Lending Club Case Study

Sai Mohan Reddy Neelam

Contents

Introduction
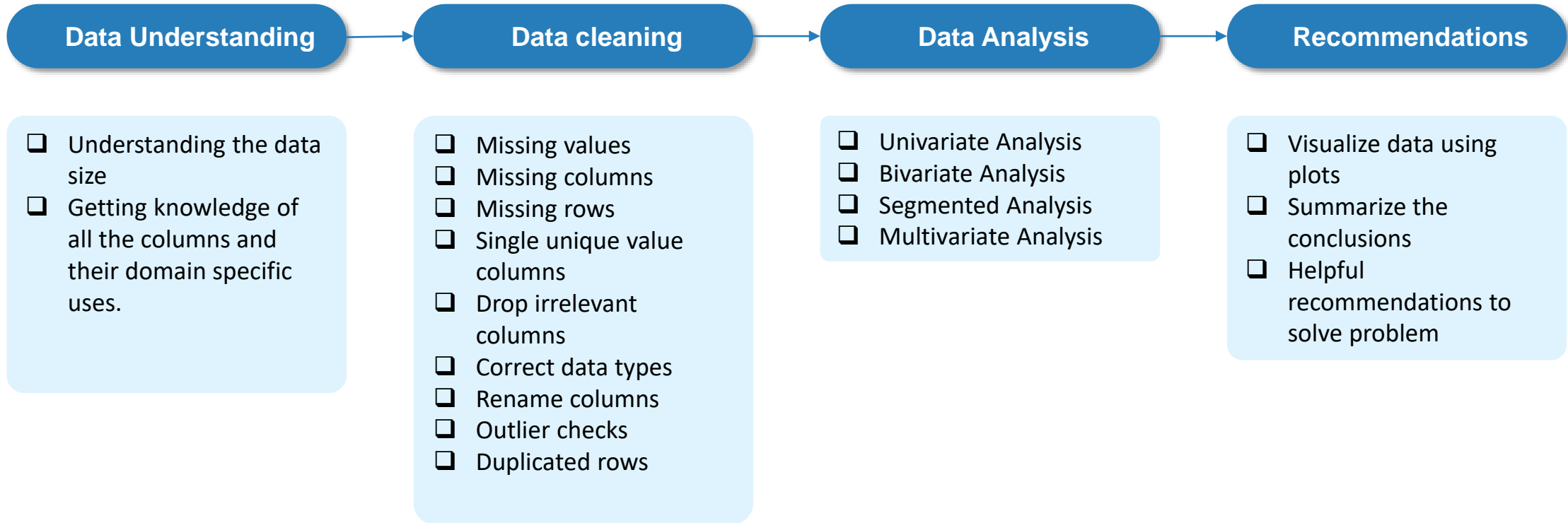
- ❑ Background
- ❑ Problem solving process

Background

❑ This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

❑ Borrowers can easily access lower interest rate loans through a fast online interface.

❑ 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.

LOAN DATASET
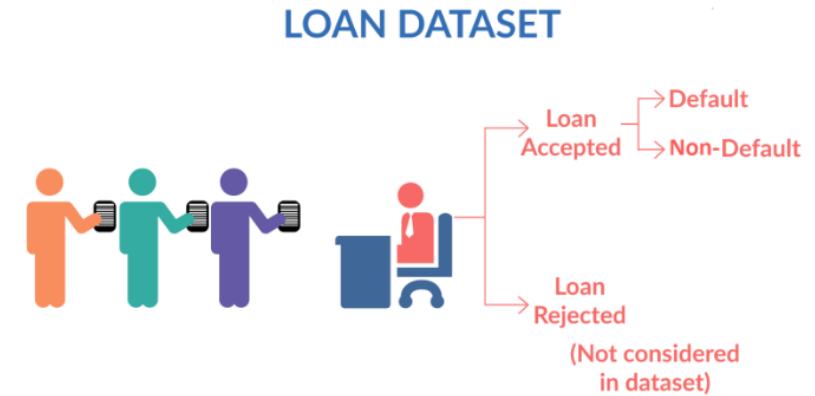
❑ When the company receives a loan application, the company must decide for loan approval based on the applicant's profile.

❑ The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as:

    ➢ denying the loan

    ➢ reducing the amount of loan,

    ➢ Lending (to risky applicants) at a higher interest rate

Objective: Understanding the driving factors behind loan default

Problem solving Process

Data Understanding → **Data cleaning** → **Data Analysis** → **Recommendations**

Data Understanding
- ❑ Understanding the data size
- ❑ Getting knowledge of all the columns and their domain specific uses.

Data cleaning
- ❑ Missing values
- ❑ Missing columns
- ❑ Missing rows
- ❑ Single unique value columns
- ❑ Drop irrelevant columns
- ❑ Correct data types
- ❑ Rename columns
- ❑ Outlier checks
- ❑ Duplicated rows

Data Analysis
- ❑ Univariate Analysis
- ❑ Bivariate Analysis
- ❑ Segmented Analysis
- ❑ Multivariate Analysis

Recommendations
- ❑ Visualize data using plots
- ❑ Summarize the conclusions
- ❑ Helpful recommendations to solve problem

Data Understanding

- ❑ Loan dataset Information
- ❑ Loan status information

Loan data set information

Important variables

Customer details

- ❑ Employee title
- ❑ Employee length
- ❑ Zip code
- ❑ Home ownership
- ❑ Annual income
- ❑ Address state

Loan Details

- ❑ Loan Description
- ❑ Loan Purpose
- ❑ Loan Amount
- ❑ Funded amount
- ❑ Loan term
- ❑ Interest rate
- ❑ Loan installment
- ❑ Loan grade
- ❑ Loan sub grade
- ❑ Loan Verification status
- ❑ Loan issue date
- ❑ Loan Purpose

Customer Behaviour

- ❑ Loan status
- ❑ Debt to income ratio
- ❑ Delinquency in 2 years
- ❑ Earliest credit line opened
- ❑ Inquiries in last 6 months
- ❑ Number of open accounts
- ❑ Number of total accounts
- ❑ Public records
- ❑ Revolving balance
- ❑ Revolving balance utilization
- ❑ Public record of bankruptcies

Original data has 39717 rows and 111 columns

Loan Status

Labels description

- ❑ **Fully paid**: Applicant has fully paid the loan.
- ❑ **Current**: Applicant is in the process of paying the instalments, i.e., the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
- ❑ **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e., he/she has **defaulted** on the loan. **this is item of interest in this analysis**

14% of the data has default/charged off cases.

