Lending Club Case Study-EDA

Name: Akshat Rohitkumar Mehta

Batch:September-2022

Introduction

- Solving this case study 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.

Problem Statement

- This firm facilitates personal loans, corporate loans, and finance for medical operations. It is the largest online lending marketplace.
- Through a quick internet interface, borrowers can readily get loans with cheaper interest rates. Like most other lenders, the biggest cause of loss is giving loans to "risky" borrowers (called credit loss).
- When a borrower defaults on a loan or flees with the money owed, the lender suffers a credit loss.
- In other words, defaulting borrowers result in the biggest loss for lenders. The "defaulters" in this instance are the consumers who have been marked as "charged-off."

Target

- By identifying the contributing variables to loan default, EDA approaches can help reduce credit losses.

End Goal

- Reduce the risk of losing money while lending to customers by using data.

Steps-EDA of Lending Club Case Study

1.Data Understanding

2.Data Cleaning

3.Data Analysis and Data Visualization

4.Conclusion and Recommendations

1.Data Understanding

It can be done by below mentioned steps

- ❑ Import Needed Libraries for EDA.
- ❑ Import warnings for ignore warnings in notebook
- ❑ Import Data Dictionary and Read it.
- ❑ Import Load Dataframe and Read First 10 Entries.
- ❑ Check Shape and Information about Loan Dataframe
- ❑ Check Id have unique Values
- ❑ Check Data Types of all Fields in Loan Dataframe
- ❑ Describe Dataframe with Numeric Fields.

2.Data Cleaning

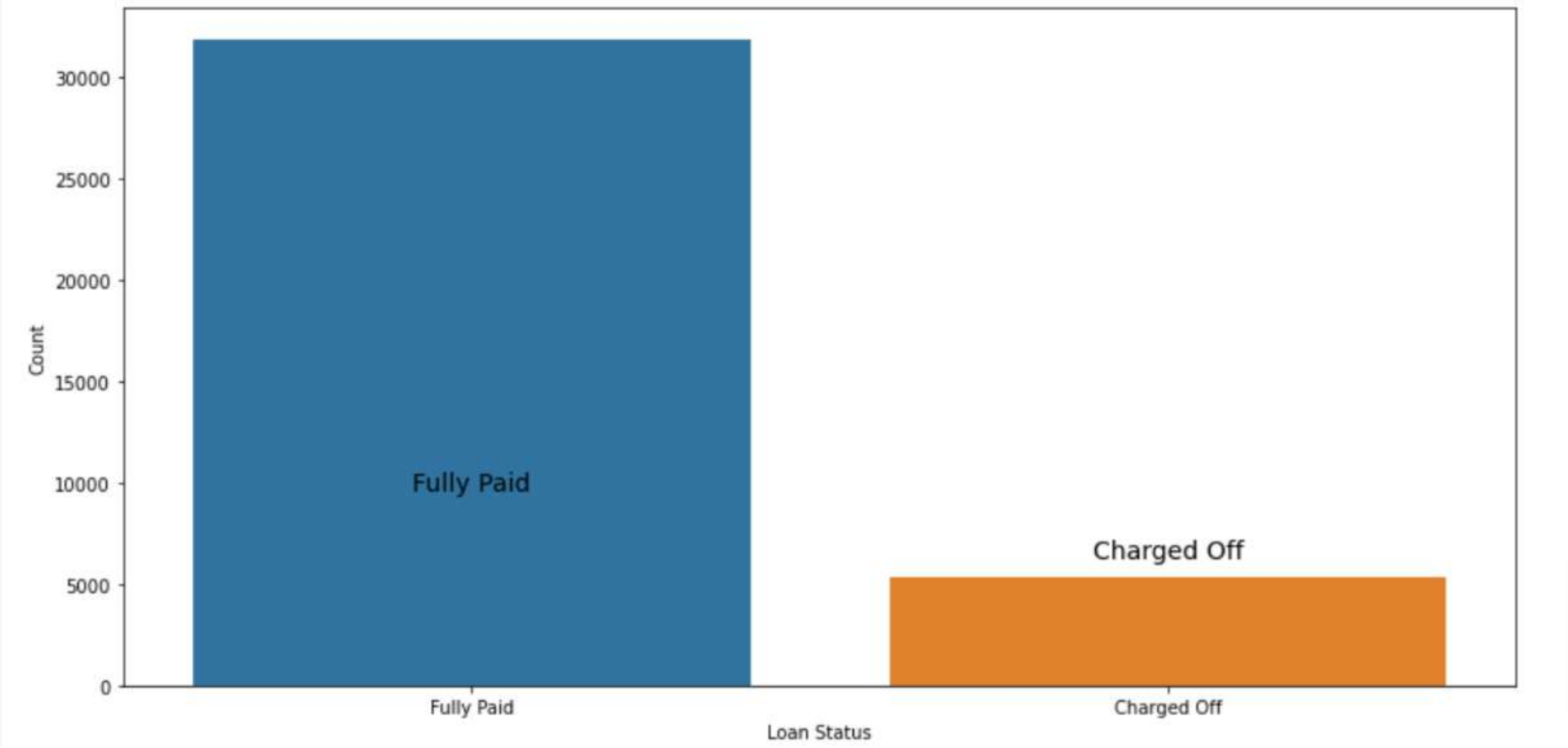
It Can be done by below mentioned points

- ❑ Fix row and columns with Null Values
- ❑ Remove Columns which is not in EDA.
- ❑ Correct Datatypes in Dataframe
- ❑ Derive new Columns(month and year)
- ❑ Remove Columns with Current Loan Status
- ❑ Removing Outliers from Annual Income Field

3a) Univariate Data Analysis

Field: Loan Status

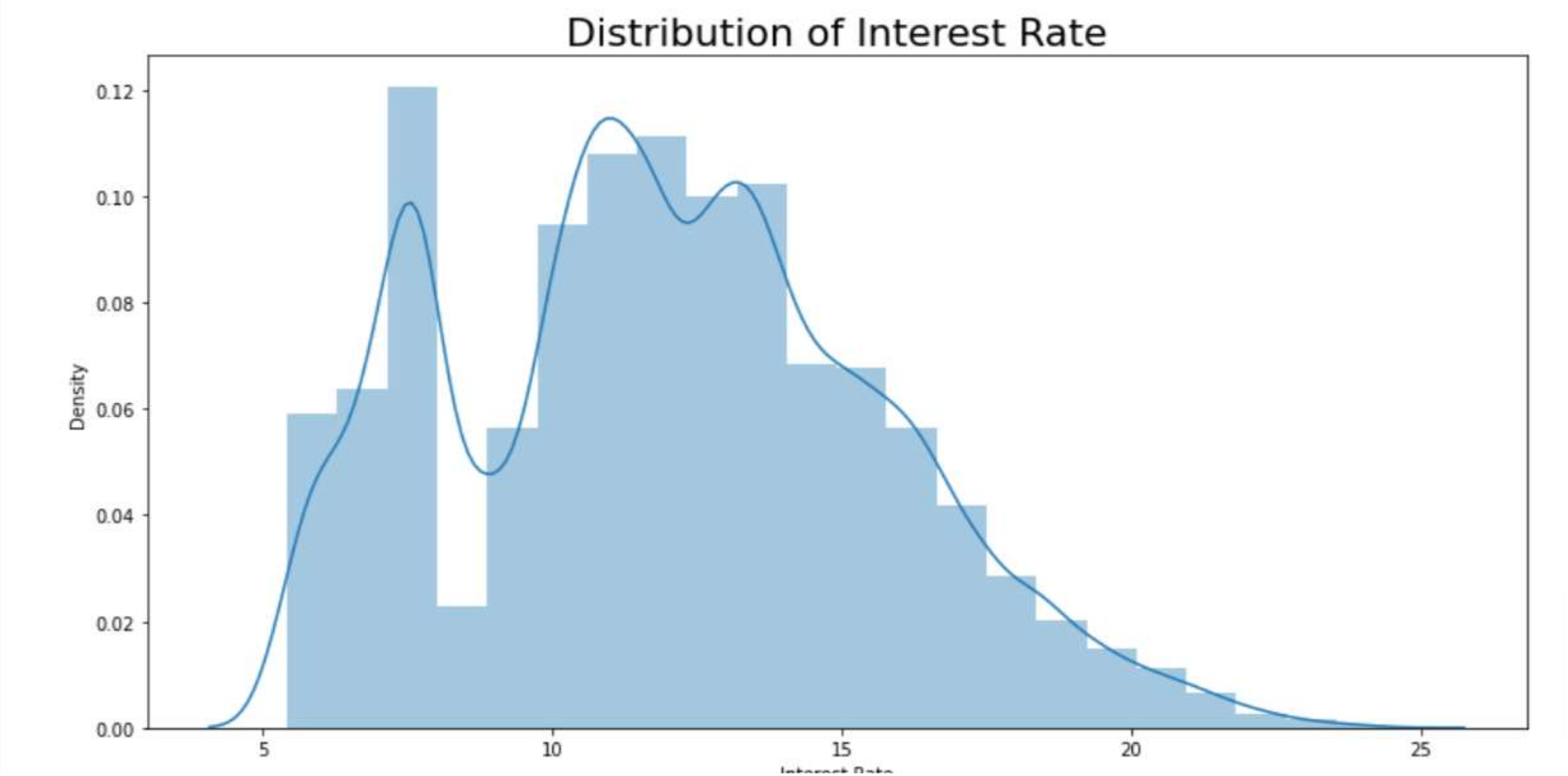
Conclusion: Fewer loans have defaulted than have been fully paid.



3a) Univariate Data Analysis

Field: Interest Rate

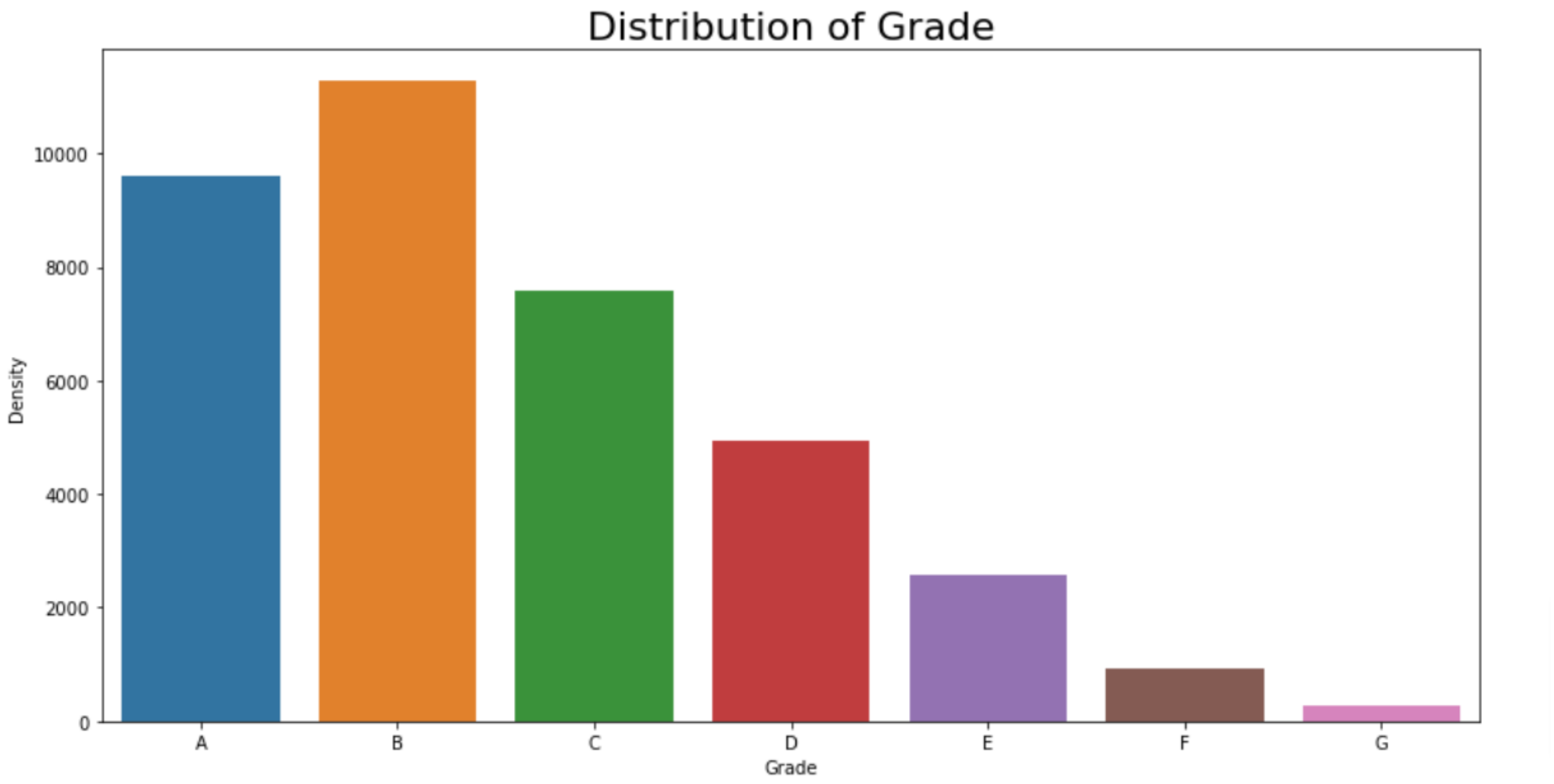
Conclusion: The interest rate is lower around 10 and more congested between 5 and 10 and 10-15.



