Case Study Submission- Lending Club Case Study

SUBMITTED BY

POONAM PARATE

PRASAD NARALA

Problem Statement

▶Solving this assignment will give you an idea about how real business problems are solved using EDA. In this case study, apart from applying the techniques you have learnt in EDA, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Business Understanding

You work for a consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

▶ If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company

▶ If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company
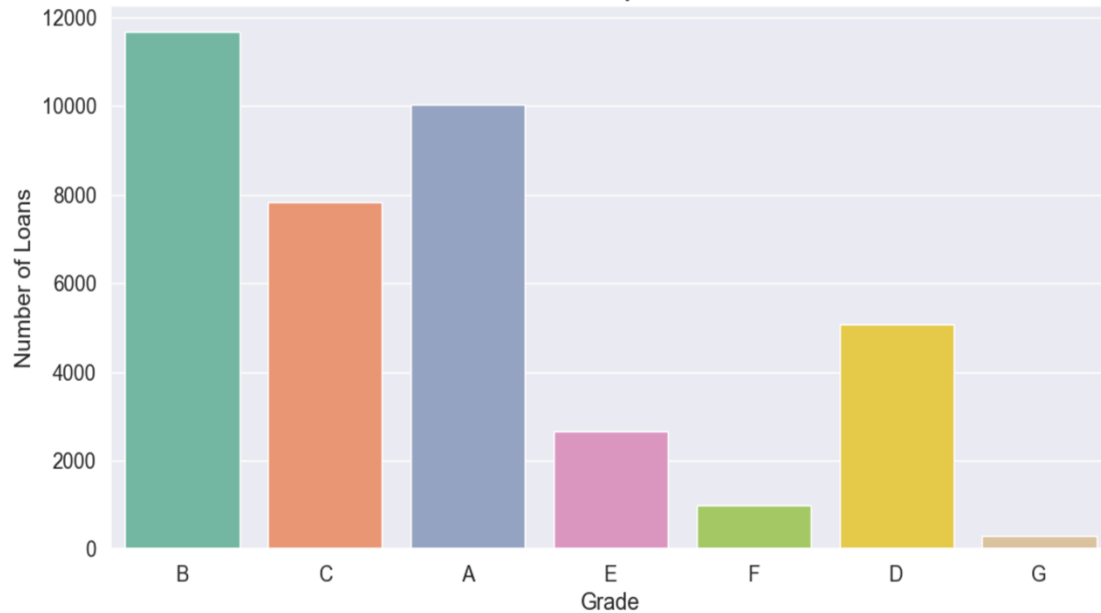
Univariate Analysis

