



Lending Club Case Study

Submitted by:

Imran Mohammad

Saurabh Shauran



Problem Statement

- **Lending Club**, a consumer finance marketplace focused on providing diverse loan options to urban clientele, encounters a significant challenge in overseeing its loan approval process. In assessing loan applications, the organization is required to make prudent decisions to reduce financial losses, which primarily arise from loans granted to applicants deemed "risky."
- Financial losses, known as Credit Losses, arise when borrowers do not fulfil their loan repayment obligations or default on their loans. In more straightforward terms, borrowers classified as **Charged-Off** contribute the largest share of losses for the company.
- When a loan application is submitted to the company, it must evaluate the applicant's profile to determine whether to approve the loan. This decision involves two types of risks:
 - a. If the applicant is deemed capable of repaying the loan, failing to approve it could mean a missed business opportunity for the company.*
 - b. Conversely, if the applicant is assessed as unlikely to repay the loan, approving it could result in a financial loss for the company.*
- Lending Club seeks to pinpoint high-risk loan applicants to minimize the issuance of such loans, ultimately reducing potential credit losses. Our objective is to uncover the key factors that contribute to loan defaults.

Data Evaluation Process

Data Understanding & Comprehension Of Data :

- The Loan Data Set encompasses the comprehensive information for all loans issued between 2007 and 2011.
- The Data Dictionary provides definitions for each variable included in the Loan Data Set.

Two types of decisions that could be taken by the company:

- Loan Acceptance Outcomes consist of three possible scenarios:
 - 1. Fully Paid - This category includes applicants who have successfully repaid both the principal amount, and the interest associated with the loan.
 - 2. Current - Applicants in this group are actively making loan payments, indicating that the loan term is still ongoing. These individuals are not classified as 'defaulted.'
 - 3. Charged-off - This classification applies to applicants who have not made timely payments for a significant duration, leading to a 'default' status on the loan.
- Loan Rejection - When a loan application is denied by the company, typically due to the applicant not meeting the necessary criteria, there is no transactional history for these individuals. As a result, this information is not available to the company and is excluded from the dataset.

Data Cleaning(addressing missing values, eliminating unnecessary columns, etc.):

- We eliminated columns with over 60% missing values, those with only a single unique value, and columns that do not impact our target variable, which is loan status.
- We are operating under the assumption that columns indicative of customer behaviour, such as Delinquency and revolving balance/utility, are not pertinent to our analysis.
- We have excluded all records where the loan status is "Current," as this information does not contribute to our analysis.

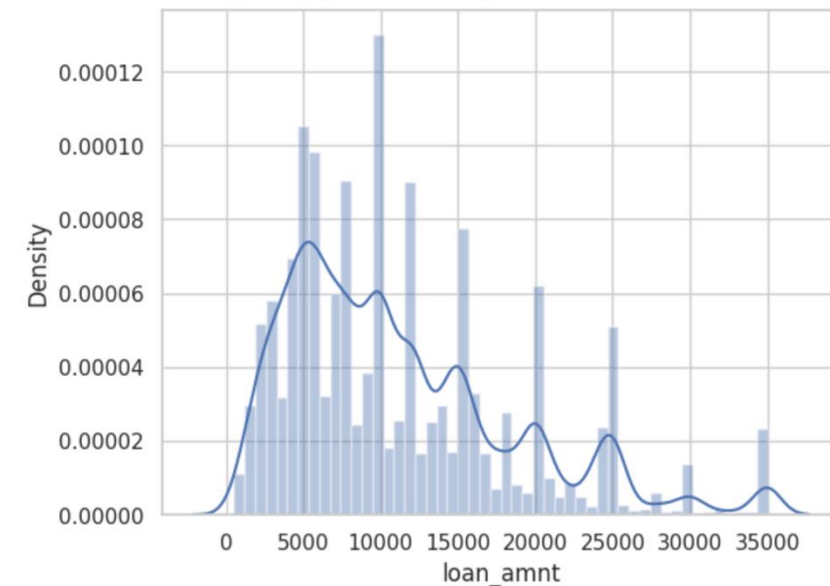
Categorical Variables Vs the Percentage of Charged Off:

- The primary objective of this analysis is to identify the factors that most significantly influence Loan Status, specifically whether an applicant will successfully fulfil the loan term or default. To achieve this, we will introduce the term Percentage of Charged Off, defined as follows:
- $\text{Percentage of Charged Off} = (\text{Number of Applicants Charged Off} / \text{Total Number of Applicants}) * 100.$
- We will compute this percentage in relation to each variable during the bivariate analysis and examine how the Percentage of Charged Off varies across these variables.

After cleansing the data, we verify the allocation of loan amount.

```
[33] #Verifying the allocation of the loan amount.  
sns.distplot(loan.loan_amnt)
```

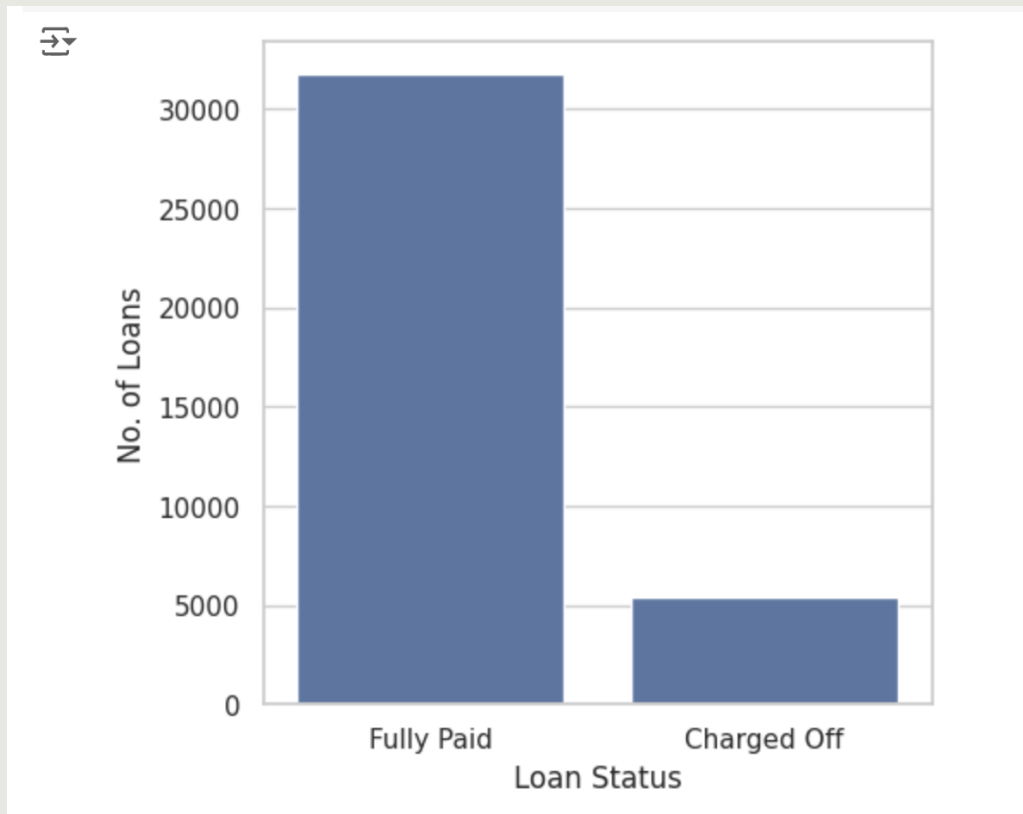
<Axes: xlabel='loan_amnt', ylabel='Density'>