3a) Univariate Data Analysis

Field: Grade

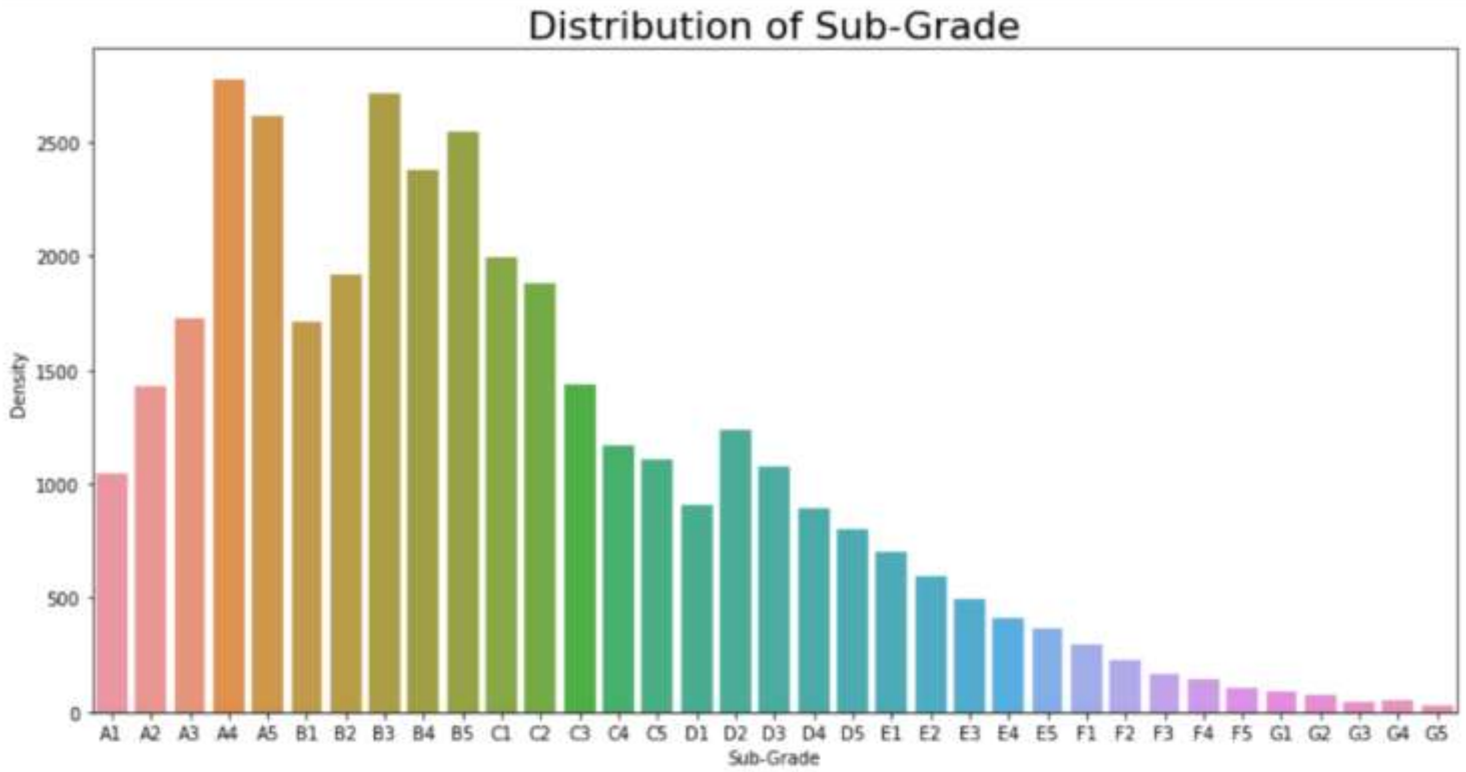
Conclusion: The majority of loans are high grade loans, as seen by the significant percentage of loans with grades "A" and "B" compared to the rest.



3a) Univariate Data Analysis

Field: Sub-Grade

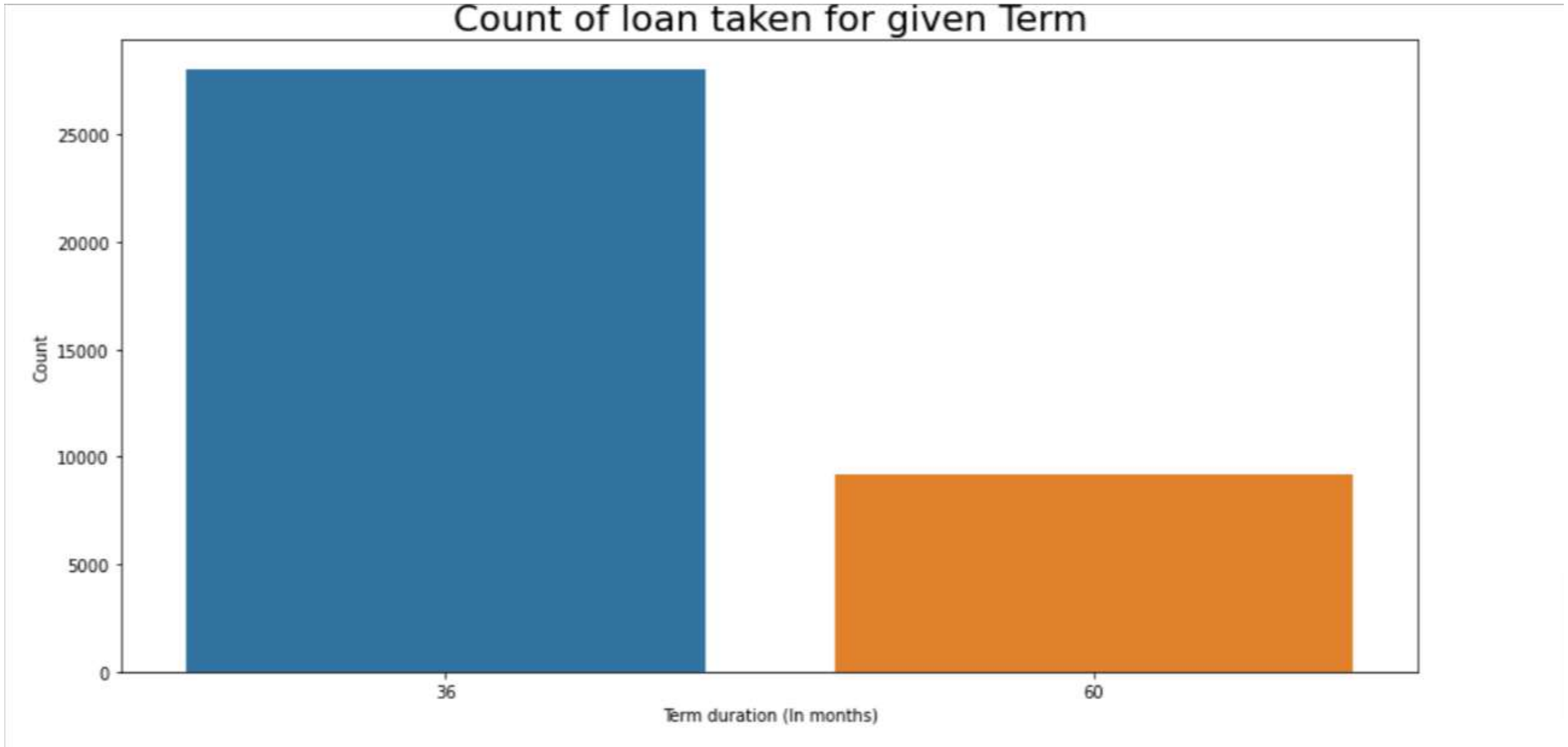
Conclusion: The majority of loans are high grade loans since a sizable portion have lower subgrades in grades "A" and "B" compared to the rest. This is consistent with the distribution of grades.



3a) Univariate Data Analysis

Field: Term of Loan

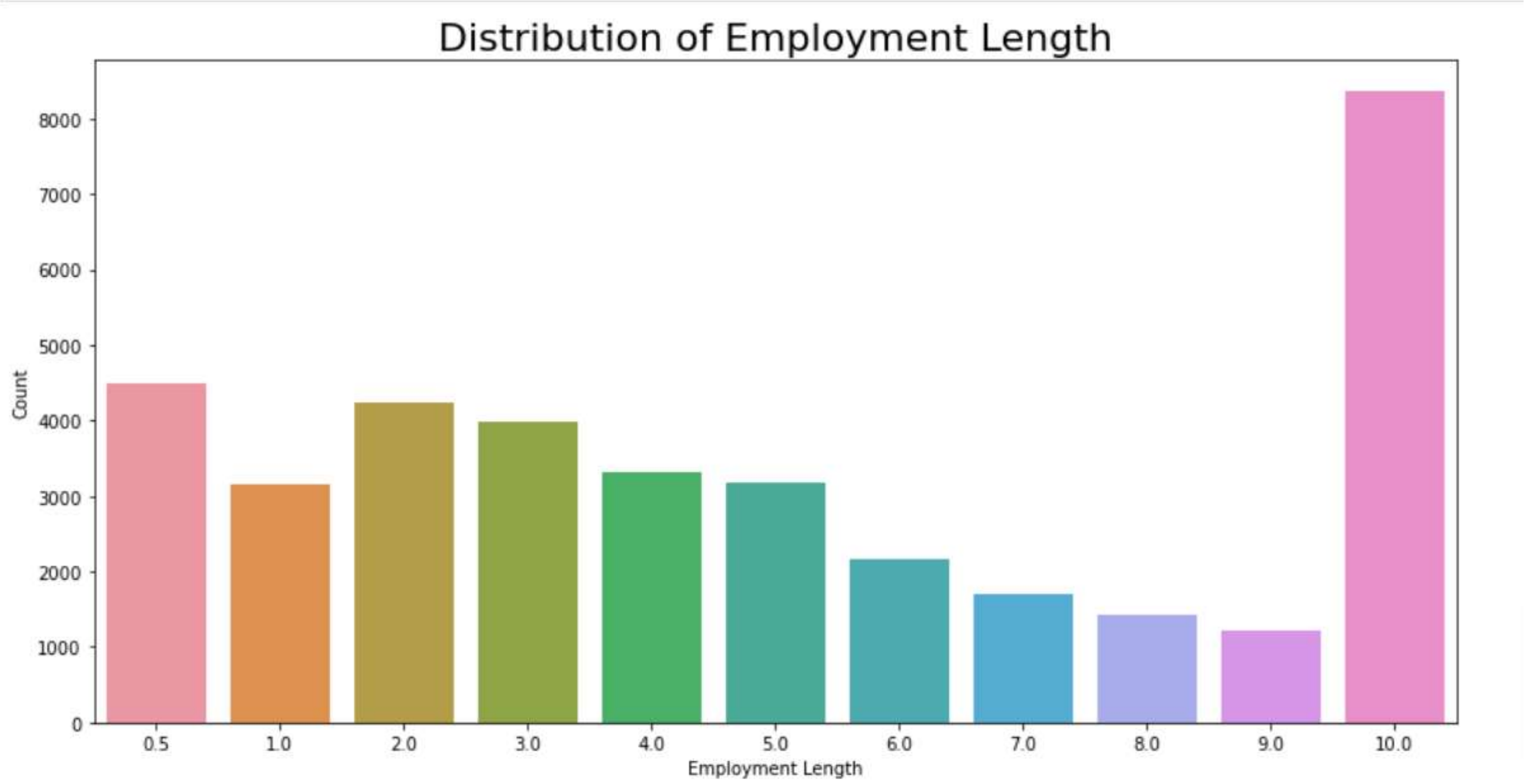
Conclusion: As opposed to 60 months, more than half of the loans taken have terms of 36 months.



3a) Univariate Data Analysis

Field: Employee Length

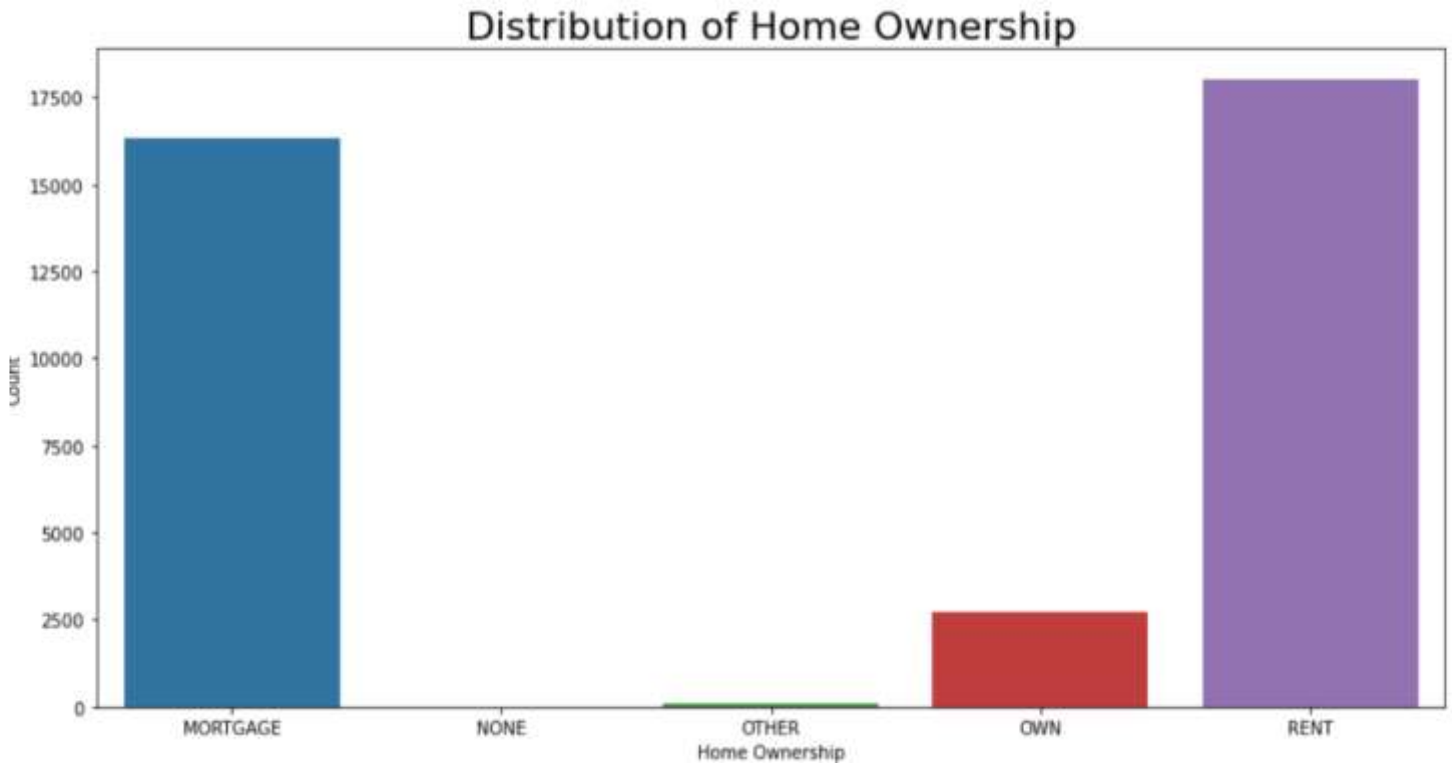
Conclusion: The majority of borrowers have more than ten years of job experience.