Number of Applicants Year-wise
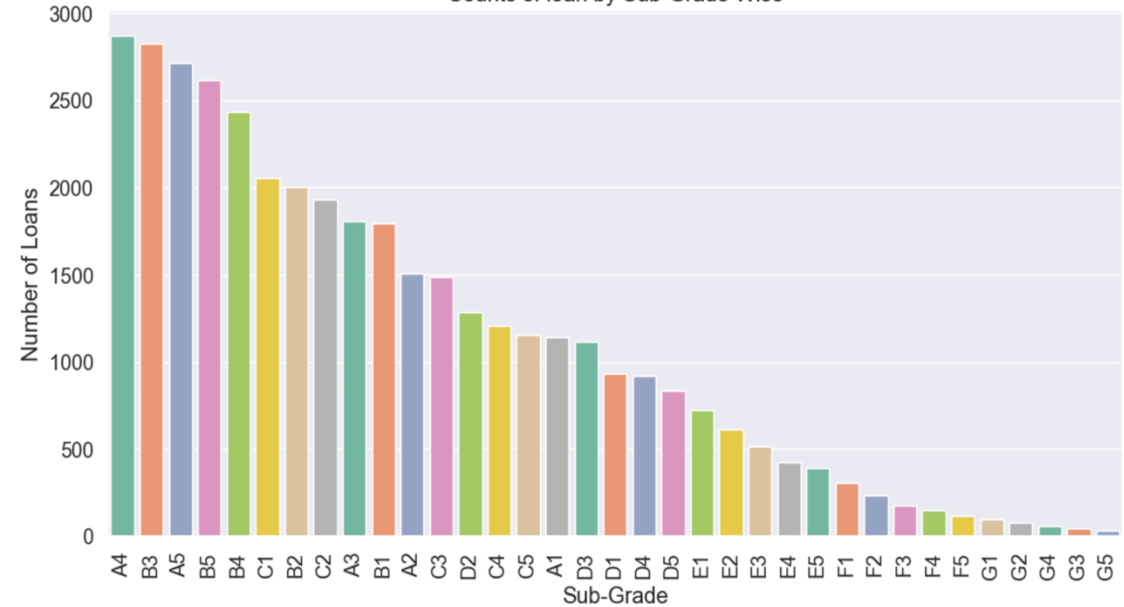

Count of Loan Purposes
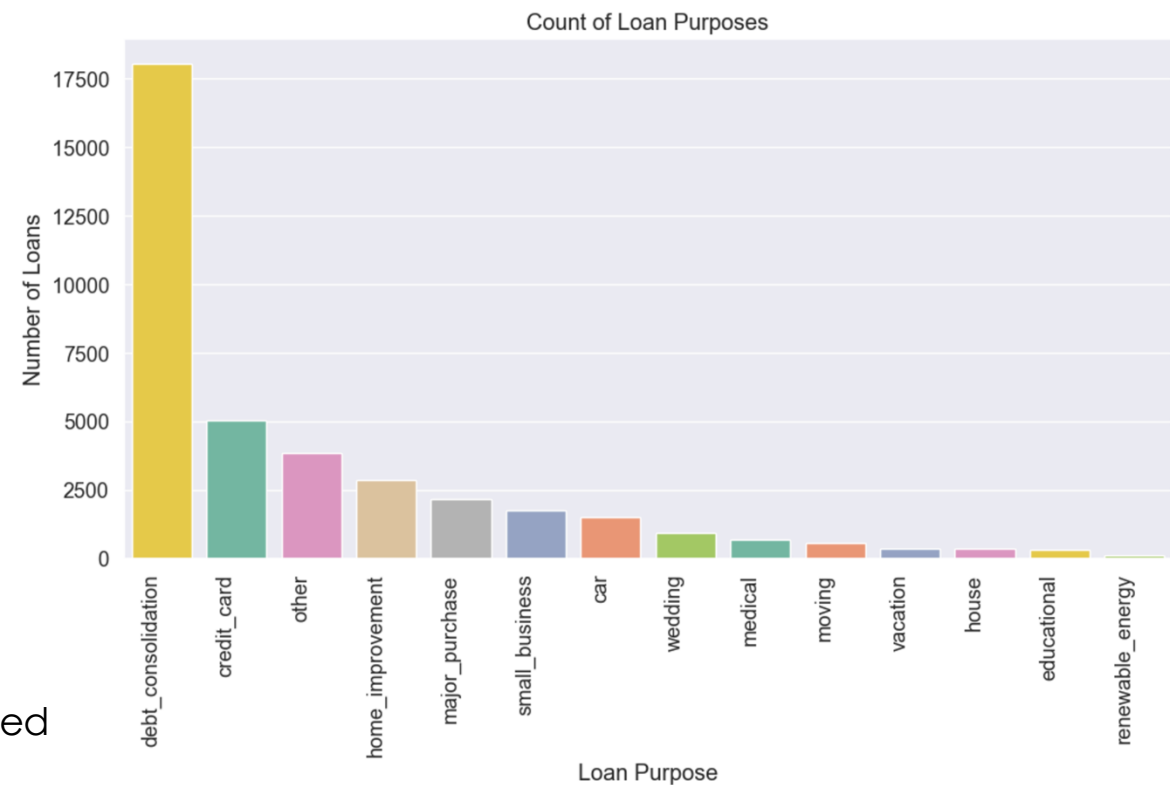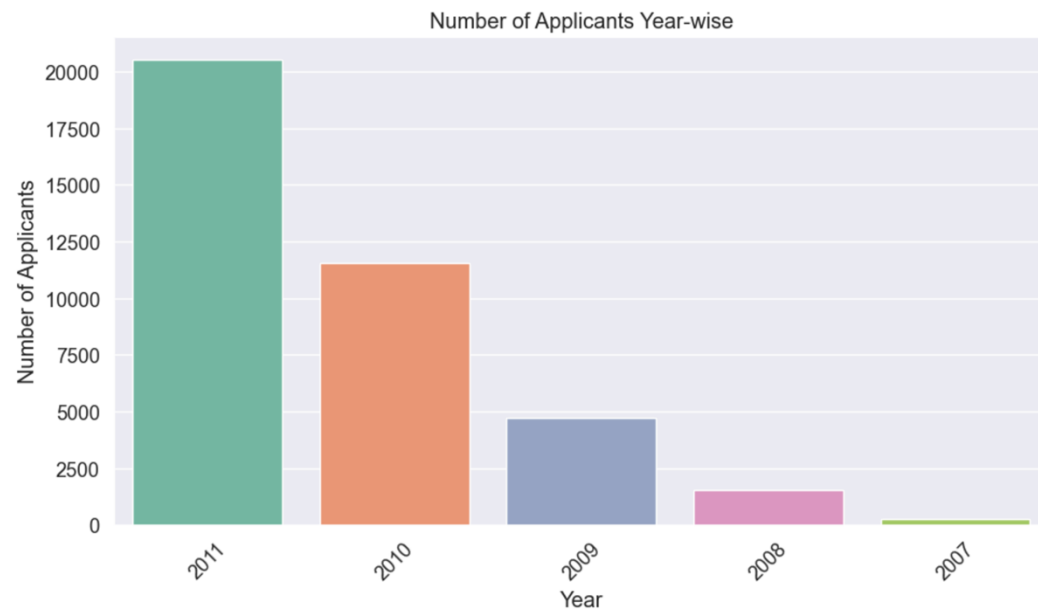
- The number of loan applicants has significantly increased each year from 2007 to 2011
- The major purposes for loans are debt consolidation, credit card, other, home improvement, etc.

Univariate Analysis
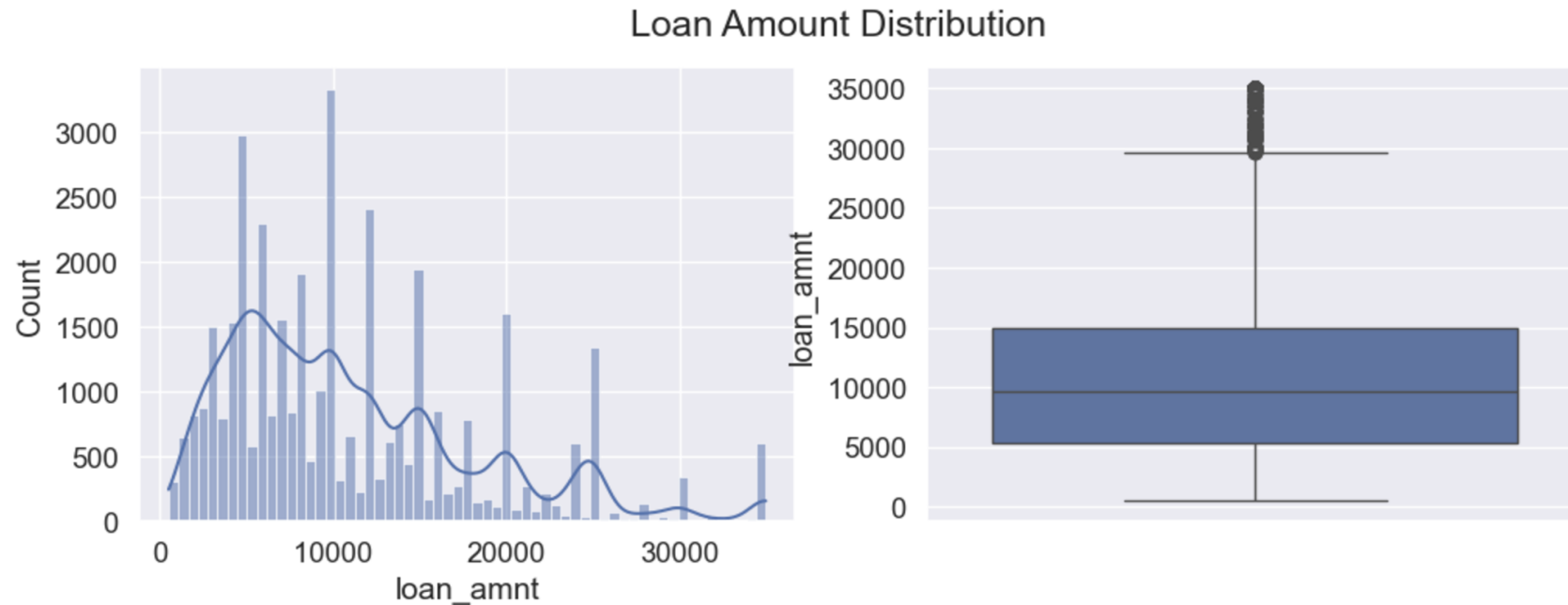

Count of Loan by Grade wise


Counts of loan by Sub-Grade Wise

- The majority of loans fall under the higher Grades (B, C, A) and specific Sub-Grades (A4, B3, A5).

