# Lending Club Case Study

USING EXPLORATORY DATA ANALYSIS

# Contents:
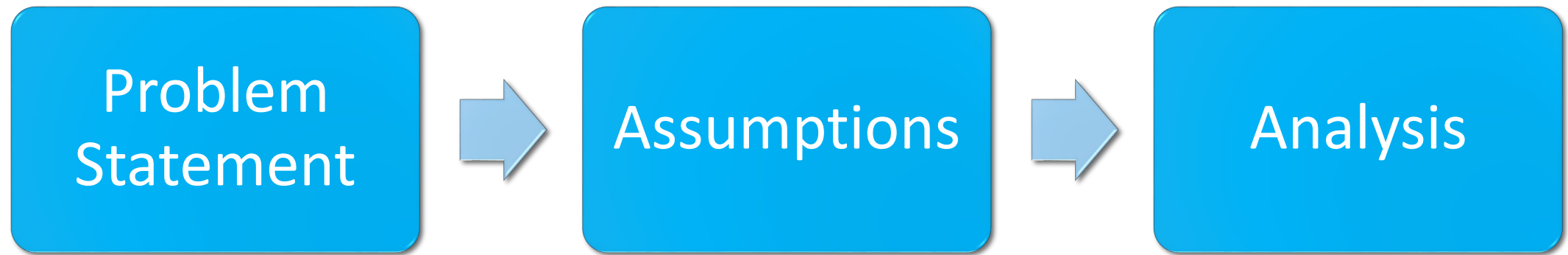
# Problem Statement:

**LOAN DATASET**

Loan Accepted → Default
             → Non-Default

Loan Rejected
(Not considered in dataset)
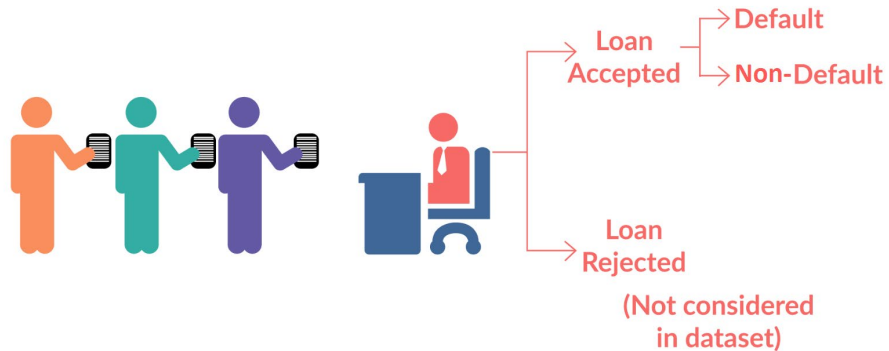
❑ Lending Club, an online loan marketplace wants to understand the '**driving factors'** behind loan default.

❑ Given a loan dataset, Using EDA we derive insights using the attributes that influence defaulters.

# Assumptions:

- Outliers in any attribute are omitted to eliminate bias in the analysis

- 'Current' applicants are not considered as the analysis is based on defaulters.

- Assumed Home Ownership of 'None' type as 'Other' as there are only three such scenarios and 'Other' best suites the case.

- Assumed empty values in Employment Title as 'Not Disclosed'

- Assumed empty values in Employment Length as < 1 year, considering them unemployed.

- Assumed description is not a required field to be available, so filled in empty values as 'Not Disclosed'

# Analysis:

## 1. DATA LOADING AND UNDERSTANDING

- Loading the data

- Understanding the business goals and the attributes provided in the Data Dictionary

## 2. PRE-PROCESSING

- Handling data issues

- Filtering the columns

- Making the data usable for analysis.

# Analysis:

## 3. HANDLING MISSING VALUES

❑Data can be of Numeric or Categorical type

❑Missing values are in rows and columns are handled based on the type of data.

## 4. OUTLIER TREATMENT

❑Identifying and Removing biased Outliers
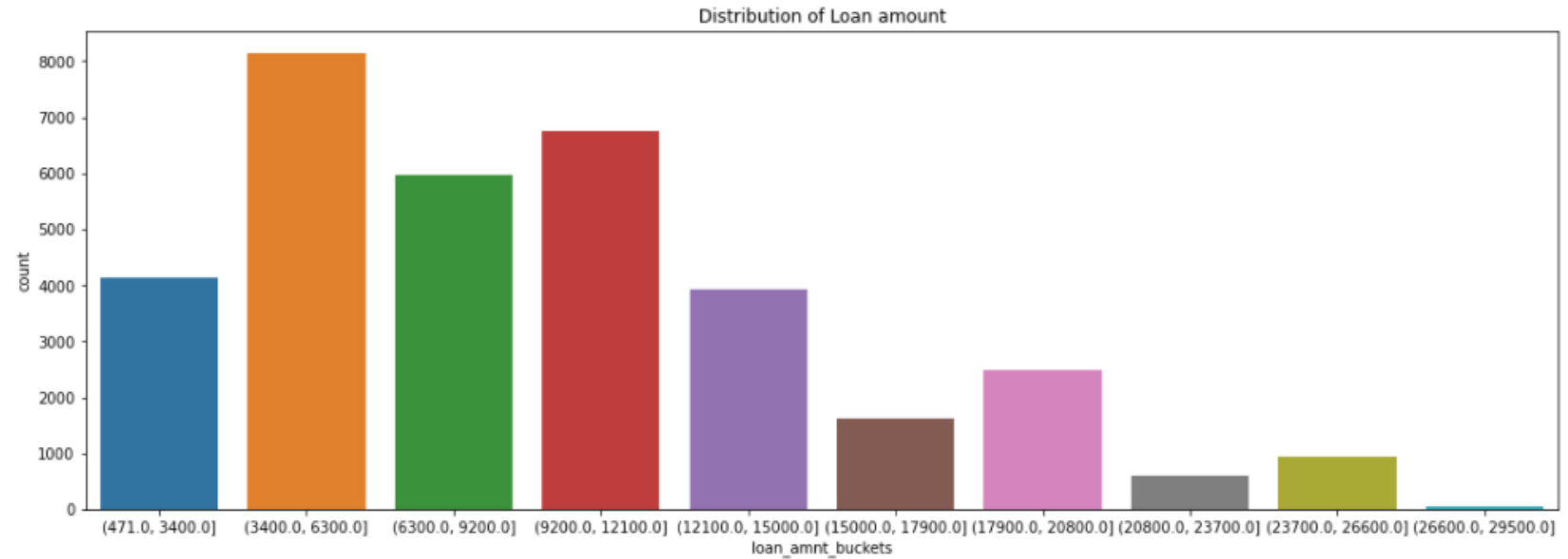
# Analysis:

## 5. EXPLORATORY DATA ANALYSIS:

➢UNIVARIATE ANALYSIS

➢SEGMENTED UNIVARIATE ANALYSIS

➢BIVARIATE ANALYSIS

# Analysis:

UNIVARIATE ANALYSIS

# Loan Amount:

Most customers have taken a loan between 5000 to 13487, specifically in the range of 3400 to 6300 but less loans were taken as amount increases from 12100



Distribution of Loan amount

# Funded Amount:

Most customers have been increasingly funded for the loan between 5000 to 13000 towards the range of 5290 to 8630, whereas there can be seen a constant decrease of funded amount after 8630
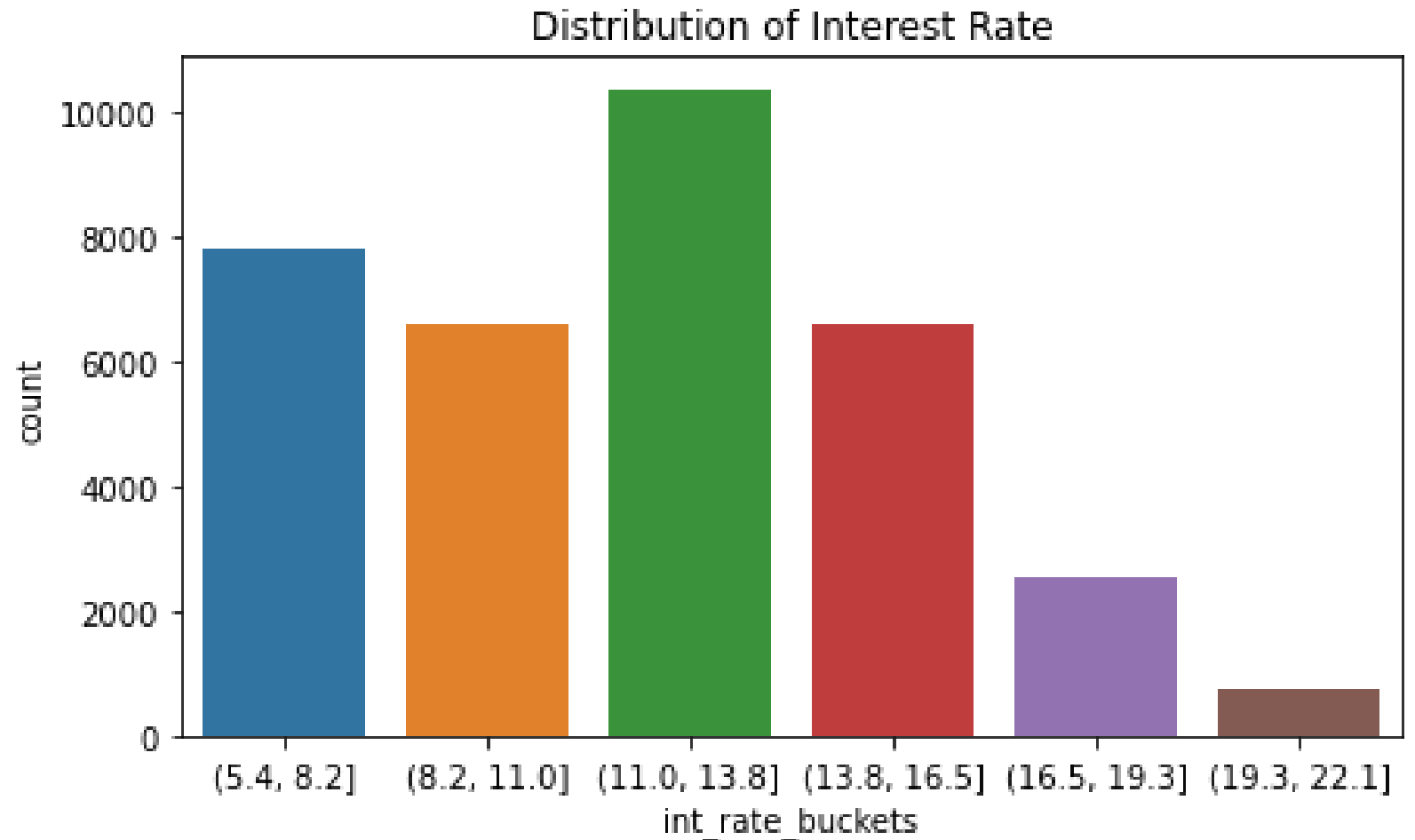


Distribution of funded amount
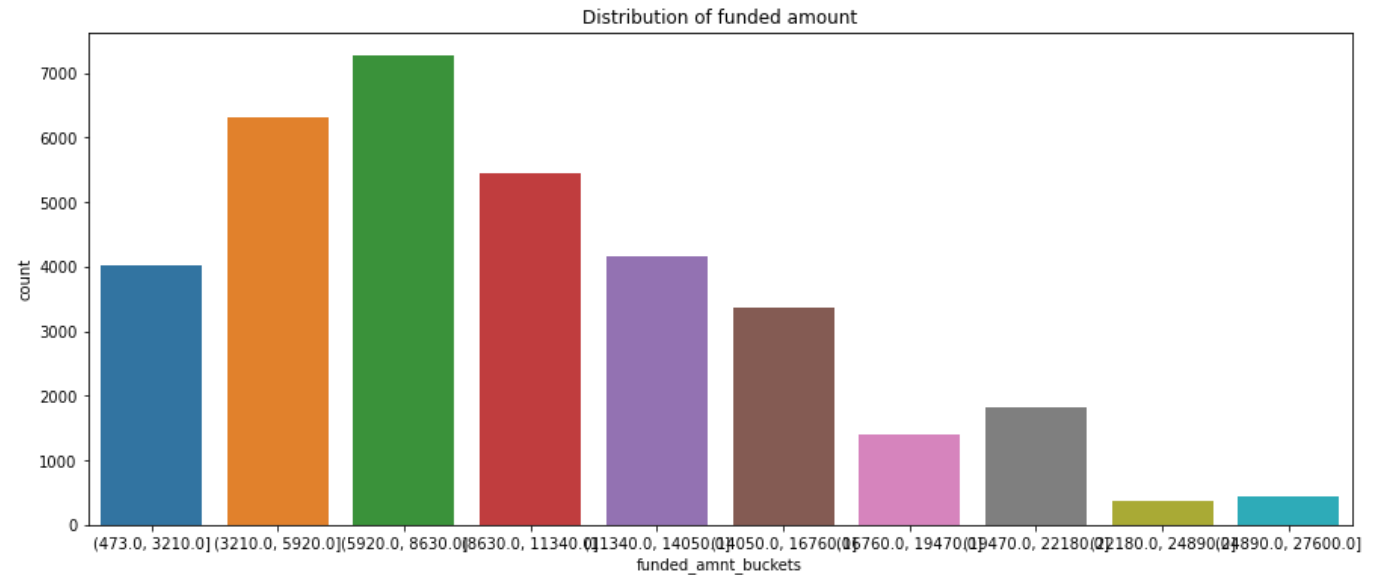
# Term:

Most of the loans are taken for a term of 36, where as loans taken for term of 60 are is about 50% less than that of 36 term loan.
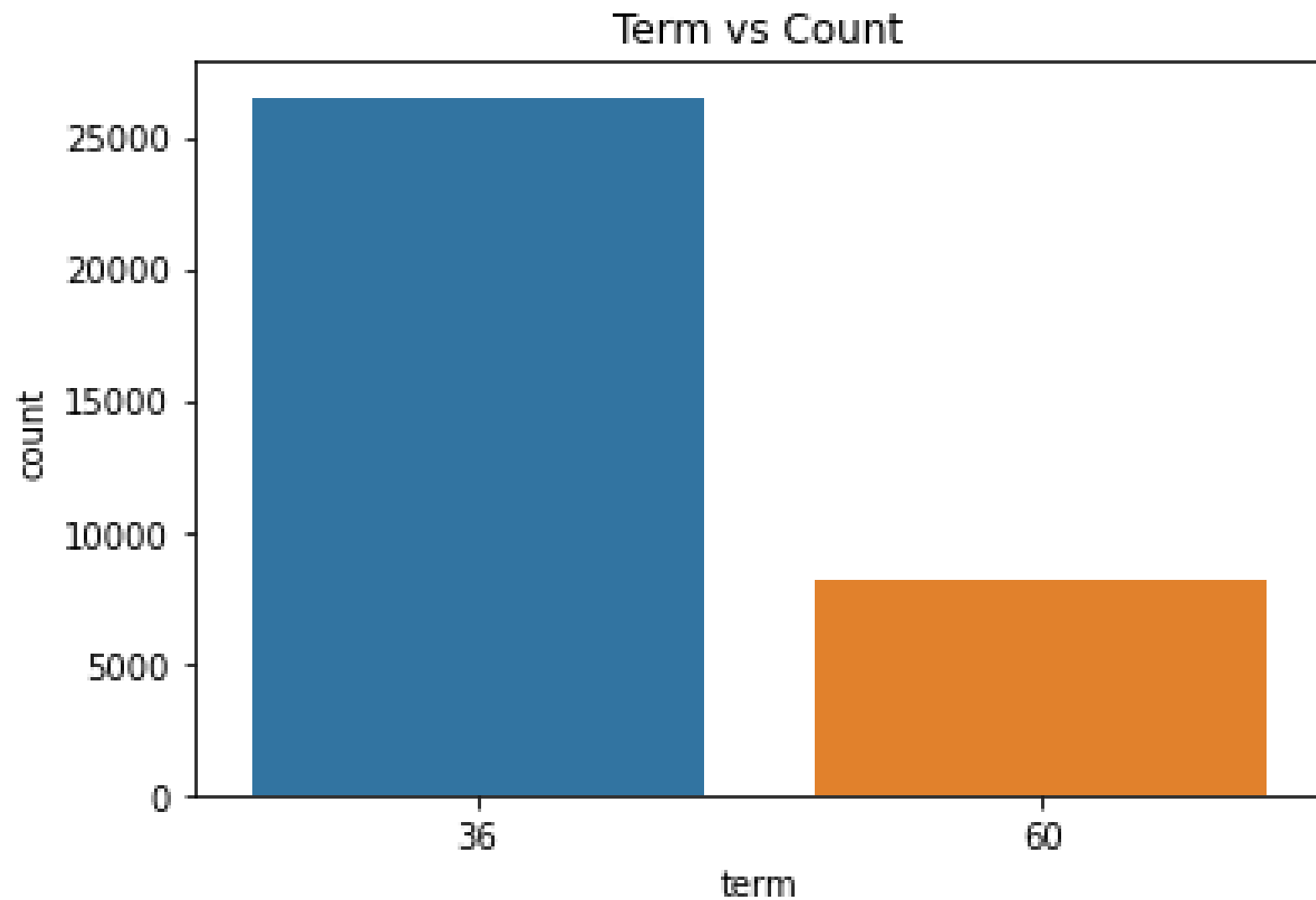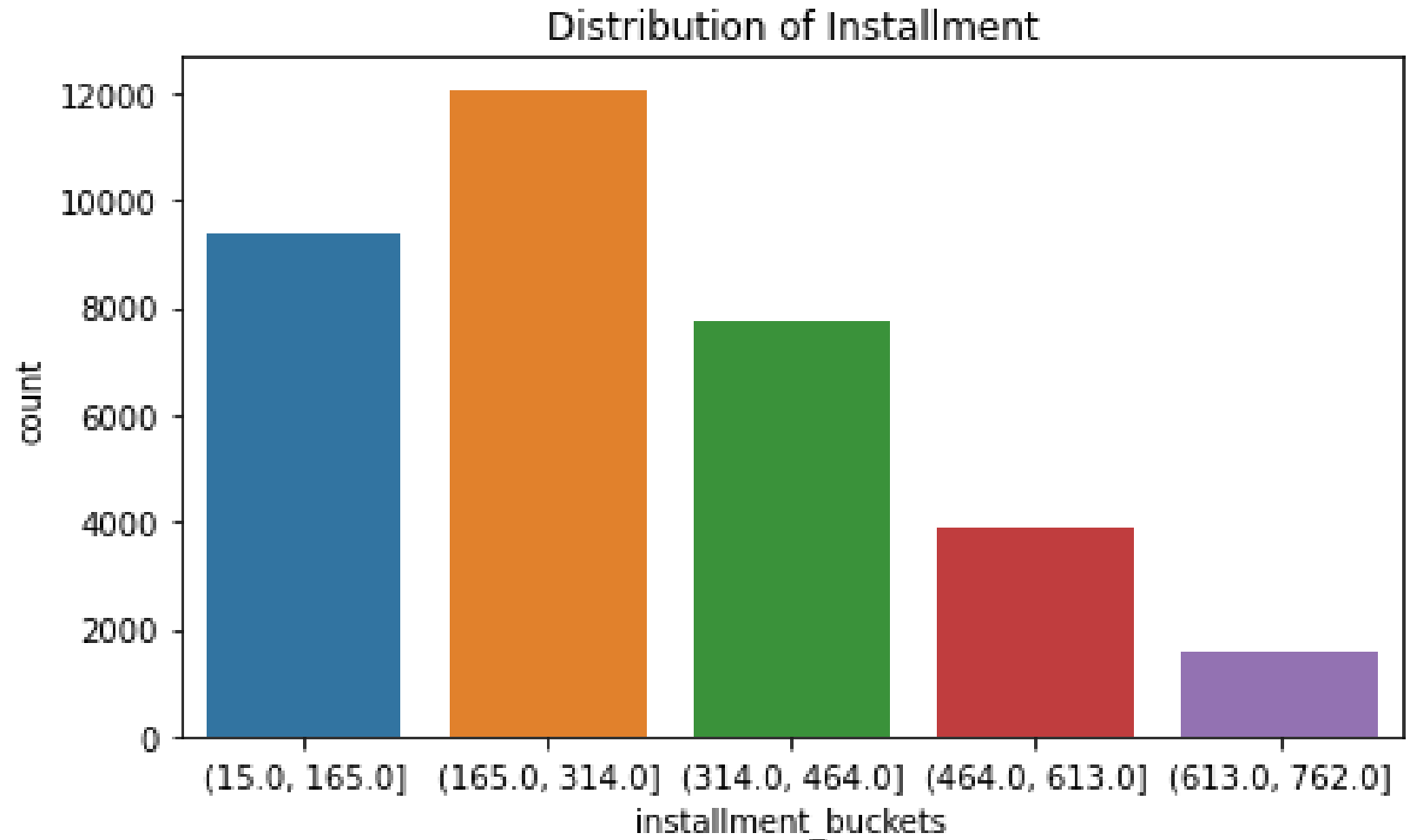
# Interest Rate:

Most customers have taken the loans with interest rate between 8.9% to 14.22%, specifically in the range of 11% to 13.8%
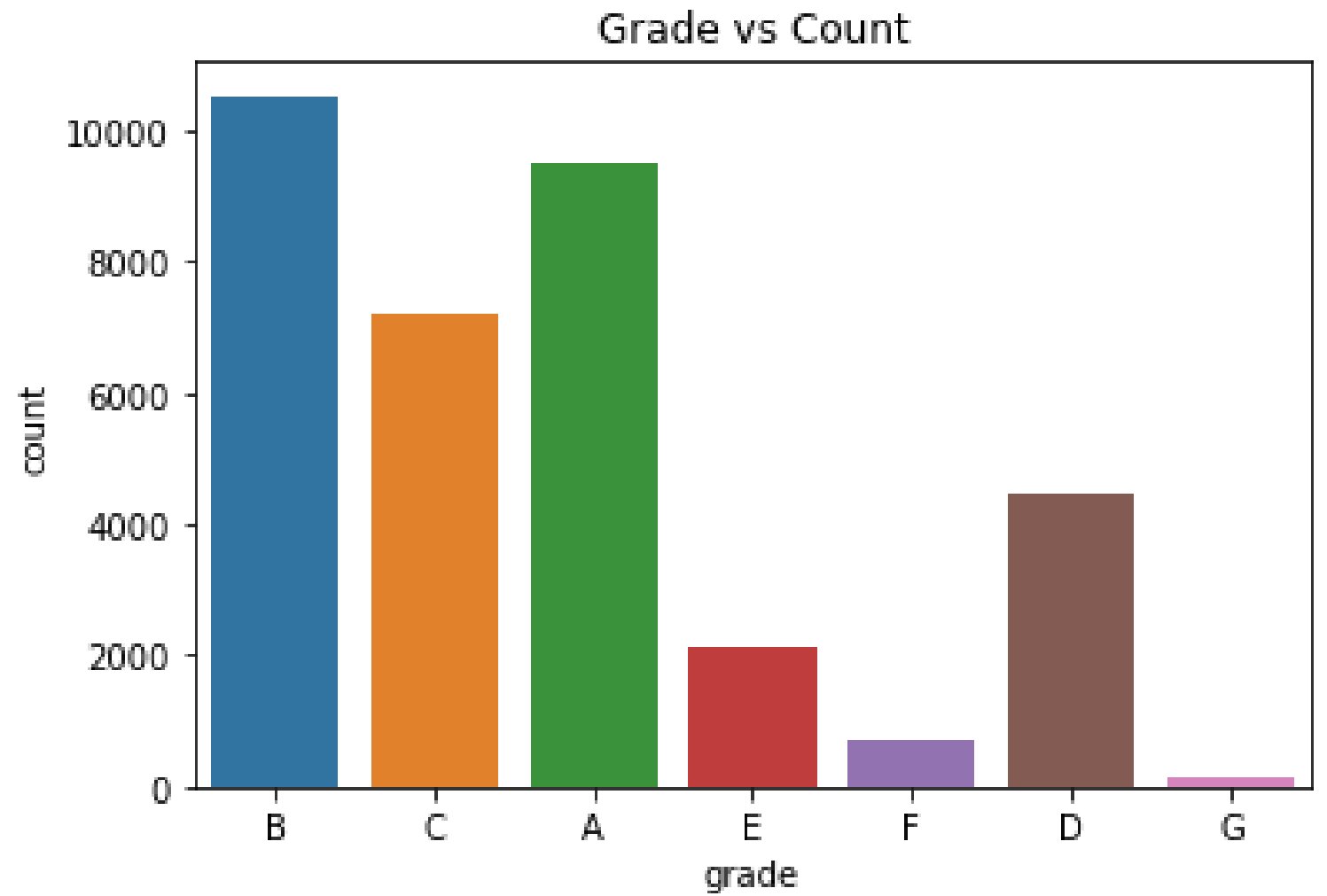


Distribution of Interest Rate

# Installments:

Installment owed by most of the customers is in the range of 158.86 to 383.66, especially between 165 to 314
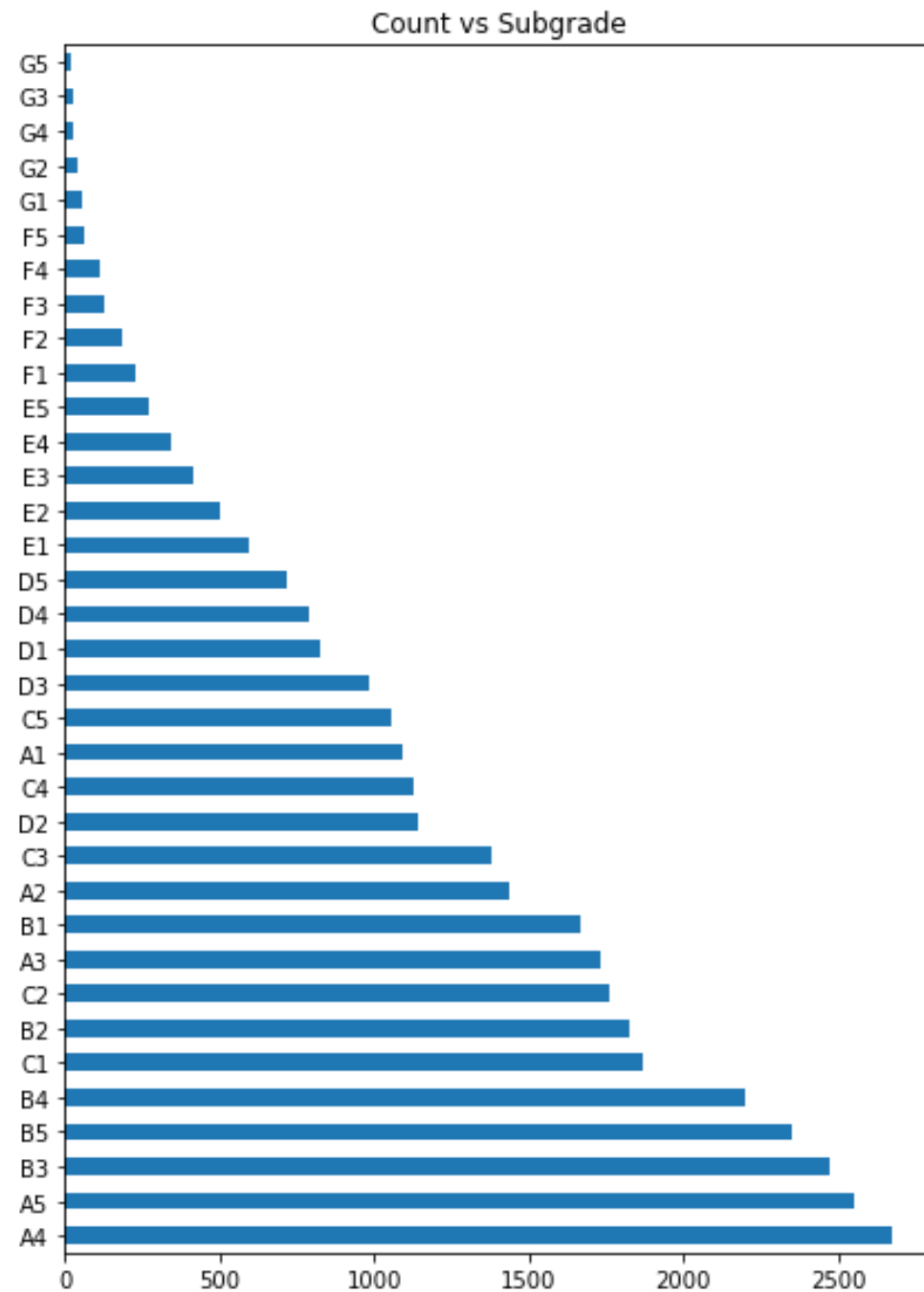


Distribution of Installment

# Grade:

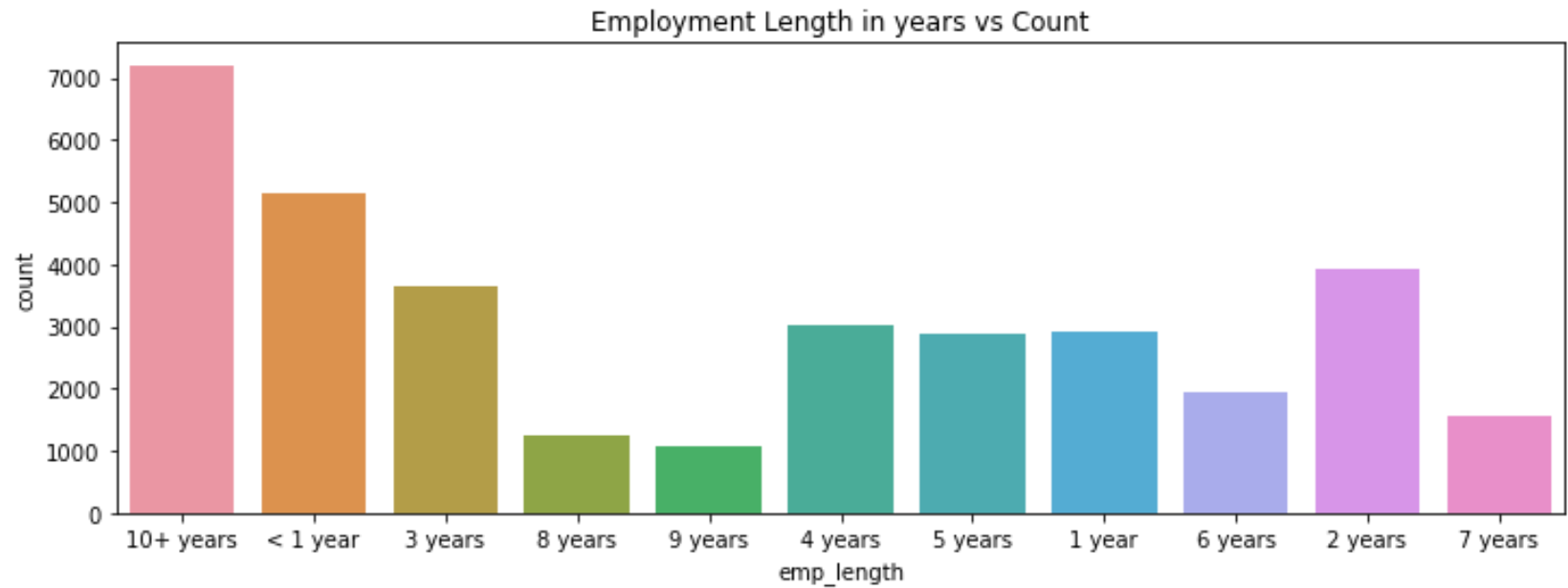Most customers were of grade 'B', followed by 'A' and 'C'

# Sub Grade:

Most number of borrowers belong to the sub grade A4.
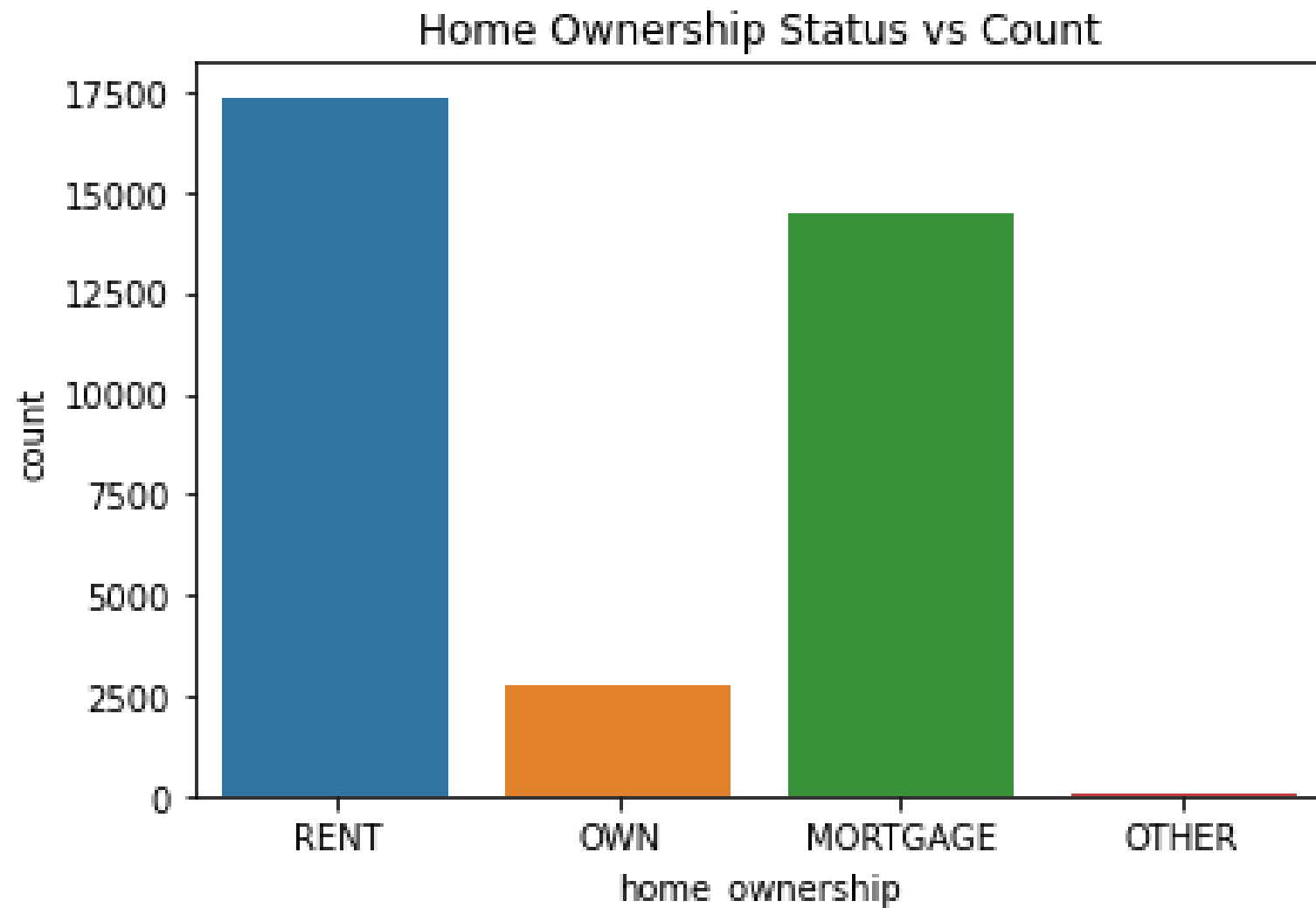


Count vs Subgrade

# Employment Length:

Majority of borrowers applying for the loan have employment length more than 10 years
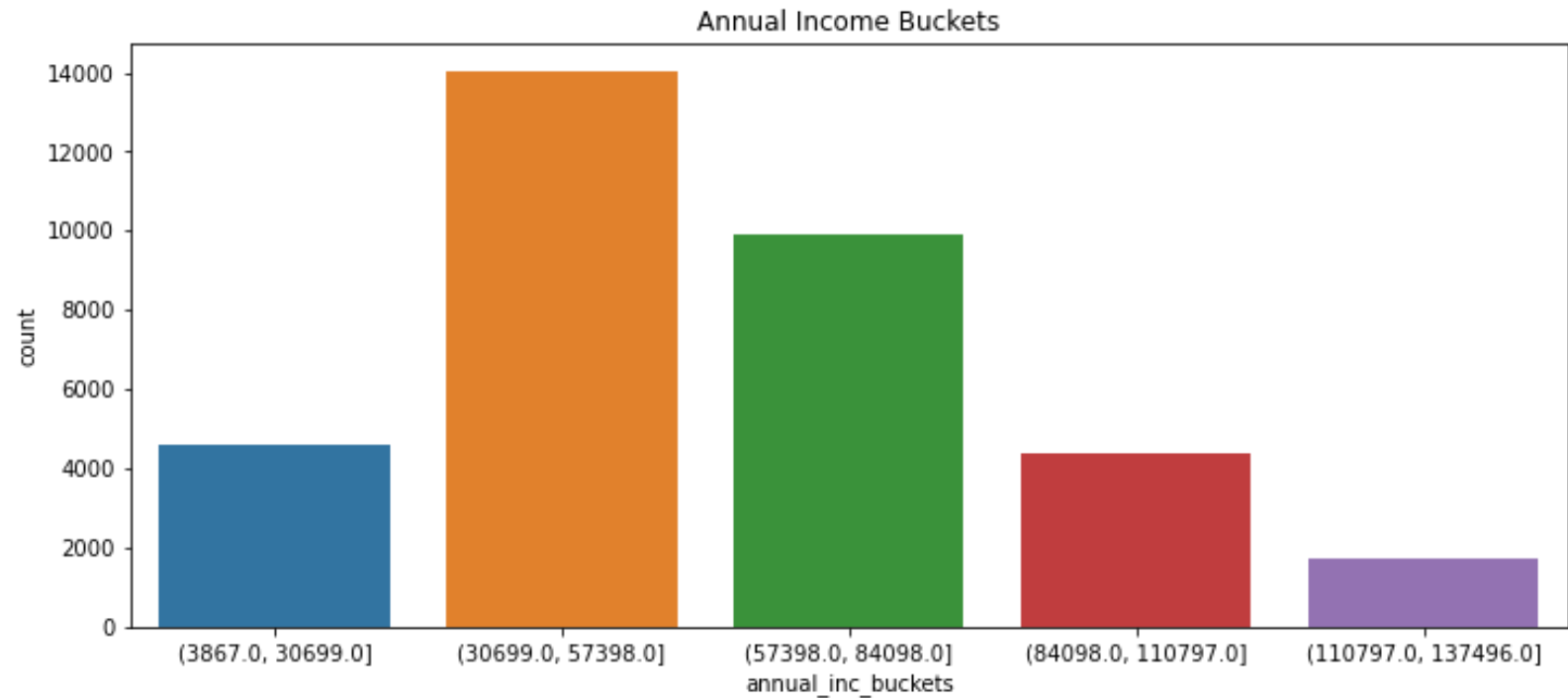
# Home Ownership:

Majority of the borrowers have a Home Ownership Status of RENT followed by MORTGAGE
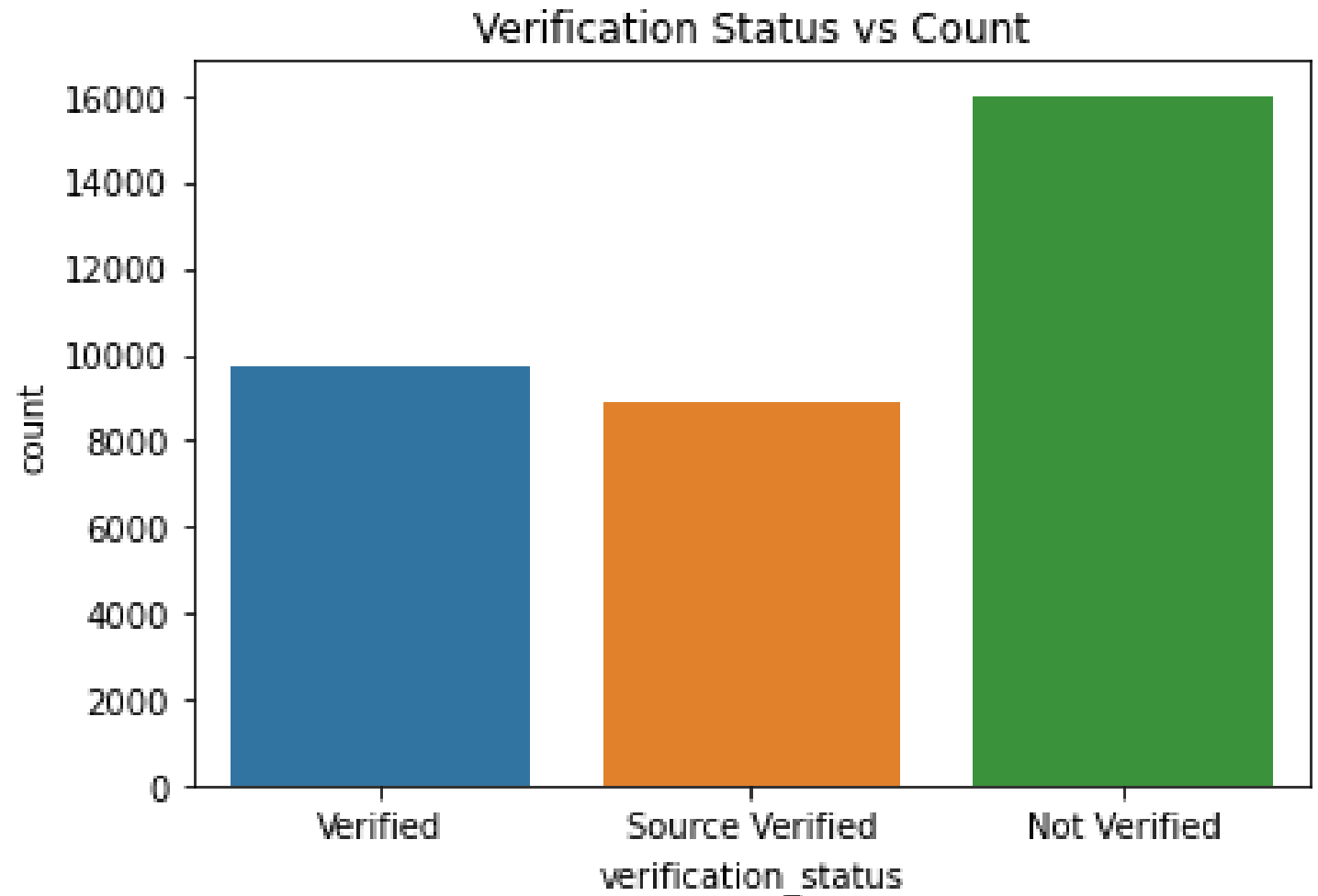
# Annual Income:

Majority of the customers have a annual income in the range of 30699 to 57398.



Annual Income Buckets

# Verification Status:

Around 50% of the customers are granted loan with income Not verified. Among the remaining customers, around 25% of them have their income verified and the rest 25% have their income source verified

# Loan Issued Year:

Most loans were issued in the year 2011. The number of loans granted are exponentially increasing by every year
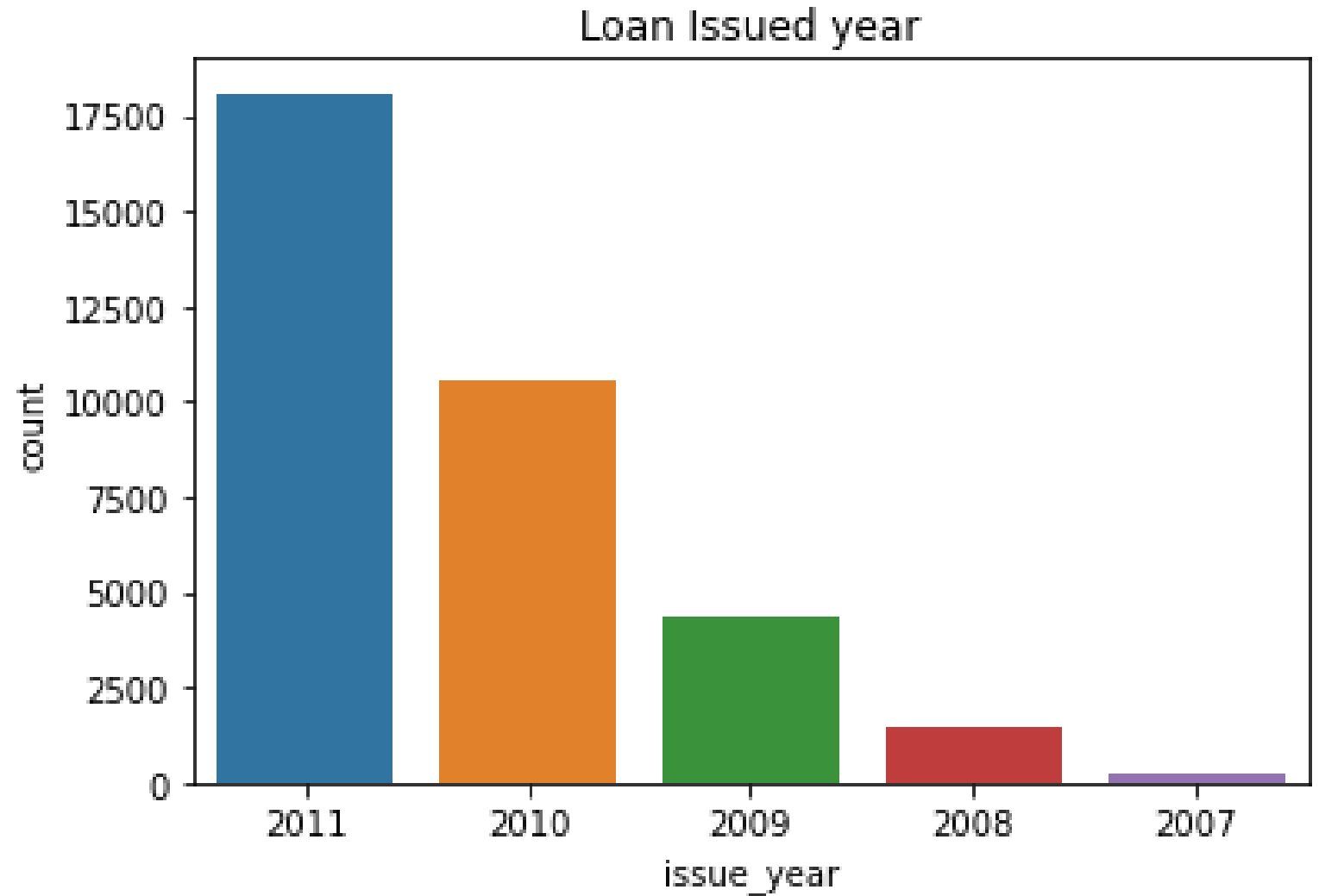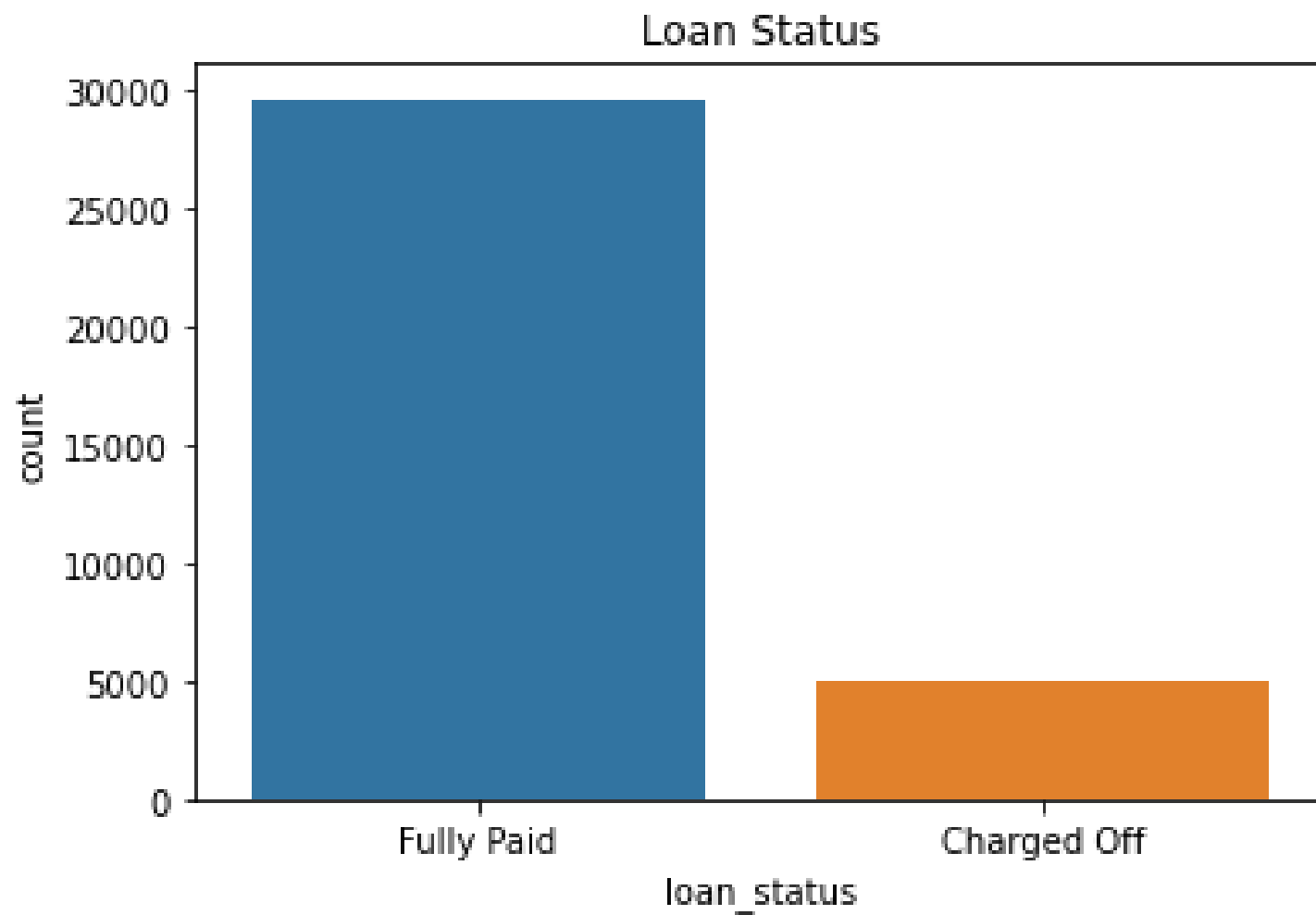


Loan Issued year

# Loan Issued month:

More loans are issued towards the end of the year compared to the year beginning, there is a constant increase in loan issue from January to December. The loan issuing gets stabilized in the middle of the year, around June-August.
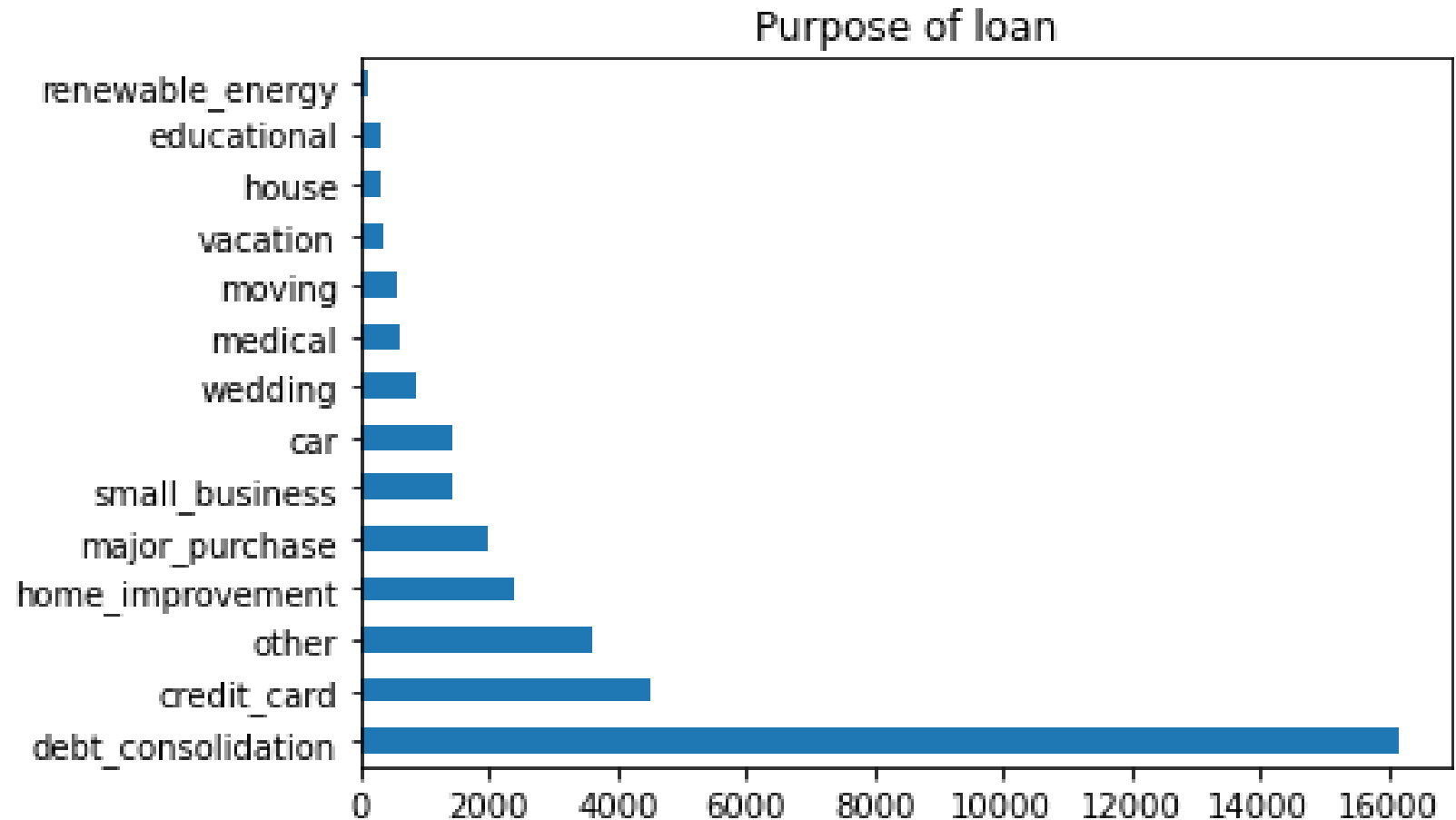


Loan Issued Month

# Loan Status:

The Charged off customers are about 14% of the total customers.

# Loan Purpose:

Majority of the loans were taken with the purpose of debt Consolidation and credit card. Renewable energy, education, housing and vacation are the lowest probable reasons to attract a loan.



Purpose of loan

# Borrower's state:

Majority of the loans were taken from borrowers belonging to California state, followed by New York, Florida and Texas. Maine, Idaho, Nebraska, Iowa, Tennessee, Mississippi almost have negligible amount of borrowers.
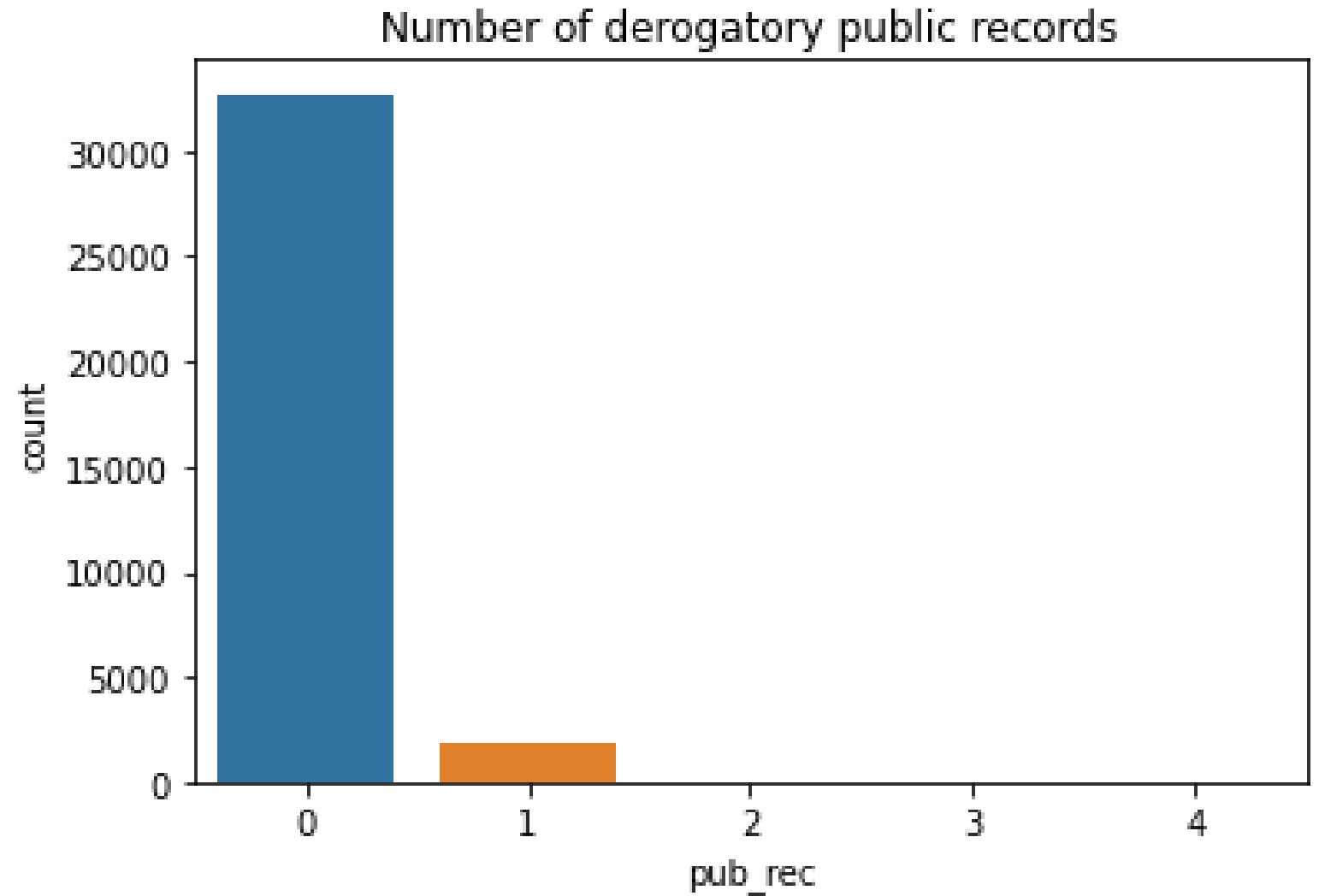


State of the Borrower
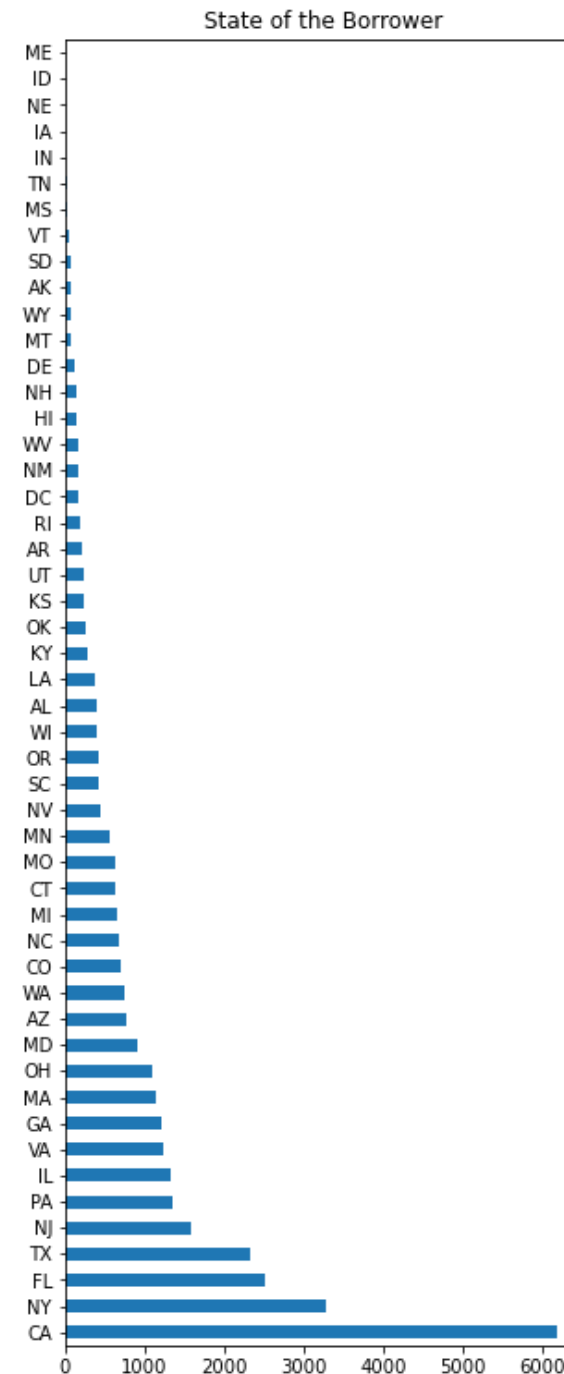
# Inquiries in the last 6 months:

No inquiries excluding auto and mortgage inquiries were made majorly in the last 6 months. Although there were 1-3 inquiries made in some cases.