- ❑ It has decent amount of charged off cases for making inferences.
- ❑ Default and charged off terms will be used in this presentation interchangeably.
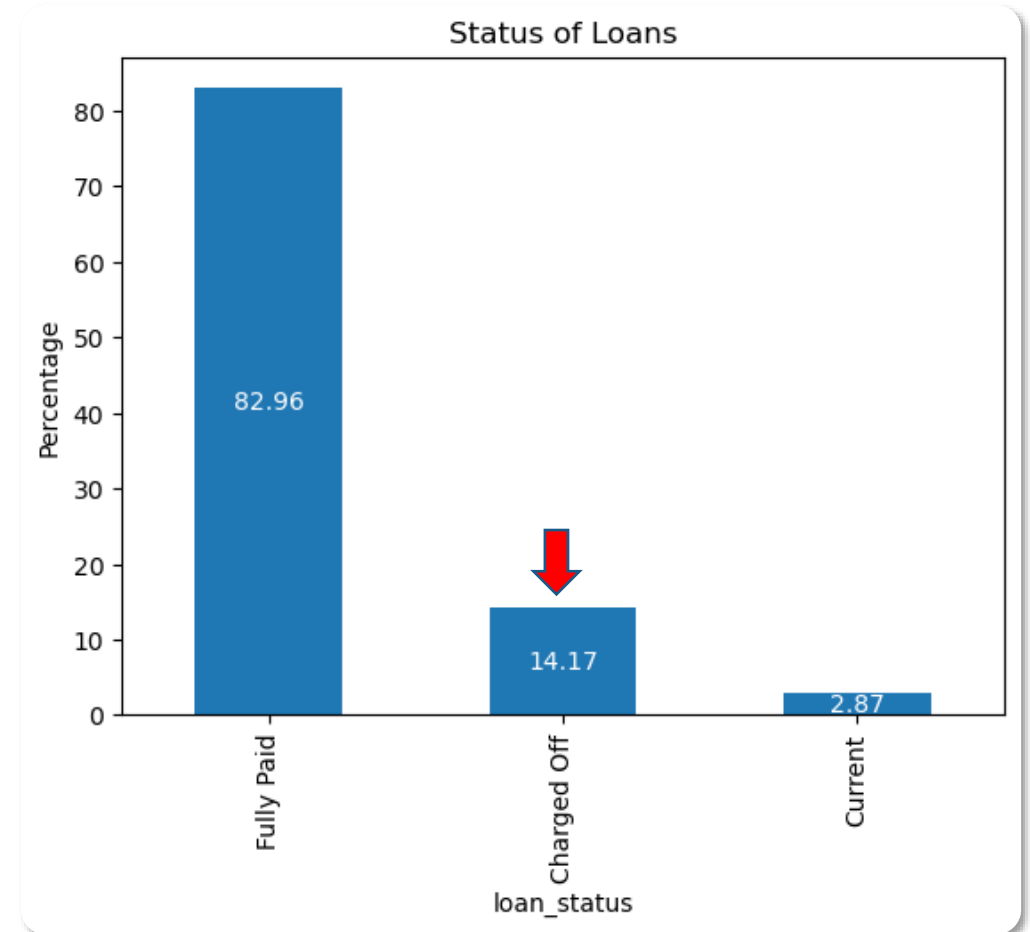
Status of Loans — bar chart showing Percentage by loan_status: Fully Paid 82.96, Charged Off 14.17, Current 2.87

Data Analysis

- ❑ Univariate Analysis
- ❑ Bivariate Analysis
- ❑ Segmented Analysis
- ❑ Multivariate Analysis

Statistical inferences

Most representative values for these columns (Median values):

- ❑ **Loan amount: Rs. 10000**
- ❑ **Loan Interest rate: 12%**
- ❑ **Monthly Installment of Loan : Rs. 280**
- ❑ **Annual Income of applicant: Rs. 60000**
- ❑ **Applicant Debt to income ratio (current dent to his monthly income ratio): 13**
- ❑ **Applicant total credit revolving balance (current : Rs. 9000**
- ❑ **Applicant credit revolving utilization (credit used out of available credit): 50%**
- ❑ **Applicant total credit accounts: 20**
- ❑ **Applicant current open credit accounts: 9**

	count	mean	std	min	25%	50%	75%	max
loan_amnt	38537.0	11043.919480	7345.344493	500.00	5300.00	9600.00	15000.00	35000.00
funded_amnt	38537.0	10780.985806	7087.032648	500.00	5200.00	9525.00	15000.00	35000.00
funded_amnt_inv	38537.0	10219.550468	7019.612638	0.00	5000.00	8725.00	14000.00	35000.00
int_rate	38537.0	11.932719	3.691611	5.42	8.94	11.71	14.42	24.40
installment	38537.0	322.378369	208.534515	15.69	165.74	277.86	425.55	1305.19
annual_inc	38537.0	67622.815809	43532.969574	4000.00	40000.00	58800.00	82000.00	667680.00
debt_to_income	38537.0	13.284026	6.666168	0.00	8.15	13.38	18.56	29.99
revol_bal	38537.0	13282.795703	15861.565836	0.00	3650.00	8761.00	16903.00	149588.00
revol_util	38487.0	48.707753	28.365340	0.00	25.20	49.10	72.30	99.90
total_acc	38537.0	22.049096	11.424068	2.00	13.00	20.00	29.00	90.00
open_acc	38537.0	9.274853	4.401232	2.00	6.00	9.00	12.00	44.00

- ❑ **Loan amount, funded amount by LC, funded amount by investors looks similar.**
- ❑ **By comparing Medians, Invested amount (Rs. 8725) is 90% of is requested loan amount (Rs. 9600).**

Univariate – Loan status

Status of Loans
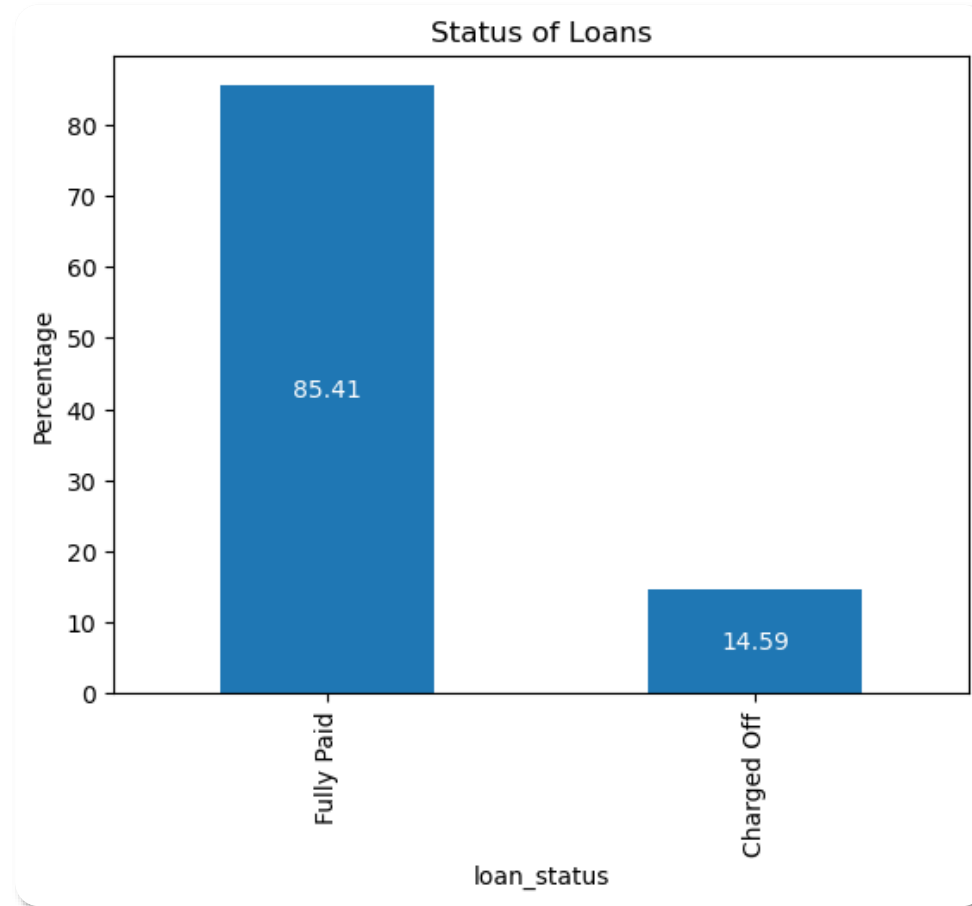
After cleaning data, and dropping current loans, Loan default rate is 14.5%

Univariate – Distribution Analysis

Observations

❑ loan amount, funded amount and funded amount invested distributions looks same.

❑ Majority of the loan amounts are round numbers like 5000, 10000, 15000, 20000.

❑ Most of them are right skewed distributions.

