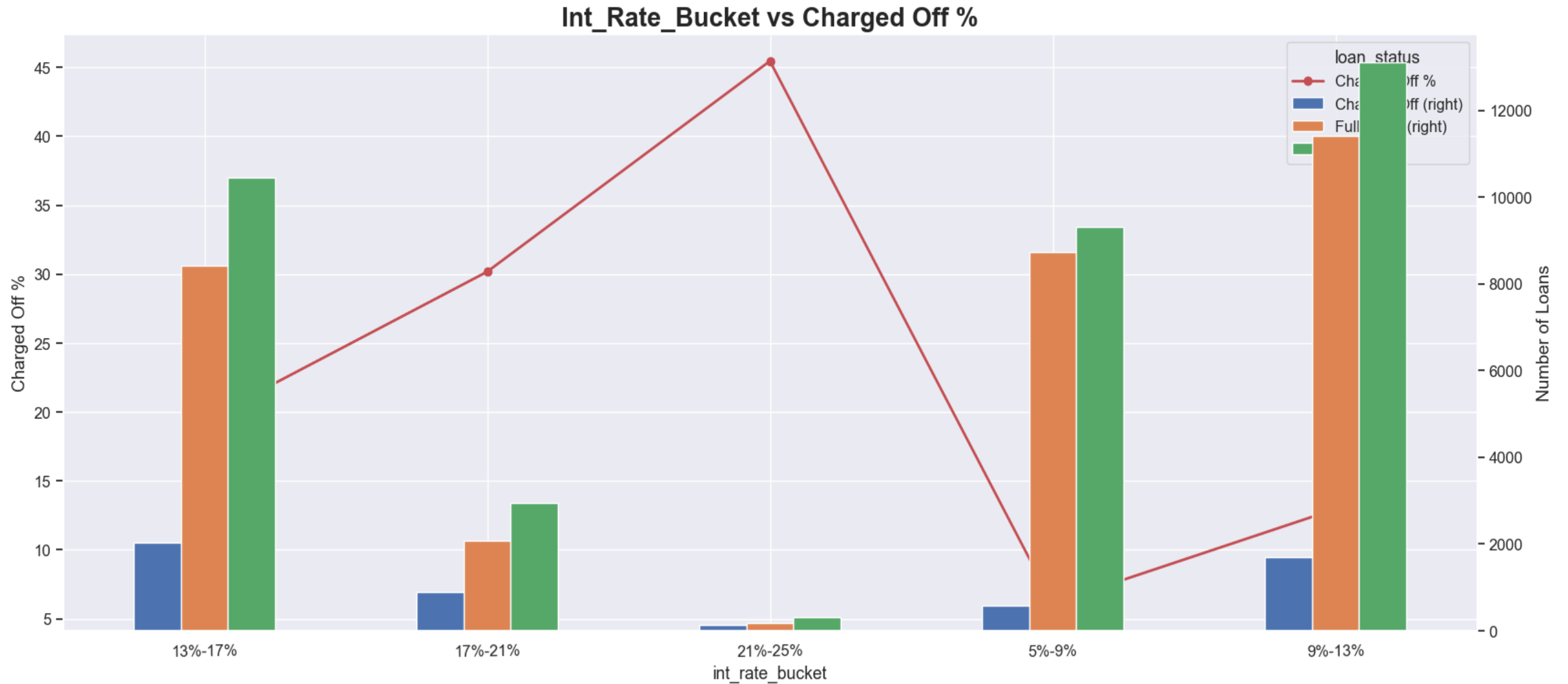
The borrowers who are in lower income groups have maximum tendency to default the loan and it generally decreases with the increase in the annual income.