Inquiries excluding auto and mortgage inquiries in the past 6 months
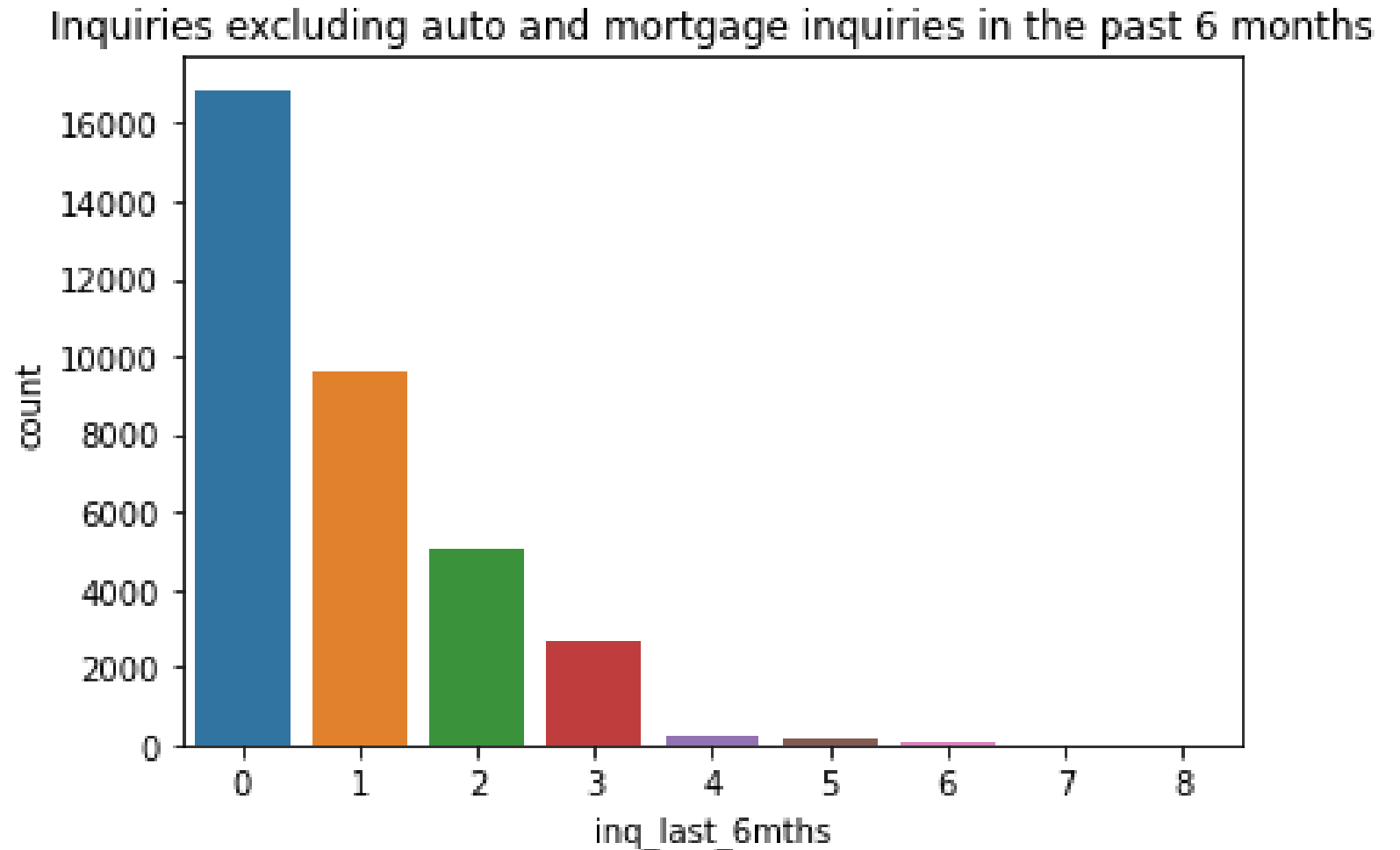
# Derogatory public records:

Only around 6% of the borrowers have derogatory public records
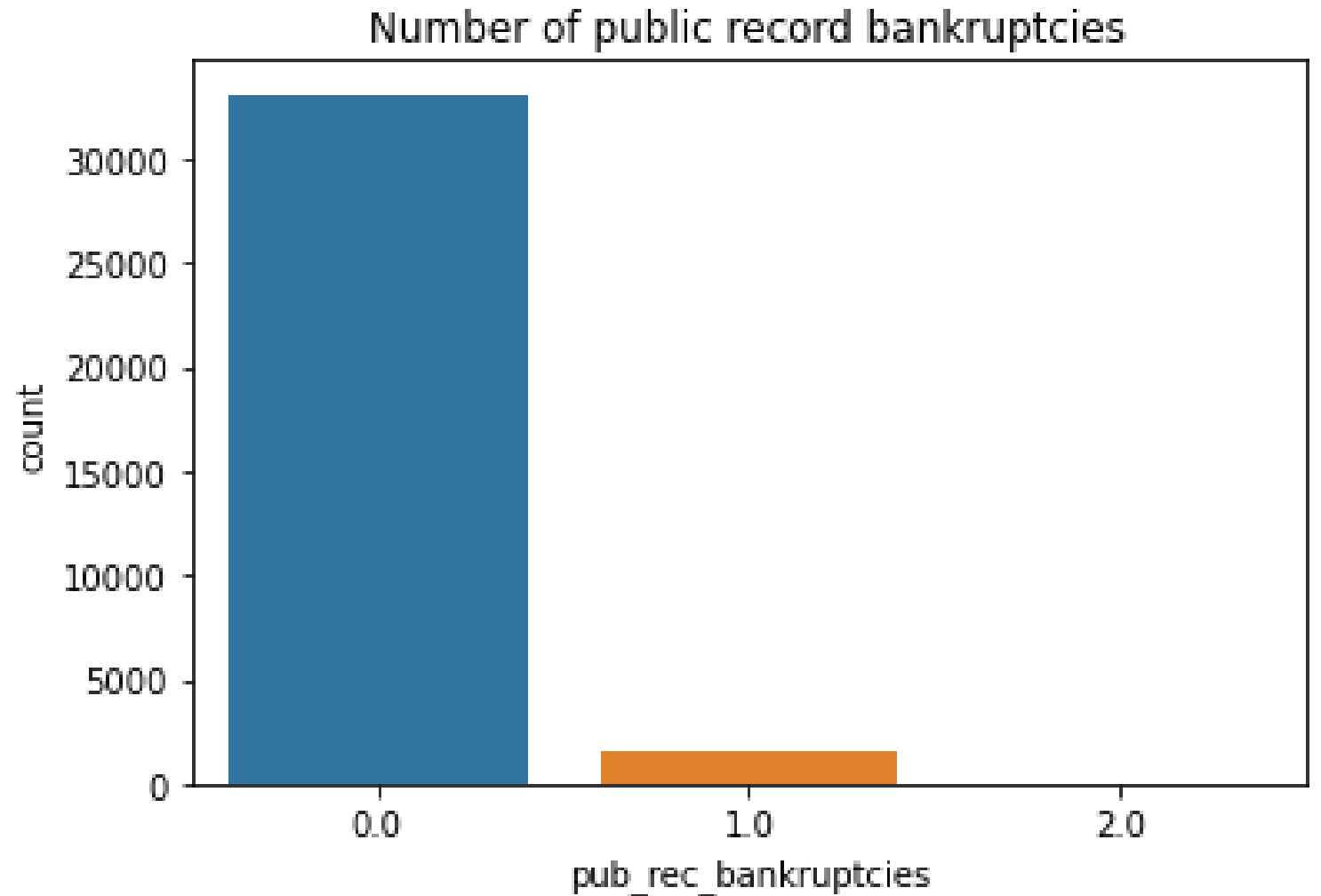


Number of derogatory public records

# Public record bankruptcies:

Only around 5% of the borrowers have public record bankruptcies



Number of public record bankruptcies

# Analysis:

SEGMENTED UNIVARIATE ANALYSIS

# Loan Amount

1. There is a 29% chance that a loan can be charged off for the loan amount applied between 26600 and 29500.

2. Most number of Charged Off loans are applied for amounts between 5000 and 15000.

3. Even though less number of loans are taken in the range 20800 to 29500, On an average there is more than 23% chance that a loan can be charged off among the loans applied for this amount.



Loan Amount vs Charged Off %
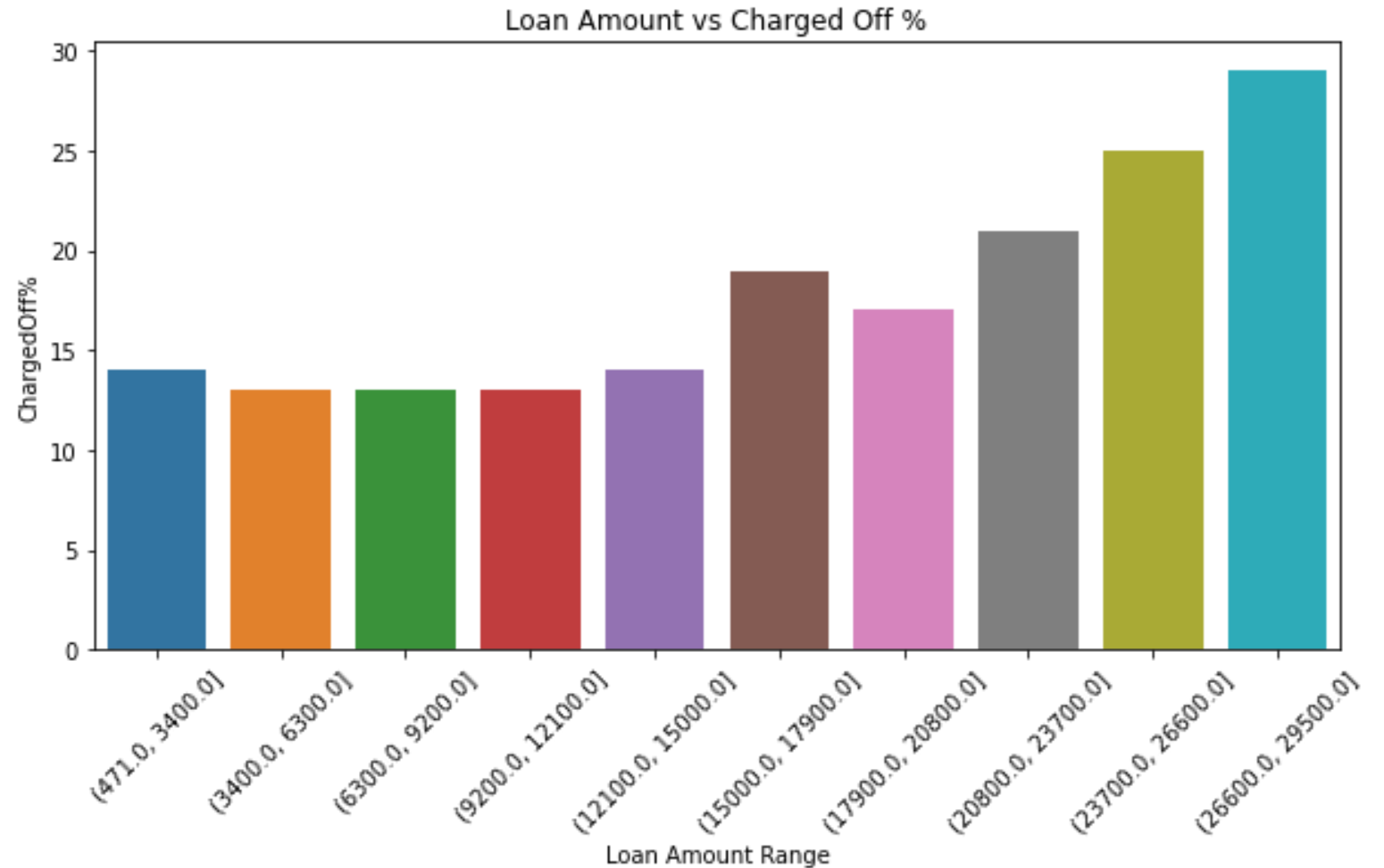
# Funded Amount

1. There is a 27% chance of charge off if the funded amount is in the range 24890 to 27600

2. Most number of charged off loans are funded between the range 900 to 15000.

3. Although less number of loans are taken, there is a greater chance of charge off for amounts above 16760
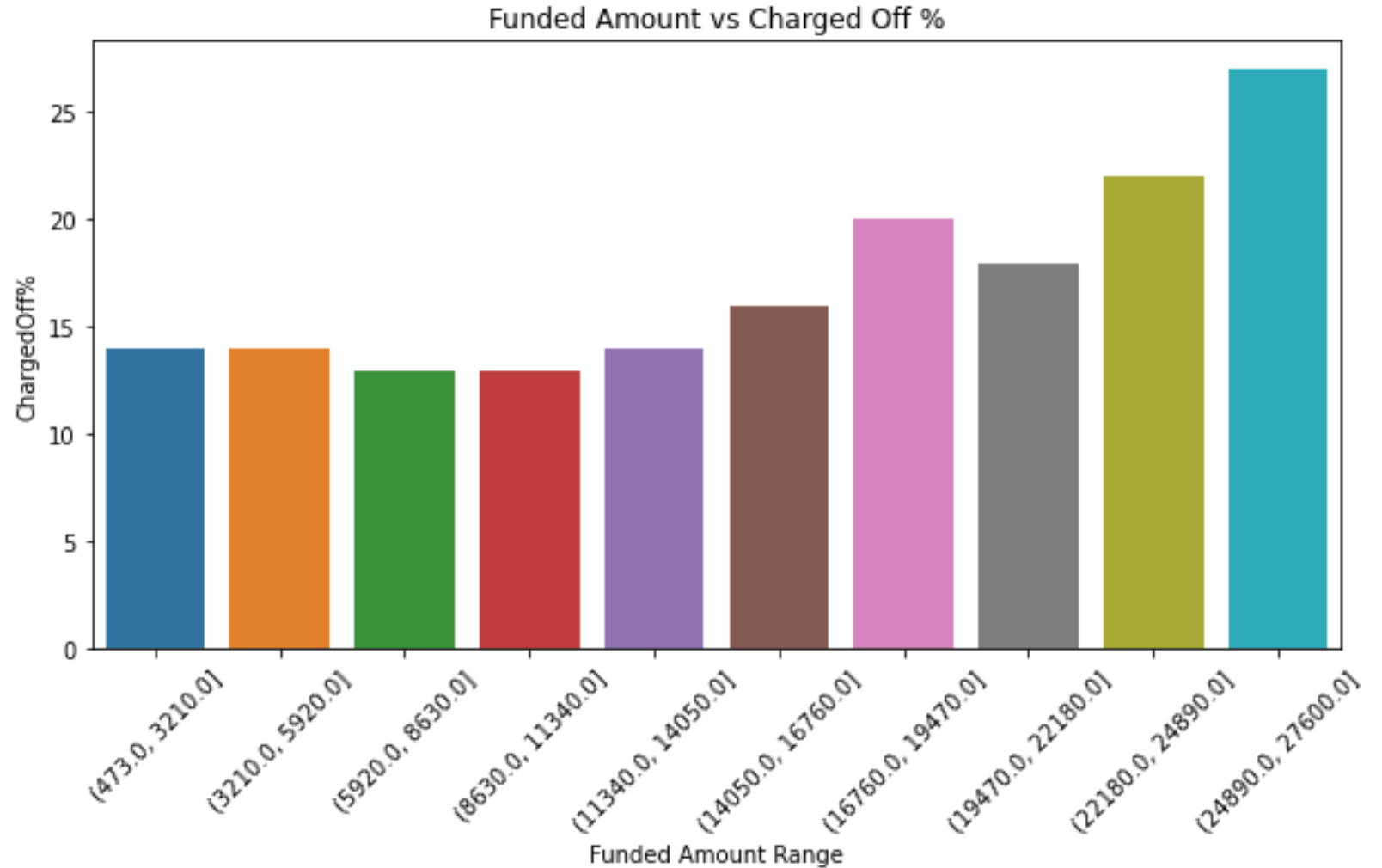
# Term

There is a 25% chance that a loan can be charged off for a term of 60 months whereas there is a 11% chance of charge off for a 36 month term.