Univariate Analysis

Loan Amount Distribution

- The median loan amount is close to 10,000, similar to our calculation (Median Loan Amount: 9,600). Loan amounts range from 0 to 35,000, with 25% at 5,000, 50% around 10,000, and 75% at 15,000. Outliers are above 30,000.
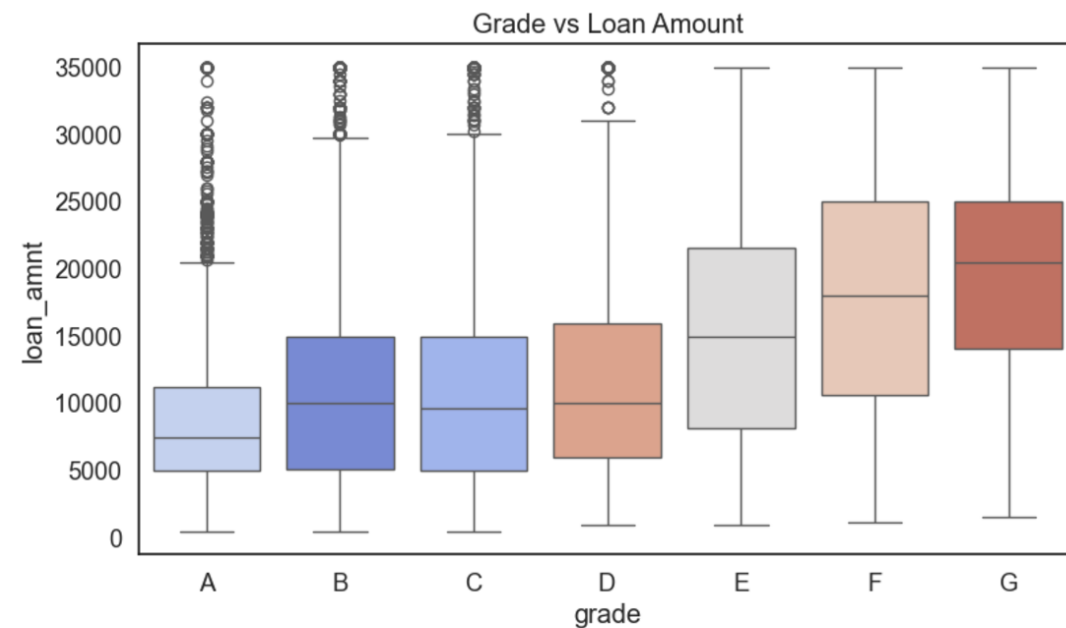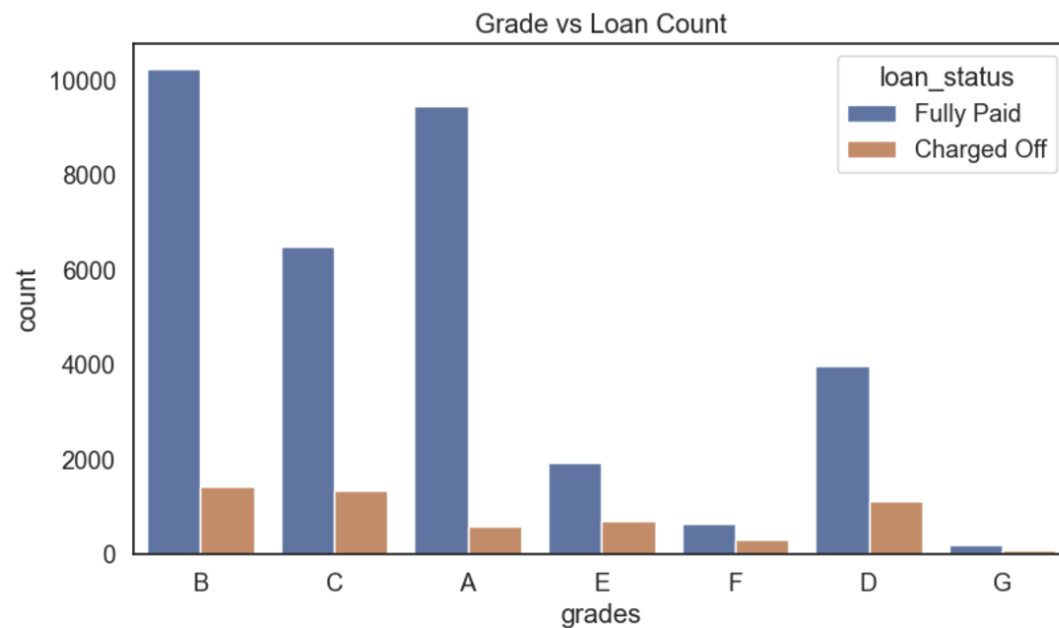
Univariate Analysis
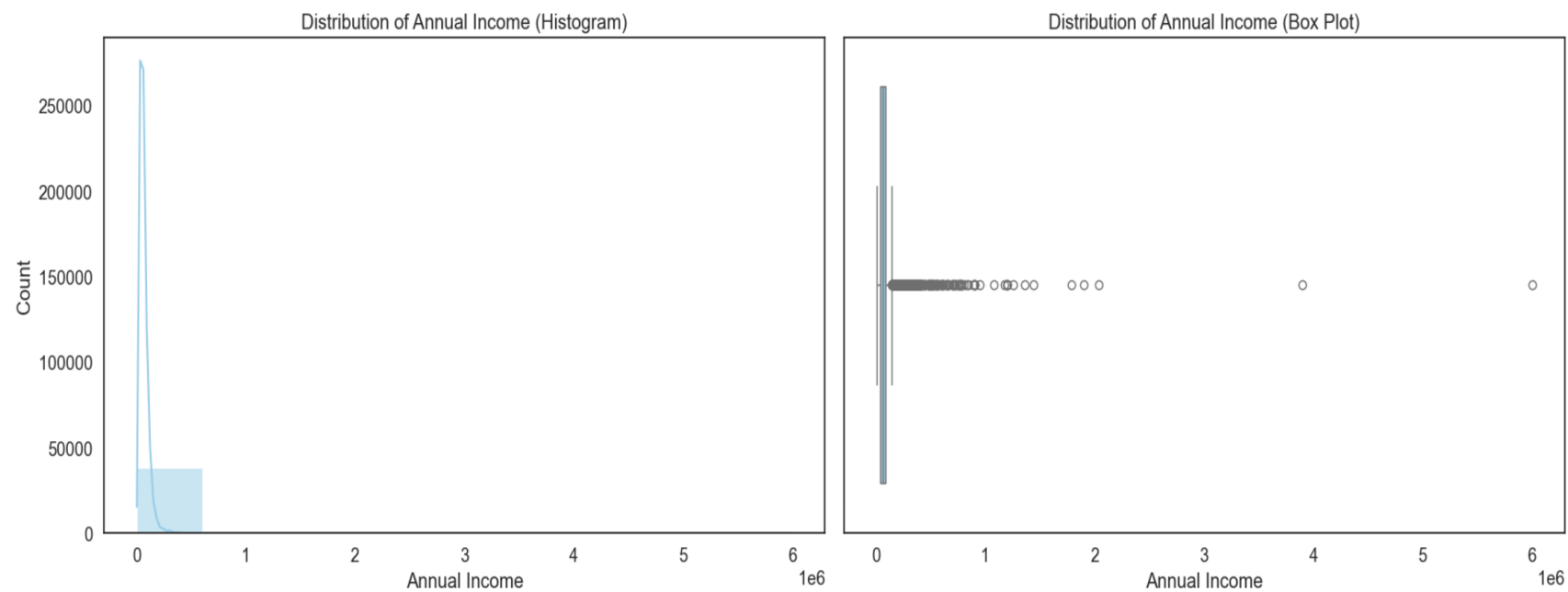
- The histogram on the left shows that the majority of individuals have low annual incomes, with a rapid decline as income increases.
- The box plot on the right reveals a high concentration of income at the lower end, along with numerous high-income outliers, highlighting a strong skew in the data.