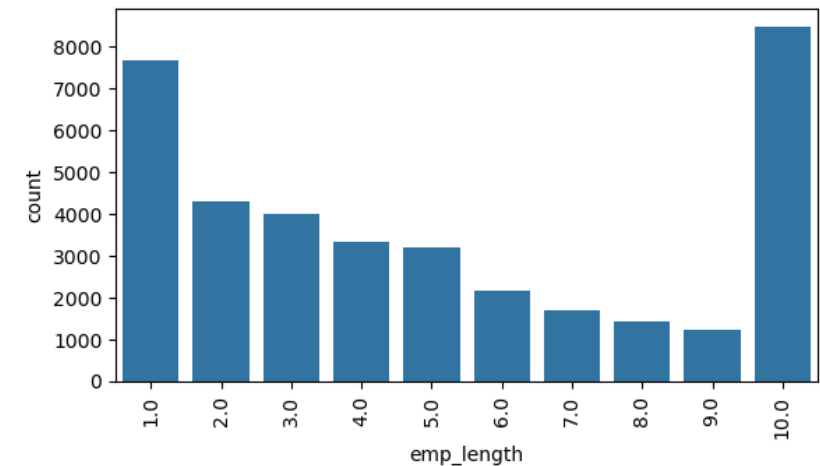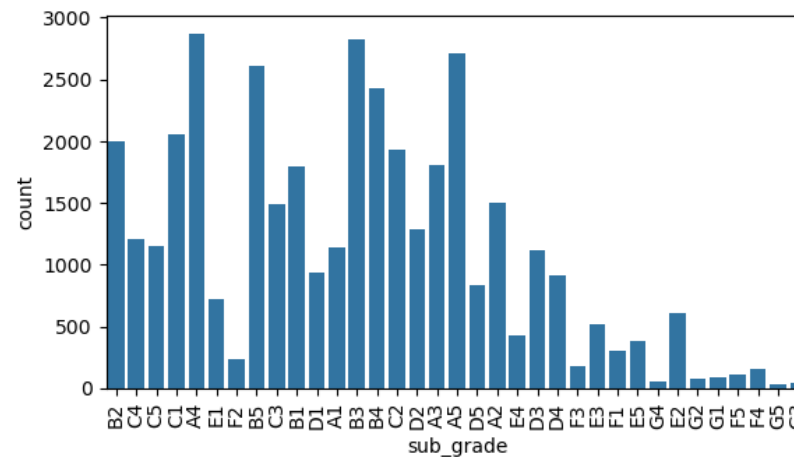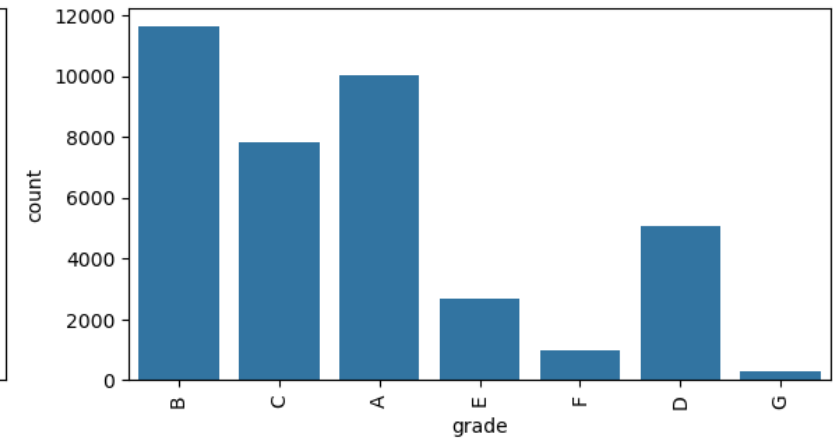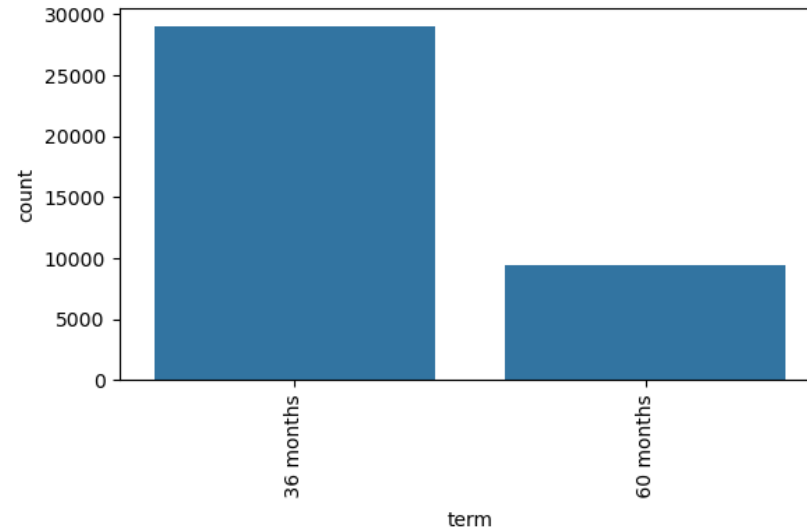❑ Applicant revolving credit utilization looks like uniform distribution.

Most of them are right skewed distributions.
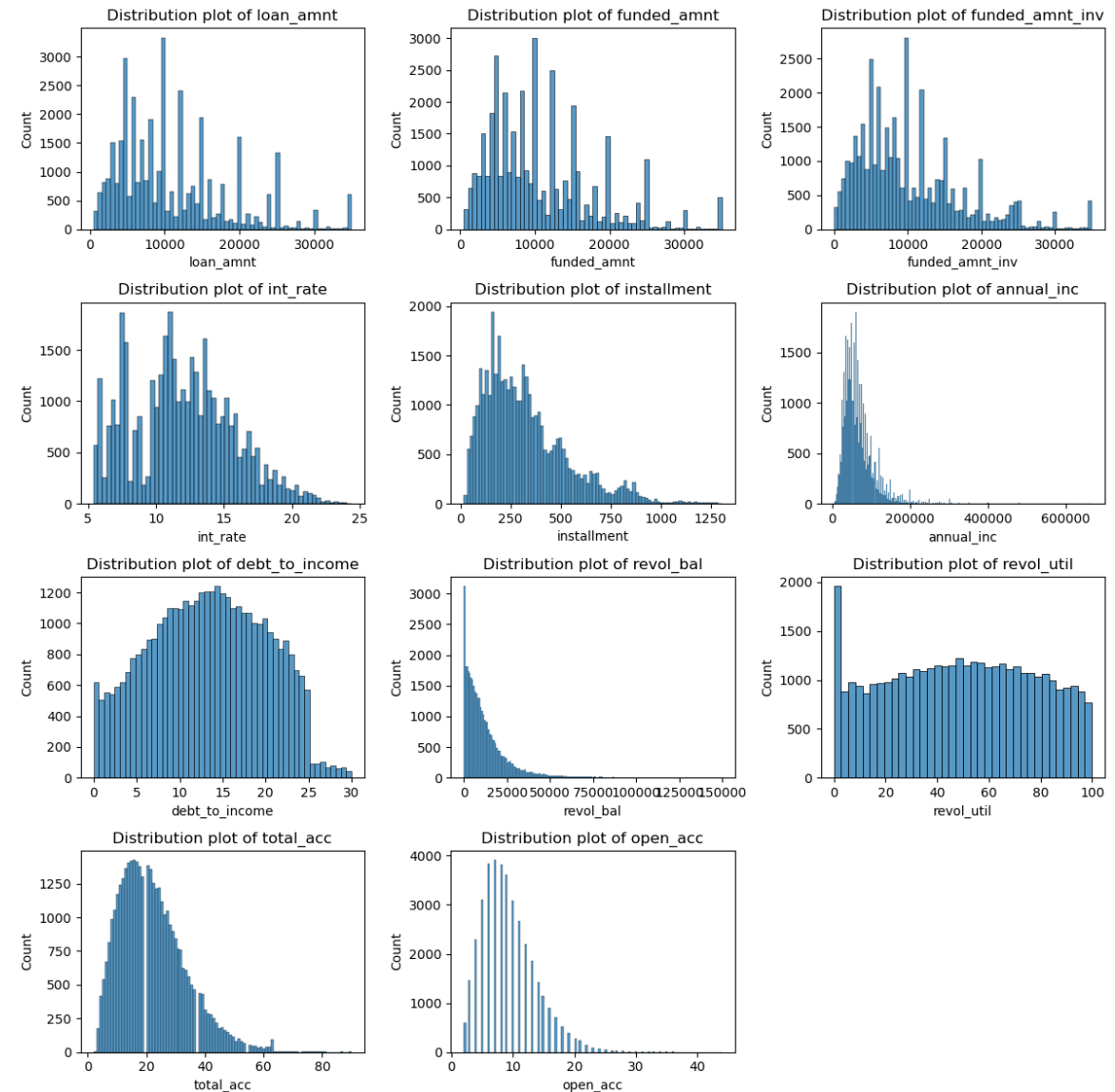
Univariate – Categorical variables Analysis

Observations

❑ Loan Term:

   36 months is higher

❑ Loan Grade:

   A,B,C,D are major grades

❑ Applicant Employment

   length:

   10+ years is major one,

   next is 1 years.