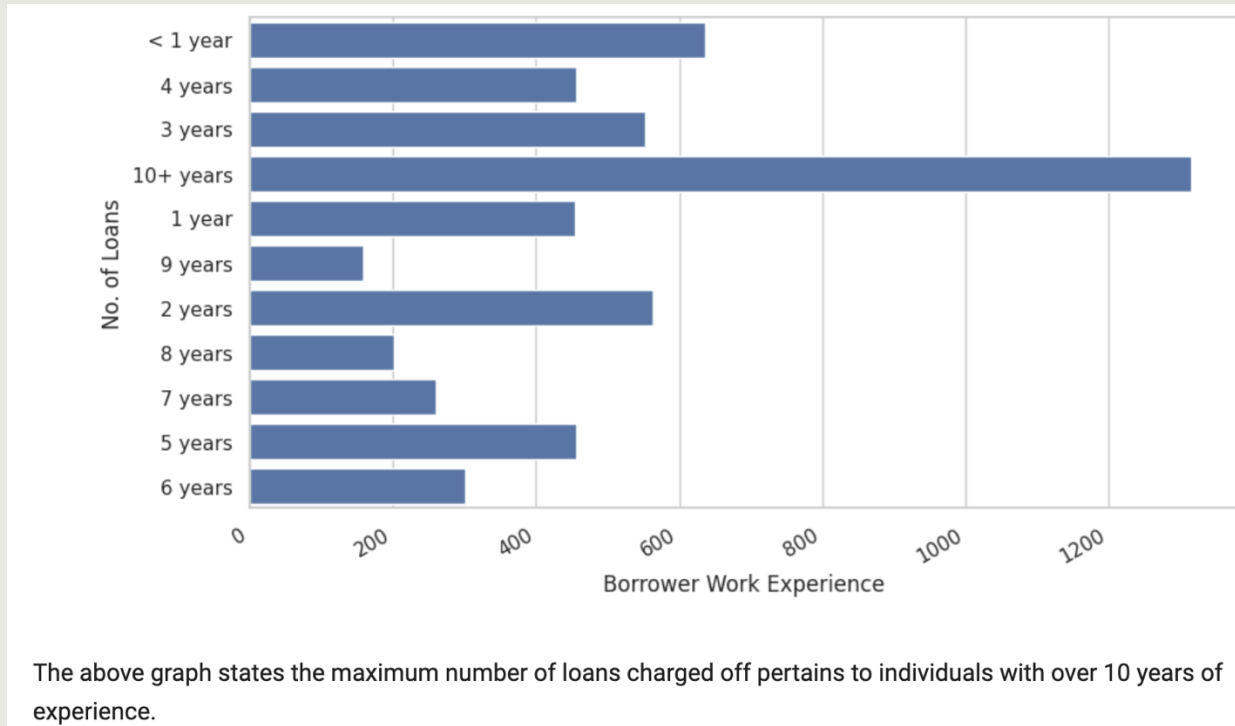
Loan Analysis

- **Univariate Analysis:**

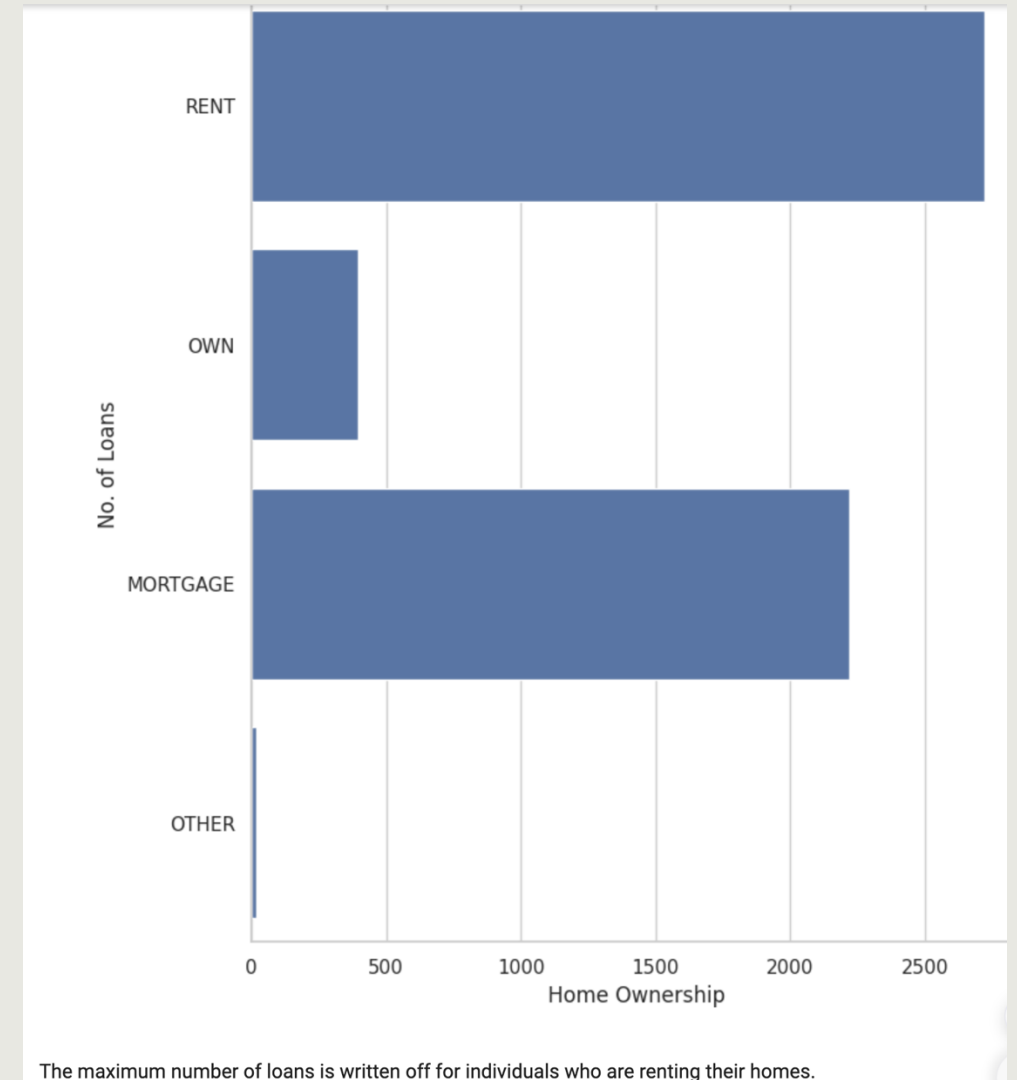
- Univariate analysis is a statistical technique employed to examine and summarize data sets that focus on a single variable. This method concentrates on analysing one variable at a time, allowing for insights into its distribution, central tendency, and variability.
- Visualization of Loan Status Distribution is shown below :