Buckets of Annual Income vs Percentage of Charged-off Loans

Multivariate Analysis

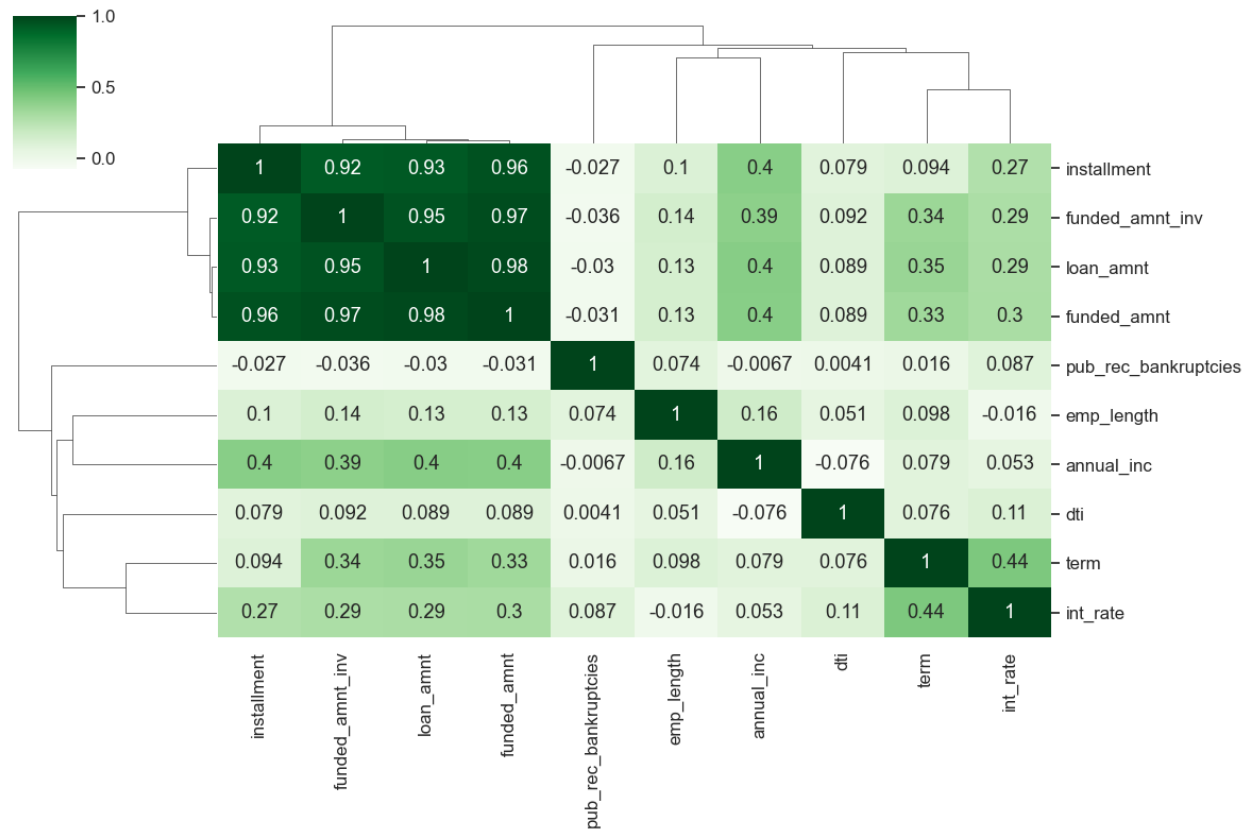
The tendency to default the loan is increasing with increase in the interest rate.



Buckets of Interest Rate vs Percentage of Charged-off Loans

Correlation Analysis

The tendency to default the loan is increasing with increase in the interest rate.



Correlation Matrix among variables namely installment, funded_amnt_inv, funded_amnt, loan_amnt, pub_rec_bankruptcies, annual_inc, emp_length, dti, int_rate , term

Correlation Analysis

Inferences from Correlation Metrics

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

Conclusion

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.

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.

Conclusion

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.

High Loan Amounts: 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.

DTI and Interest Rates: 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.

Low Annual Income: 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.

Recommendations

- **Strengthen Verification Processes:**

- Action: Enhance verification procedures for loan applicants.
- Rationale: By reducing the number of "Not Verified" applicants, default risks can be significantly mitigated.

- **Focused Risk Assessment:**

- Action: Pay special attention to mid-range loan amounts, lower annual incomes, higher DTI ratios, and elevated interest rates.
- Rationale: These factors are critical indicators of increased default risk and warrant closer scrutiny.

- **Develop State and Purpose-Based Restrictions:**

- Action: Implement targeted policies for high-risk geographic regions and specific loan purposes, especially debt consolidation.
- Rationale: Tailored policies address the unique risks associated with different geographic areas and loan purposes.

- **Comprehensive Risk Evaluation:**

- Action: Integrate various risk factors, such as employment length, loan grade, and term duration, into the evaluation process.
- Rationale: A holistic evaluation approach provides a more accurate assessment of default risk.

Thank you

MANIKANTA GRANDHI

VISHAL LOKE

A solid orange horizontal bar spanning the width of the slide, located at the bottom.