3a) Univariate Data Analysis

Field: Home Ownership

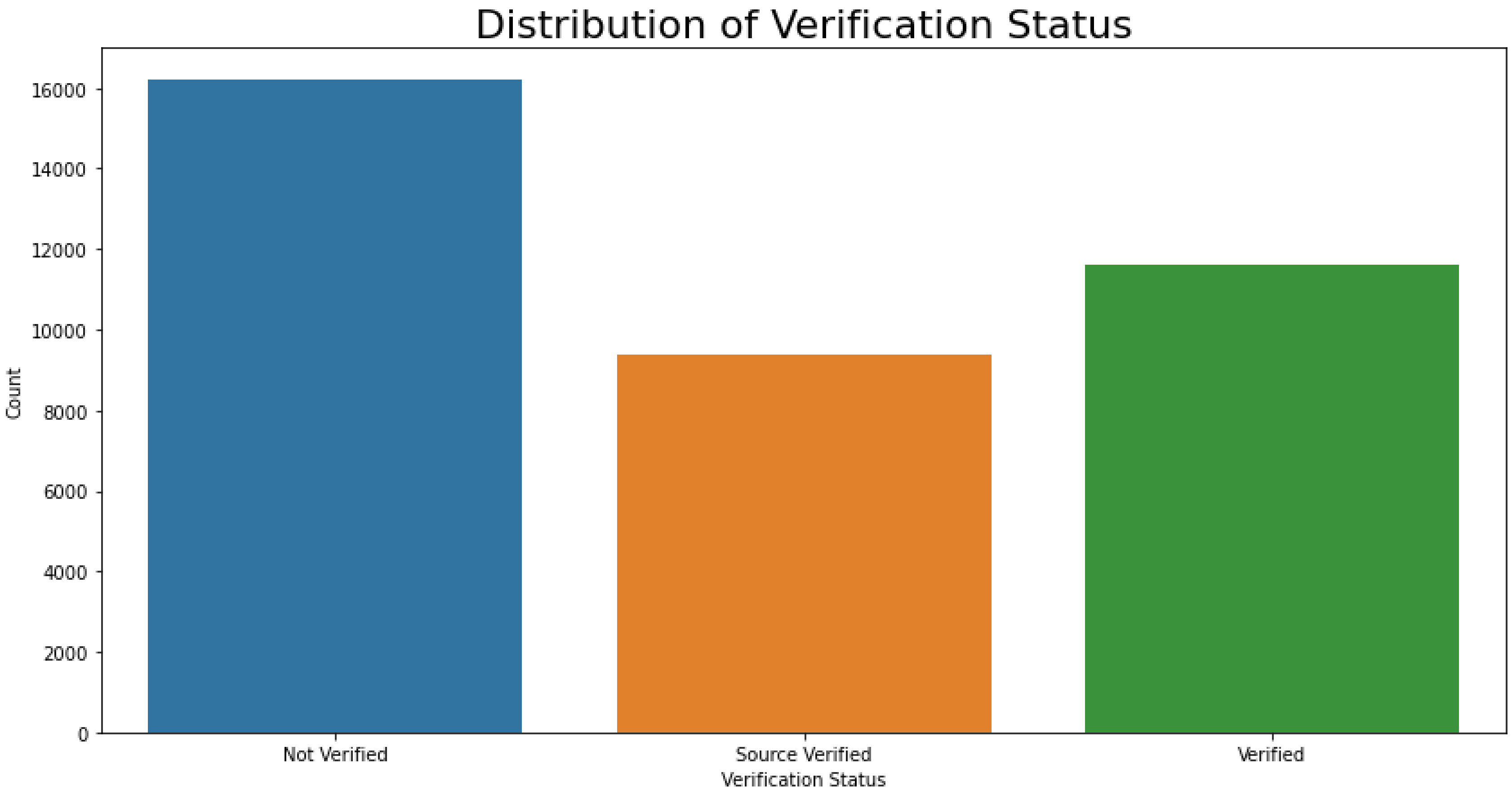
Conclusion: The majority of borrowers do not own property and are either renting or paying a mortgage.



3a) Univariate Data Analysis

Field: Verification Status

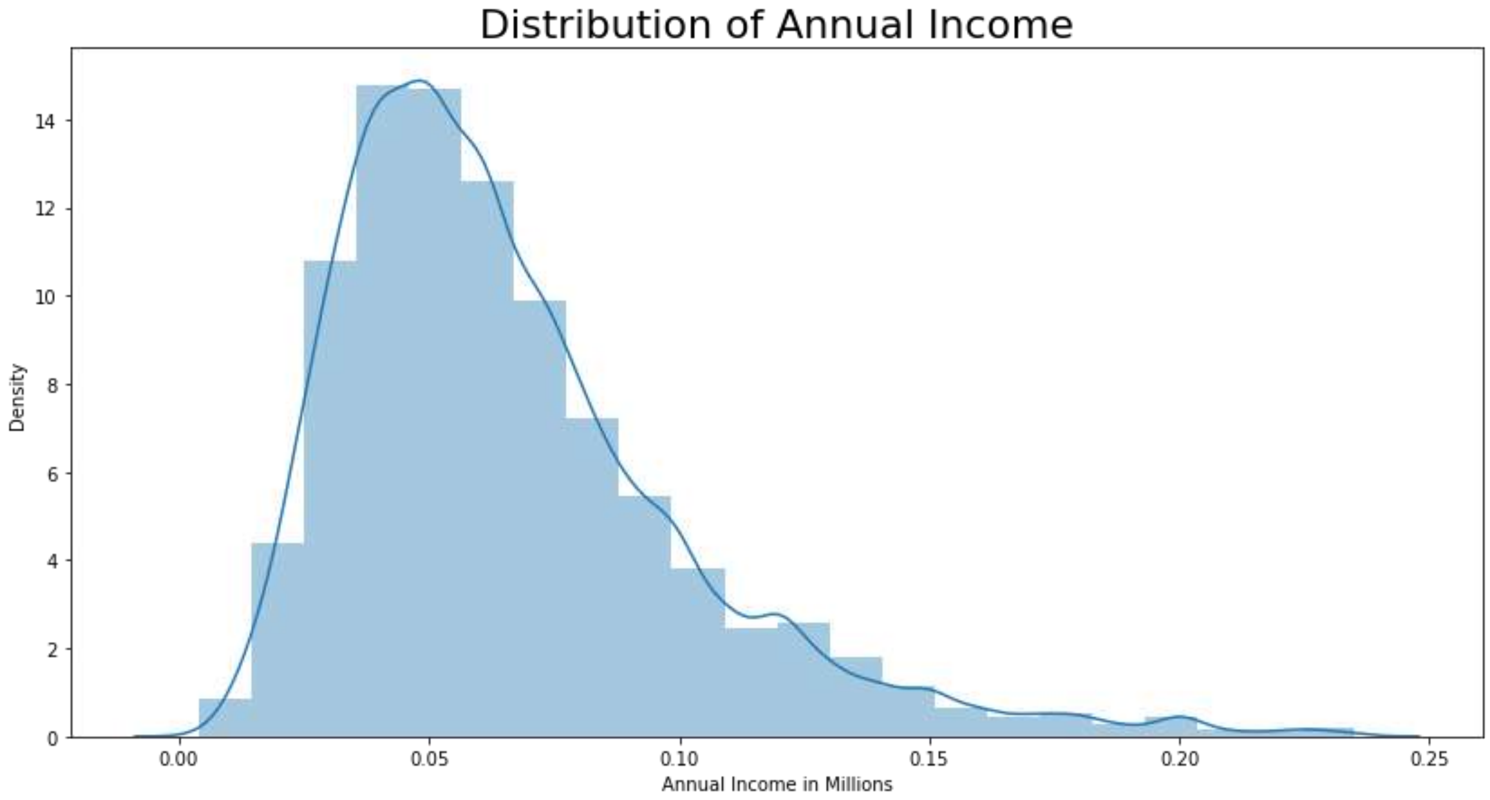
Conclusion: The company verifies or has a source verify about 50% of the debtors..



3a) Univariate Data Analysis

Field: Annual Income

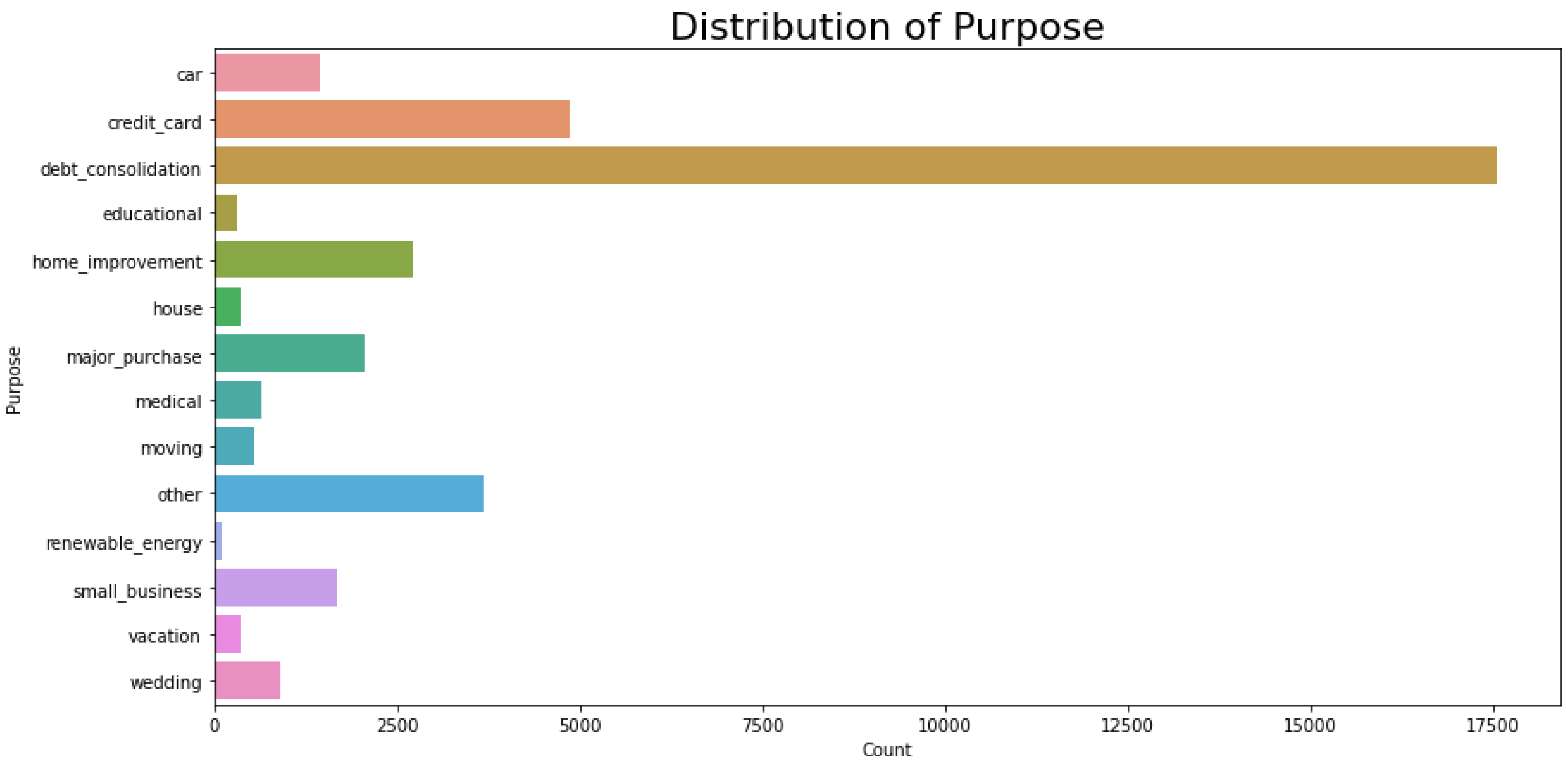
Conclusion: Since the yearly income has a left-skewed normal distribution, the bulk of burrowers have extremely low annual incomes in comparison to the rest of the population.



3a) Univariate Data Analysis

Field: Purpose

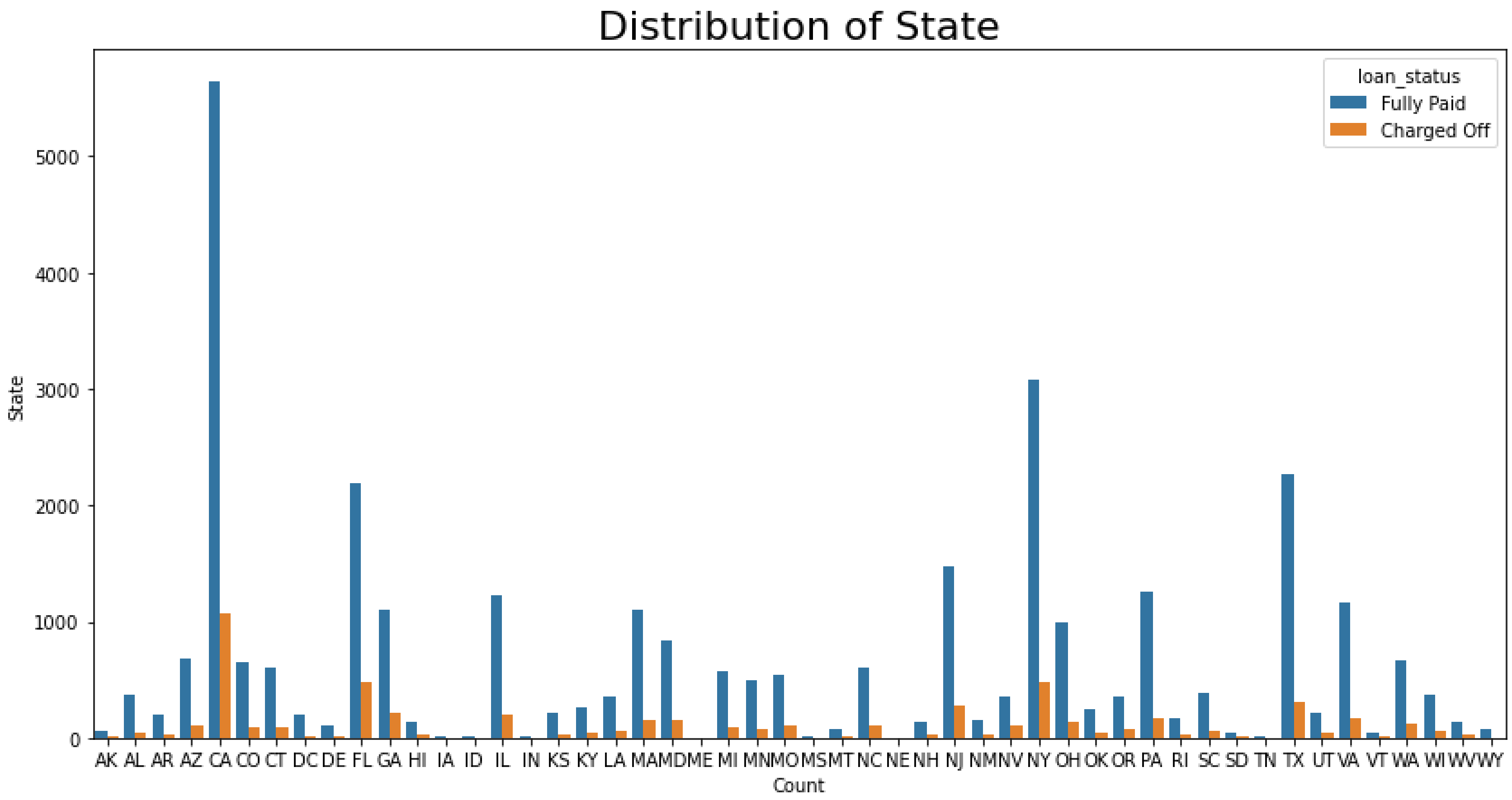
Conclusion: Credit cards are used the most frequently for debt consolidation loans..