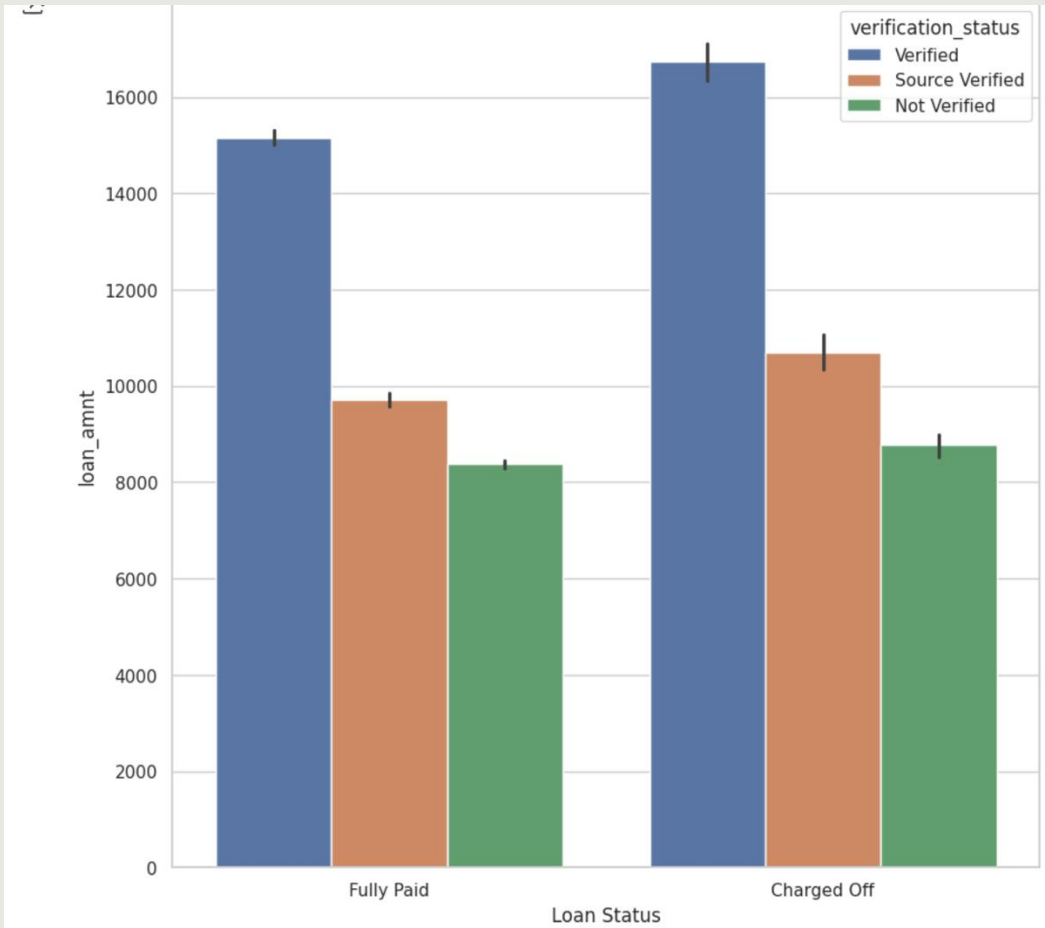
- Visualization of data focused on work experience of borrower :



- Visualization of data focused on Home Ownership of borrower :

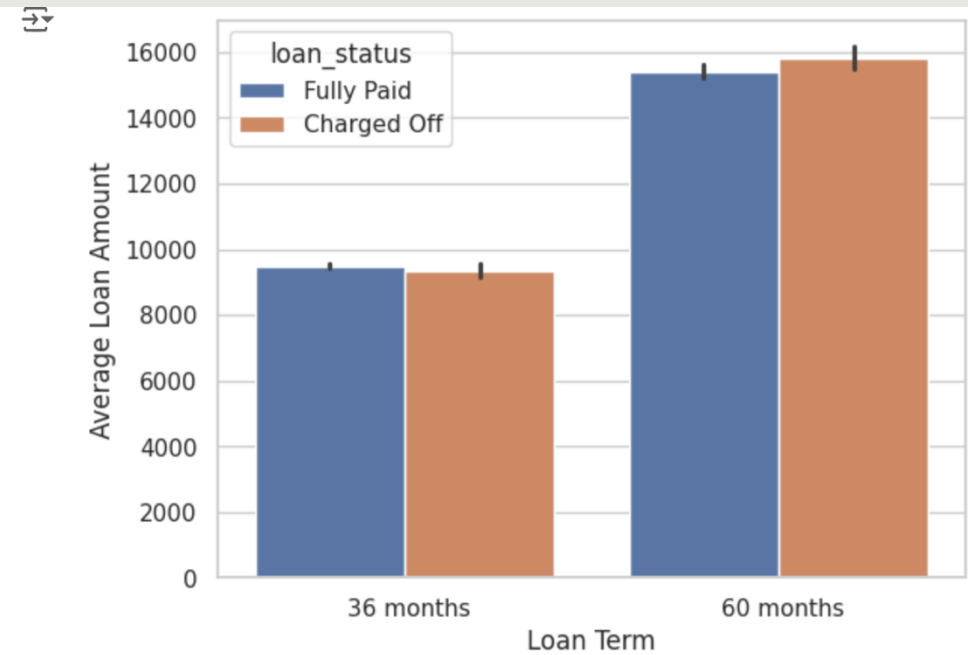


Visualization of data focused on Verification Status of borrower :



The above graph states the maximum number of loans is charged off when the verification status is confirmed as Verified.

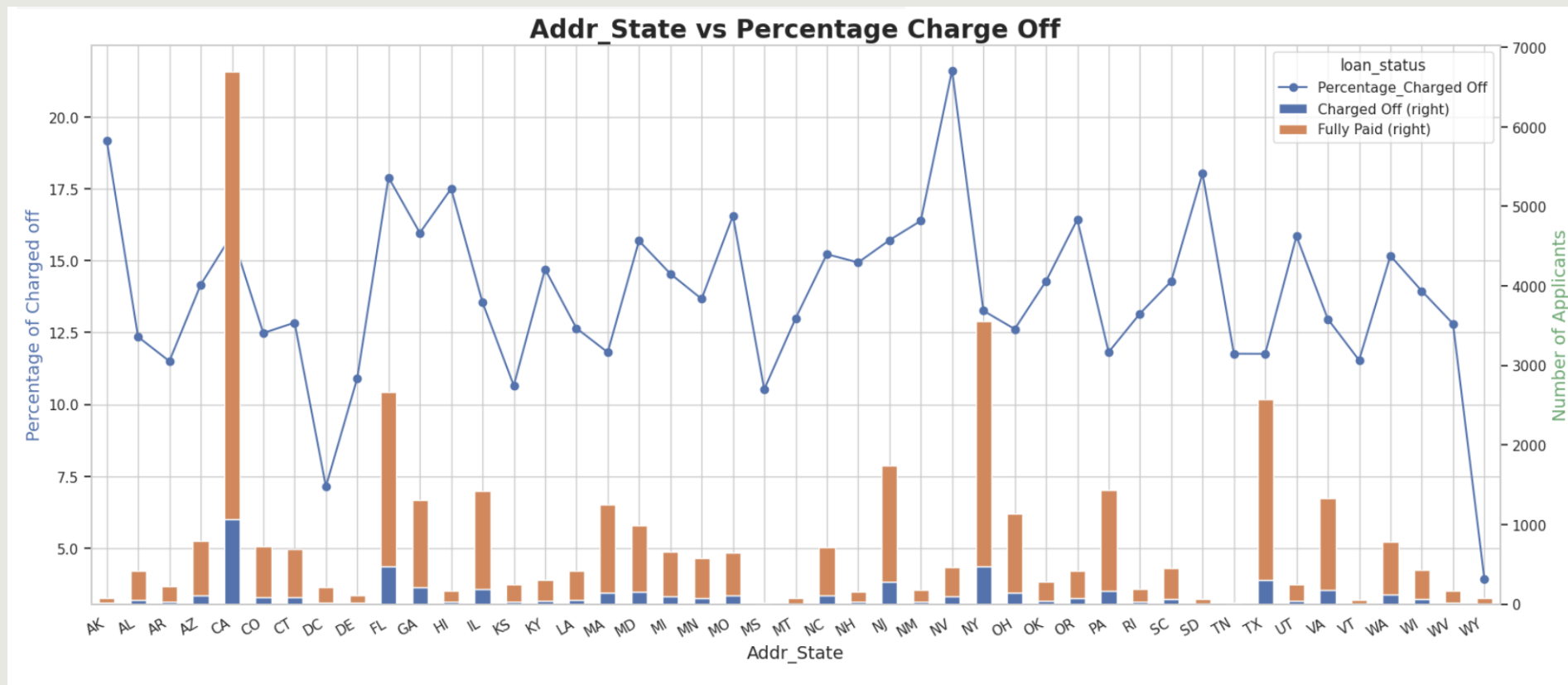
Visualization of data focused on Loan Term of borrower :