Term vs Charged Off %
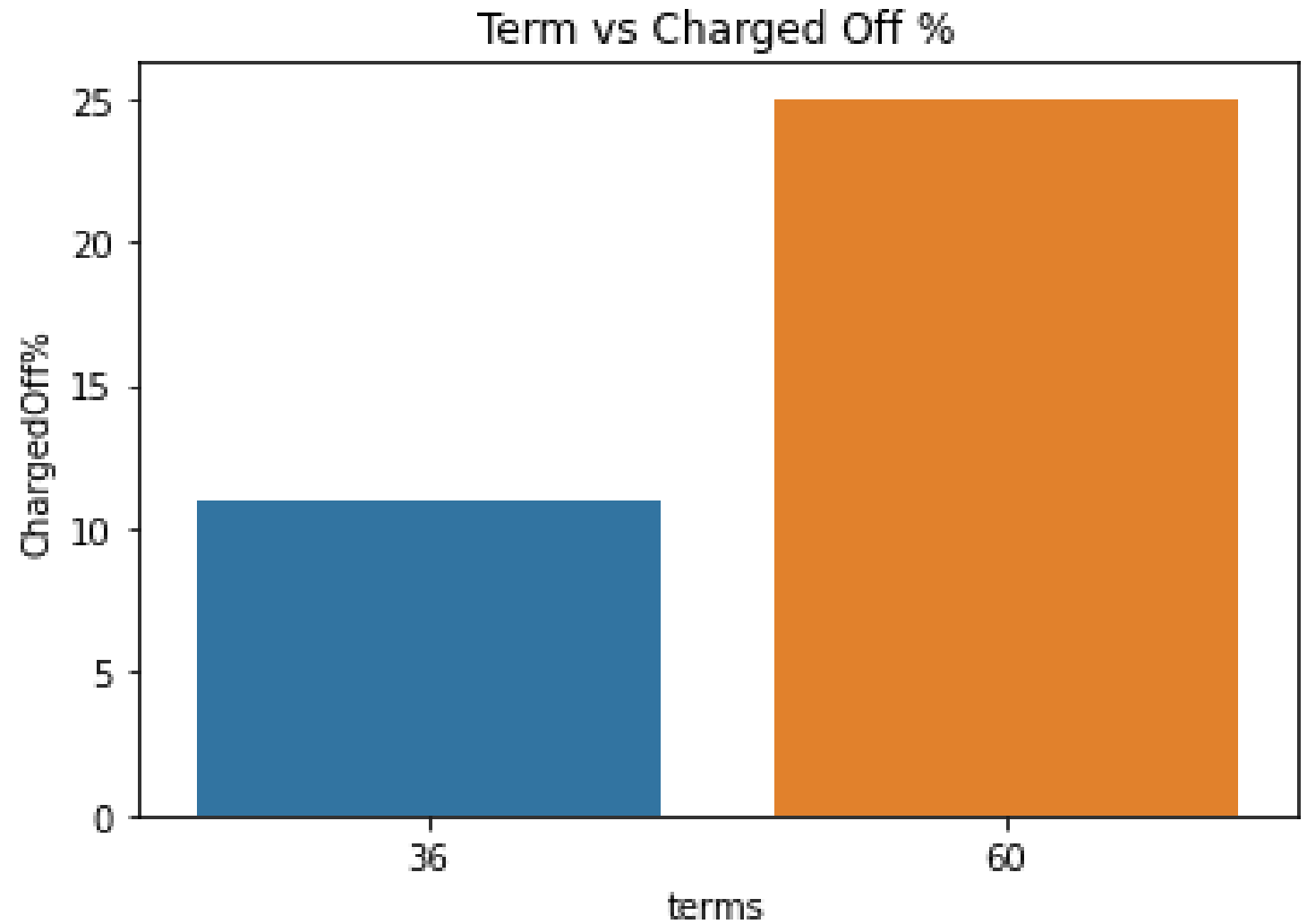
# Interest Rate

1. Most number of charged off loans are taken at a interest rate between 11.11 and 16.00

2. There is a 39% chance of charge off if the interest rate is between 19.3 and 22.1

3. Although less number of loans are taken at a interest rate above 19.3, there is a greater chance of charge off in this case.

4. Median of interest rate in Charged off loans is 13.5% which is way above than that of interest in fully paid loans(11%)



Interest Rate vs Charged Off %
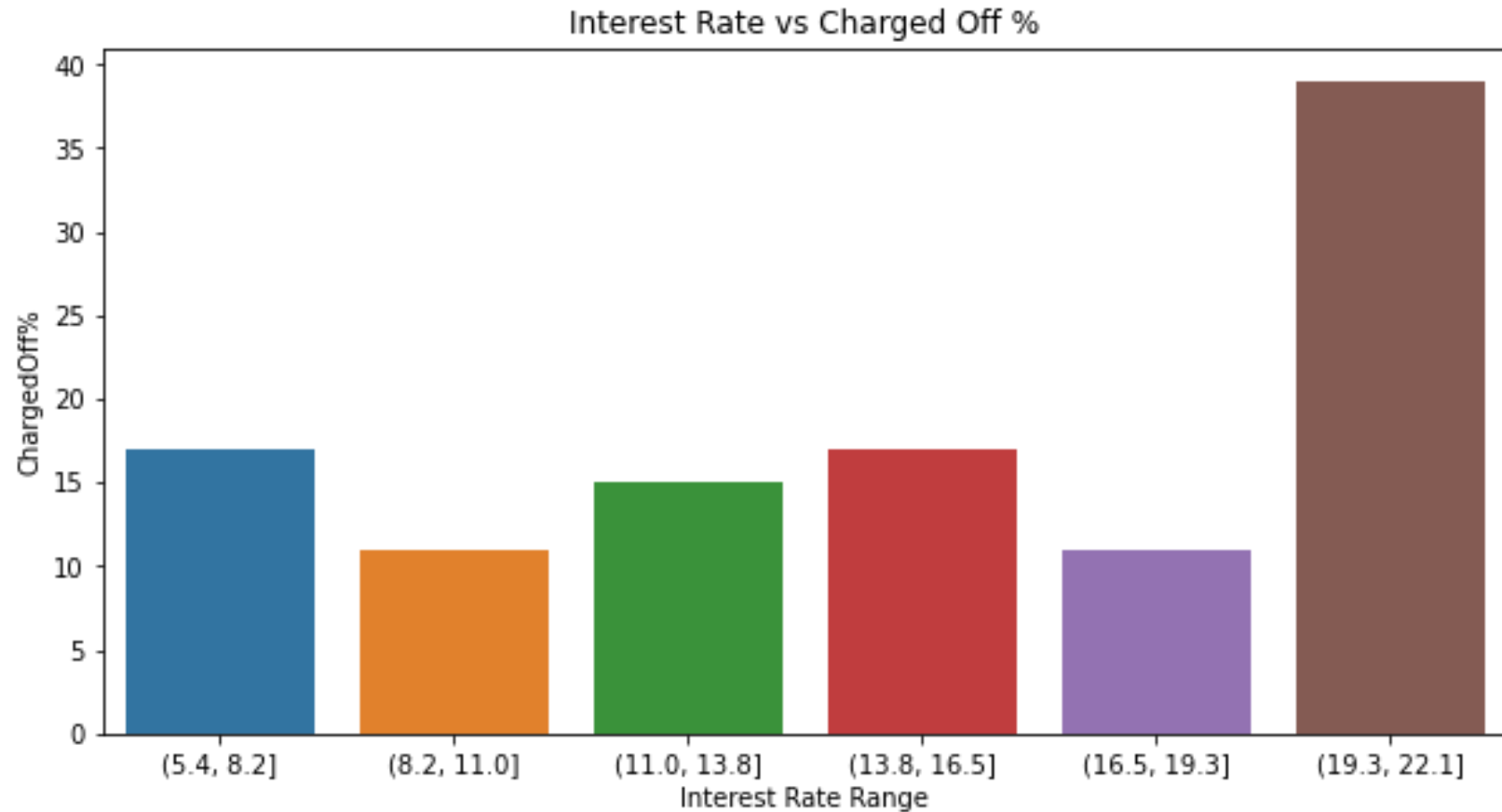
# Grade

1. A borrower with a grade G has the highest chance to go default.

2. The risk can be inferred according to the grade as **G>F>E>D>(C=A)>B,** Borrower with grade 'G' having the highest chance to go default and borrower of grade 'B' with relatively lower chances.



Grades vs Charged Off %
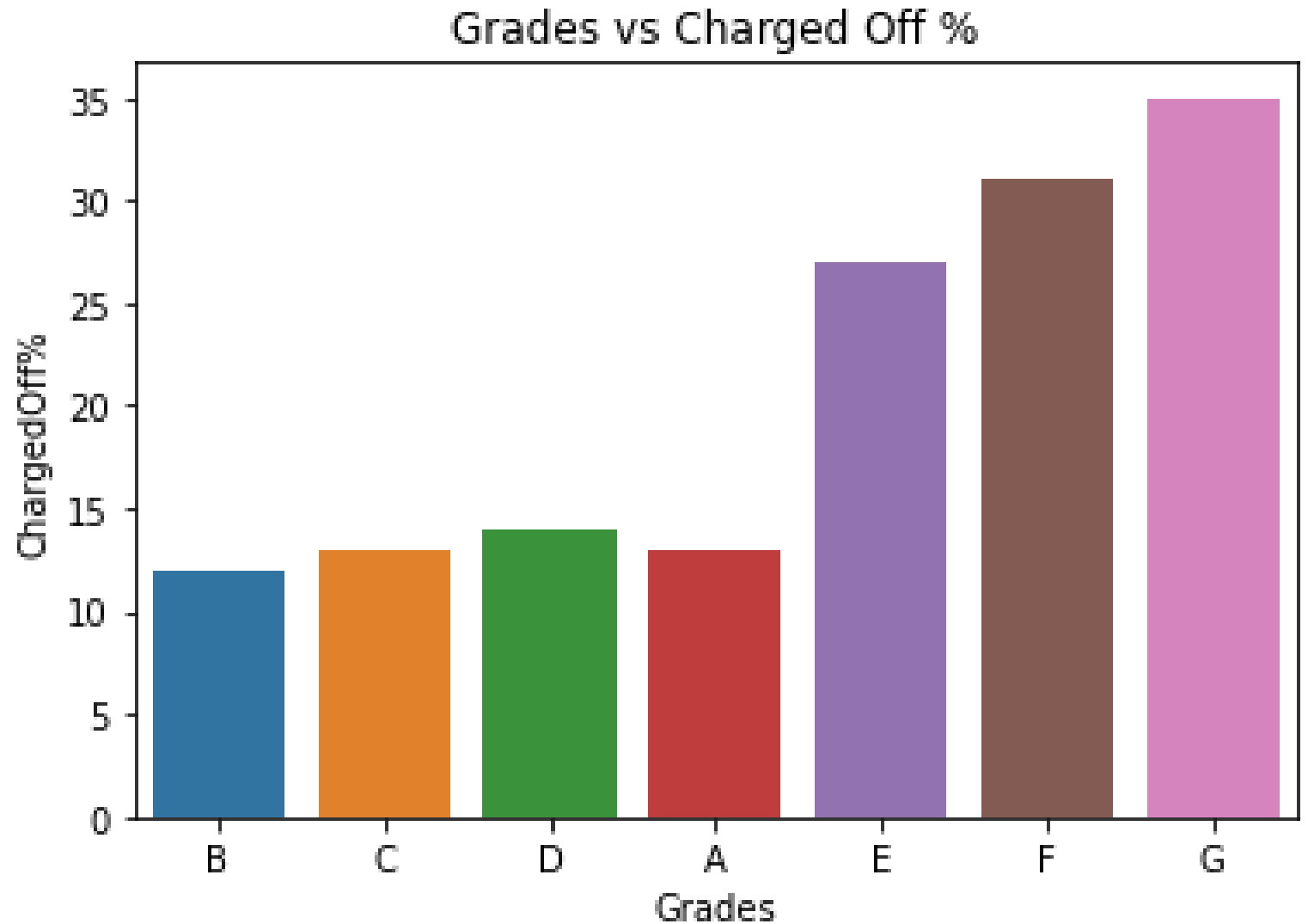
# Employment Length

Borrowers with employment length either above 10 years or <1 year have the highest chance to be charged off.
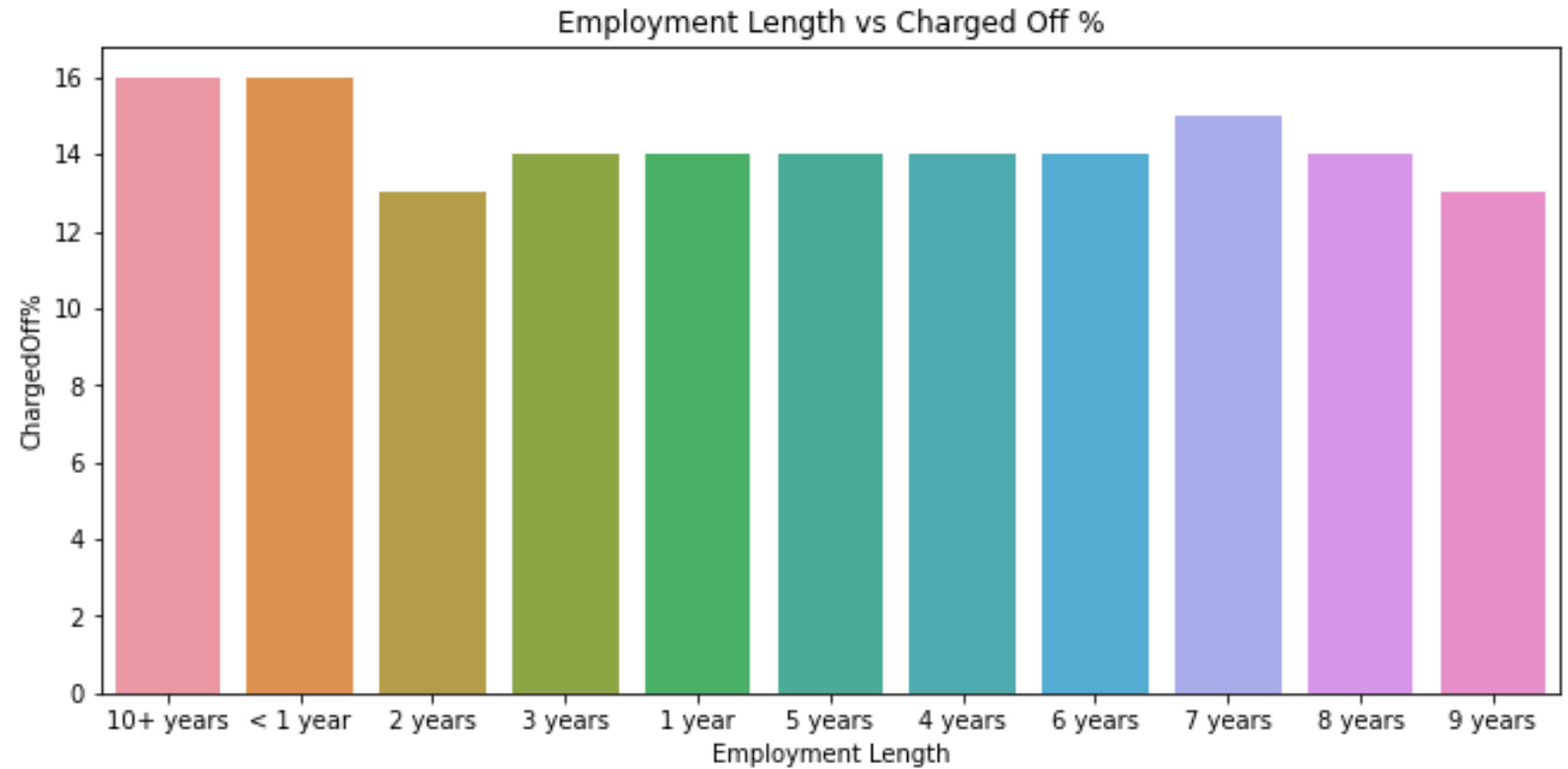
# Annual Income

1. Most number of charged off loans are applied by the borrowers with annual income between 36000 and 67074

2. There is a high chance of charge off if the borrower's income is below 30699

3. There is a low chance of charge off if the borrowers income is above 84100

4. The median of charged off borrower's income falls in the range 30699 and 57398, which has a bit higher chance of charge off.



Annual Income vs Charged Off %
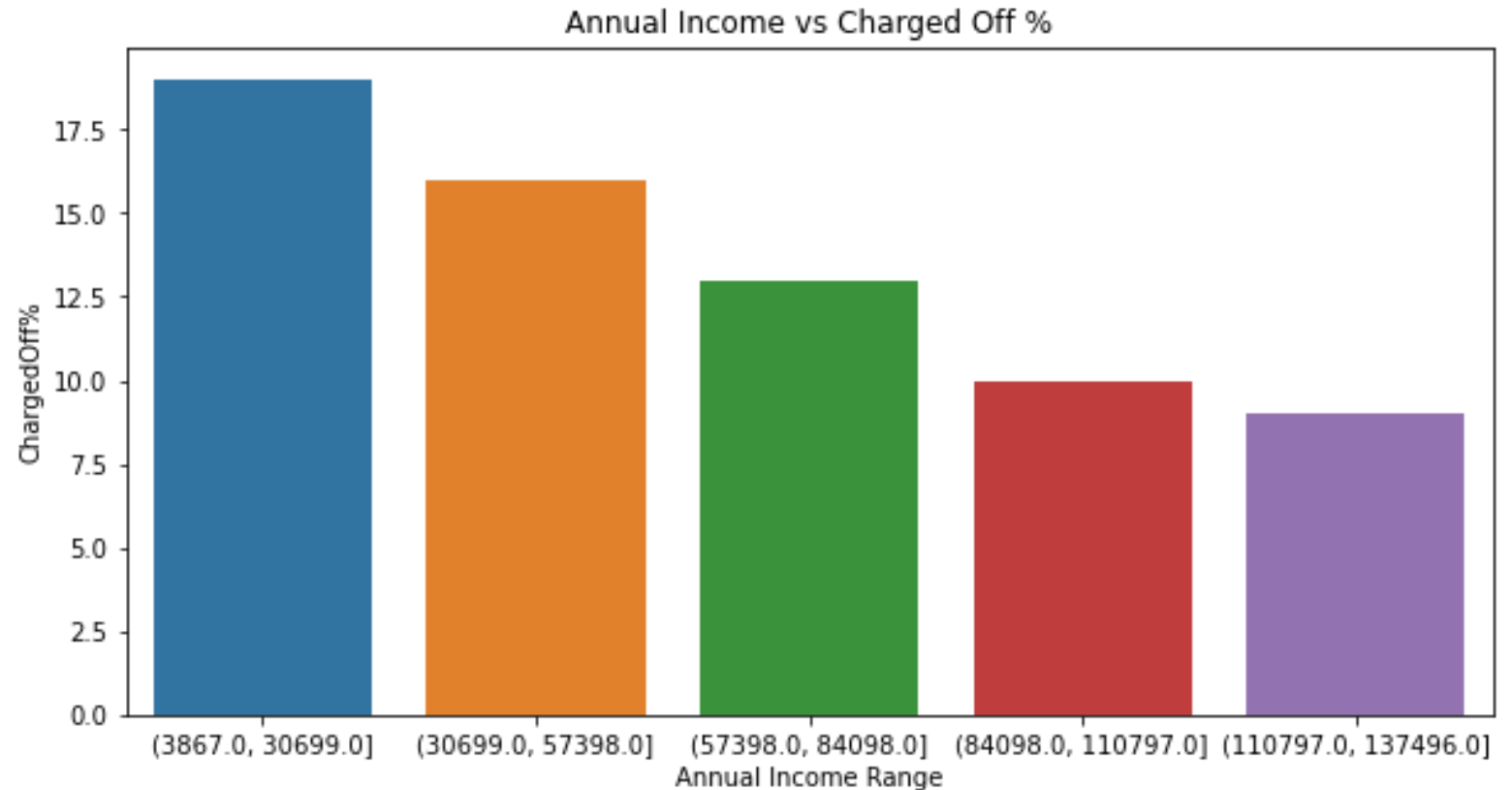
# Loan Issued Year