# Univariate Analysis

Verification Status of Loan Applicants
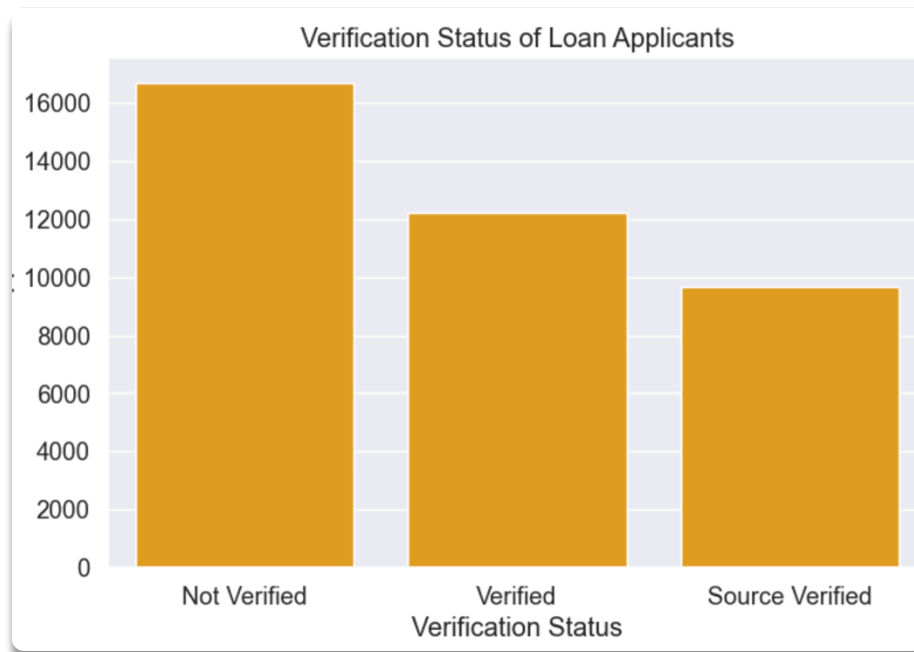
▶ The bar plot shows that "Not Verified" applicants make up the largest portion of the dataset, suggesting that most applicants have not undergone verification.

▶ In contrast, other categories like "Verified" or "Source Verified" have significantly lower counts, highlighting a drop-off in applicants completing verification.

Segmented Univariate Analysis

- Grades B and C represent the majority of loan takers, indicating a higher demand for loans among individuals in these categories.
- Although Grades F and G have fewer loans, they account for the highest loan amounts.

Segmented Univariate Analysis
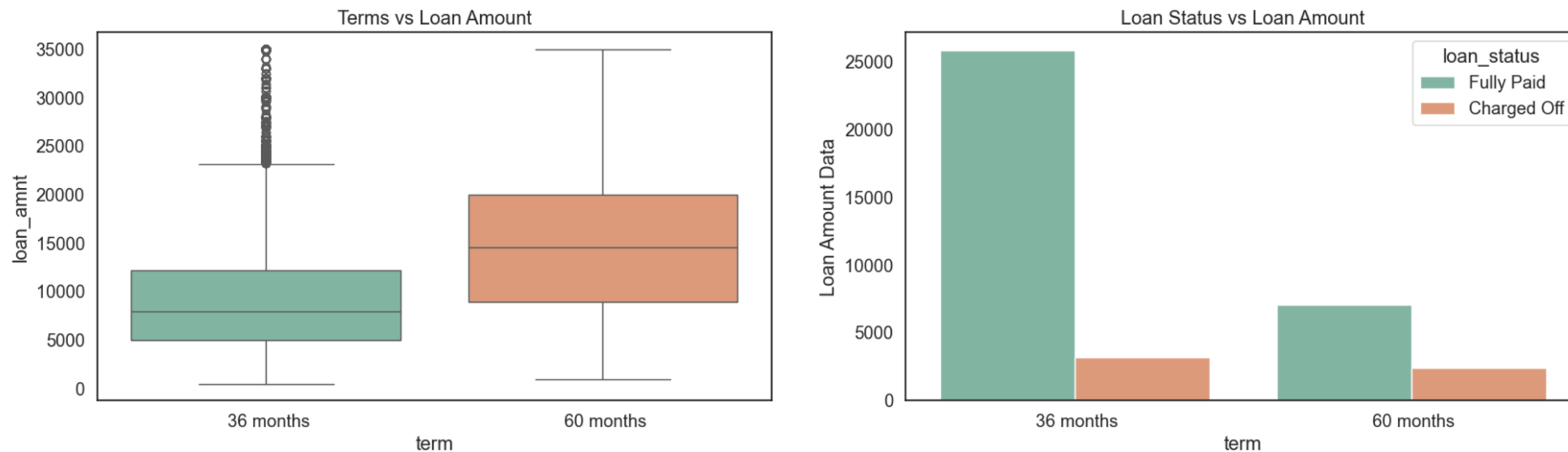
- **Loan Term vs Amount**: 36-month loans have lower loan amounts with a narrower range, while 60-month loans show higher amounts with a broader distribution.
- **Loan Status vs Amount** : 36-month loans have significantly more "Fully Paid" loans, whereas 60-month loans have a higher proportion of defaults ("Charged Off").
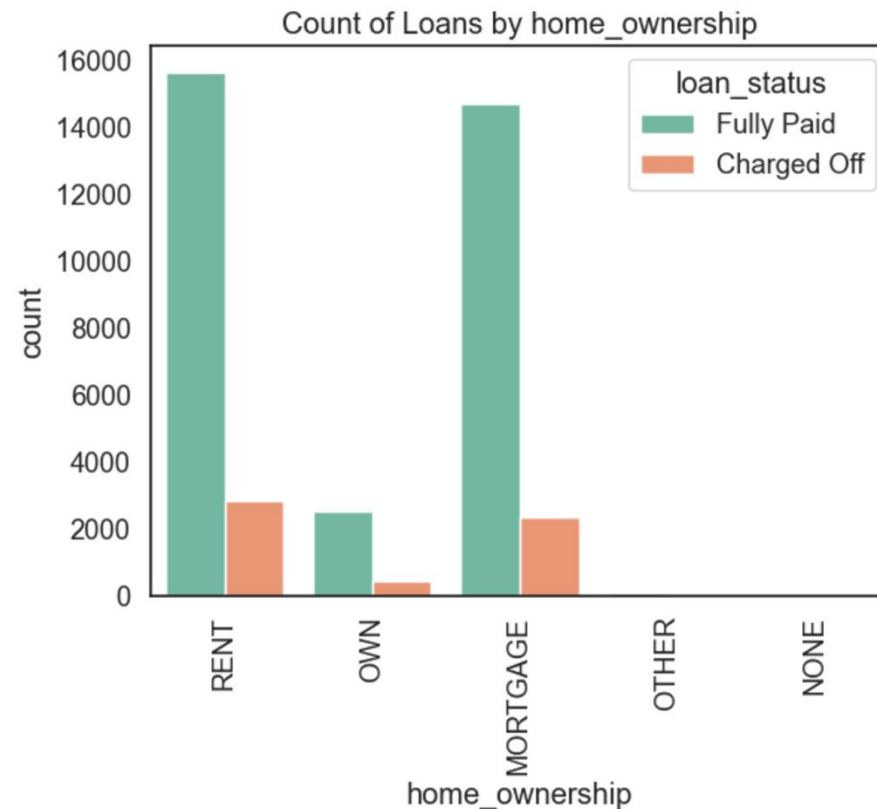
Segmented Univariate Analysis