Univariate – Categorical variables Analysis

Observations

- ❑ Applicant Home ownership:

  Rent, Mortgage are frequent ones.

- ❑ Applicant Income verification:

  Majority of them are Not verified

- ❑ Loan Purpose:

  Top items are debt consolidation,

  credit card, home improvement,

  Major Purchase, Small Business.

- ❑ Top states of applicants:

  CA - California, NY - new York, FL -

  Florida, TX - Texas
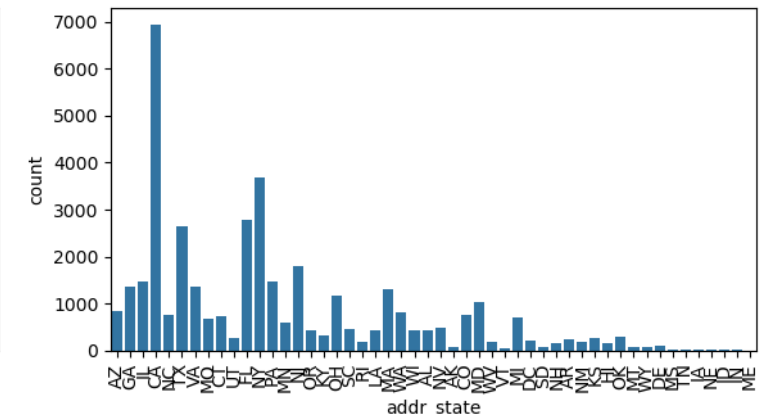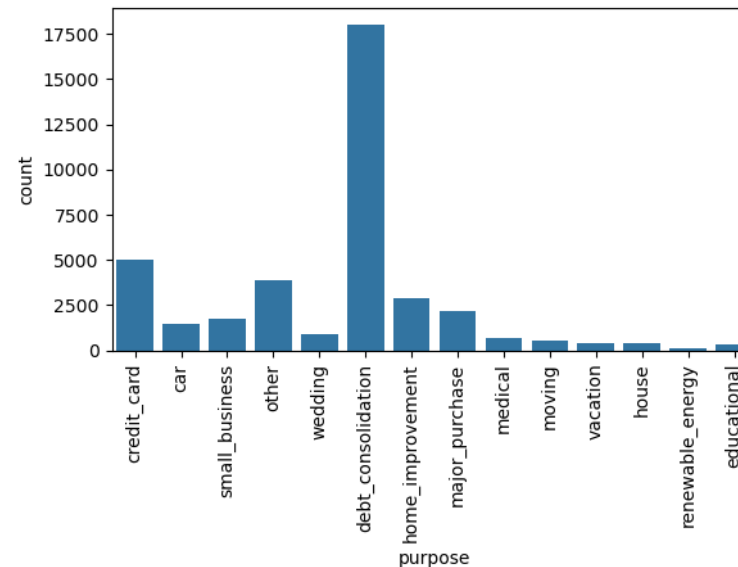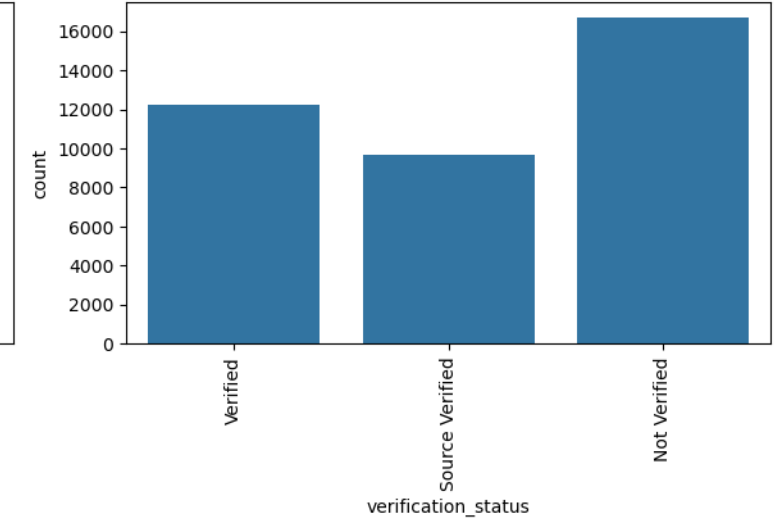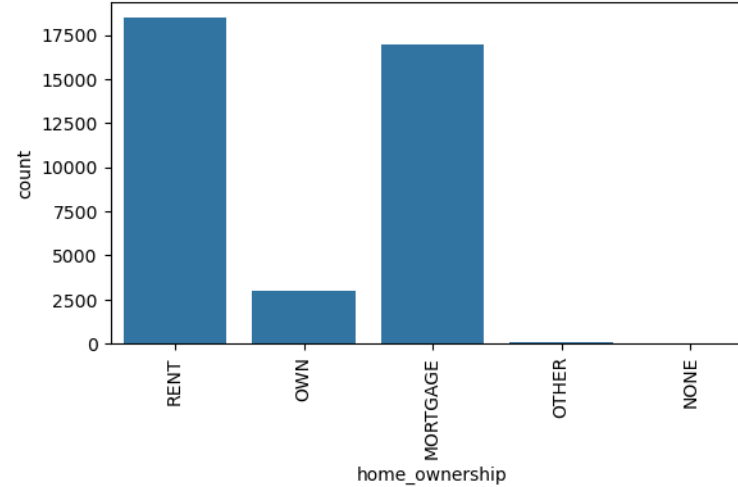
Univariate – Categorical variables Analysis

Observations

❑ Delinquency (Number of due incidences of for the past 2 years):

  Max 11, but after 1, count is negligible.

❑ Inquiry_last_6_months –

  max inquiries are 8, but after 3, count is negligible

❑ Number of derogatory public records:

  Max 4 records, but after 1, count is negligible

❑ Number of public record bankruptcies:

  Max 2 bankruptcies, but, after 1 count is negligible
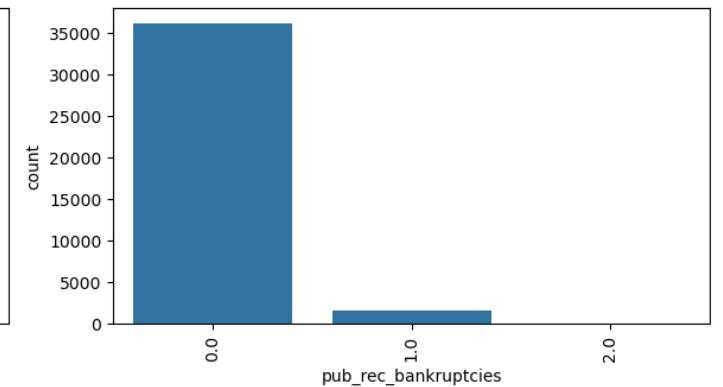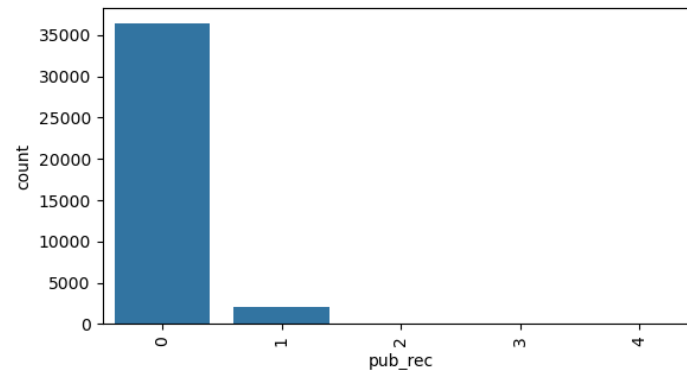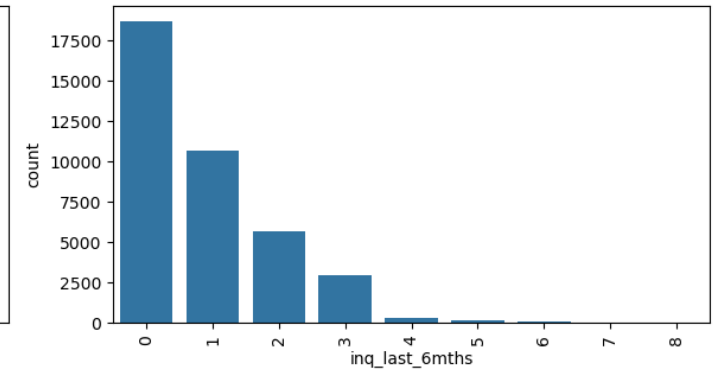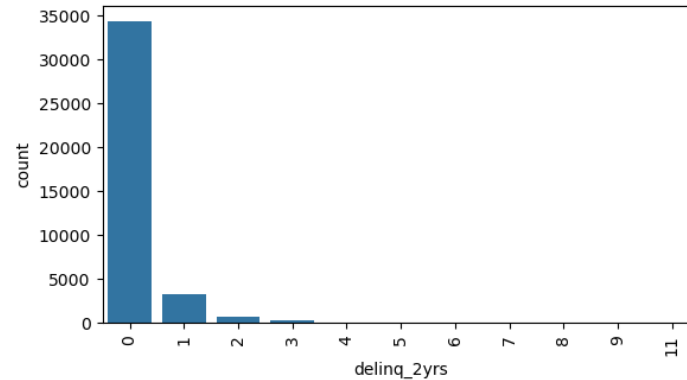
Bivariate – Loan status vs important variables Median