3a) Univariate Data Analysis

Field: Address State

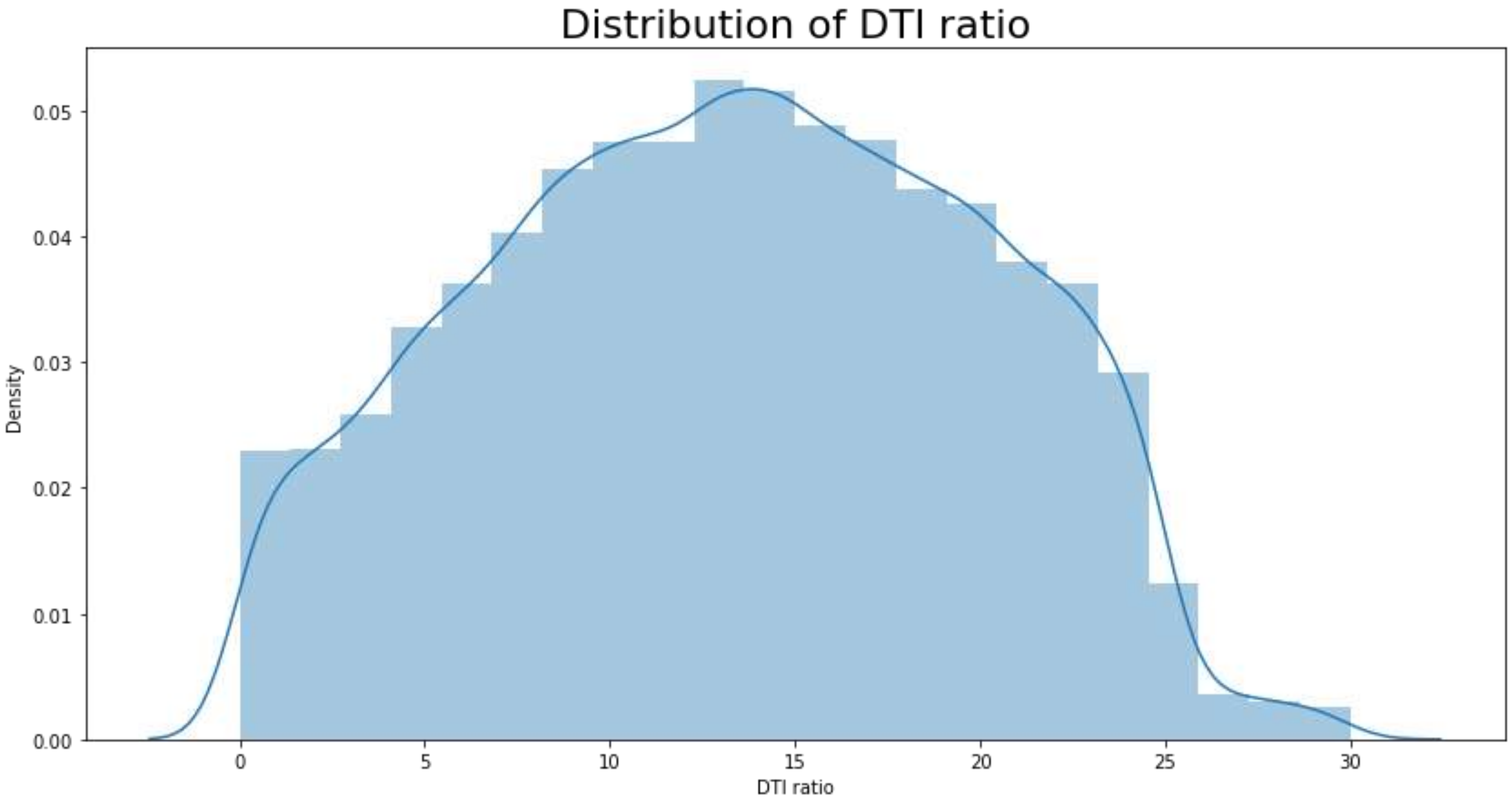
Conclusion: Majority of the borrowers are from the large urban cities like california, new york, texas, florida etc



3a) Univariate Data Analysis

Field: DTI

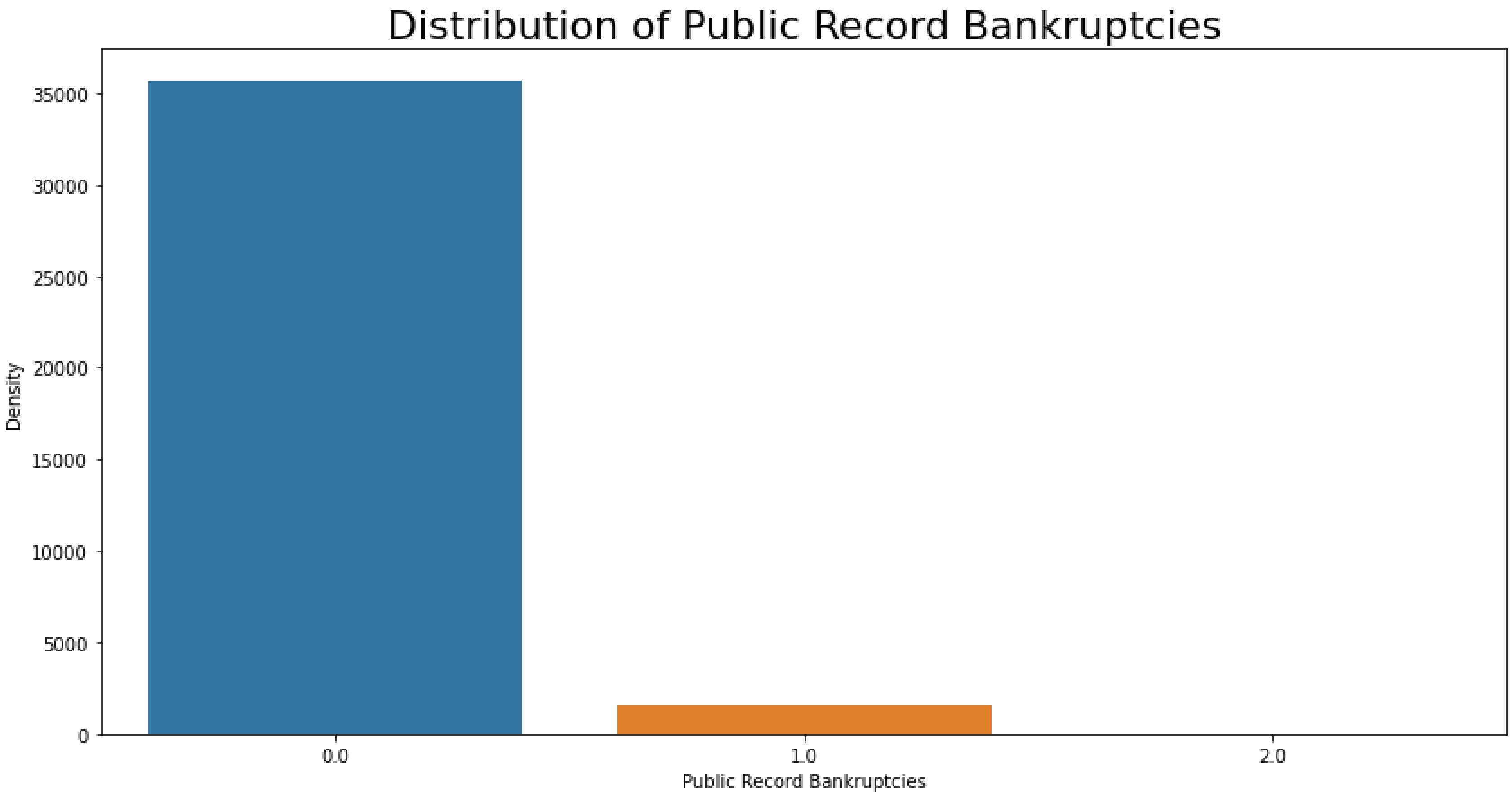
Conclusion: The majority of borrowers have DTI ratios between 10-15, which is a relatively high level of debt relative to income



3a) Univariate Data Analysis

Field: Public Record Bankruptcies

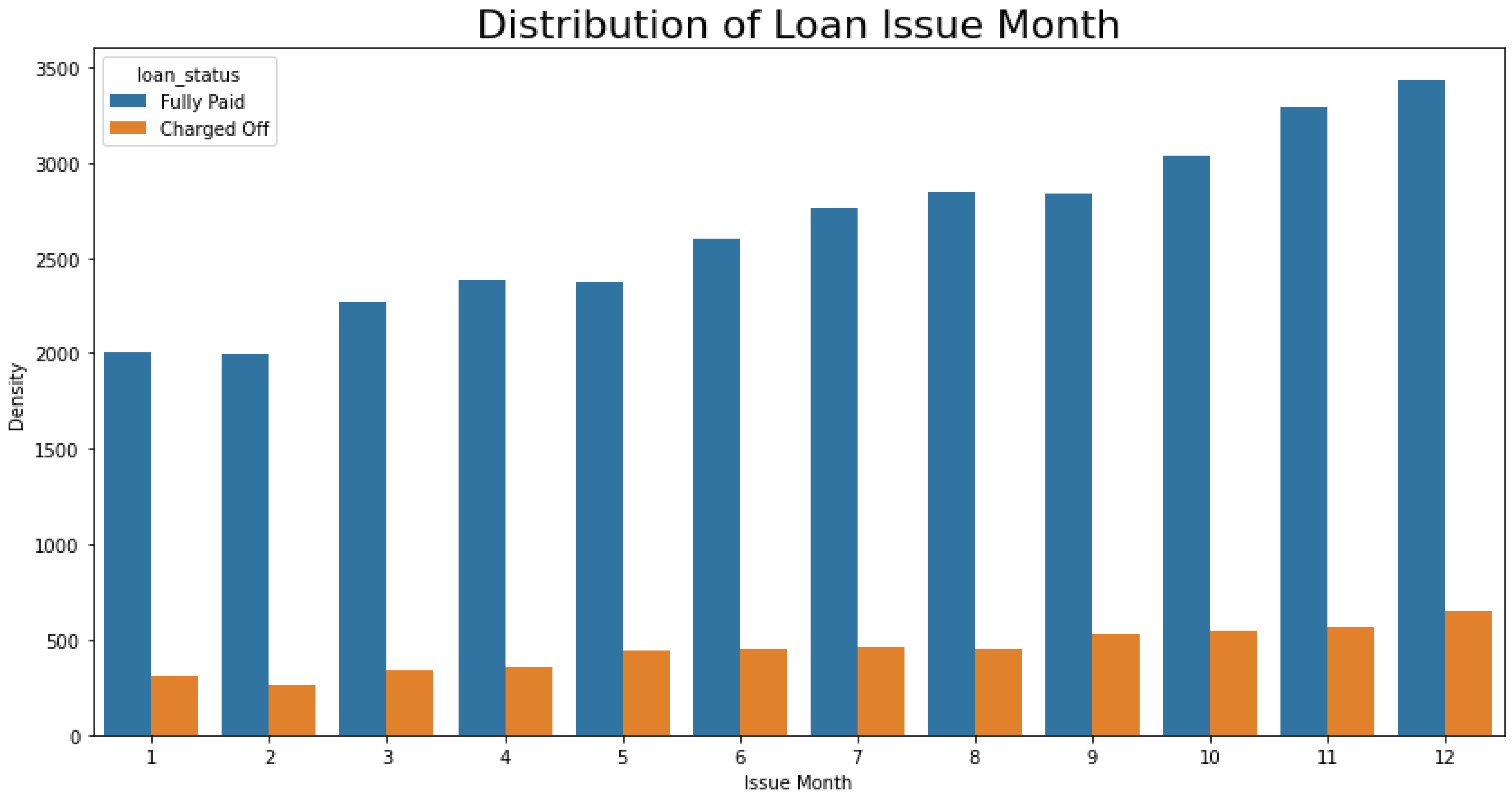
Conclusion: The majority of debtors don't have any Publicly Recorded Bankruptcies on their records.