▶ Borrowers with "Charged Off" loans have a higher average number of public-record bankruptcies compared to those with "Fully Paid" loans.

▶ The presence of bankruptcies correlates more strongly with default (Charged Off) than with successful repayment (Fully Paid).
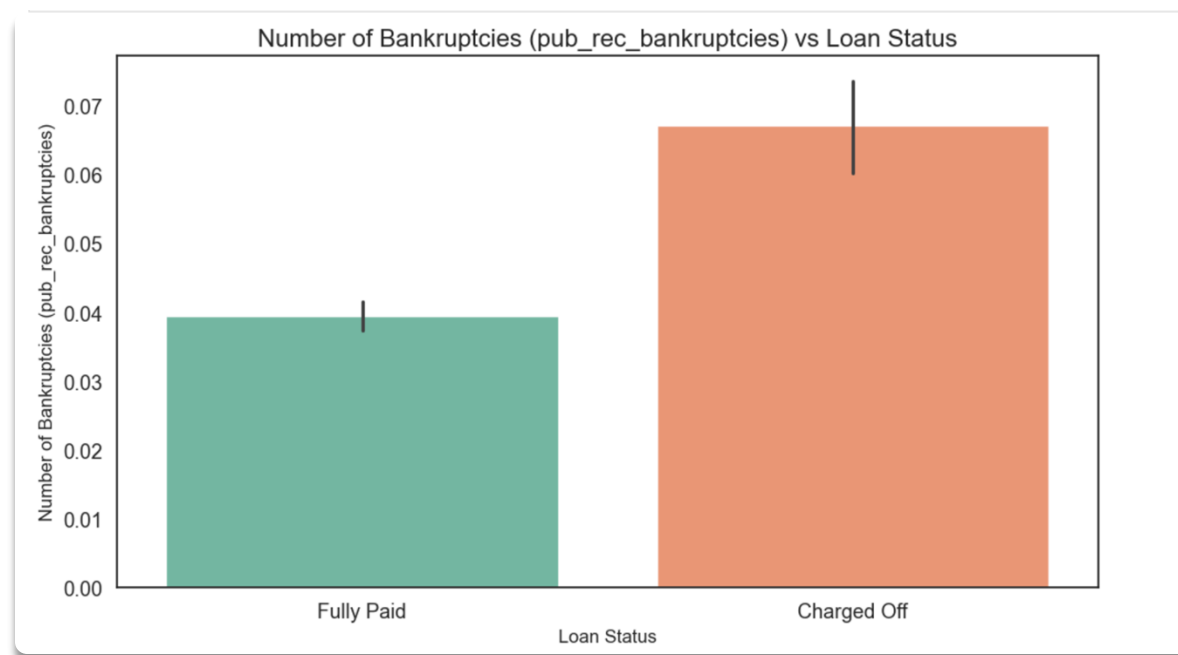
Segmented Univariate Analysis

▶ Borrowers who rent or have mortgages have the highest loan counts, with significantly more "Fully Paid" loans compared to "Charged Off" loans.

▶ Borrowers in the "OWN" and "OTHER" categories have relatively fewer loans overall, but the proportion of "Charged Off" loans remains lower than "Fully Paid."