Indicators for charged off items

❑ **Loan amount, funded amount is higher** (While difference is 400 for loan amount, but for invested amount, difference is high as 900)

❑ **Interest rate is higher by 2%.**

❑ **Annual income is lower by 7000.**

❑ **debt to income ratio is higher by 1.**

❑ **Revolving utilization percentage is higher by 10%.**

Loan Amount, Interest Rate, Annual Income, Debt to Income ratio, Revolving credit utilization are important

Median - Loan Status vs loan_amnt

Median - Loan Status vs funded_amnt_inv

Median - Loan Status vs int_rate

Median - Loan Status vs annual_inc

Median - Loan Status vs debt_to_income

Median - Loan Status vs revol_util
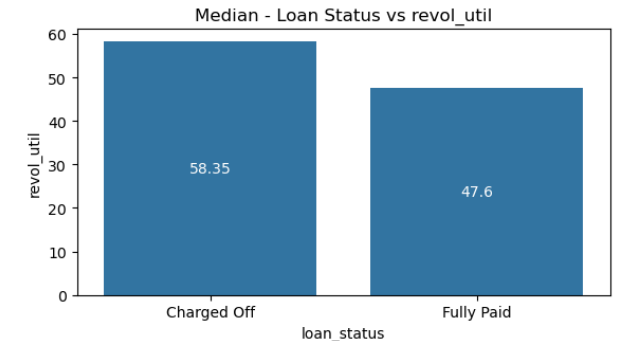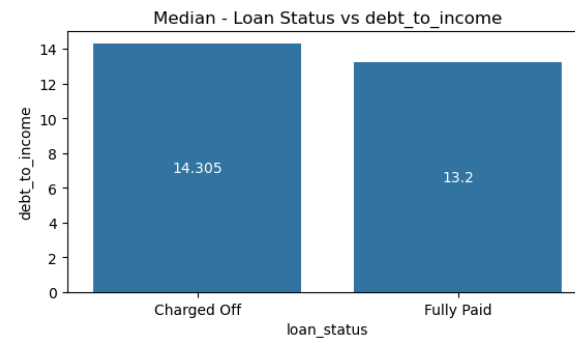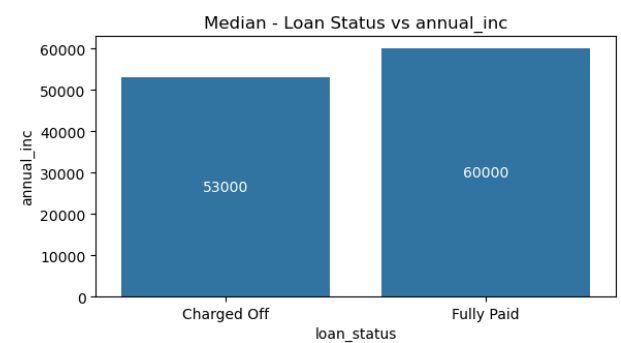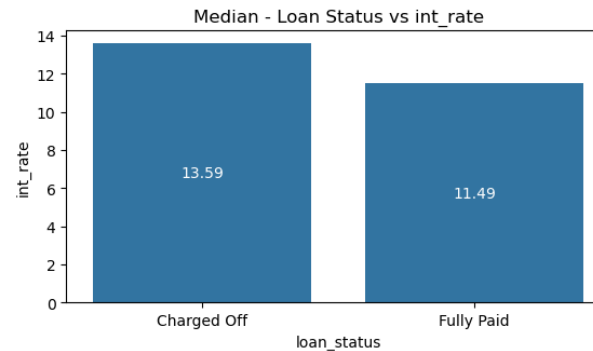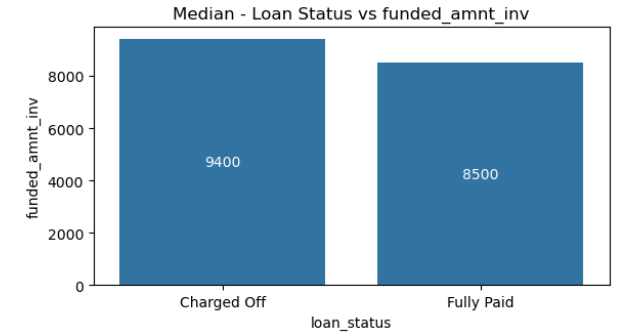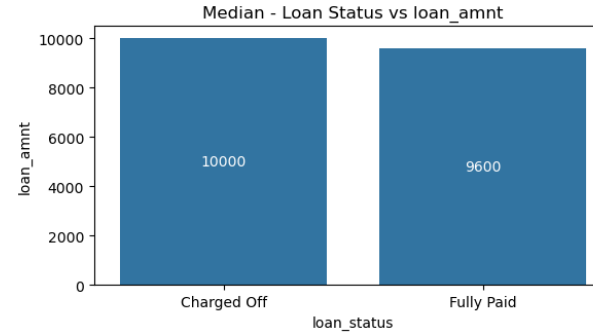
Bivariate – Loan status vs important variables 95 percentile

Indicators for charged off items

By looking at 95 percentile data, we can take corresponding values as a cutoff values.

❑ **Loan amount**: **25000** is 95 percentile value from fully paid items. **Above** this value, there is a chance of default.

❑ **Interest rate**: **17.93%** is 95 percentile value from fully paid items. **Above** this value, there is a chance of default.

❑ **Annual income**: **1,28,000** is 95 percentile value from charge off items. **below** this value, there is a chance of default.