1. The Number of charged-off loans keeps increasing each year as the number of loans issued also increased.

2. But there were more of loans that were charged off in 2007 and 2008 as well, along with 2011.

3. There were relatively lesser charged off loans issued in the years 2009 and 2010.



Issued Year vs Charged Off %
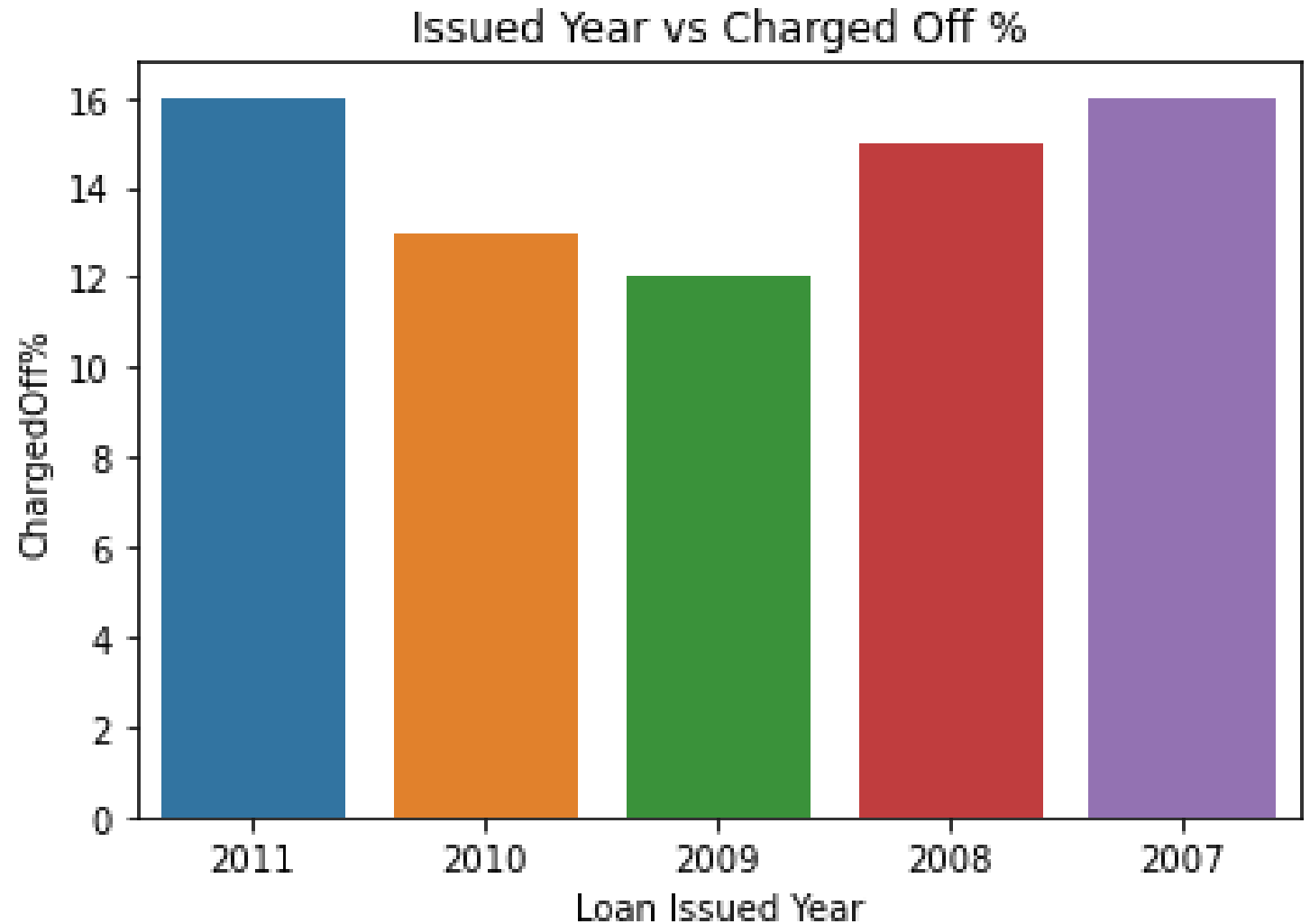
# Loan Issued Month

1. Highest loan defaults occured for the loans issued in the months of May, October and December.

2. More density of charged off loans can be observed in the last quarter of the year.

3. Lowest loan defaults are recorded for the loans issued in the month of February.



Loan issued month vs Charged Off %
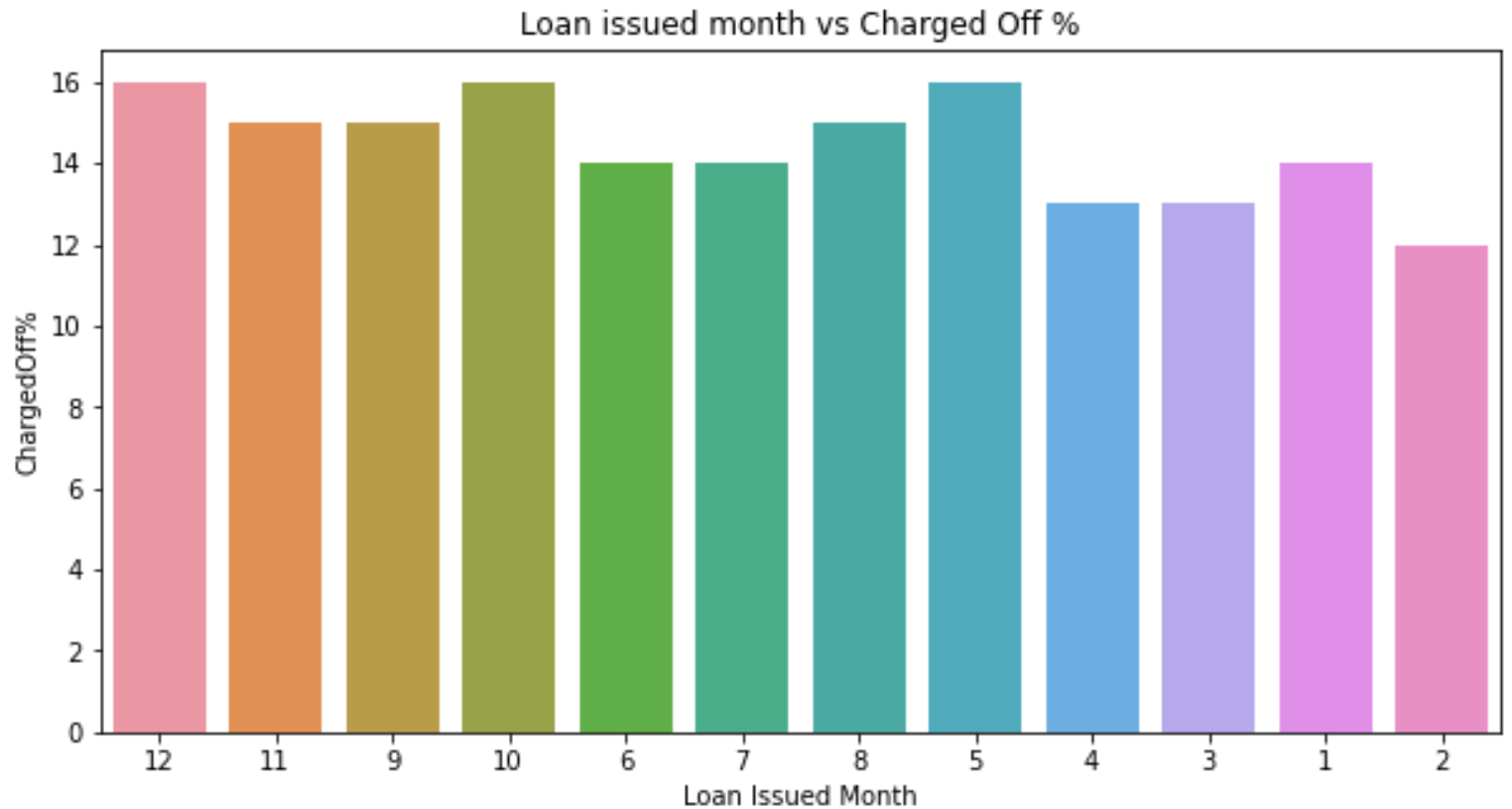
# Loan Purpose

1. Majority of the loans taken for the purpose of 'Renewable Energy' are charged off.

2. The lowest chance of defaulters were having the loan with 'car' or 'medical' purpose.



Purpose of loan vs chargedOff%
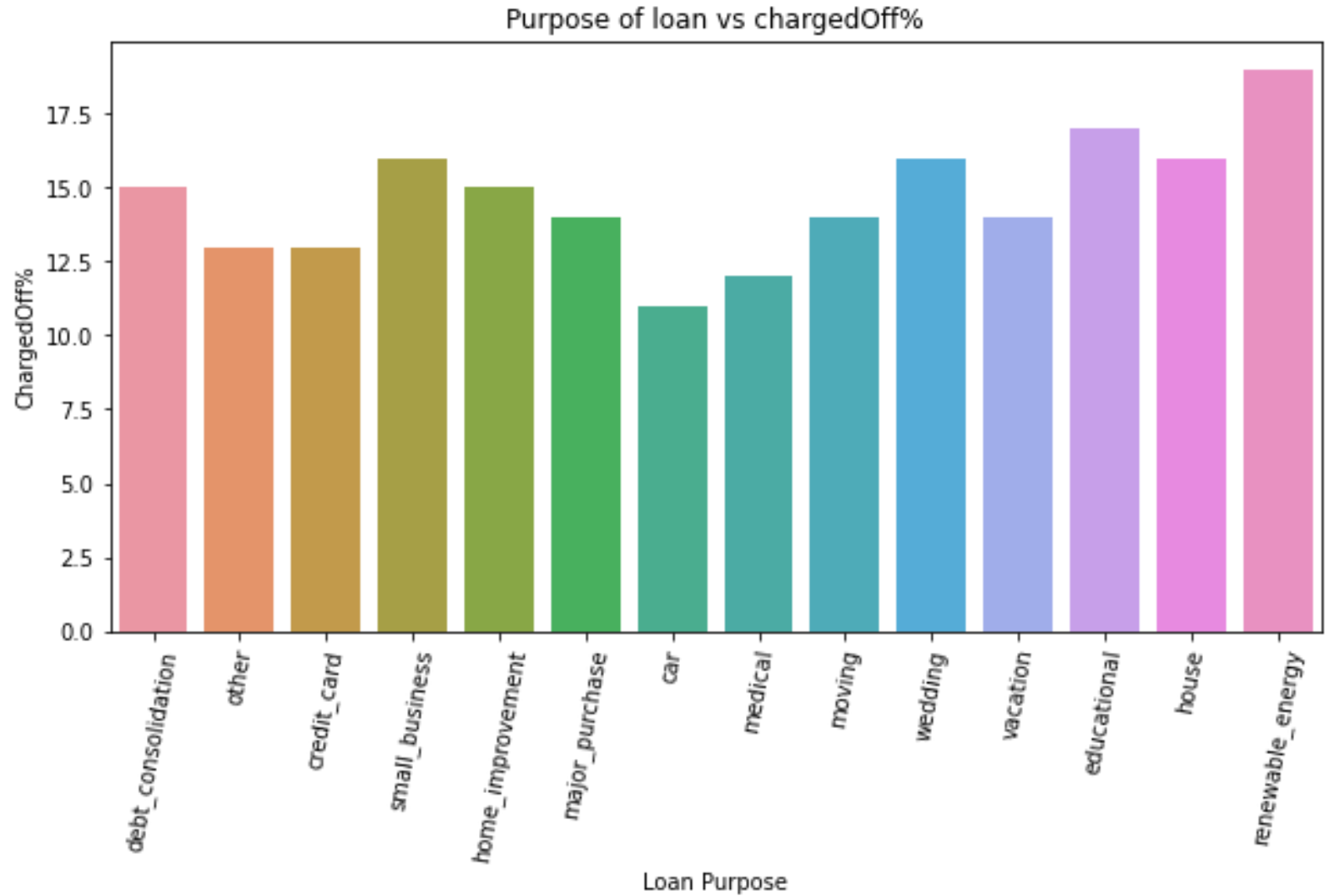
# Borrower State

1. Tennessee is the state with Highest possible chance for a loan defaulter.

2. Idaho, Minnesota are also the states with a bit more chance for defaulters.

3. Though California has more records of defaulters, the chance for loan defaults is above the average.



Borrower State vs Charged Off %
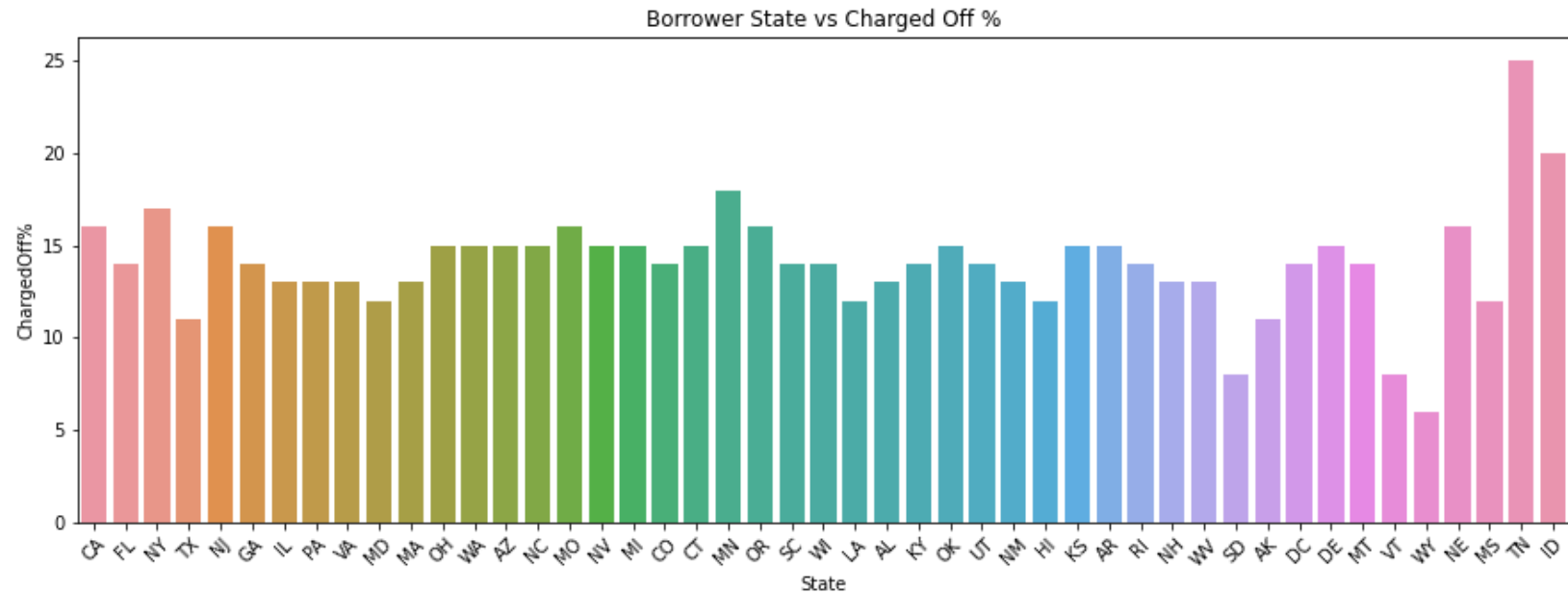
# Number of inquiries

1. As the number of inquiries increases, the chance of charge off also increases, except for 4-5 inquiries, where there was a small dip.

2. If the inquiry requests made in the last 6 months are greater than 6, such loans of borrowers have near to 30% chance of being charged off.



Number of Inquiries vs Charged Off %
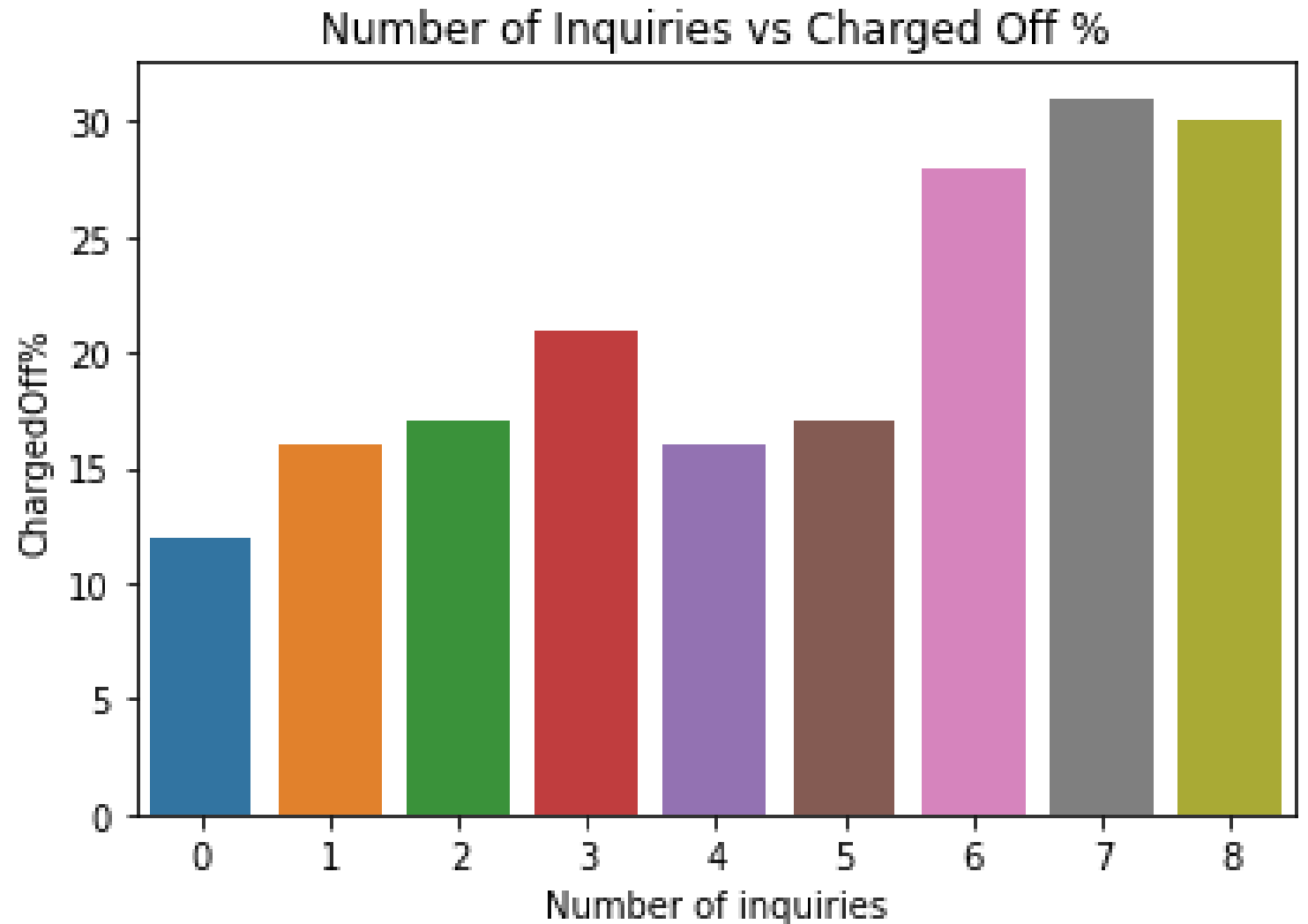
# Derogatory public records

Borrowers with atleast one derogatory public record have a 22% chance of defaulting.