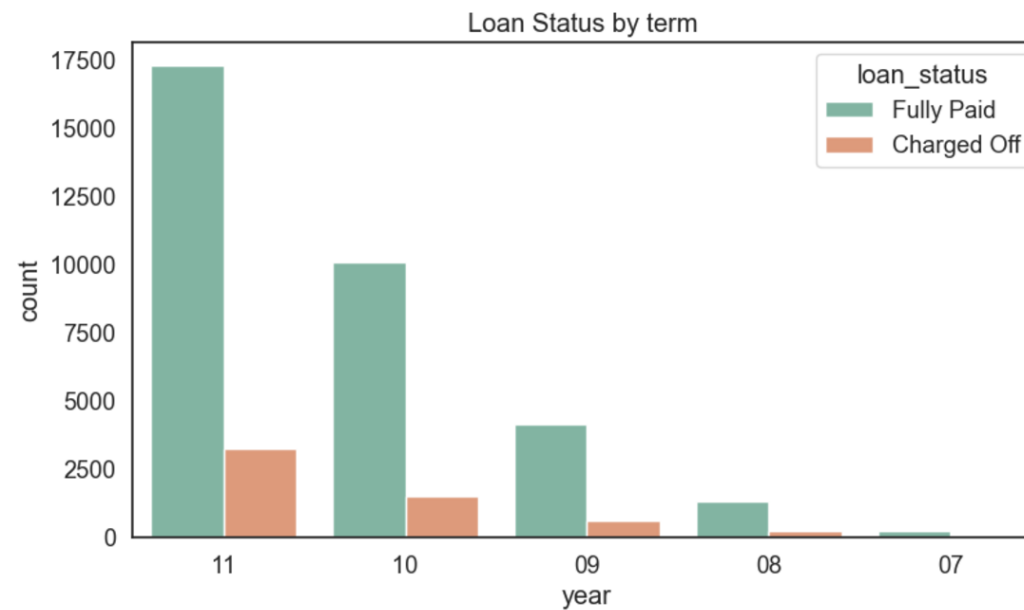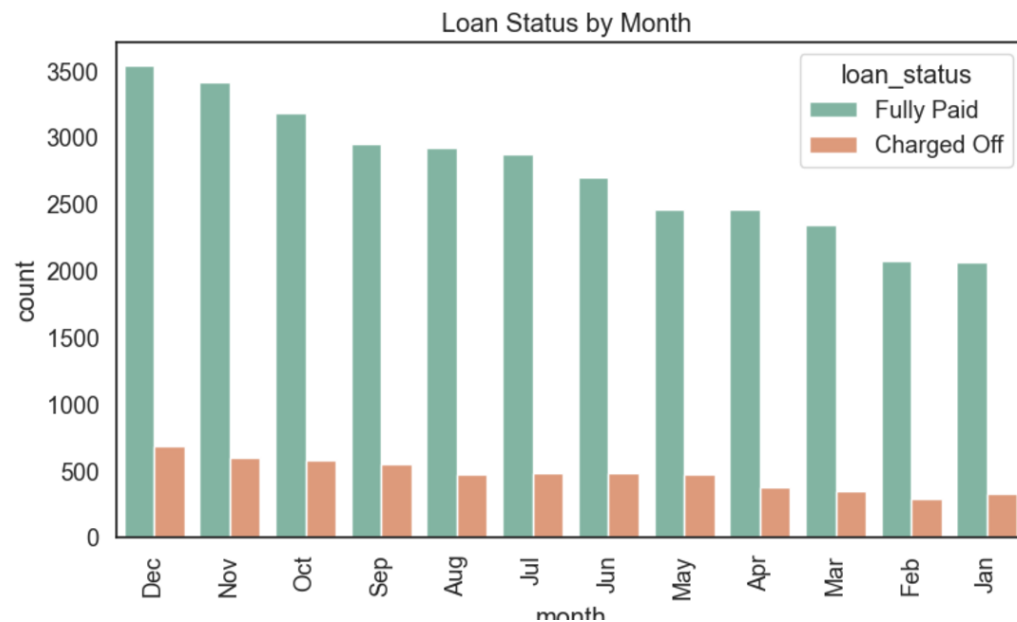Count of Loans by home_ownership

Segmented Univariate Analysis
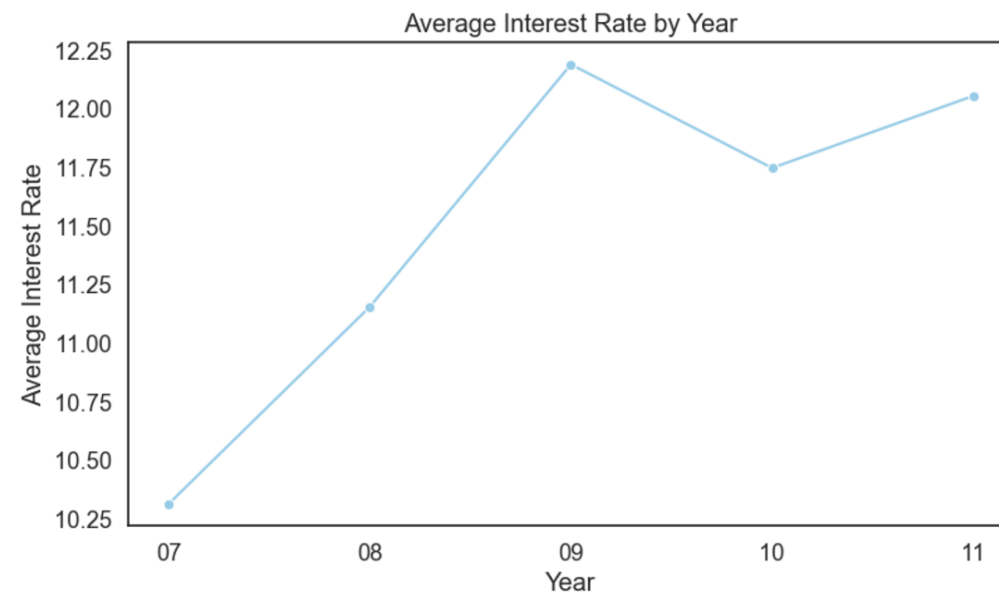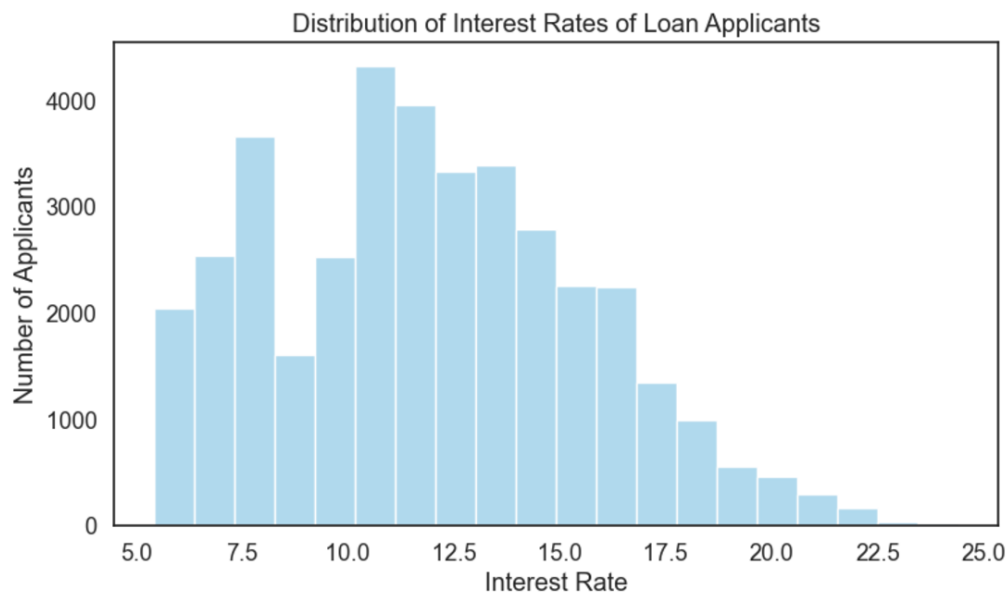
- Interest rates were higher from 2007 to 2009, decreased in 2010, and increased again in 2011.
- The most common interest rates fall within the 10-15% range.