95 Percentile - Loan Status vs loan_amnt

95 Percentile - Loan Status vs int_rate

95 Percentile - Loan Status vs annual_inc
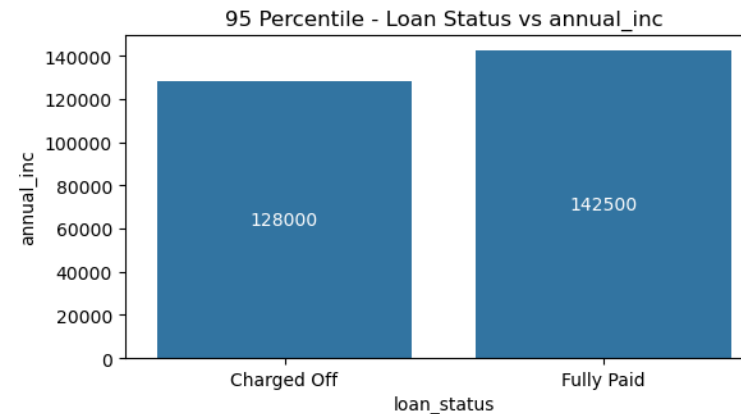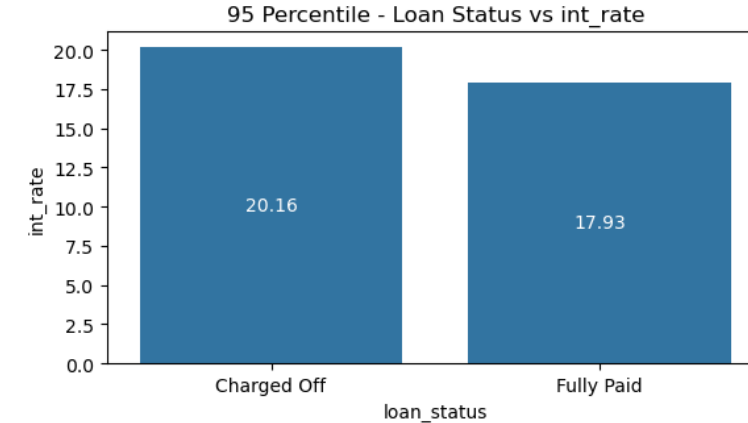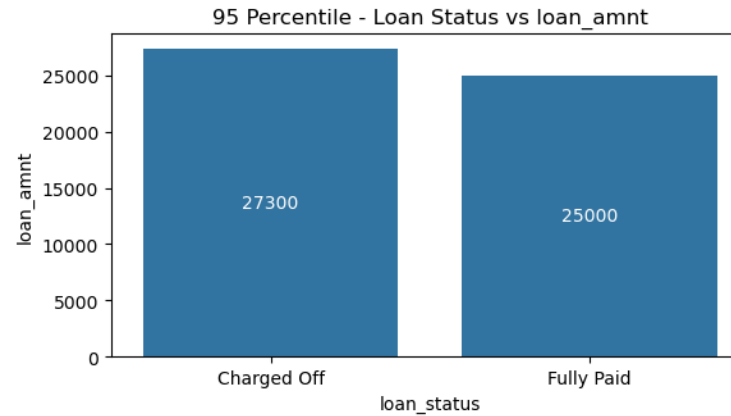
Bivariate – Charged Off Proportions - Loan status vs Loan term
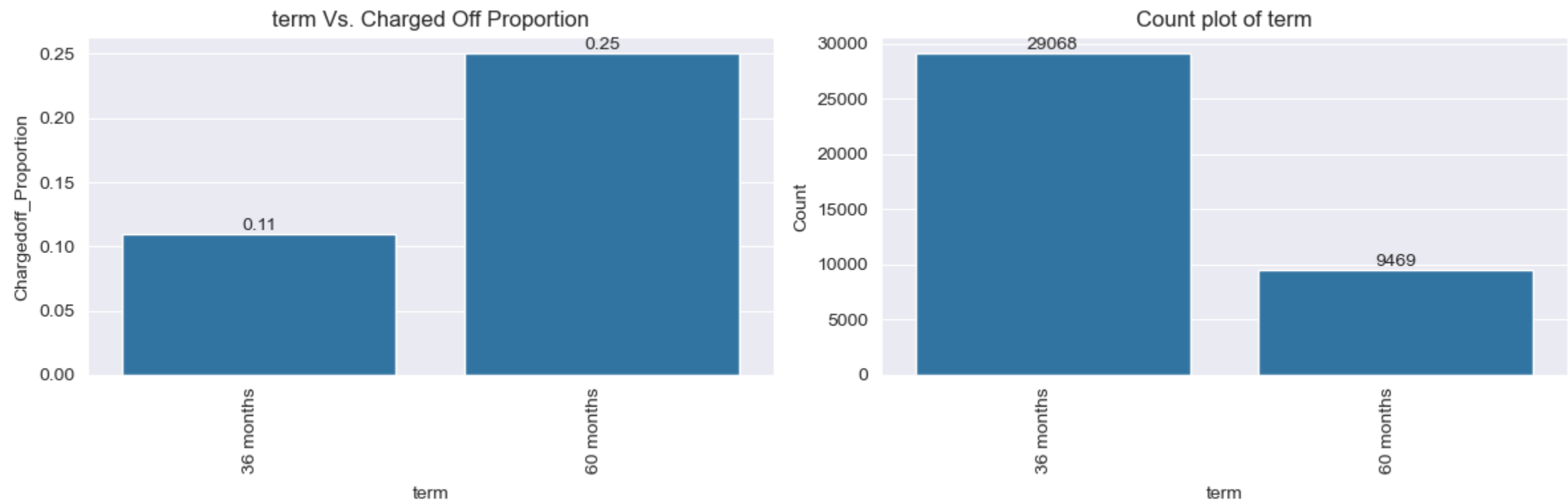
Loan Term of 60 months has higher probability of default