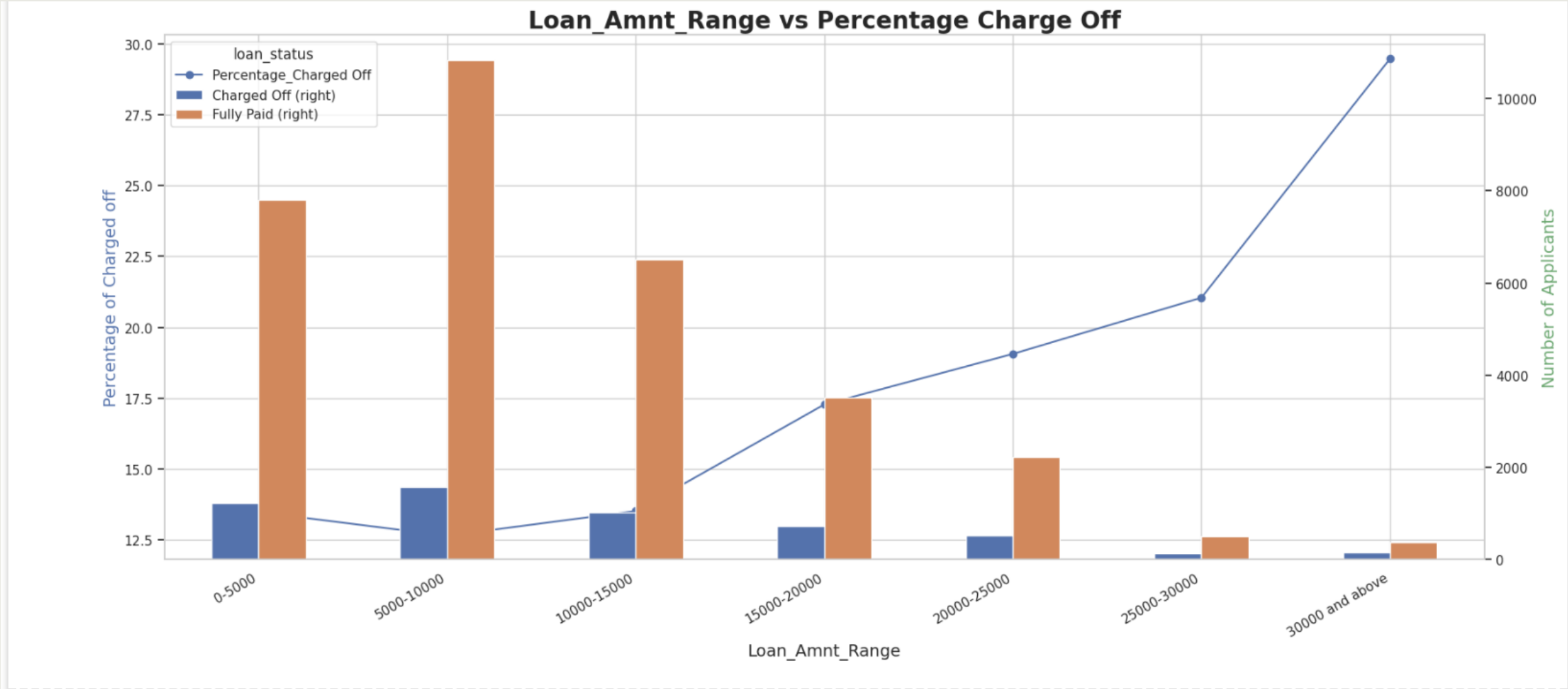
The data presented in the graph indicates that the average loan amount remains consistent across both terms for each loan status. However, it is noteworthy that the average loan amount is greater for the 60-month term.

- **Bivariate Analysis:**

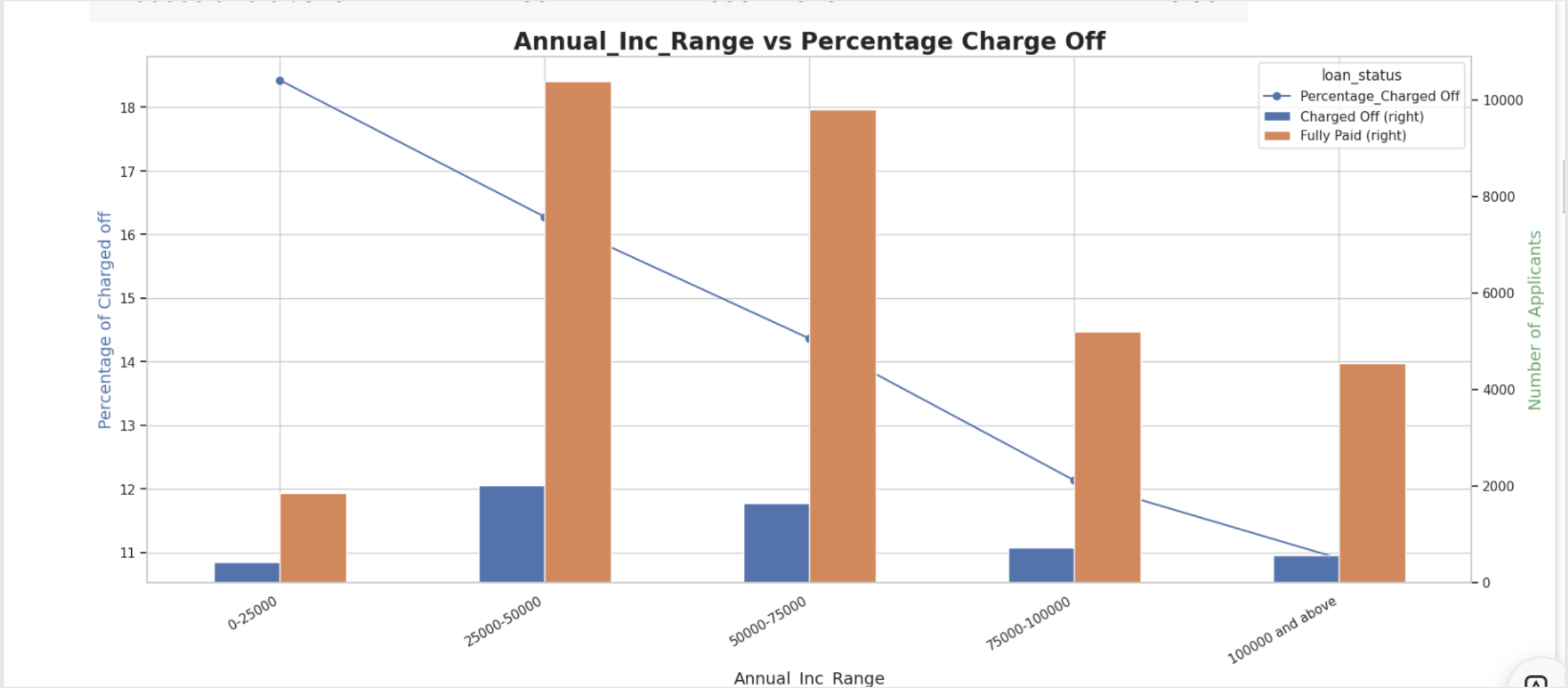
- Bivariate analysis is a statistical technique that examines two variables concurrently. Its purpose is to uncover the empirical relationship that exists between these variables. This method can be employed to test hypotheses, recognize patterns, or investigate the connections between the variables.
- The highest number of loan applications comes from individuals in California.
- The greatest number of loan charge-offs occurs among residents of NV.