Bivariate Analysis

- The highest number of loan applications occurs in December Month for every year.
- The number of loan applicants has significantly increased each year.

Bivariate Analysis

► The median loan amount across all employee experience levels (emp_length) ranges between 7,000 and 12,000, with minimal variation.

► The highest number of outliers in loan amounts is observed for employees with less than 1 year of experience.


Loan Amount by Employee Length

Bivariate Analysis
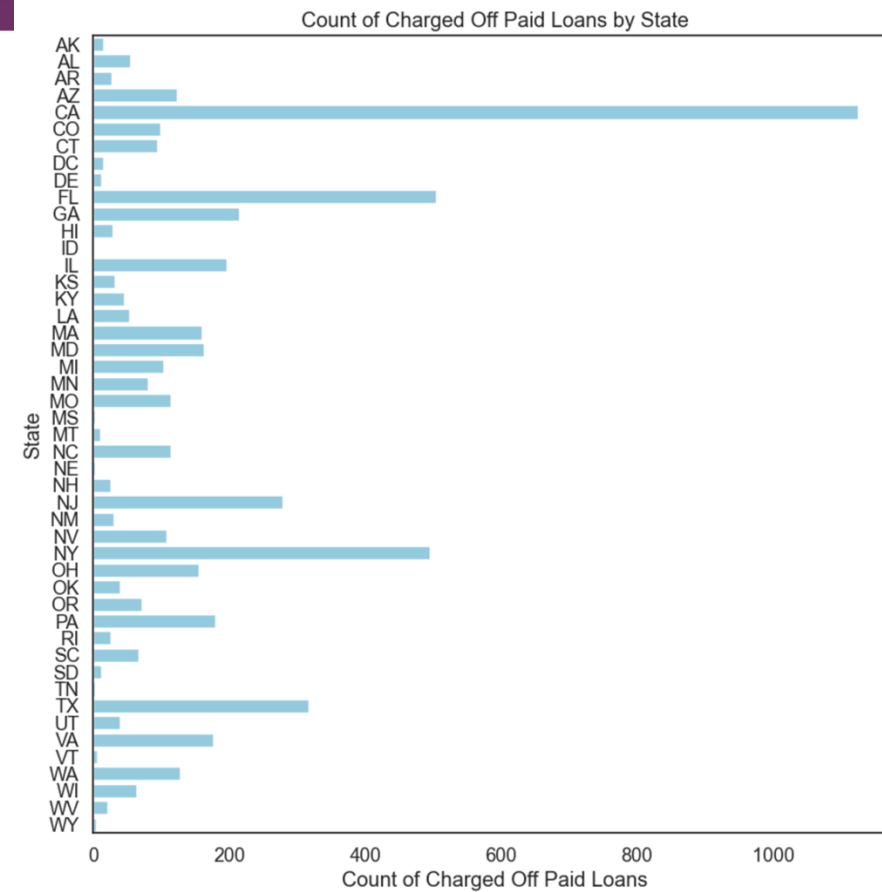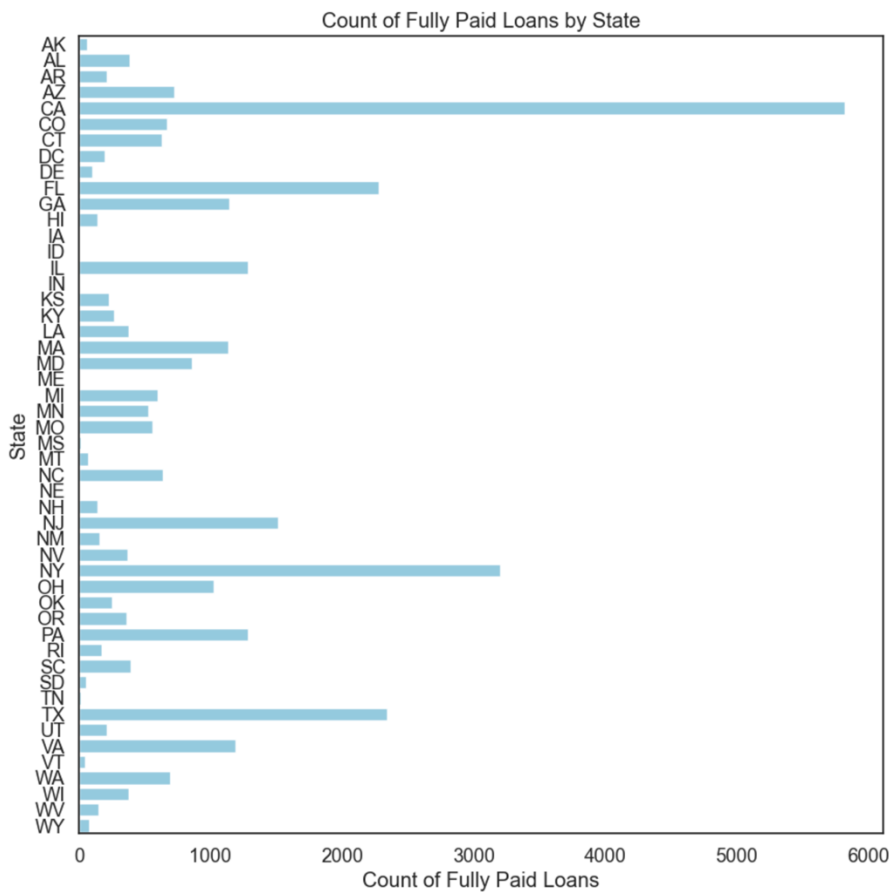
California (CA) has the highest number of loan applications in both categories.