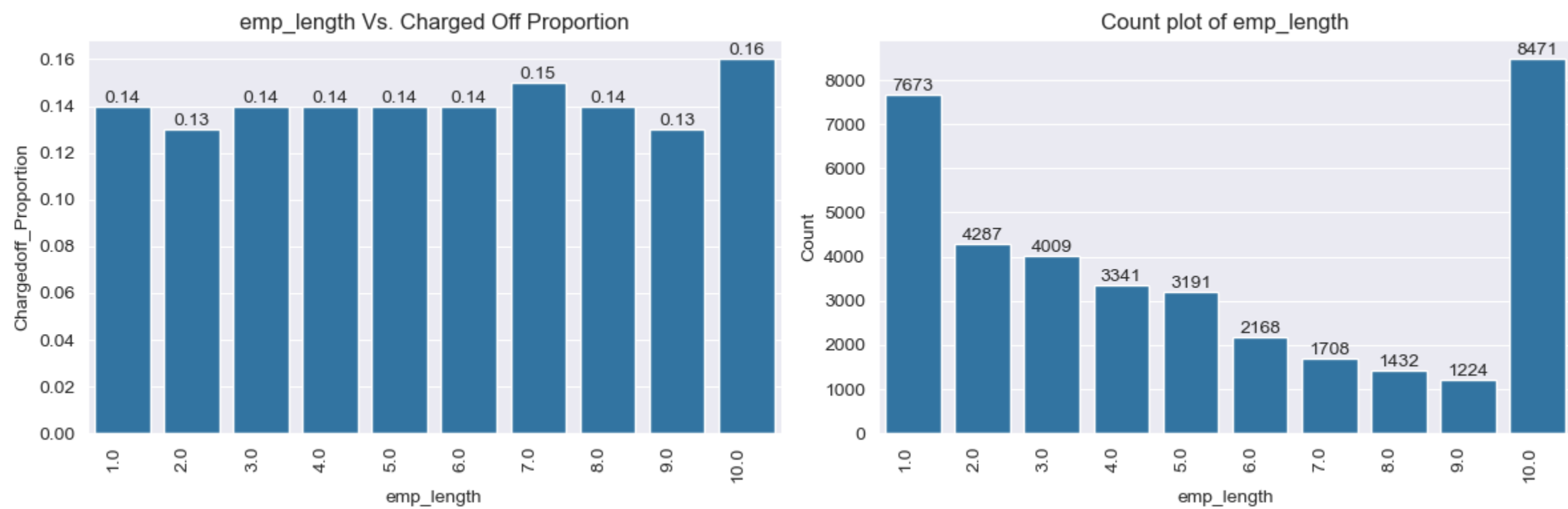# Bivariate – Charged Off Proportions - Loan status vs Loan Grade
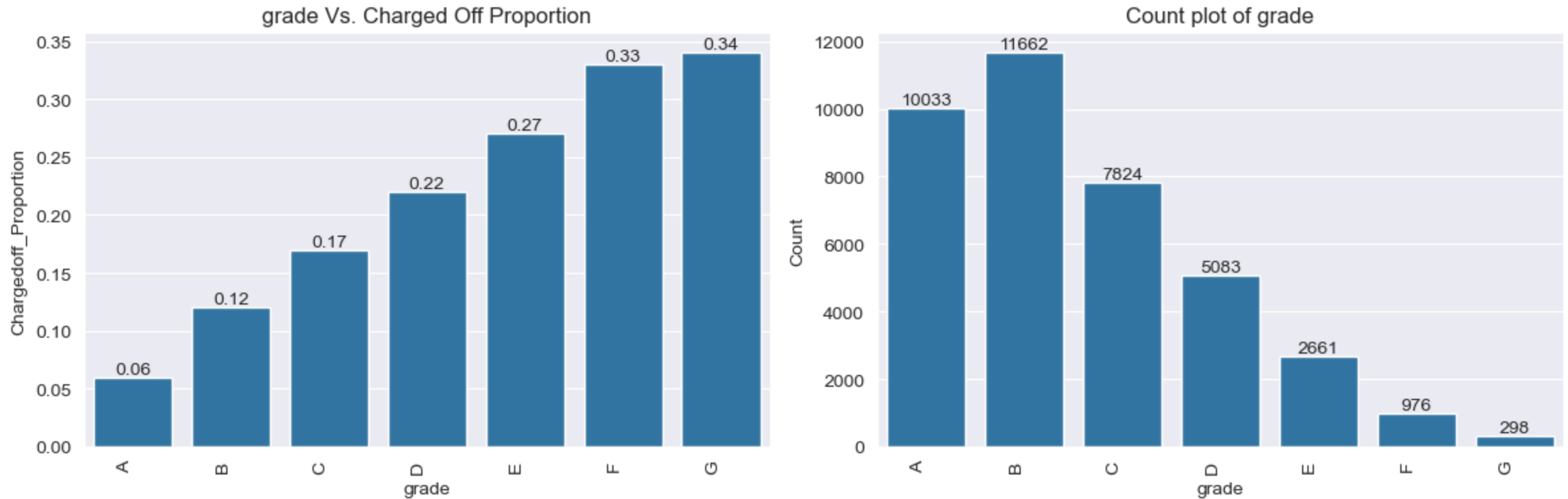
**Probability of default increases as grade increases from A to G.
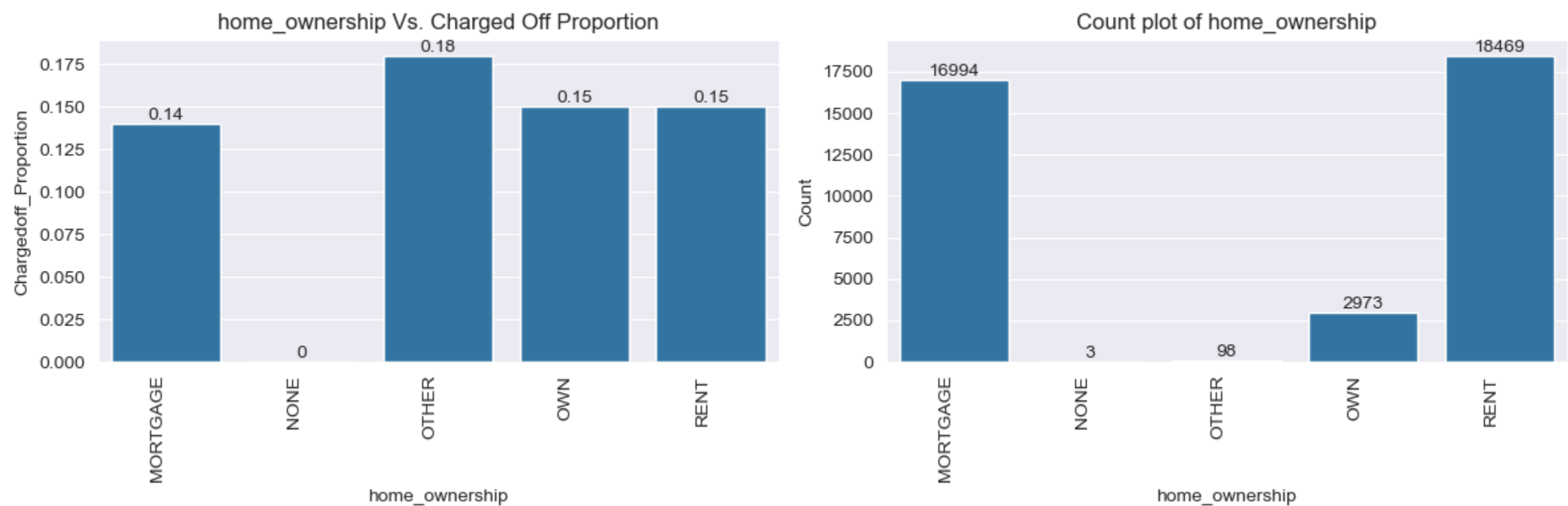Lending Club has done good job in deciding grades based on risks.**
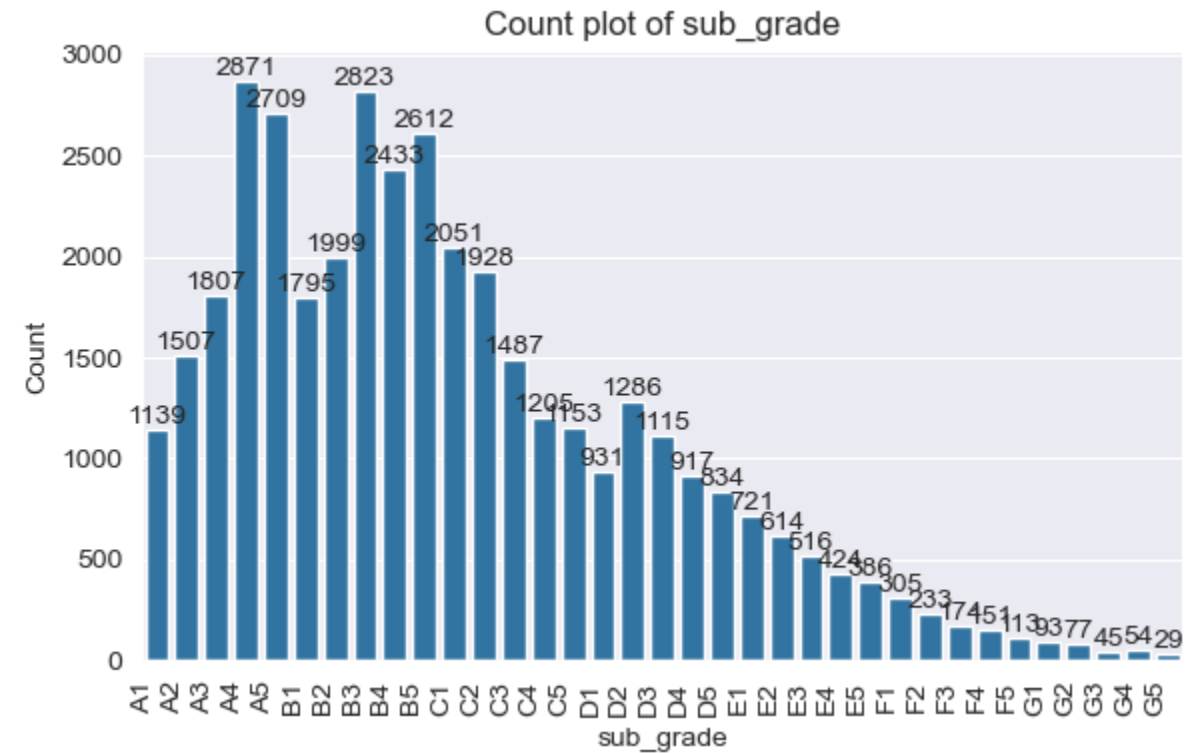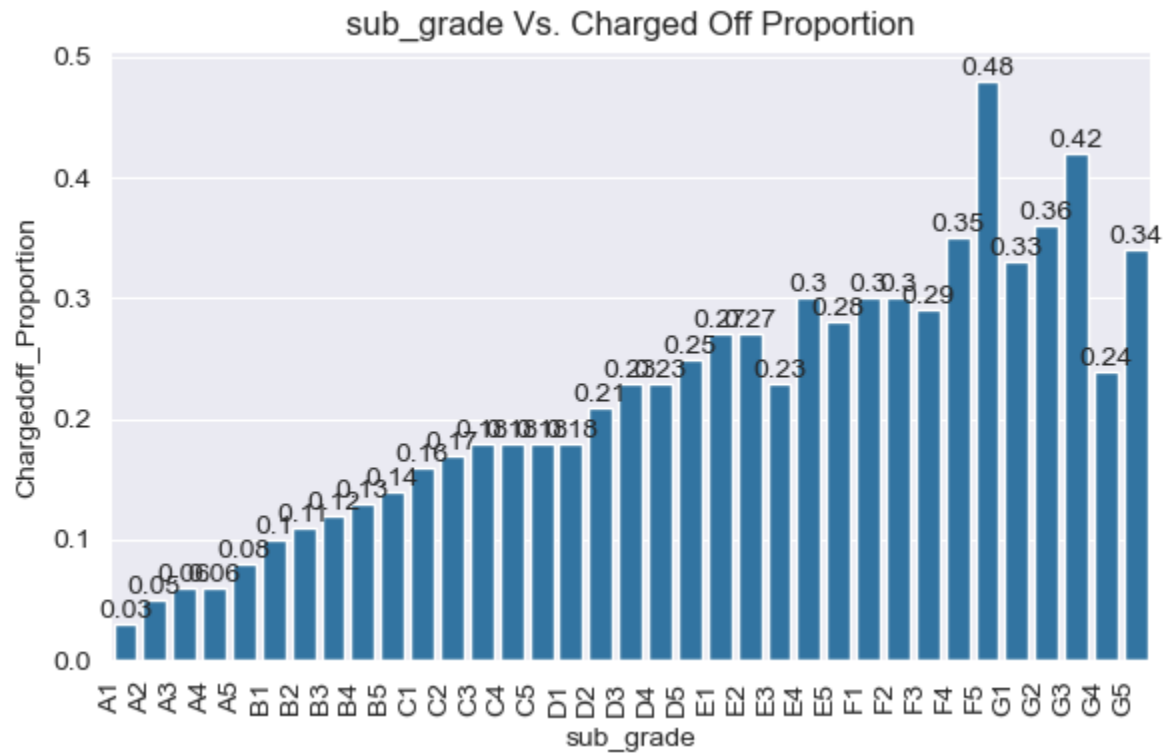
Same trend is observed in sub grades as grades. Subgrades makes sense, continue them. G grade does not need subgrades.