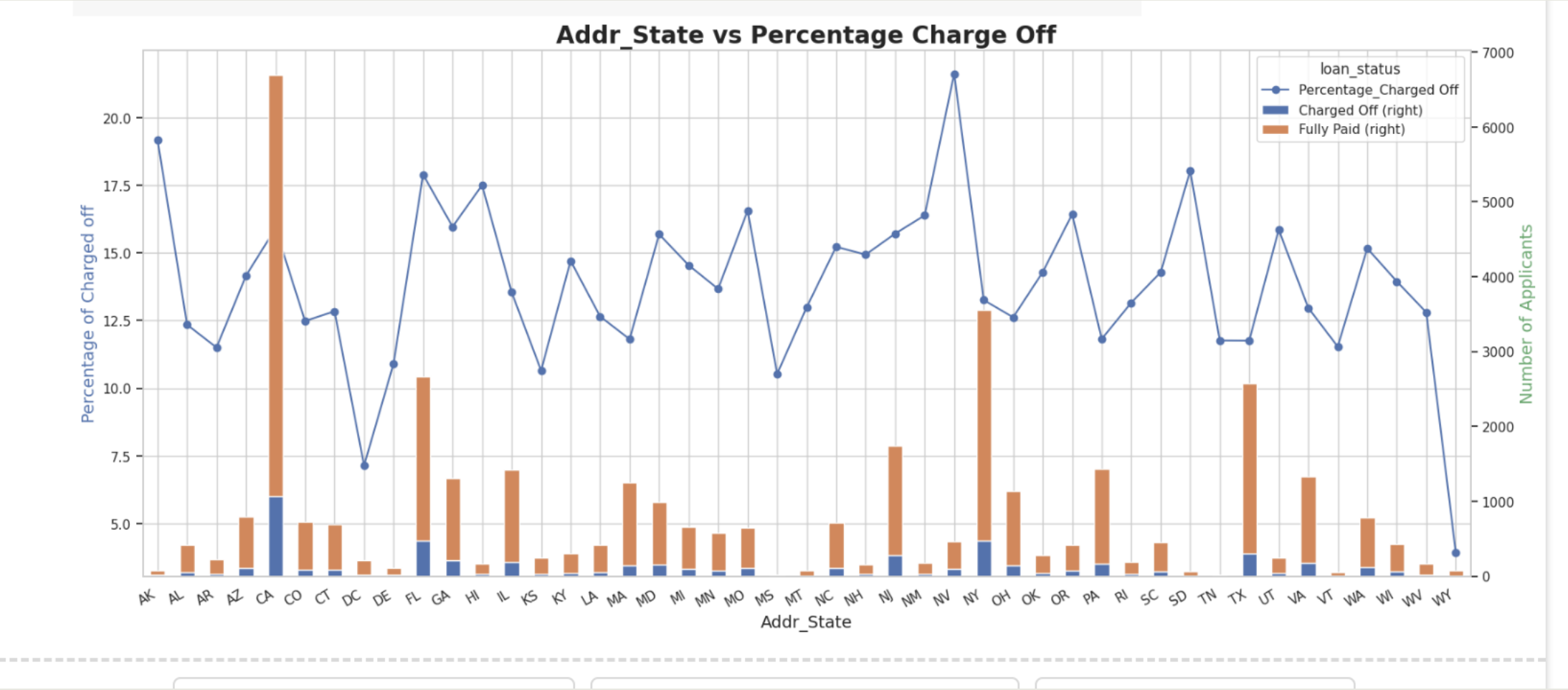
Visualizing the loan amount in relation to the percentage of loans charged off and the number of loan applicants.



Analysing the annual income range in relation to the percentage of loans charged off and the number of loan applicants.

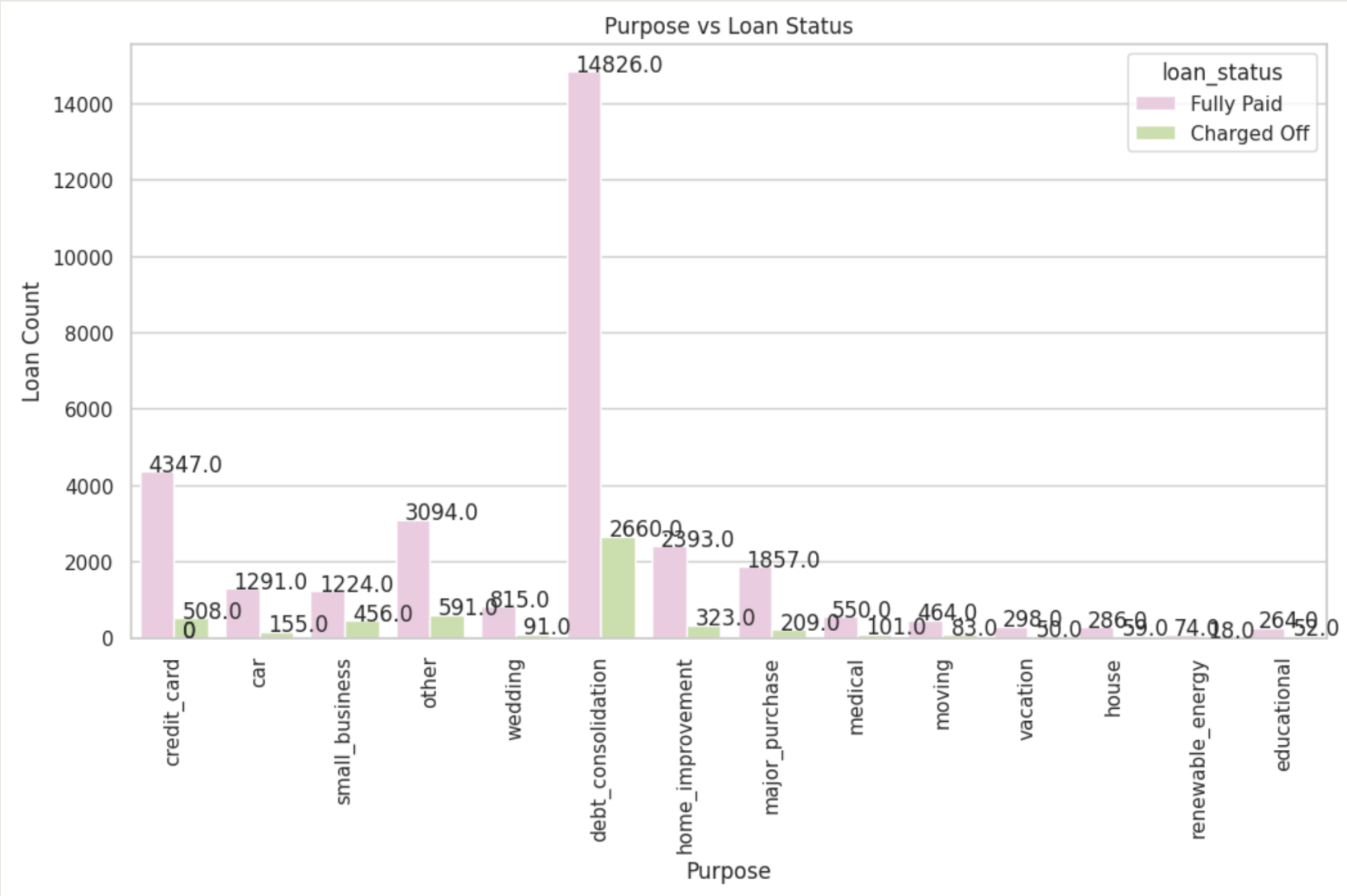


The highest number of loan applications comes from individuals in California. Conversely, the greatest number of loans charged off pertains to residents of NV.