Whereas the chance of a borrower with no derogatory public records is 14%
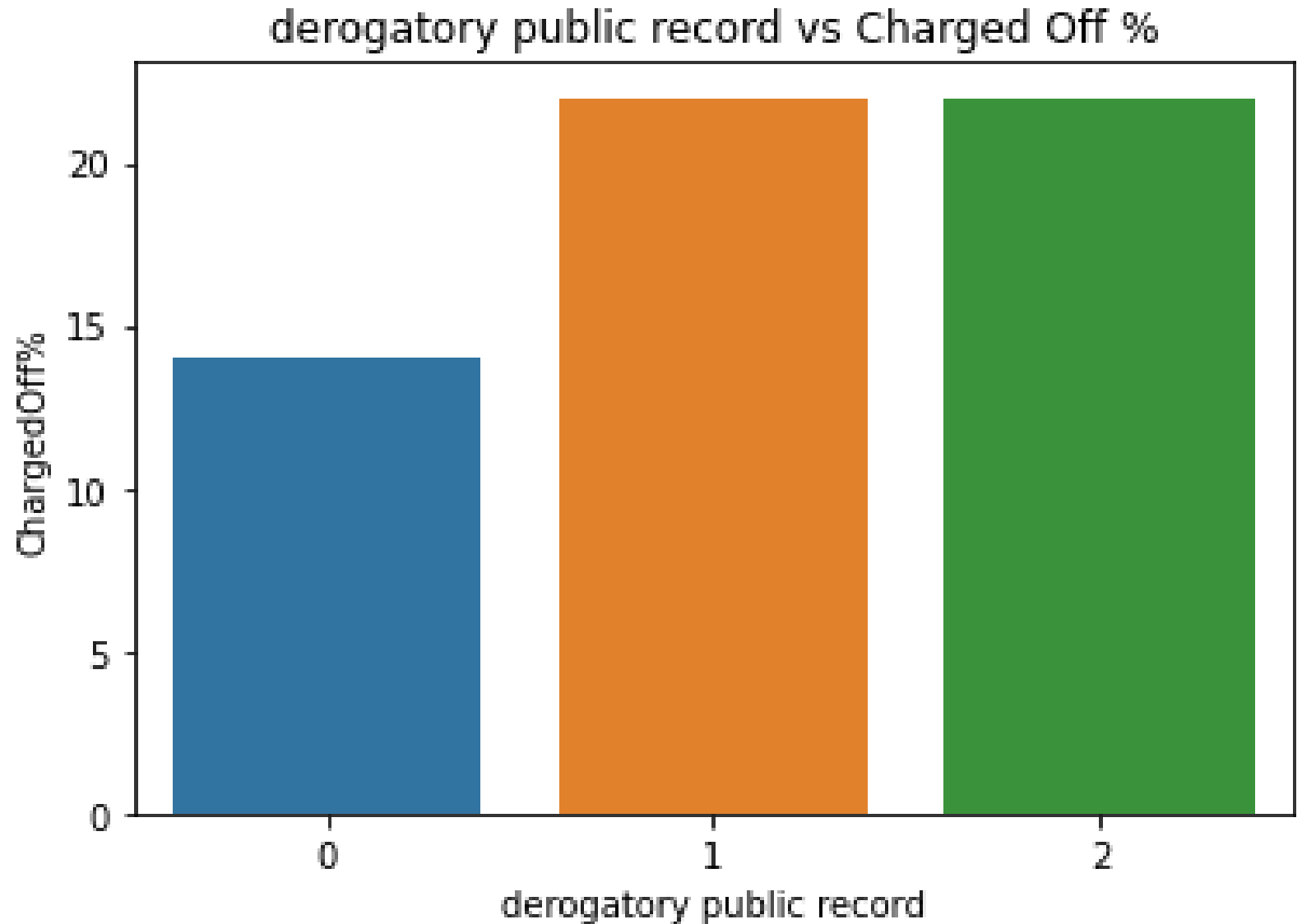
# Public record bankruptcies

Borrowers with 2 public record bankruptcies pose a risk of 40% chance of being defaulted.

Whereas the borrowers with one public record bankruptcy have a 22% chance of their lone being charged off.
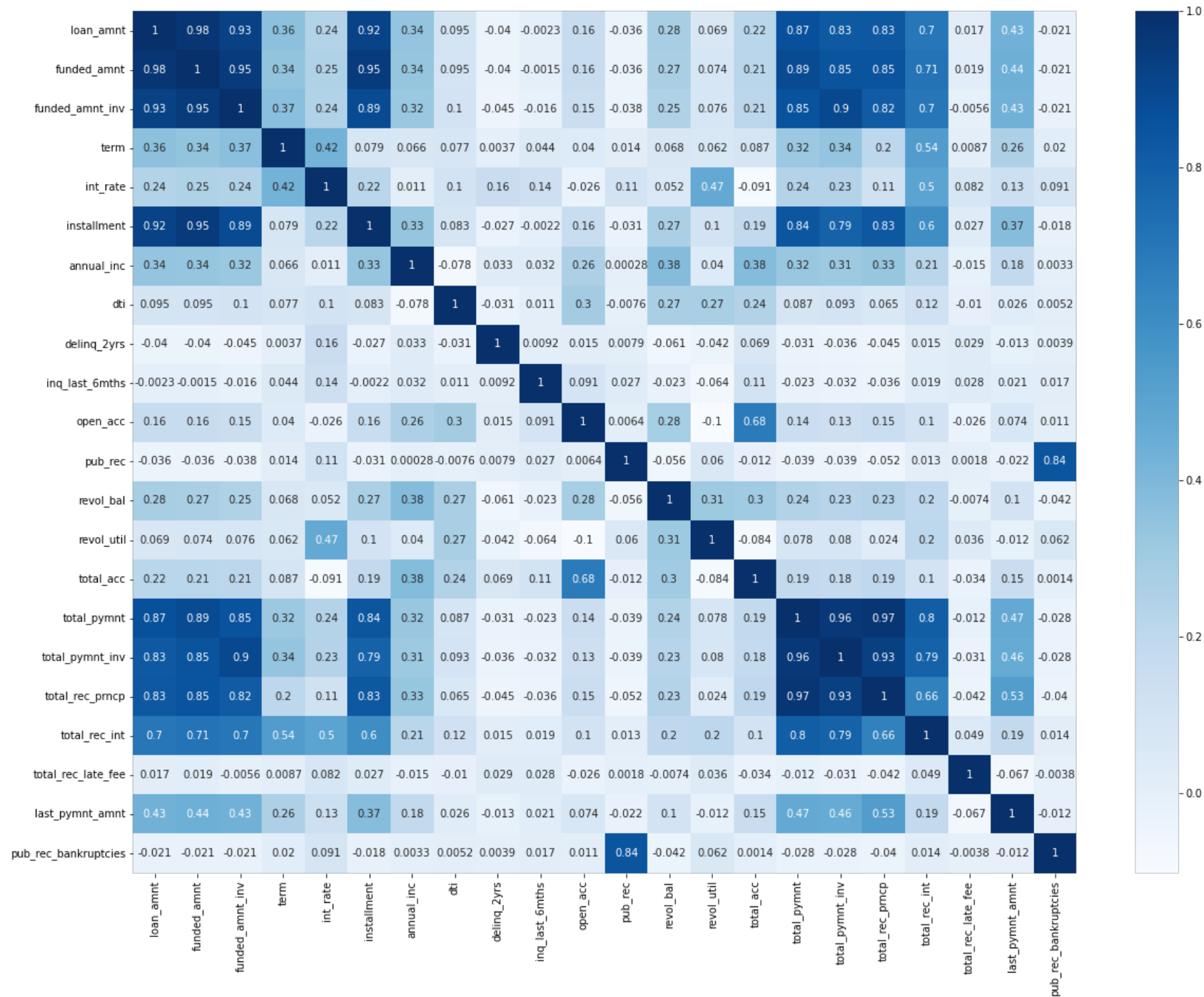


public record bankruptcies vs Charged Off %
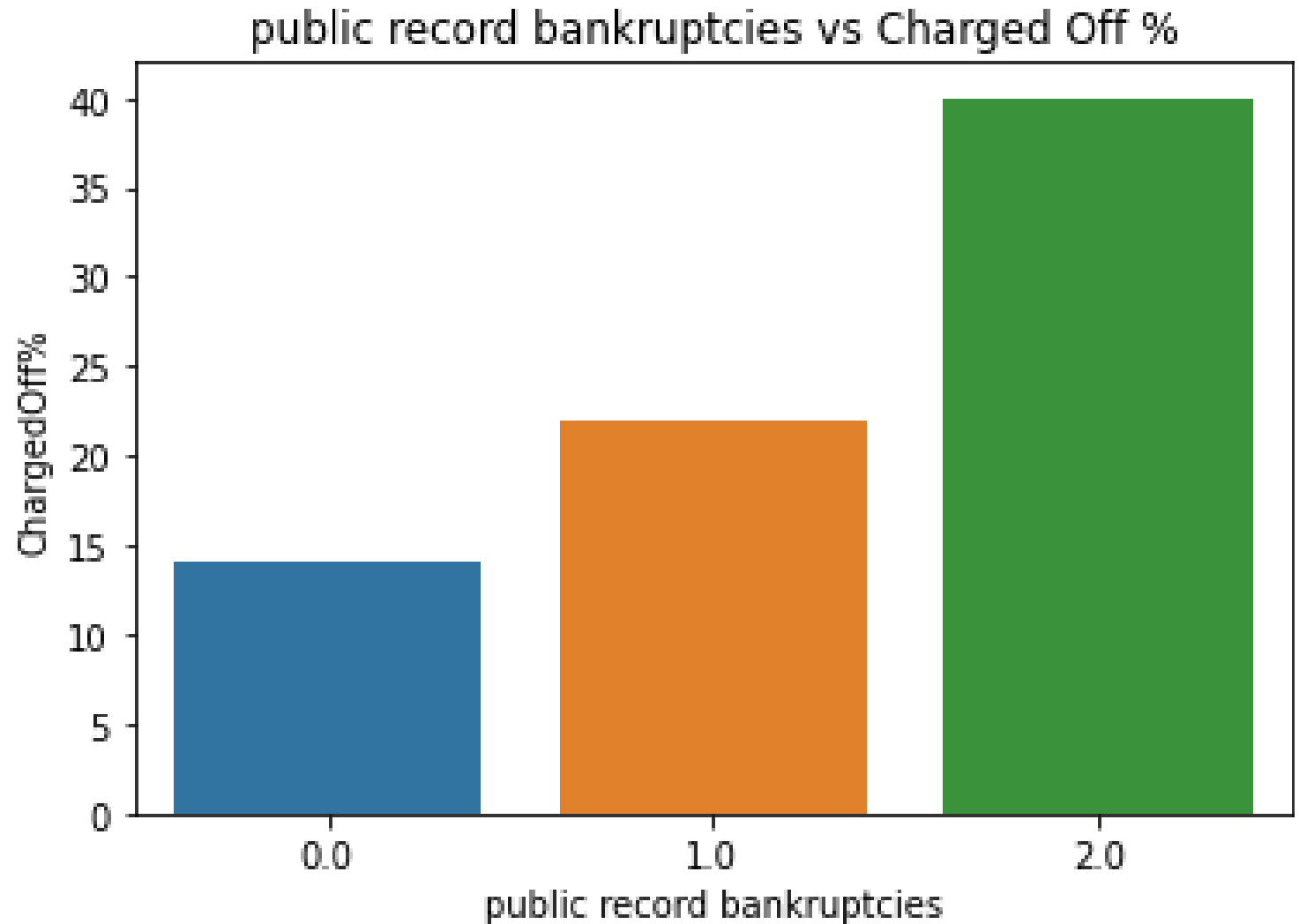
# Analysis:

BIVARIATE ANALYSIS

# Heat map

We can observe that total payment, installment follow the same trend as loan amount, which is obvious. As loan amount increases, total amount to be paid, installments etc., also go along the lane.
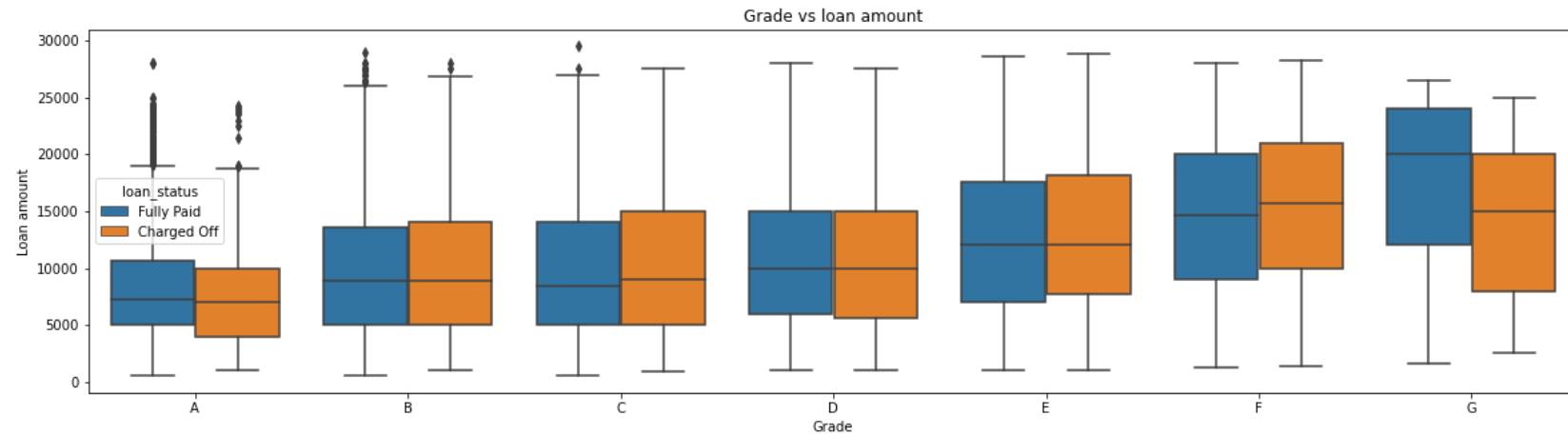
# Loan Amount
# vs
# Interest Rate

With increasing loan amount, interest rate also increased. Around loan amount of 15000, the interest rate varied greatly from 11 to 16.5, and this is where defaulters were more compared to fully paid borrowers.

Borrowers with high interest rate in their seems to be charged off in every loan ammount range.
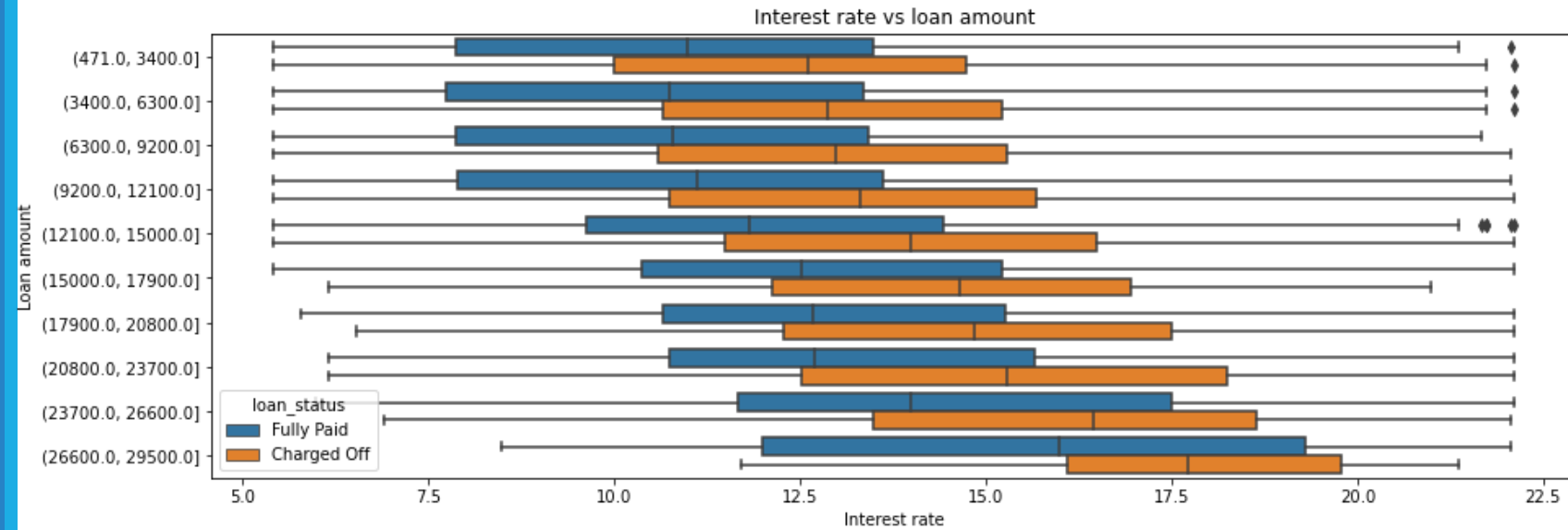
# Grade
# vs
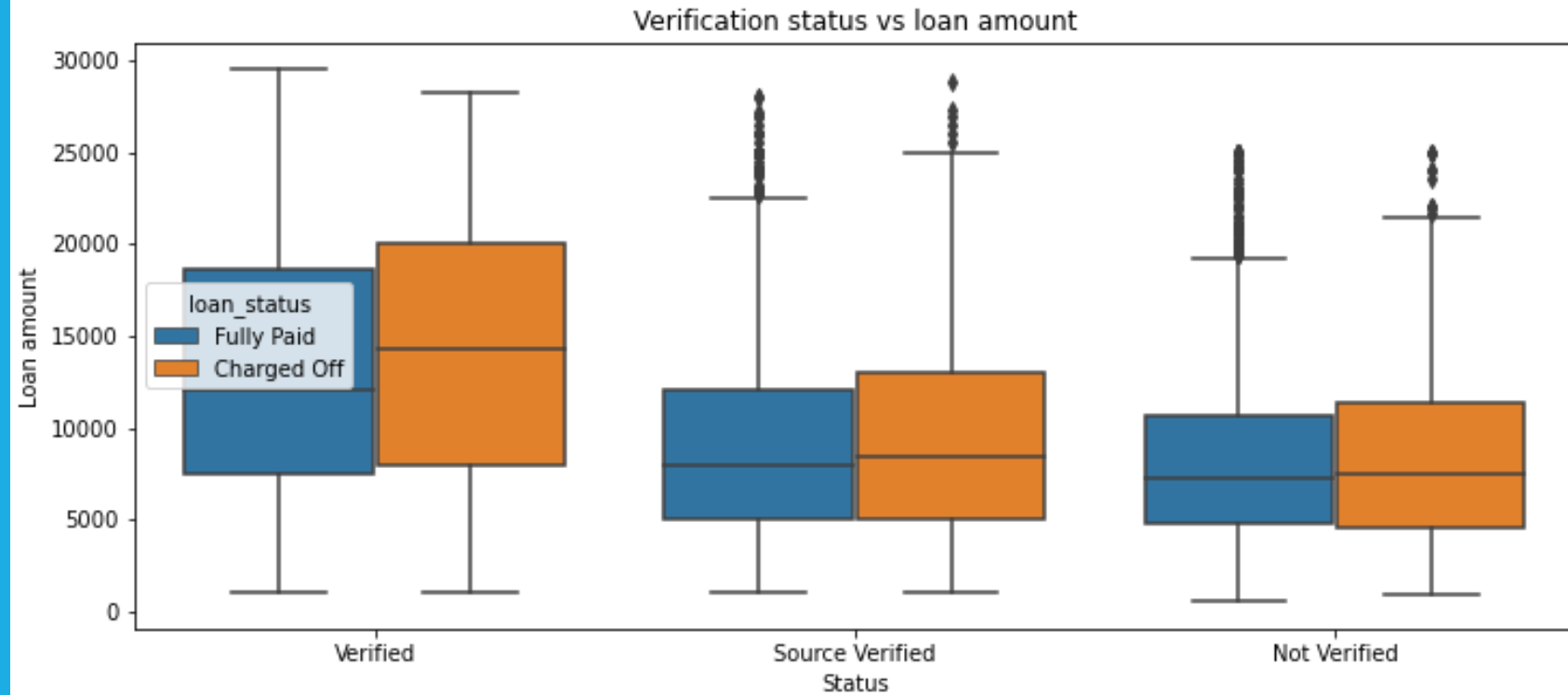# loan Amount