3a) Univariate Data Analysis

Field: Issue Month

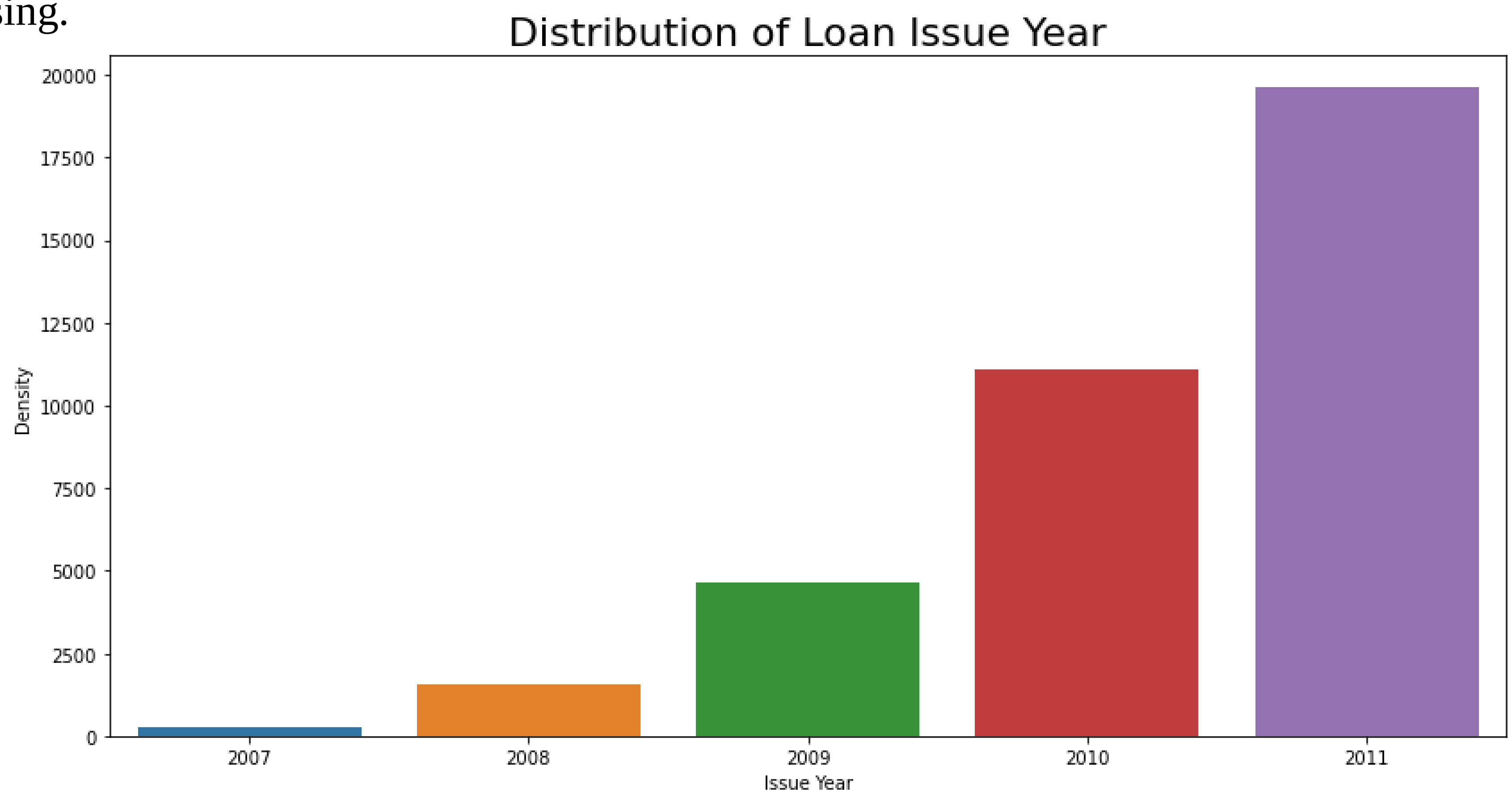
Conclusion: The last quarter of the year is when the majority of loans are disbursed.



3a) Univariate Data Analysis

Field: Issue Year

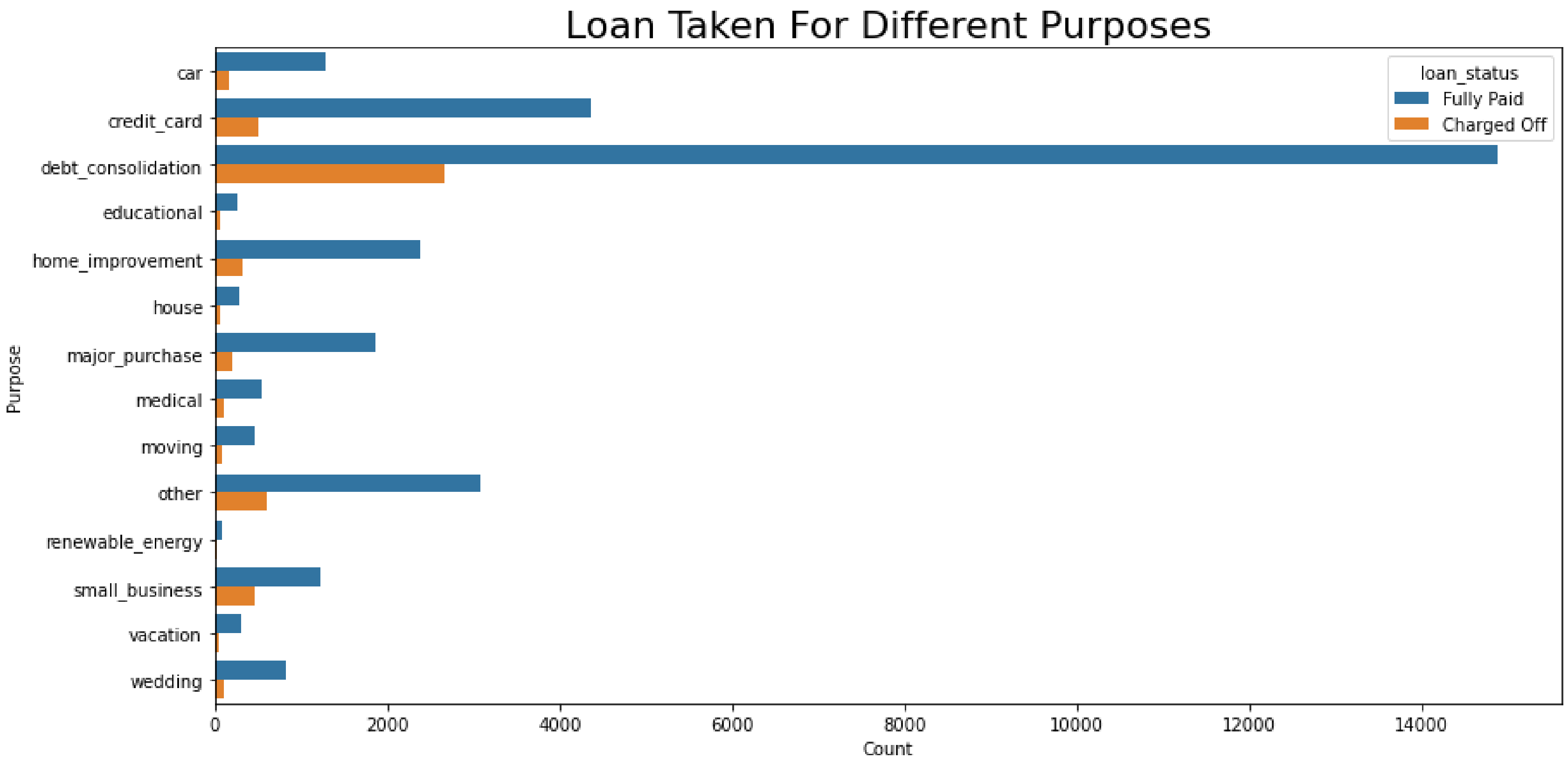
Conclusion: Since the number of approved loans grows exponentially with time, we can conclude that the rate of loan approvals is rising.



3b) Segmented Univariate Data Analysis

Field: Purpose based Loan status

Conclusion: The most common loan purpose and the one with the greatest number of fully paid and defaulted loans is debt consolidation.

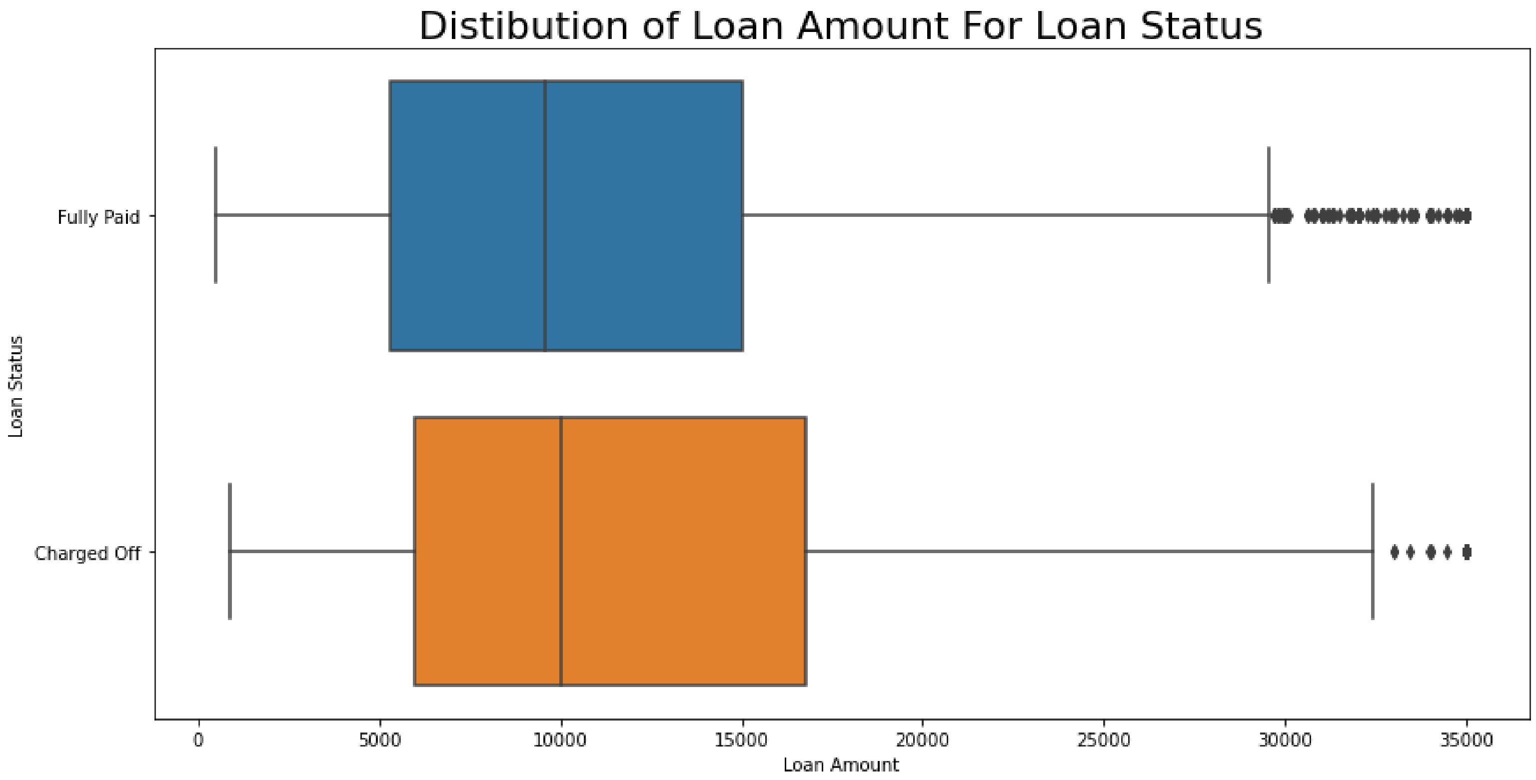


3b) Segmented Univariate Data Analysis

upGrad

Field: Loan Amount based Loan status

Conclusion: While the means and 25% percentiles for both loans are the same, the defaulted loan's 75% percentile is bigger, indicating that large loans have a higher likelihood of defaulting.