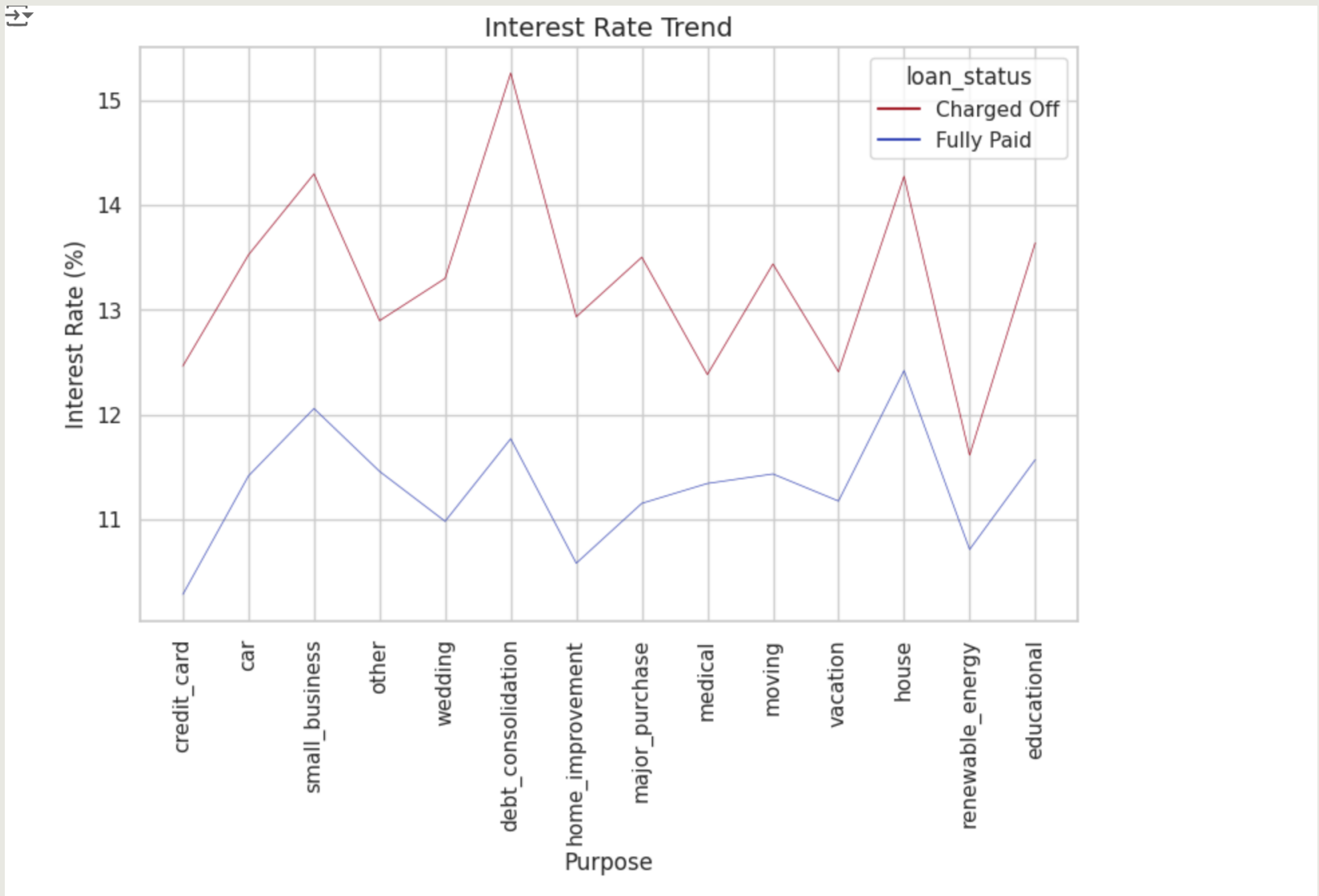
Purpose Vs Loan status

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



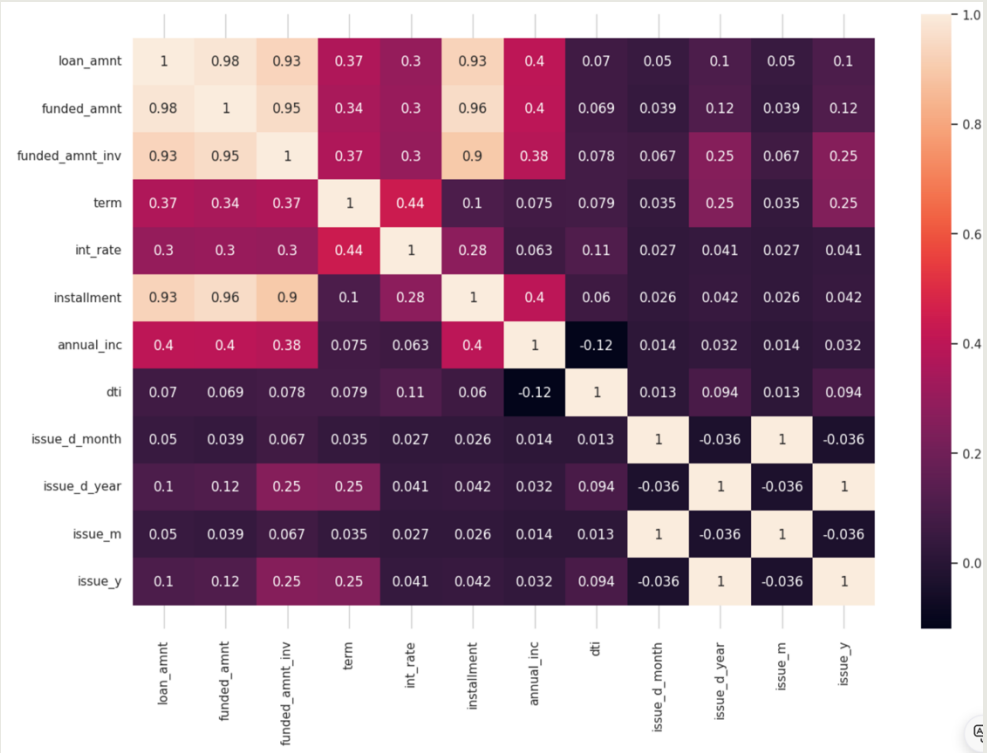
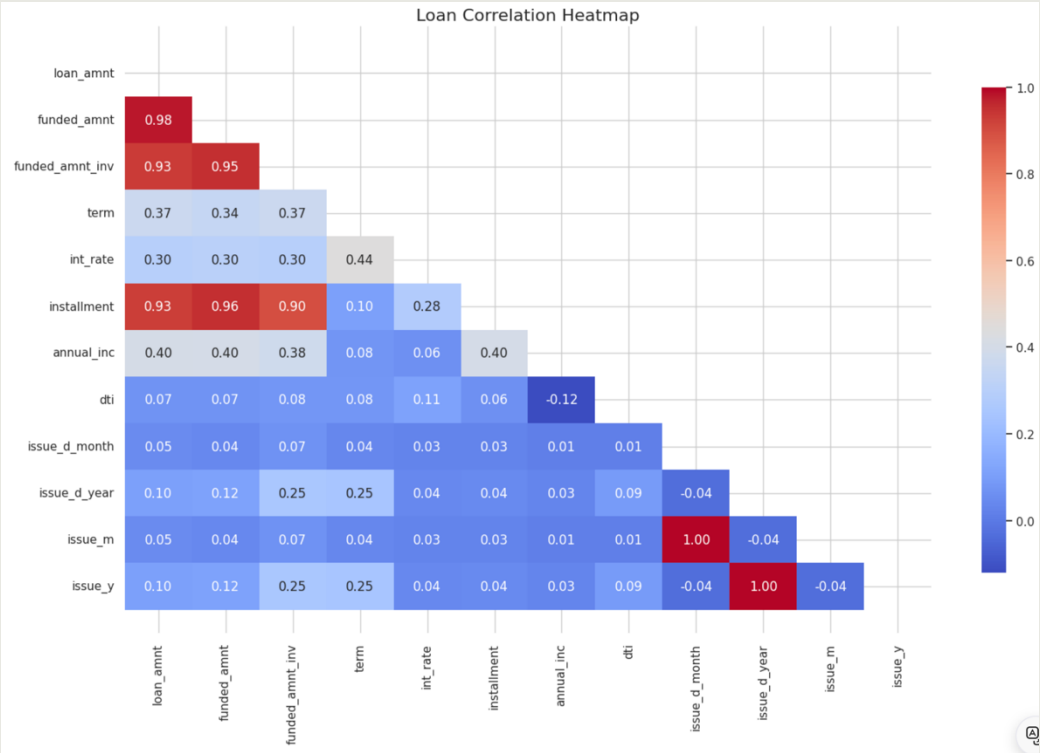
Purpose Vs Loan status