Risk Analysis of Defaulters

Risk Analysis of Charged off - Defaulters by Grade

Risk Analysis of Charged off - Defaulters by Sub-Grade
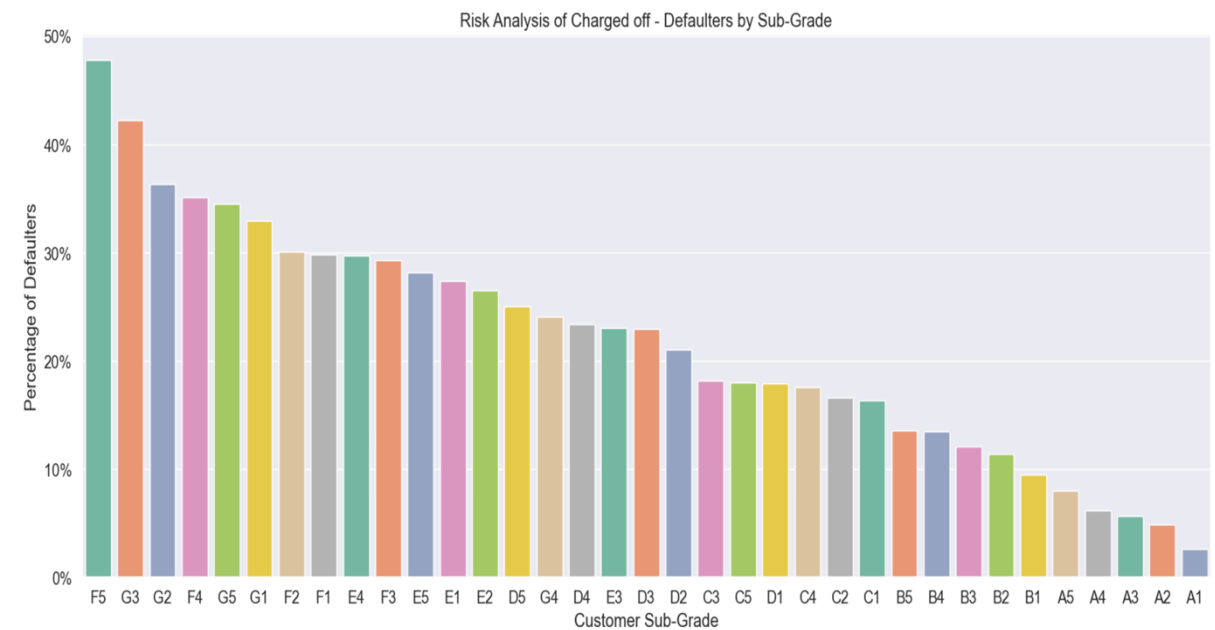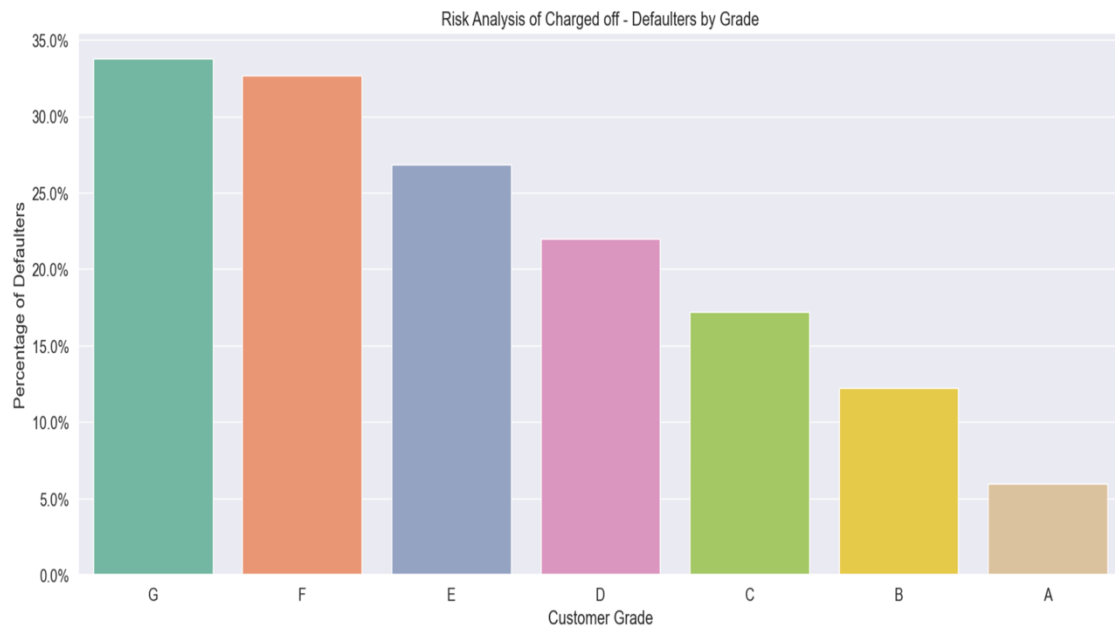
- Lower Grades has highest number of defaulters percentage. Grade G has highest risk of defaulters.
- Similarly, we are found out in sub-grade graph, sub-grade(F5, G3,G2) has highest risk of defaulters.

Risk Analysis of Defaulters

Risk Analysis of Charged off - Defaulters by Employee Length

Distribution of Defaulter Percentage Across All Interest Rates
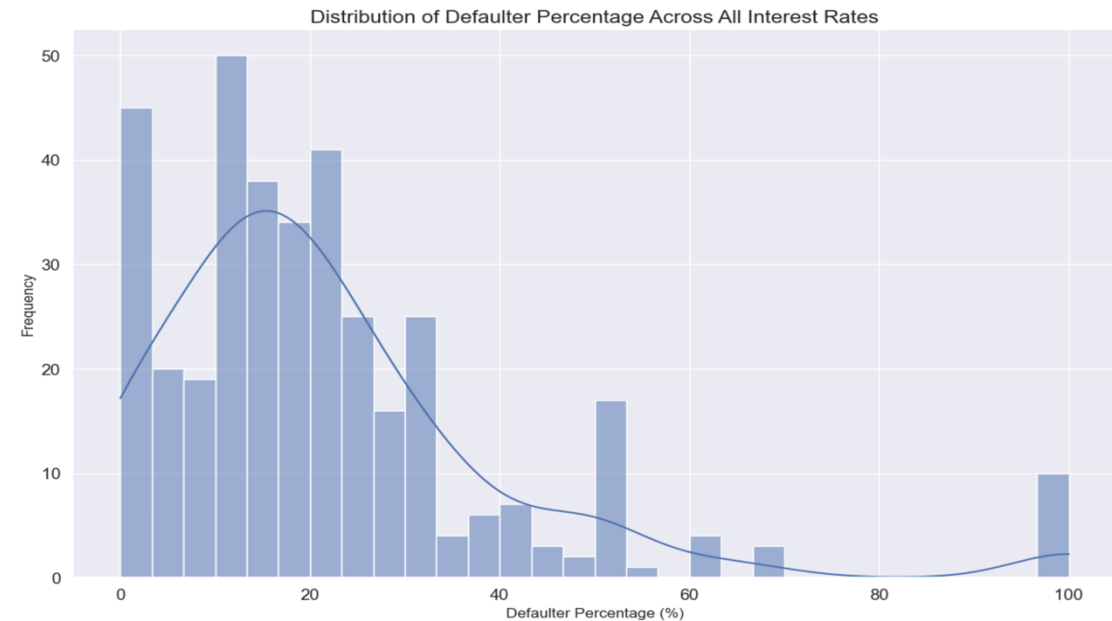
- Loan default rates are consistent across all employment lengths, averaging around 15%.
- The second graph is right-skewed, with most defaulter percentages concentrated at lower values and a median interest rate percentage of 20%.

Thank You