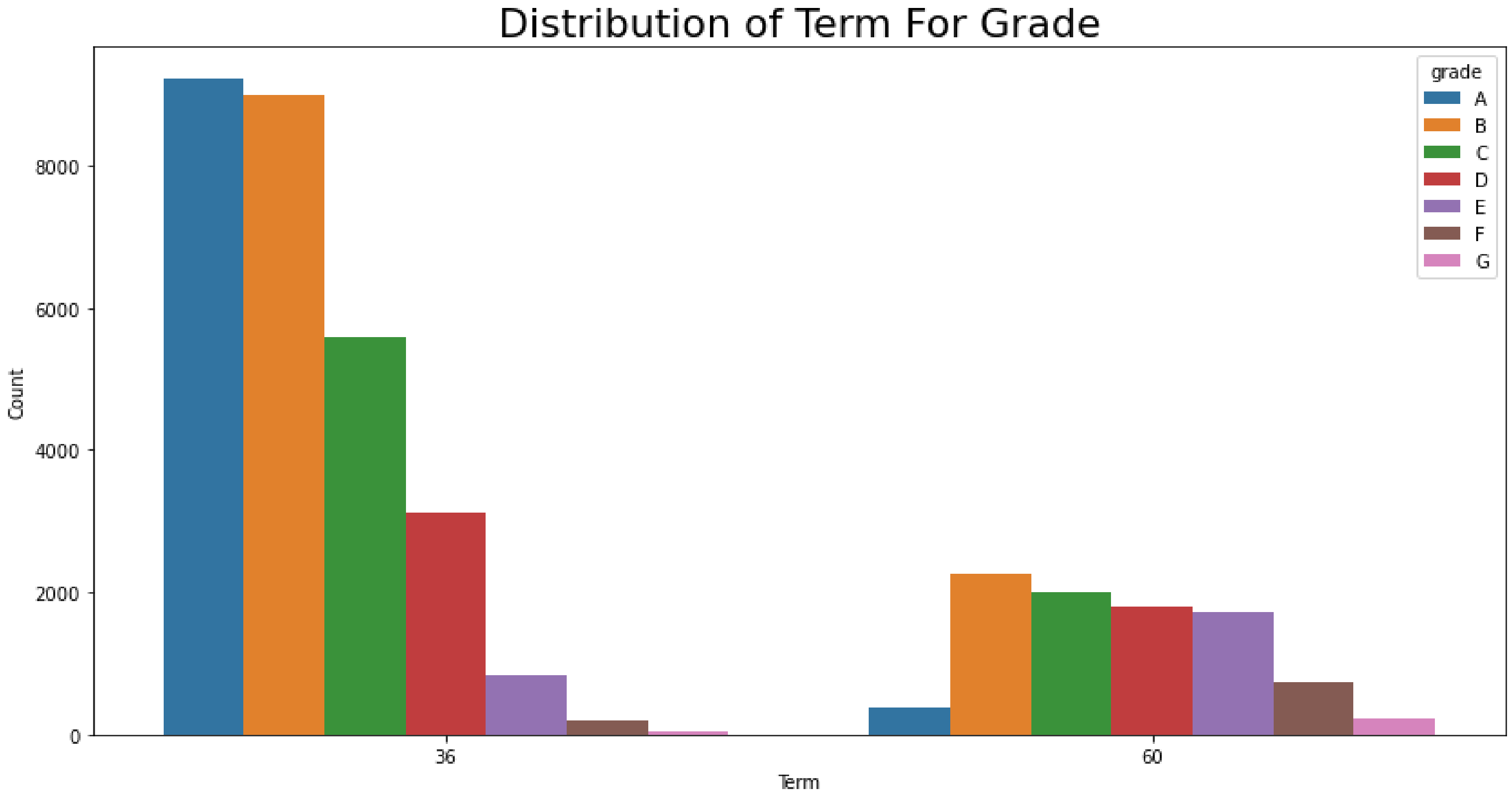


3b) Segmented Univariate Data Analysis

upGrad

Field: Term based Grade

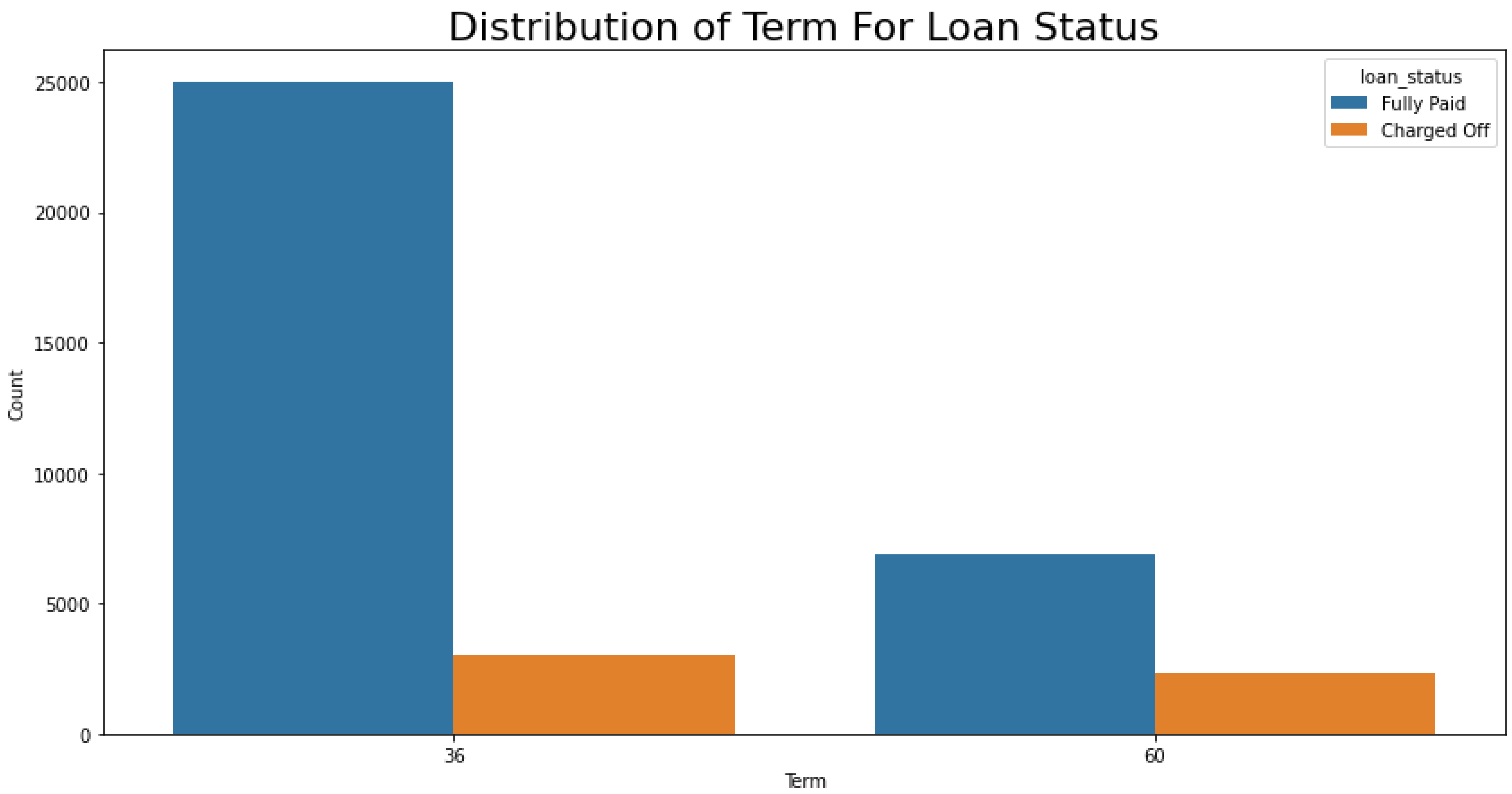
Conclusion: Grade A and B loans make up the majority of loans with a 36-month duration, while Grade B, C, and D loans make up the majority of loans with a 60-month term.



3b) Segmented Univariate Data Analysis

Field: Term based Loan status

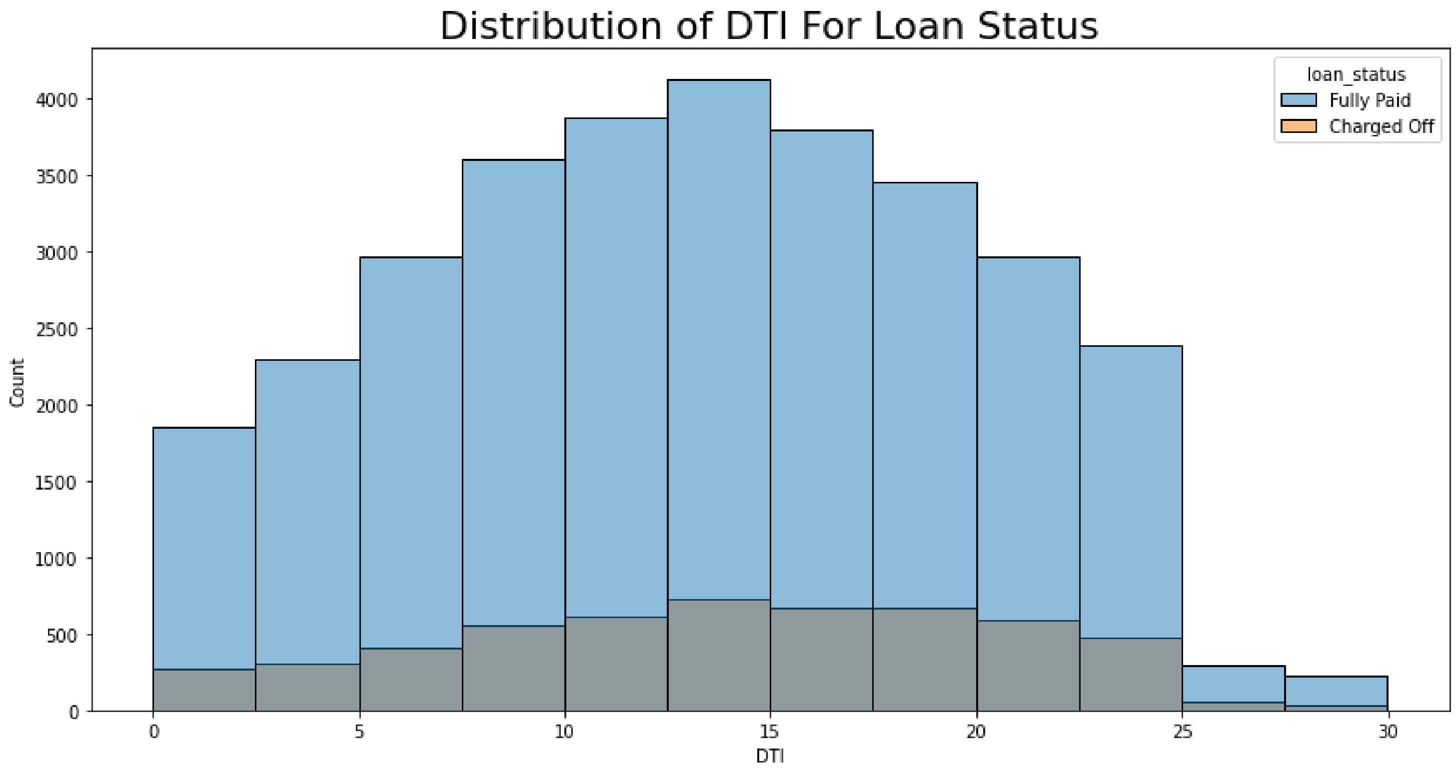
Conclusion: A loan with a 36-month term has a higher likelihood of being fully repaid than one with a 60-month period.



3b) Segmented Univariate Data Analysis

Field: DTI based Loan status

Conclusion: The Loan Status fluctuates with DTI ratio; we can observe that loans with a DTI ratio of 10-15 have a higher number of defaulted loans, but a higher DTI has a higher likelihood of default.