The interest rate is elevated for charged-off loans regardless of the purpose.



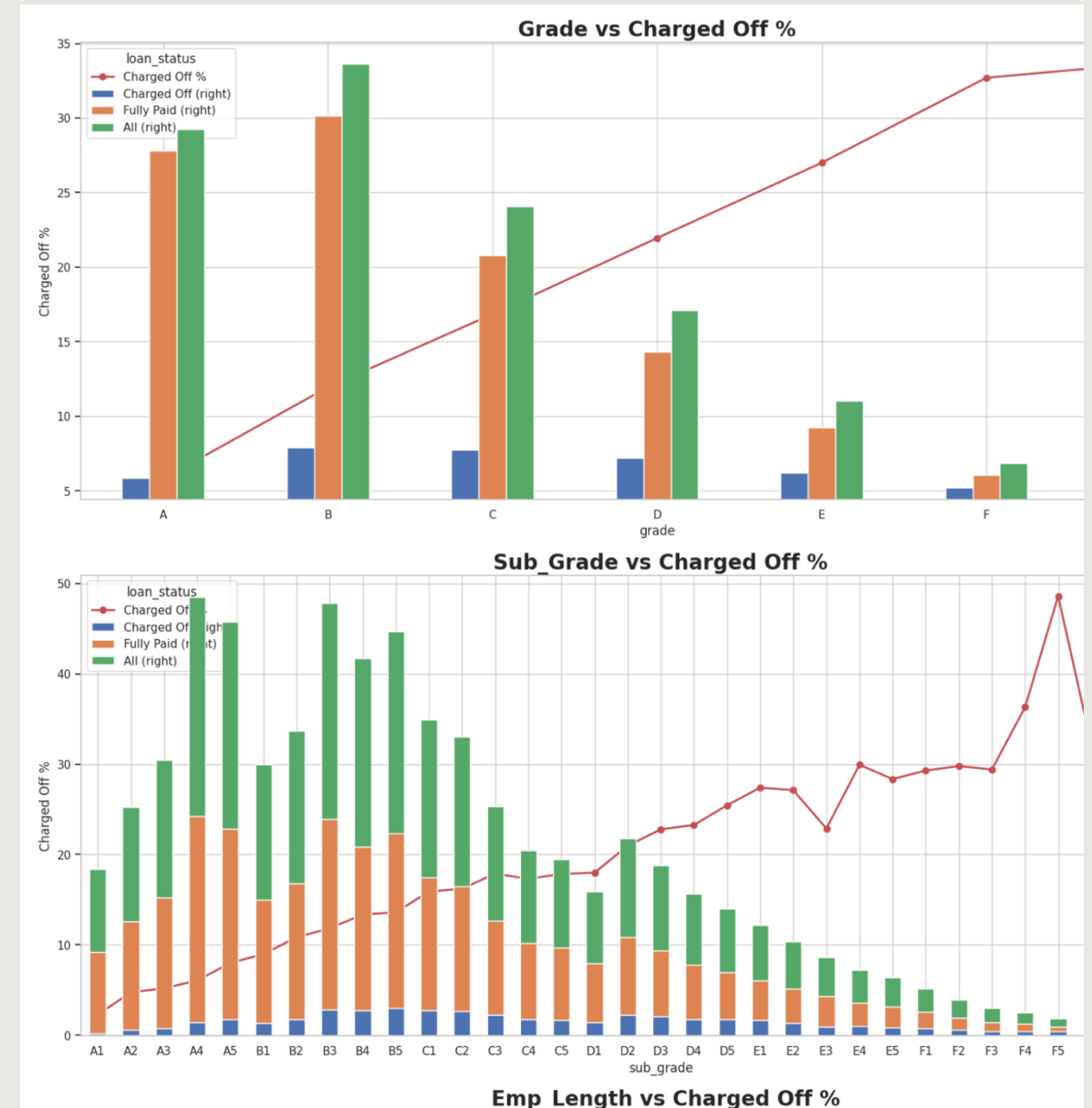
Correlation Matrix

The Loan Amount, Funded Amount, the amount invested from the Funded Amount, and the instalment payments exhibit a strong correlation with one another. Conversely, there is a negative correlation between Annual Income and Debt-to-Income (DTI) ratio.