A borrower with grade 'G' and loan amount below 20000 have the most tendency to go default.
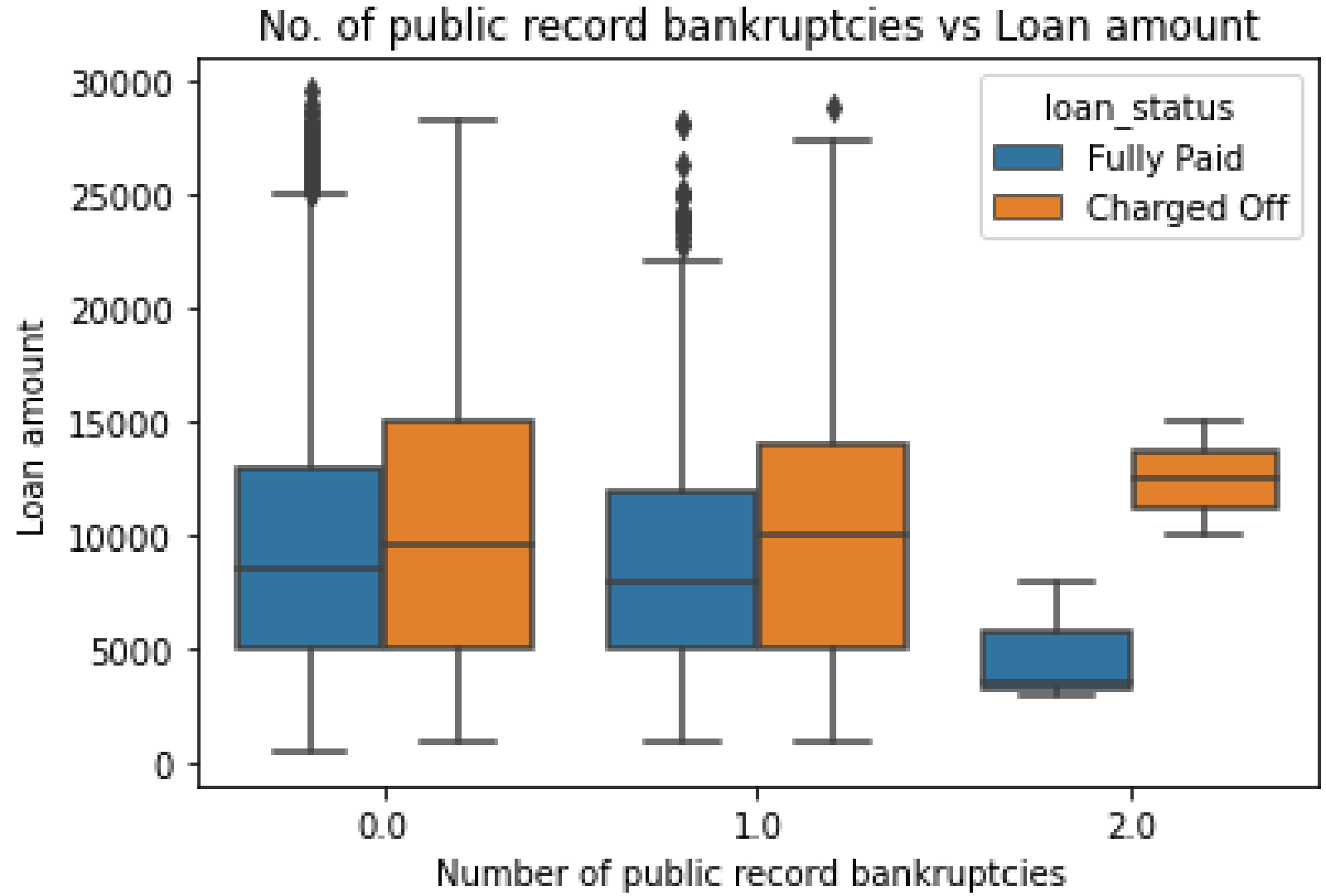


Grade vs loan amount

# Verification Status vs Loan amount

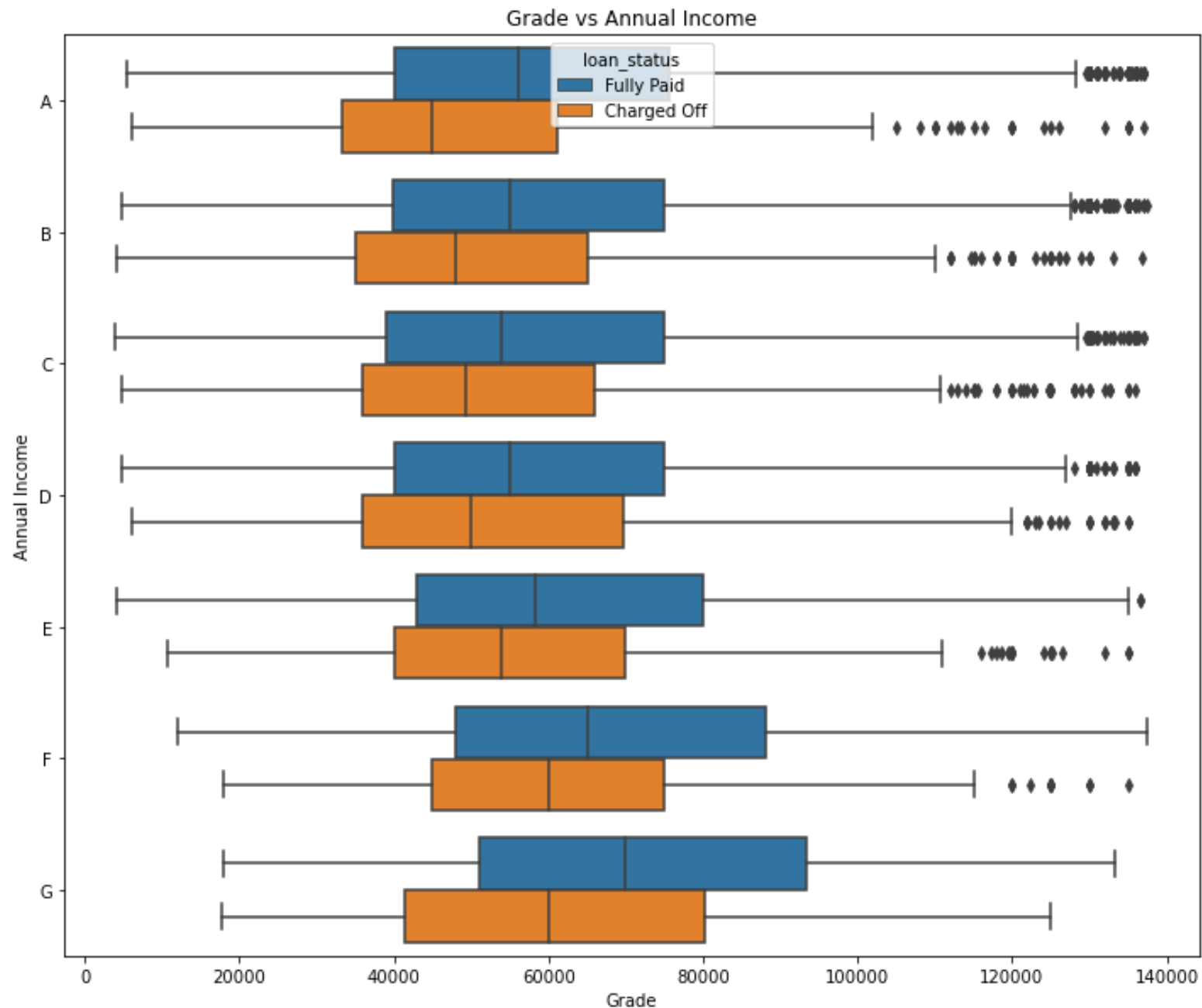Defaulters with verified income status were given loans with higher amounts

# Public record bankruptcies vs Loan amount

A loan amount greater than 10000 given to a borrower with 2 public record bankruptcies is going to be charged off
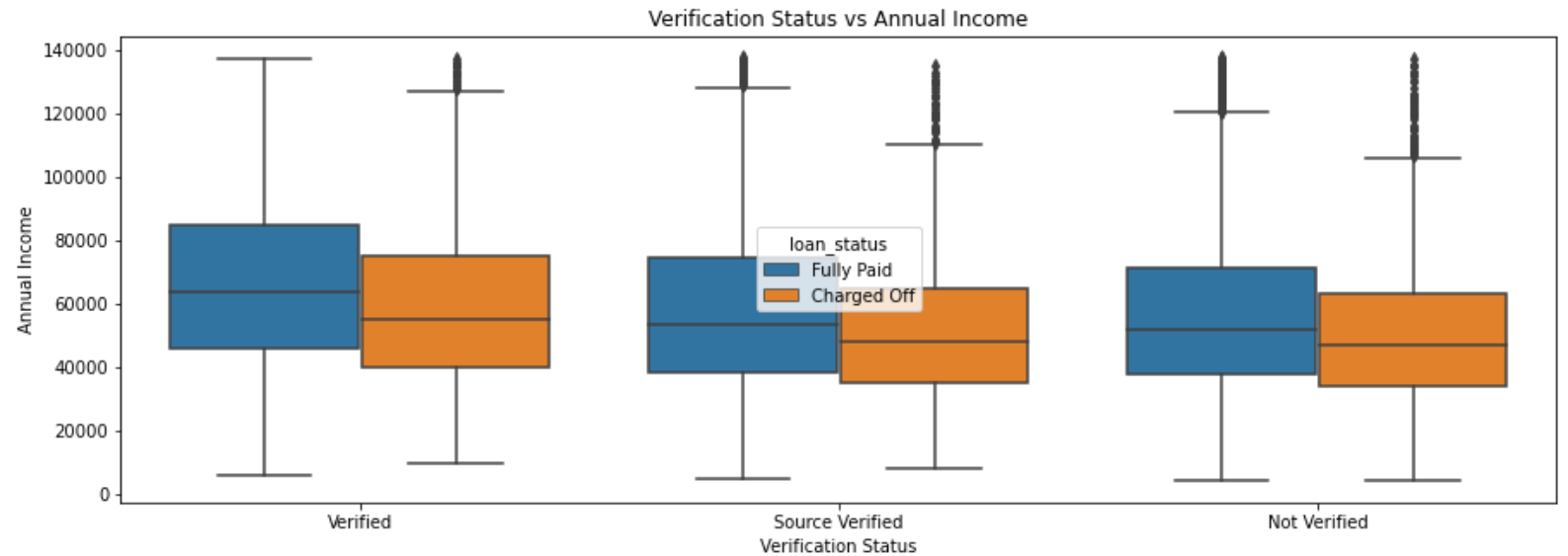
# Grade vs Annual Income

Borrowers with Grade 'G' have relatively higher annual income and also higher chance of charge off

# Verification Status vs Annual Income

All the verification status have almost the same income range, cannot interpret much from here.



Verification Status vs Annual Income

# Derogatory public record vs Annual Income

Borrowers with annual income between 70000 and 90000 with 2 derogatory public records are highly likely to be charged off.



No. of derogatory public record vs Annual Income
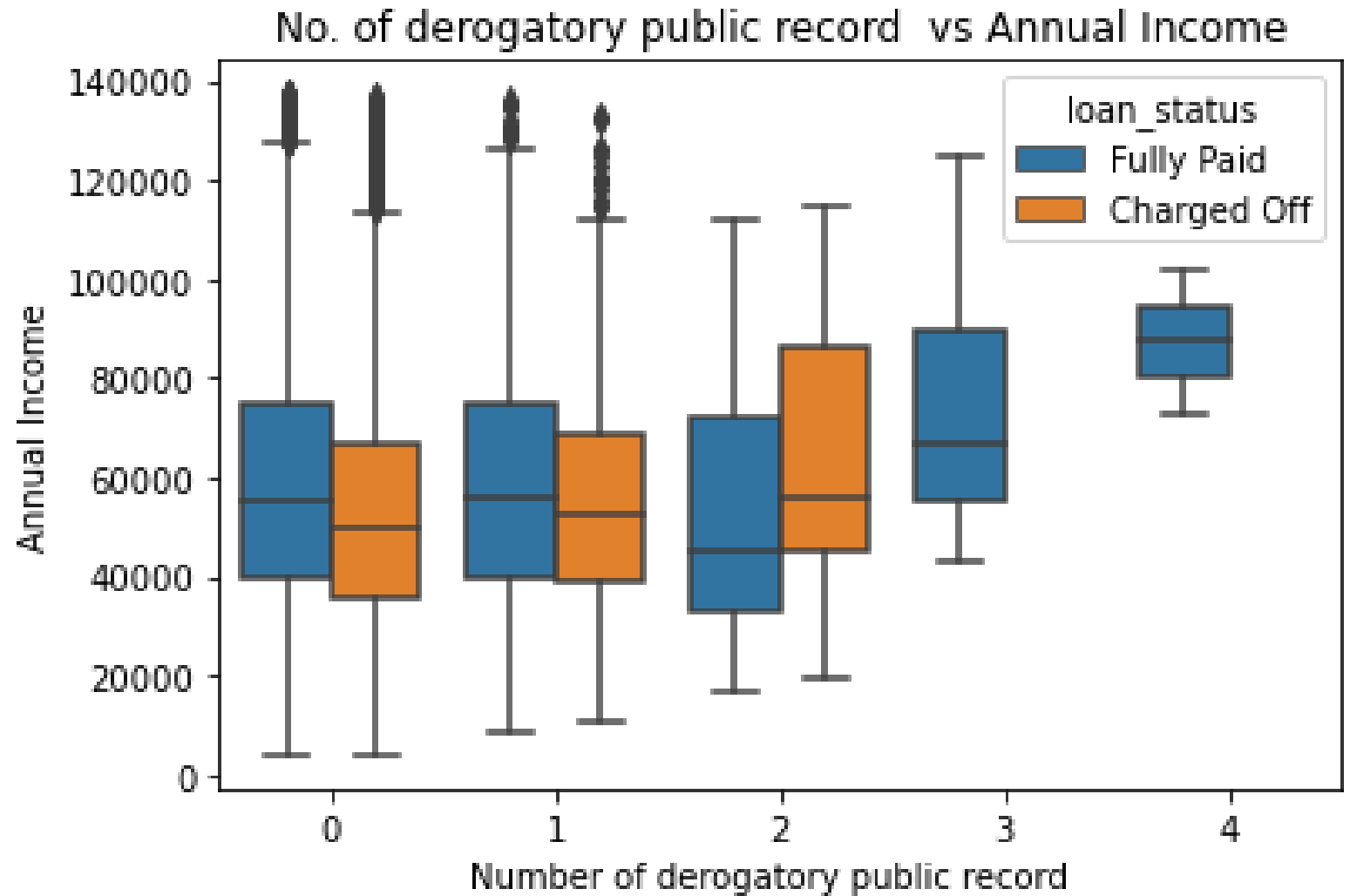
# Annual income vs Loan amount

Defaulters with annual income between 40000 to 60000 have applied for higher loan amount.



Annual income vs Loan Amount
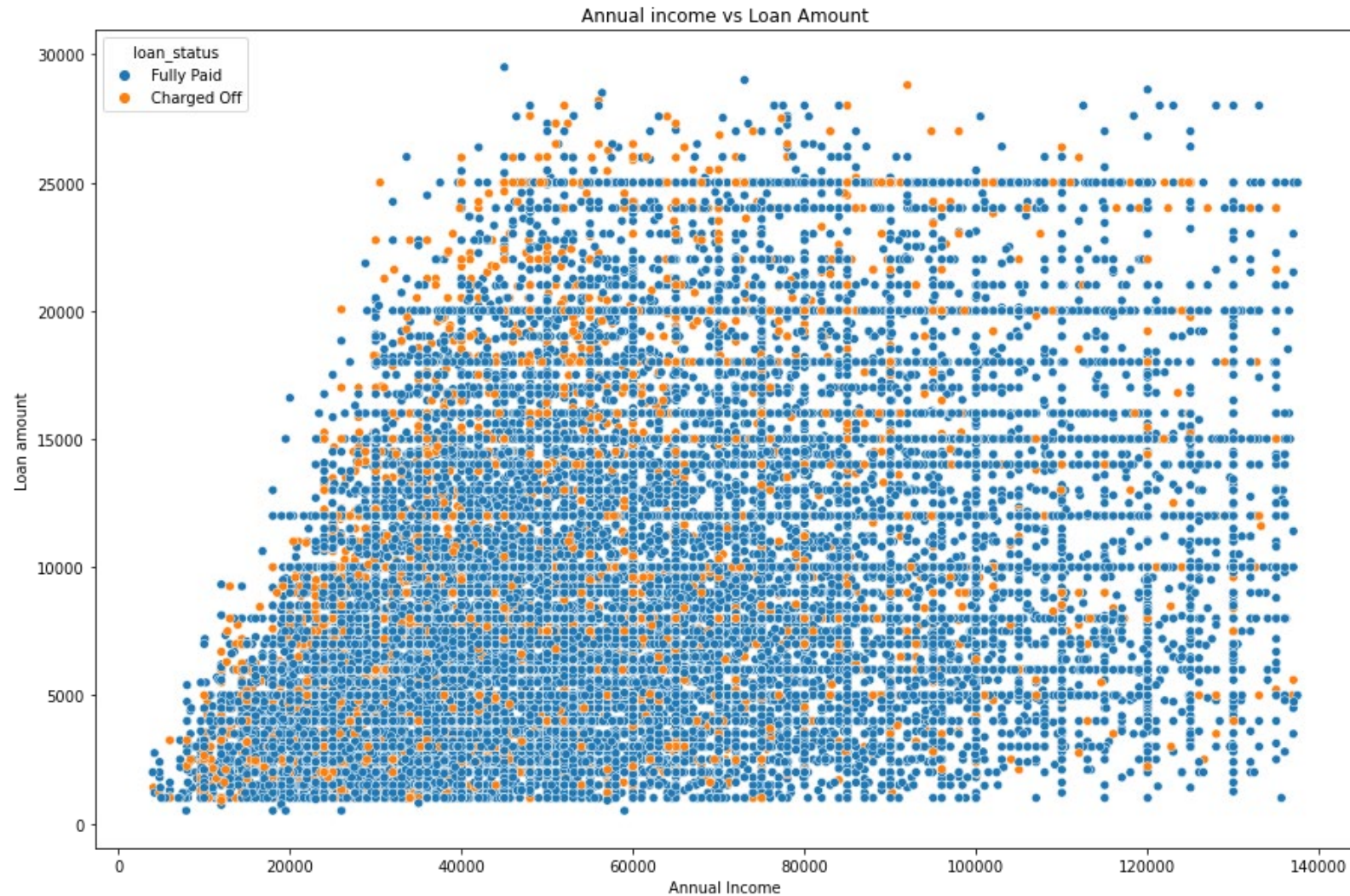
# Conclusion:

There is a high chance of default when:

1. Borrower has applied loan amount between 26600 and 29500 or has annual income between 40000 to 60000 have applied for higher loan amount.

2. Loan Is funded between 900 to 15000.

3. Has taken a 60 month term loan.

4. Loan taken with Interest rate between 19.3 and 22.1 in general or loan taken below 15000 between interest rate of 11% to 16%.

5. Has a grade G with a higher annual income or loan amount below 20000, especially G5.

6. Has a employment length less than a year or greater than ten years or with a median annual income between 50000 to 60000.

7. With Home Ownership type 'OTHER' with a annual income of 40000 to 70000

# Conclusion:

There is a high chance of default when:

8. Has a annual income below 30699 and the income status verified

9. Got the loan issued towards the last quarter of the year

10. Loan taken on any purpose with high interest rate tends to be defaulted. But loans on 'Renewable Energy' are charged off even though they were given at a lower interest rate.

11. Resident of Tennessee state with loan amount greater than 7000 or with have annual income above 60000.

12. Has a dti ratio greater than 20

13. More than 6 inquiries made in the last 6 months

14. With at least one derogatory public record or with an annual income between 70000 and 90000 with 2 derogatory public records.

15. With 2 public record bankruptcies with loan amount greater than 10000.