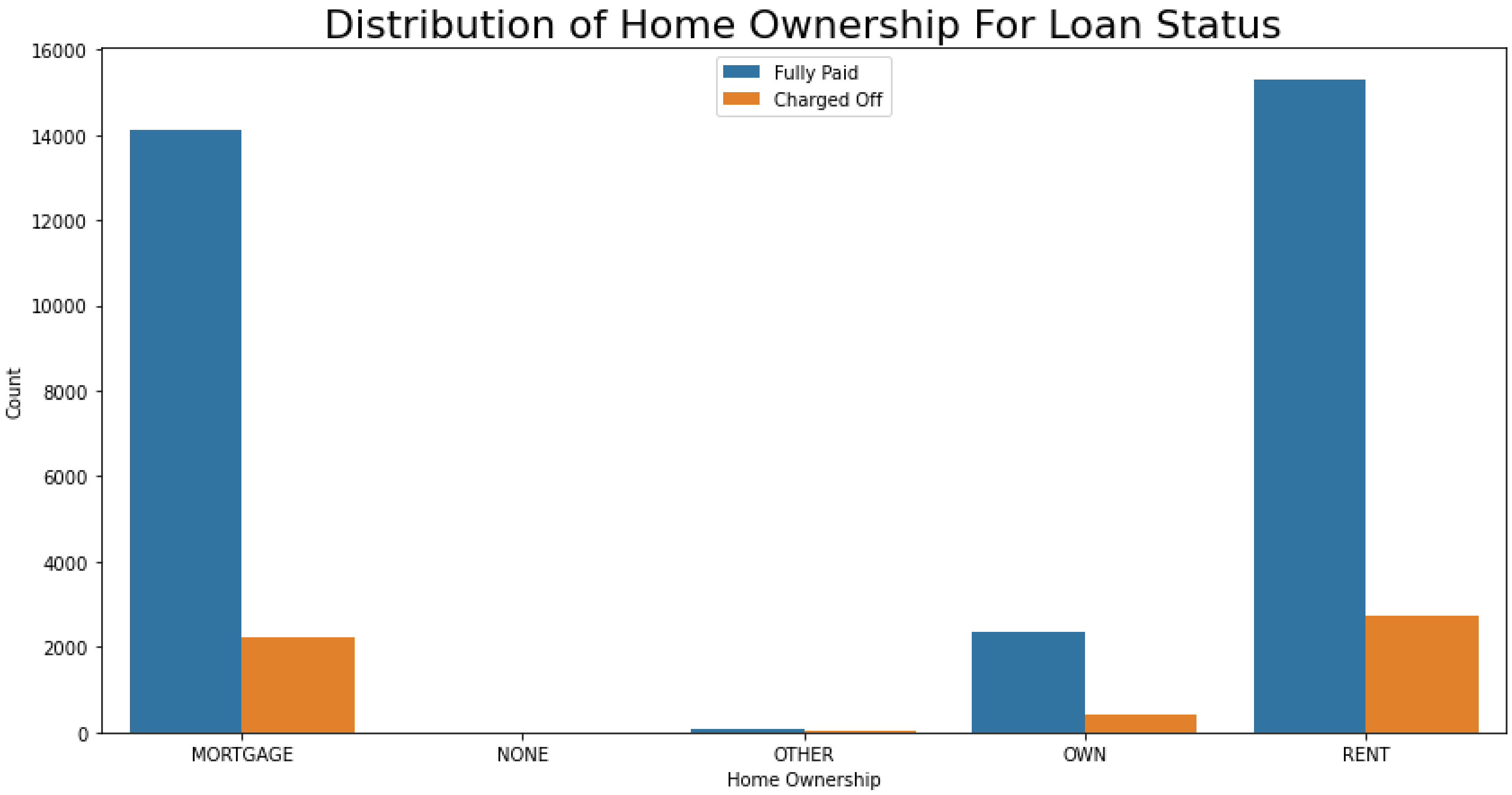


3b) Segmented Univariate Data Analysis

upGrad

Field: Home Ownership based Loan status

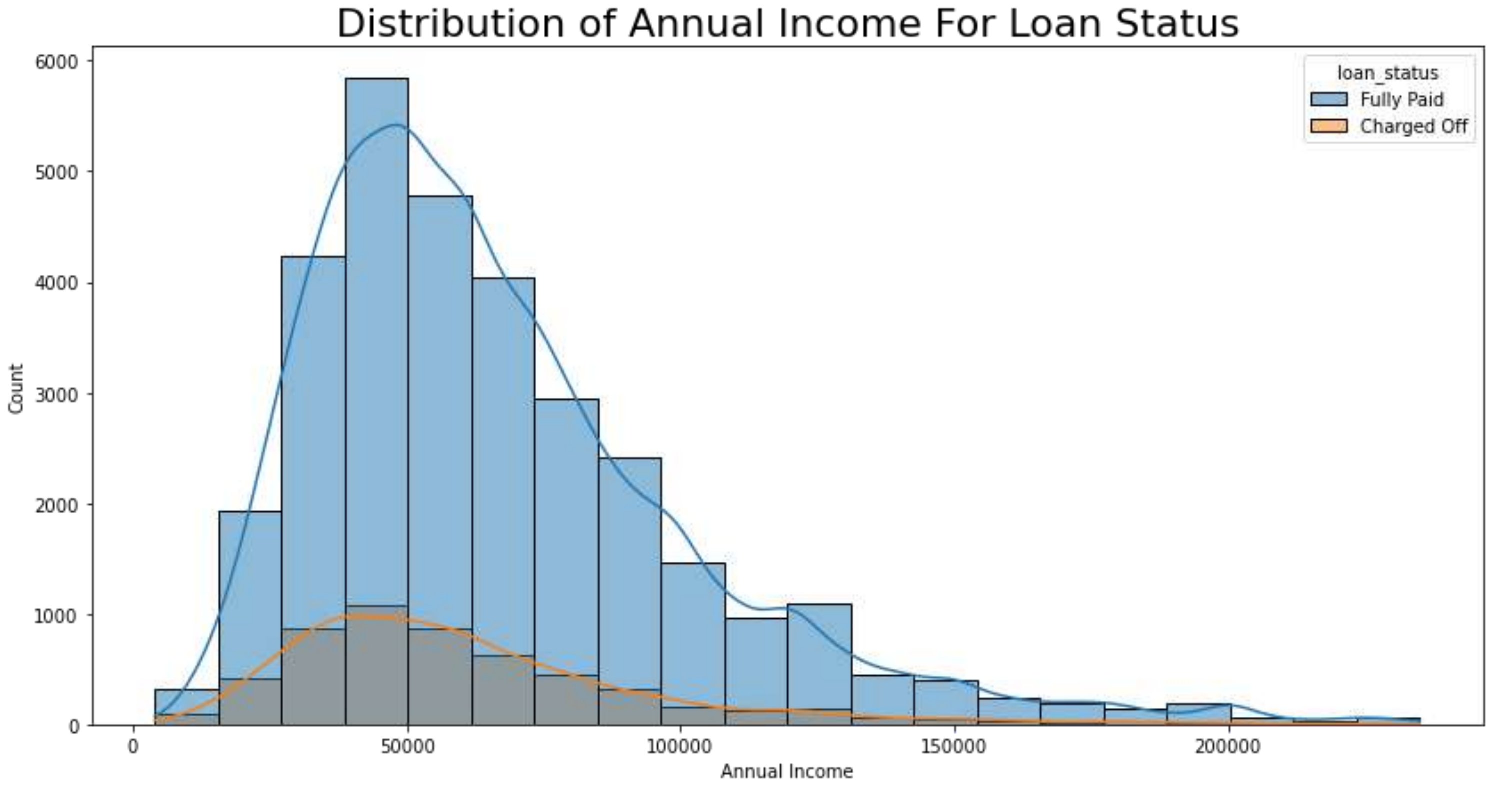
Conclusion: Borrowers who own their homes have less defaulted loans than those who are paying rent or a mortgage.



3b) Segmented Univariate Data Analysis

Field: Annual Income based Loan status

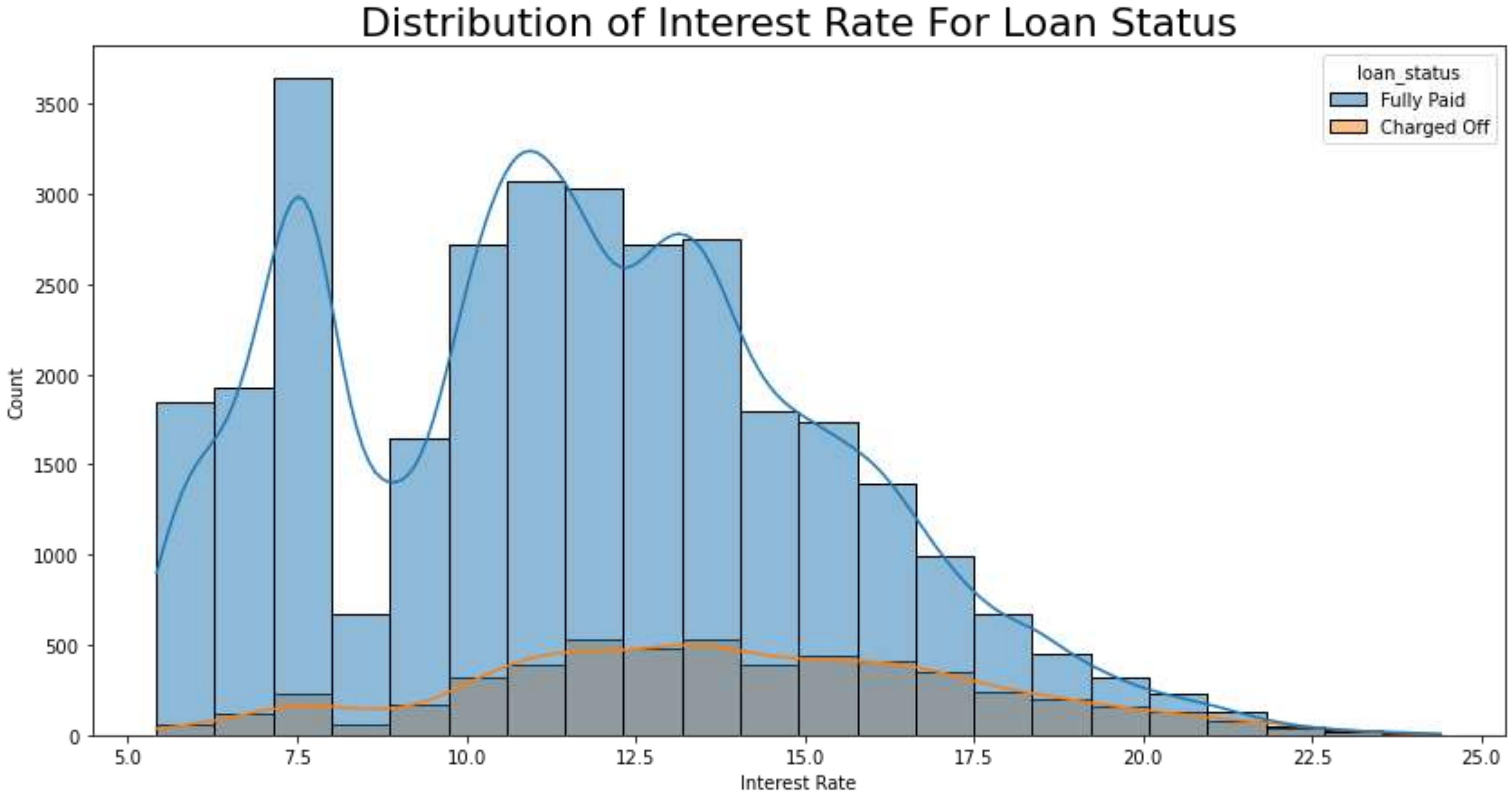
Conclusion: Borrowers with yearly incomes less 50,000 are more likely to default, whereas those with annual incomes over 50,000 are less likely to do so.



3b) Segmented Univariate Data Analysis

Field: Interest Amount Based Loan status

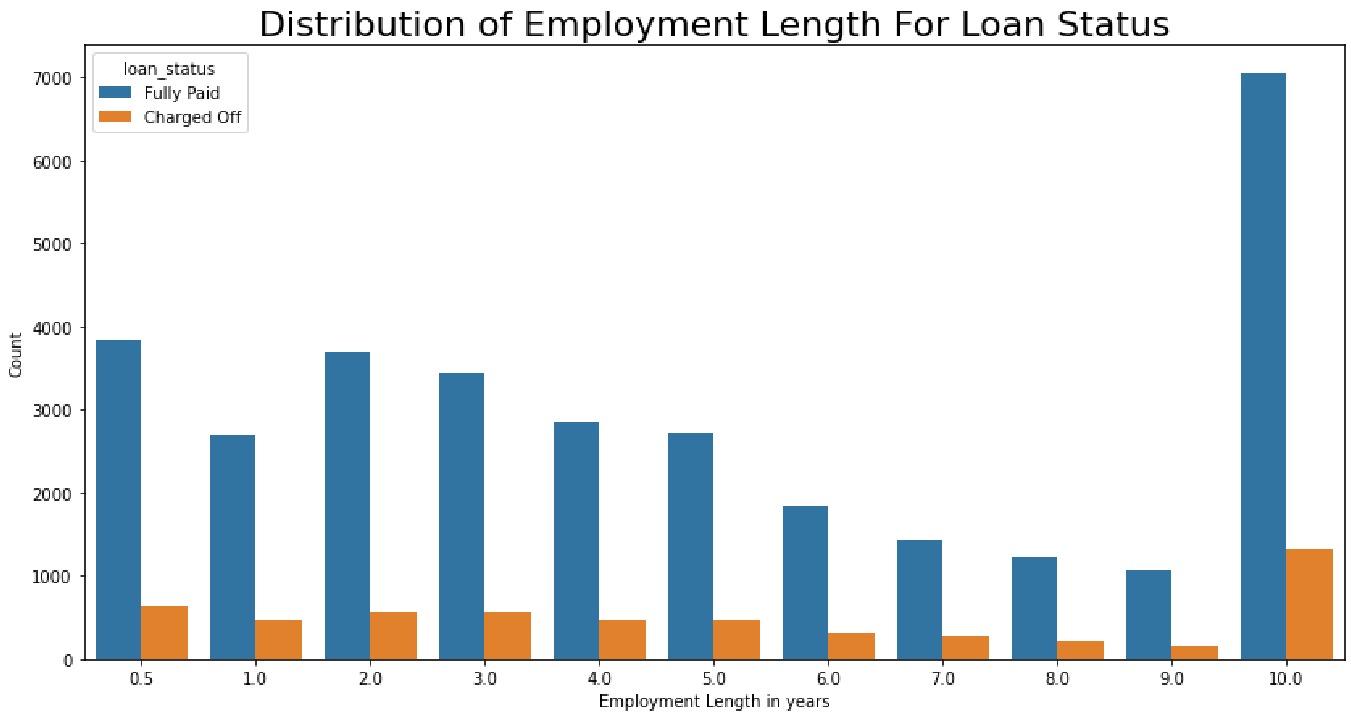
Conclusion: The default loan amount indicates a drop after a 17.5% interest rate and increases with interest rate.



3b) Segmented Univariate Data Analysis

Field: Employee Length based Loan status

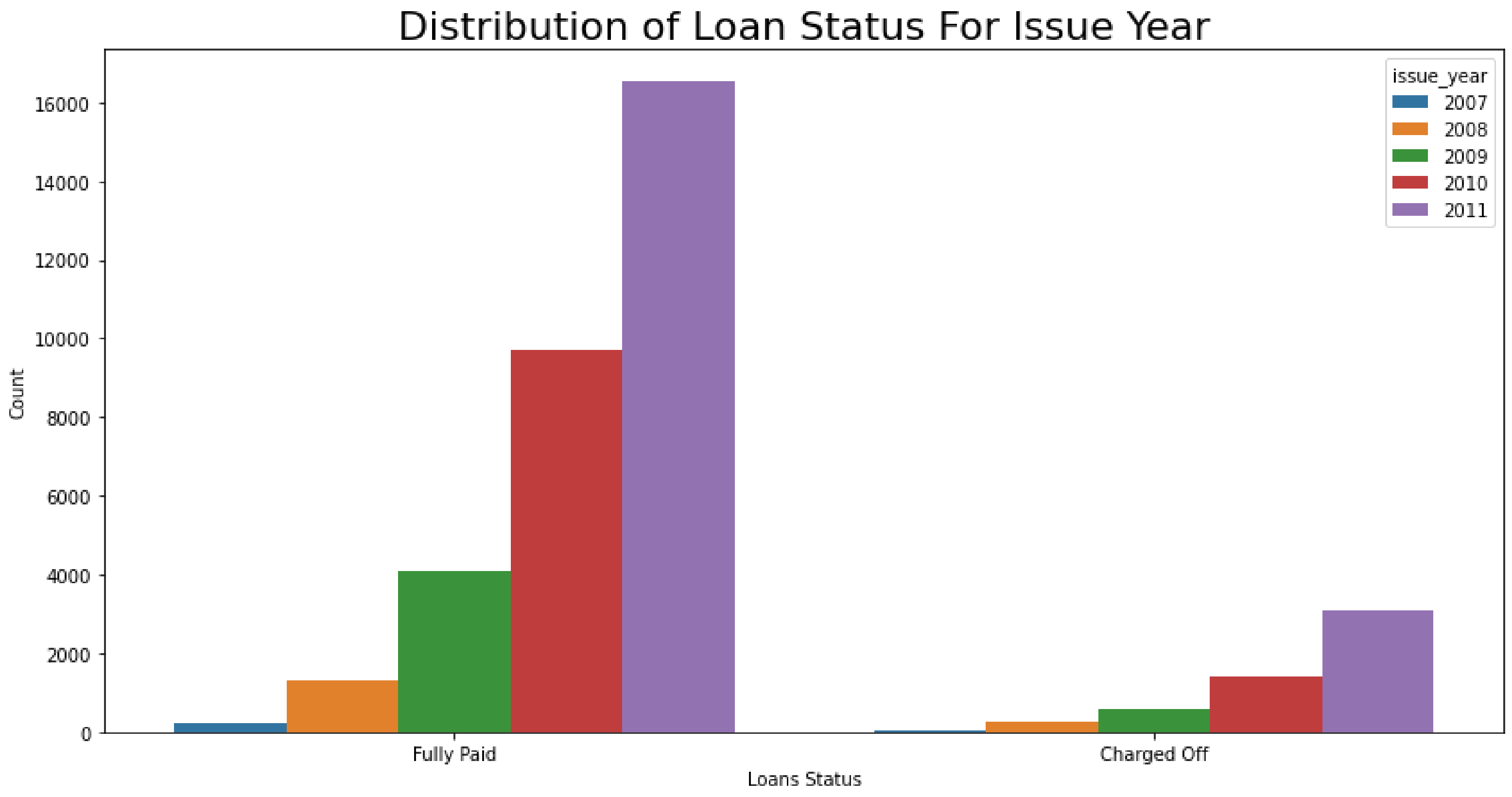
Conclusion:Employees with 10+ years of experience are more likely to fail on their loans and have a lower likelihood of repaying them in full.