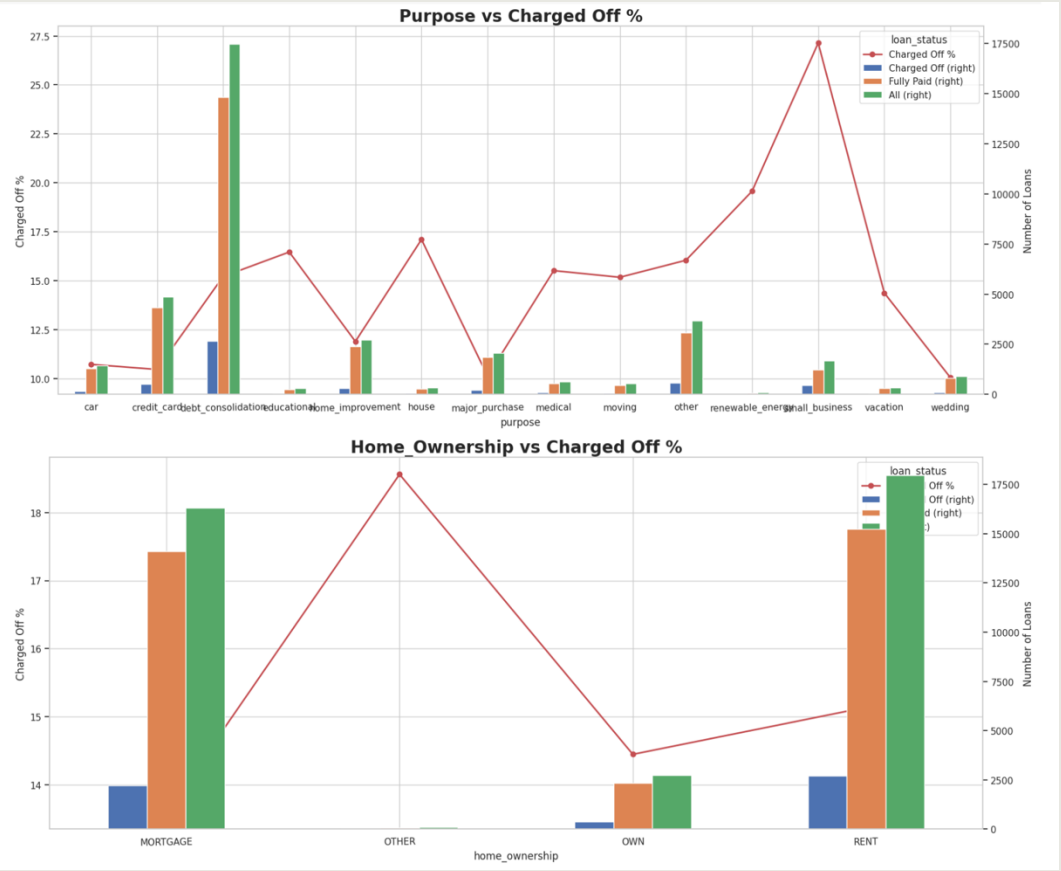
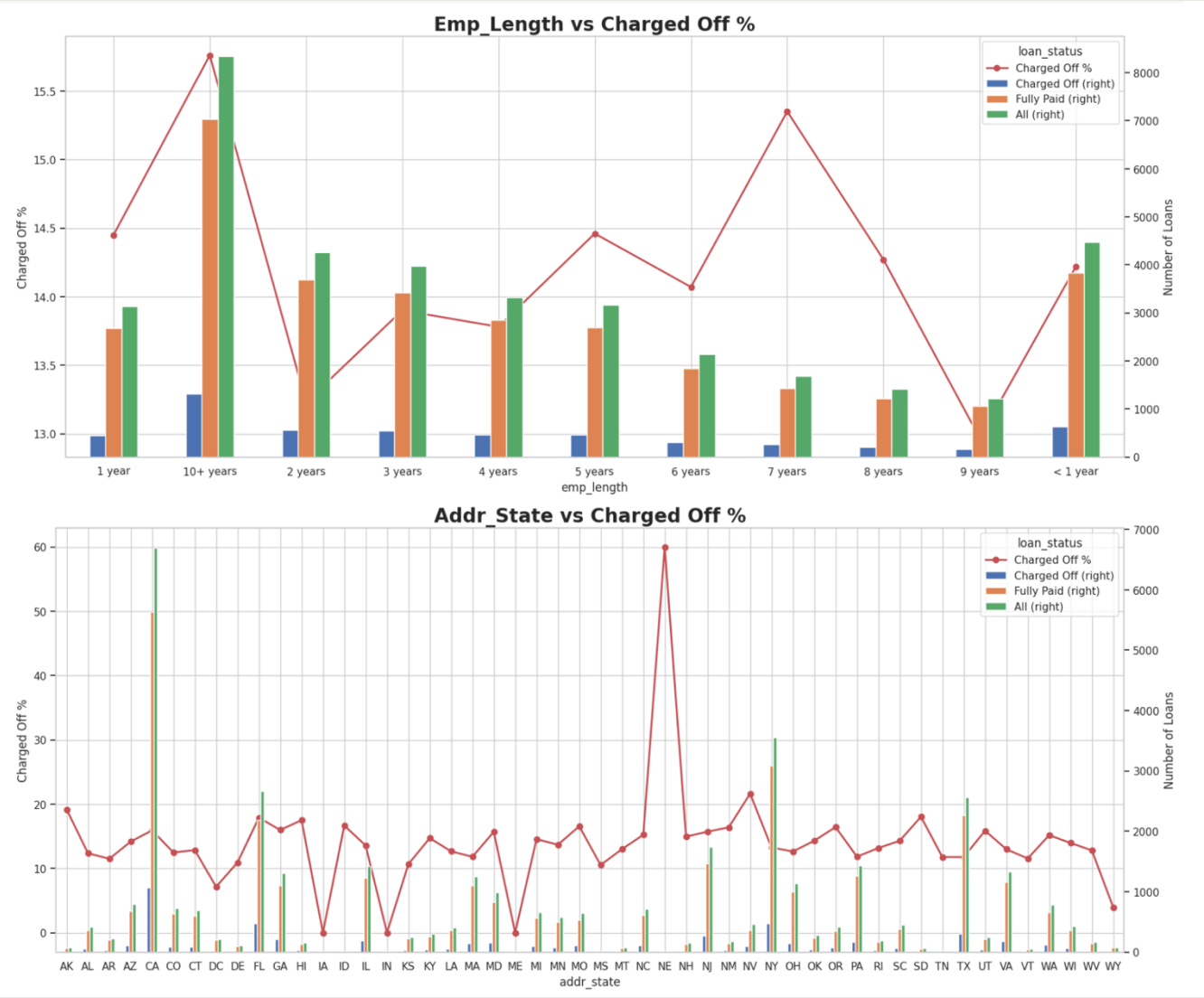


- **Multivariate Analysis:**

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



Multivariate Analysis:



- **Summary of Multivariate Analysis**

1. Loan applicants classified as grades B, C, and D exhibit a higher likelihood of defaulting on their loans.
2. Among these, borrowers in subgrades B3, B4, and B5 show the greatest propensity to default.
3. Applicants with a decade of work experience are also more prone to loan default.
4. The states of California, Florida, and New Jersey have borrowers with the highest default rates.
5. Individuals residing in rented accommodations demonstrate the highest likelihood of defaulting on loans.
6. Borrowers from lower income brackets are more likely to default, with this tendency diminishing as annual income increases.
7. There is a correlation between rising interest rates and an increased likelihood of loan default.

- **Suggestions**

1. Lending Club should exercise caution regarding loans designated for Small Business purposes, as the charge-off rate can reach as high as 27.1%. It is advisable to prioritize loans intended for Weddings, major purchases, vehicles, and credit card consolidation.
2. There is a correlation between loan amount and the likelihood of charge-offs; thus, Lending Club should focus on approving smaller loans. The highest volume of applications, totalling 12,410, falls within the 5,000to10,000 range, where the probability of charge-offs is significantly lower, indicating a reduced risk for the organization.
3. Lending Club may benefit from increasing its acceptance of loans from applicants with an annual income exceeding \$100,000, as these individuals exhibit a minimal charge-off probability.
4. It would be prudent for Lending Club to accept more loans with interest rates below 7.5%, as these loans also demonstrate a lower likelihood of charge-offs.
5. Lending Club should aim to approve a greater number of loans classified as grade A and B, while remaining cautious with loans that fall into grades E, F, and G.
6. Lending Club should consider increasing its acceptance of loans from homeowners, as they tend to present a lower risk profile.
7. The number of loan applicants is on the rise each year, which is beneficial for Lending Club, leading to an increase in both accepted loans and charge-offs. The final months of the year are particularly significant, as many loans are either fully repaid or charged off during this period.
8. Lending Club should focus on accepting more loans with a 36-month term, as the percentage of charge-offs is lower and the number of applicants is higher. Additionally, the average amount of fully paid loans has been increasing, while the average amount of charged-off loans has been decreasing over the years.



Thank You