3b) Segmented Univariate Data Analysis

Field: Loan status based on issue-year

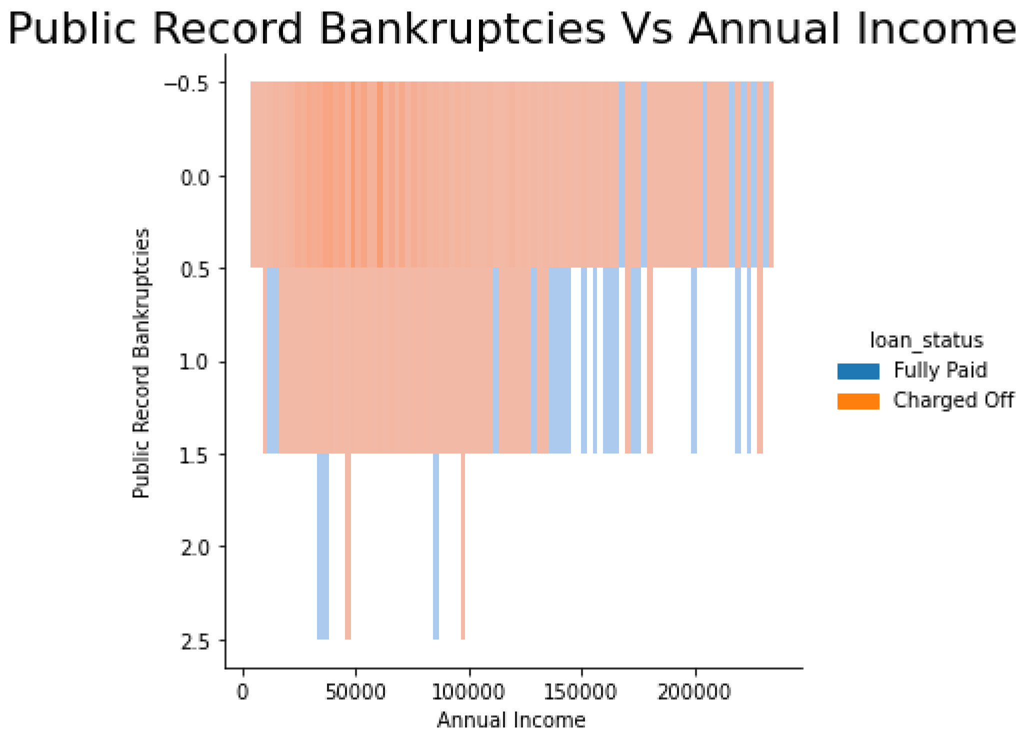
Conclusion: Fully paid loans grow exponentially over time in contrast to defaulted loans.



3c) Bivariate Data Analysis

Field:Comparison of Annual Income to Public Banktuptcy over Loan Status

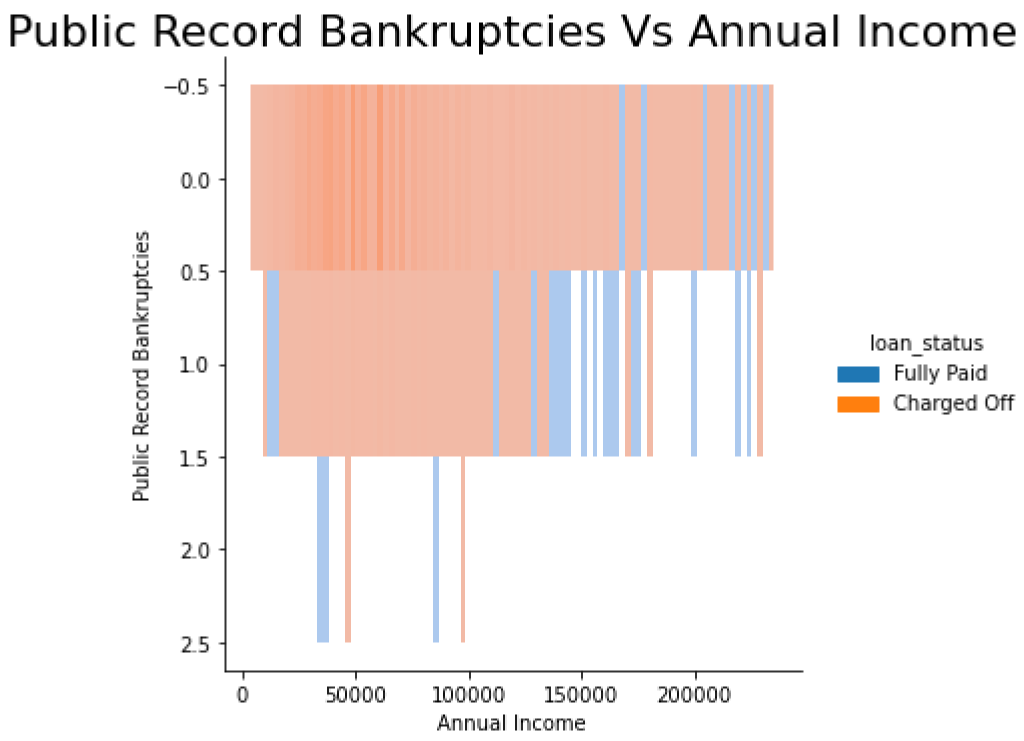
Conclusion:The borrowers are a safe choice for loan issuance because they often don't have a public record of bankruptcy.



3c) Bivariate Data Analysis

Field:Comparison of Annual Income to Public Banktuptcy over Loan Status

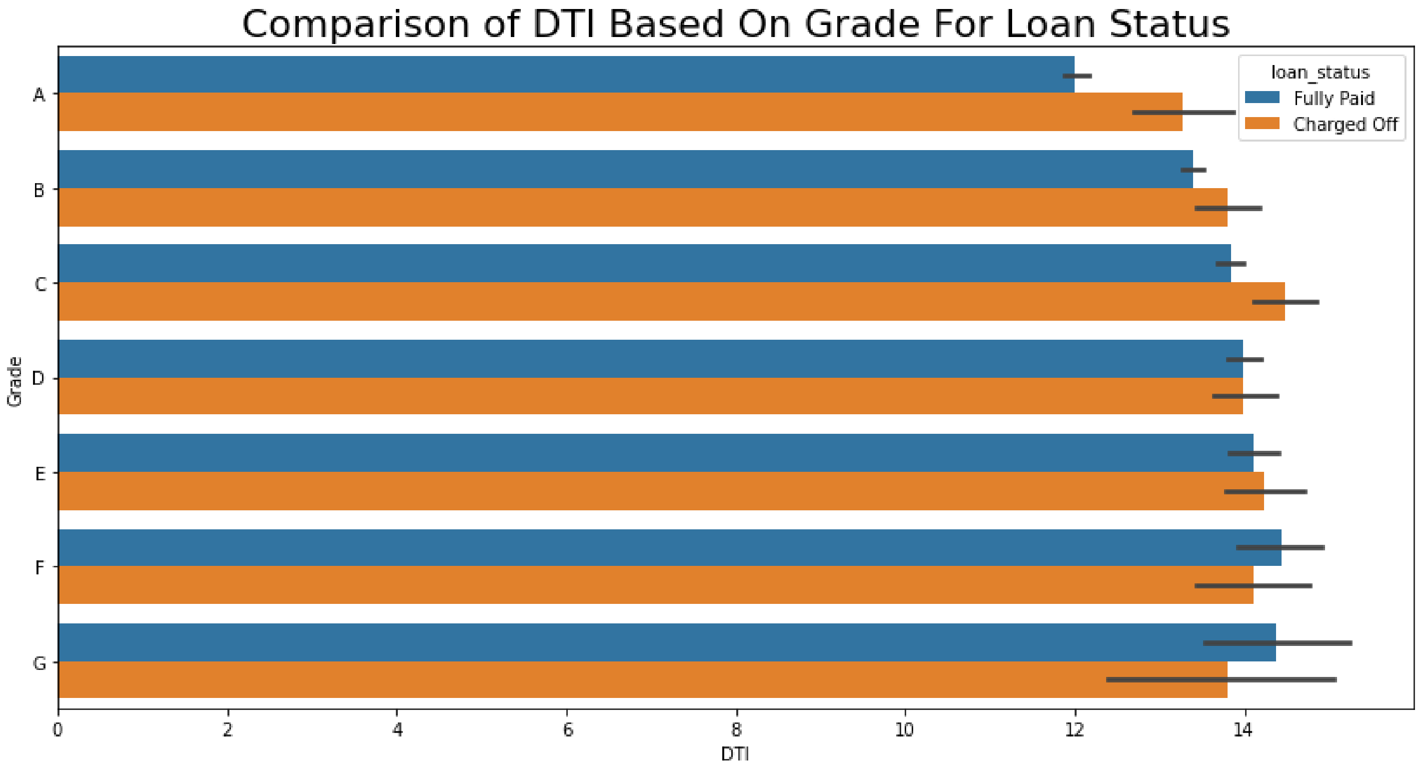
Conclusion:The borrowers are a safe choice for loan issuance because they often don't have a public record of bankruptcy.



3c) Bivariate Data Analysis

Field:Comparison of DTI to Grade over Loan Status

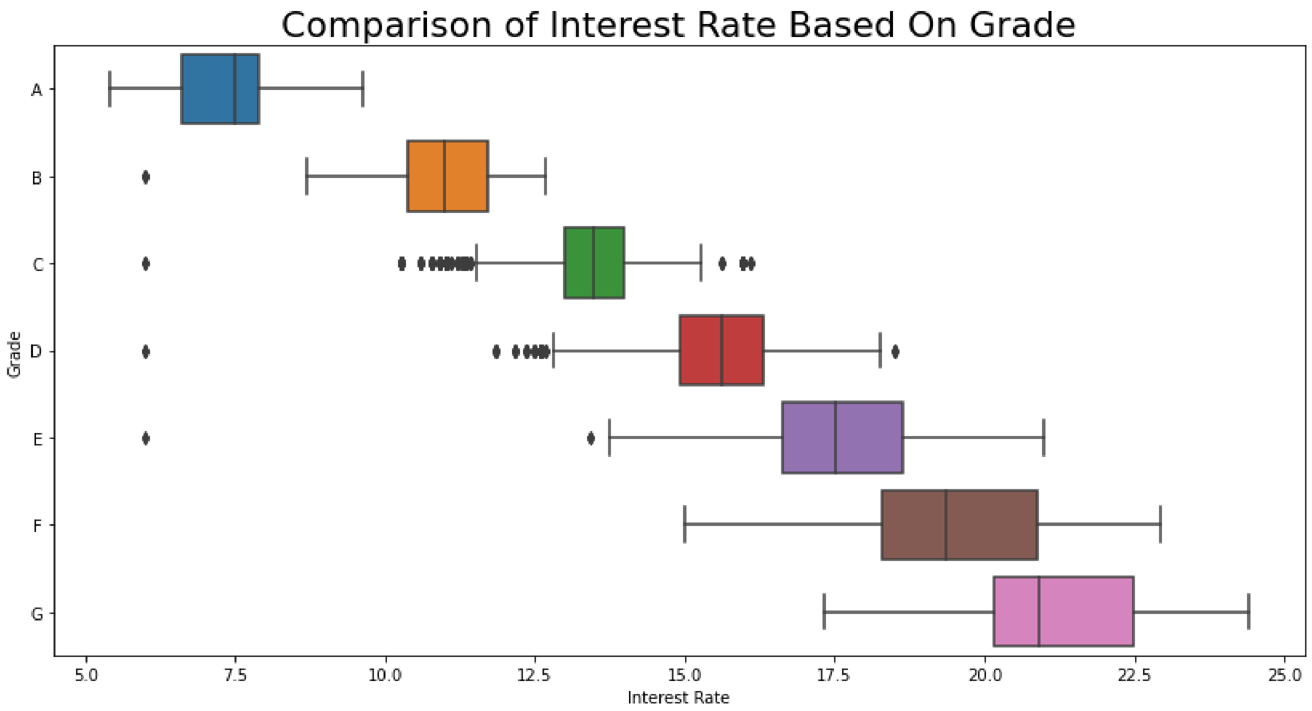
Conclusion:The Grade A, which carries the lowest risk, also has the lowest DTI ratio, indicating that higher grades have a lower default rate..



3c) Bivariate Data Analysis

Field:Comparison of Interest rate based on Grade

Conclusion:Since the grade is a risk element, we can conclude that interest rates rise as risk does.



4a) Conclusions from Univariate Analysis

1)Seven times fewer loans have defaulted than have been fully repaid.
2)The bulk of loans have terms between 36 and 60 months.
3)With a decline around 10, the interest rate is more congested between 5 and 10 and 10-15.
4)Most loans are high grade loans as seen by the significant number of loans with grades "A" and "B" compared to the rest.
5)The majority of borrowers have more than ten years of job experience.
6)The majority of borrowers are either on a mortgage or rent and do not own any property.
7)The company verifies or has a source verify about 50% of the debtors.
8)Since the yearly income has a left-skewed normal distribution, the bulk of burrowers have extremely low annual incomes in comparison to the rest of the population.
9)Debt consolidation loans account for a sizable portion of all loans, followed by credit card loans.
10)Large urban areas like California, New York, Texas, Florida, etc. are where the majority of the borrowers are from.
11)The majority of borrowers have DTI ratios between 10-15, which is a relatively high level of debt relative to income.
12)The majority of the borrowers don't have any public bankruptcy records.
13)The final quarter of the year is when the majority of loans are disbursed.
14)We can claim that the loan approval rate is rising over time because the number of sanctioned loans grows exponentially with time.

4b) Conclusions from Segemneted Univariate Analysis

1)The most common loan purpose and the one with the greatest number of fully paid and defaulted loans is debt consolidation.
2)The average and 25% are the same for both loans, but the defaulted loan has a bigger 75%, indicating that large loans have a higher likelihood of defaulting.
3)The 36 month term has a larger possibility of loan repayment in full than the 60 month term, which has a lower chance of default.
4)Grade A and B loans make up the majority of the loans with a 36-month duration, while Grade B, C, and D loans make up the majority of the loans with a 60-month term. 5)The Loan Status fluctuates with DTI ratio; we can see that loans with a DTI ratio of 10-15 have a larger percentage of defaulted loans, but loans with a higher DTI have a higher likelihood of defaulting.
6)Compared to borrowers who are paying a mortgage or rent, borrowers who own their property have lower default rates.
7)Borrowers with annual incomes under 50,000 are more likely to default, whereas those with annual incomes over 50,000 are less likely to do so.
8)Unlike defaulted loans, fully paid loans are growing exponentially over time.
9)The amount of the default loan rises with the interest rate and begins to fall after a 17.5% interest rate.
10)Employees with 10+ years of experience are more likely to fully repay the loan and are less likely to fail.

4c) Conclusions from Bivariate Analysis

upGrad

1)Since the Grade is a risk element, we may claim that interest rates rise as risk does.
2)The lowest risk grade, Grade A, also has the lowest DTI ratio, indicating that higher grades have a lower default rate.
3)The majority of the borrowers lack a public bankruptcy record, making them a secure option for loan issuance.

THANK YOU