Bivariate – Charged Off Proportions - Loan status vs Employee Length

No impact of Employment Length

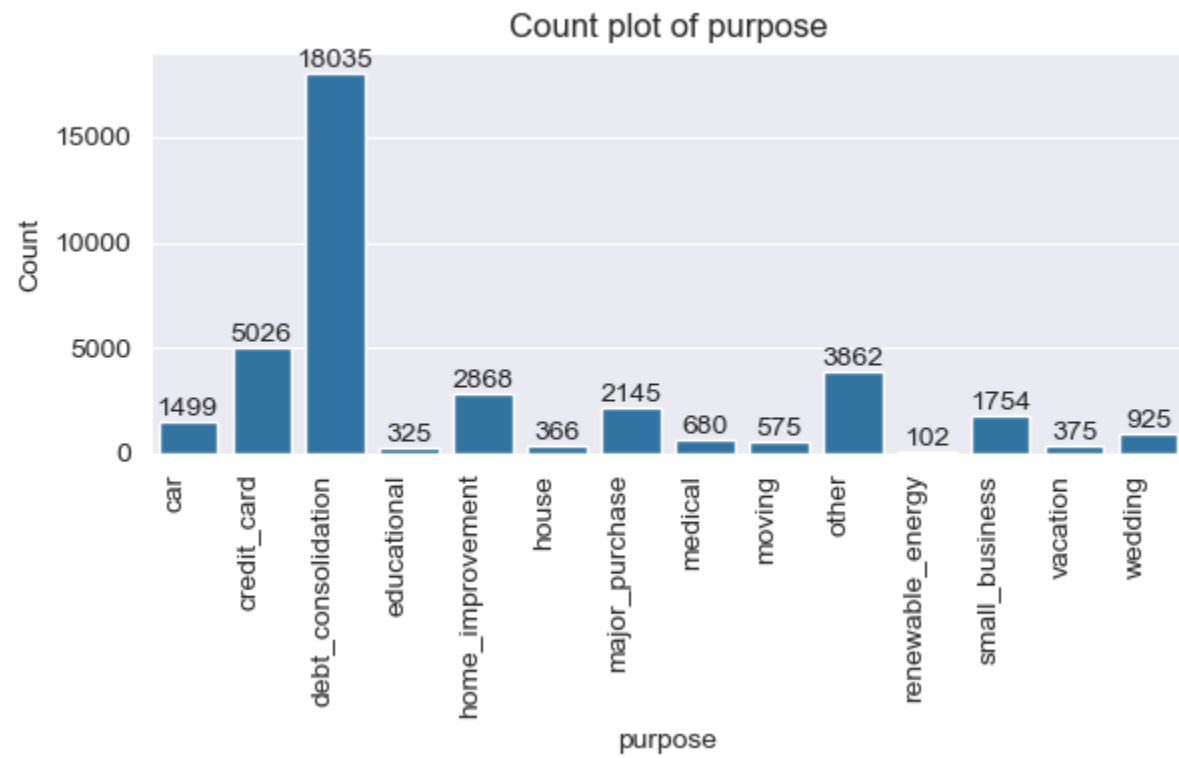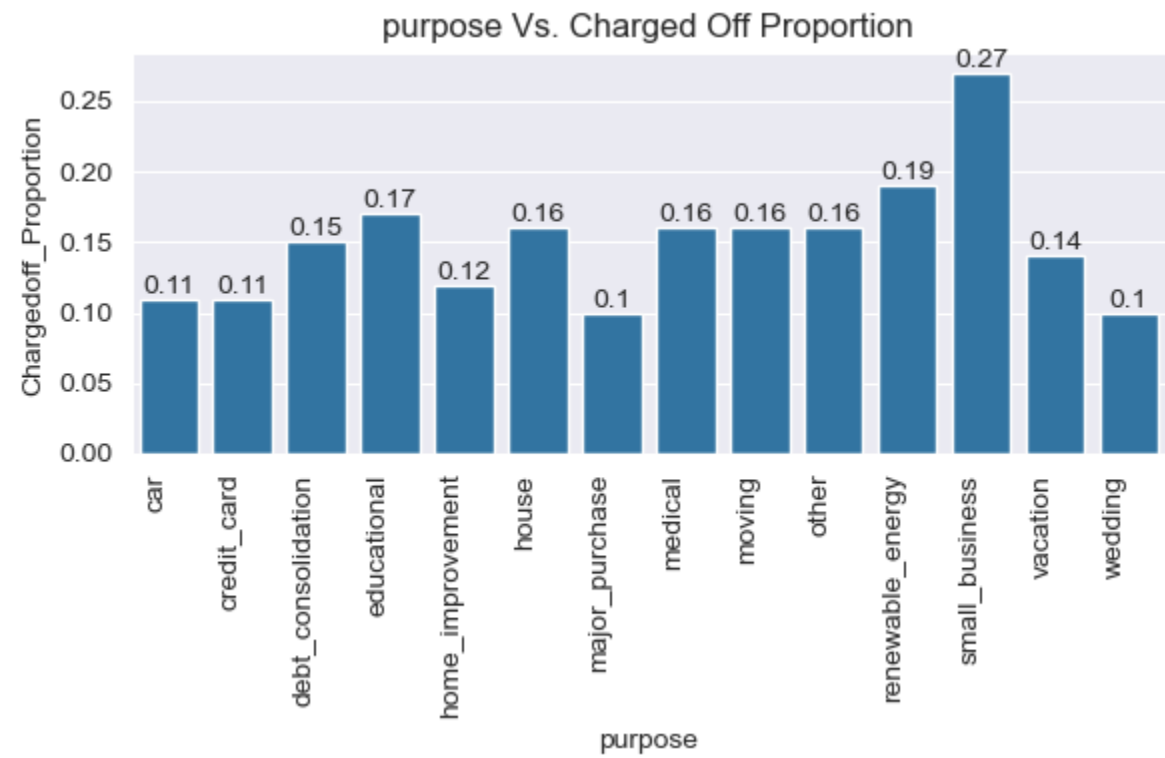# Bivariate – Charged Off Proportions - Loan status vs Home ownership

No impact of Home ownership of applicant

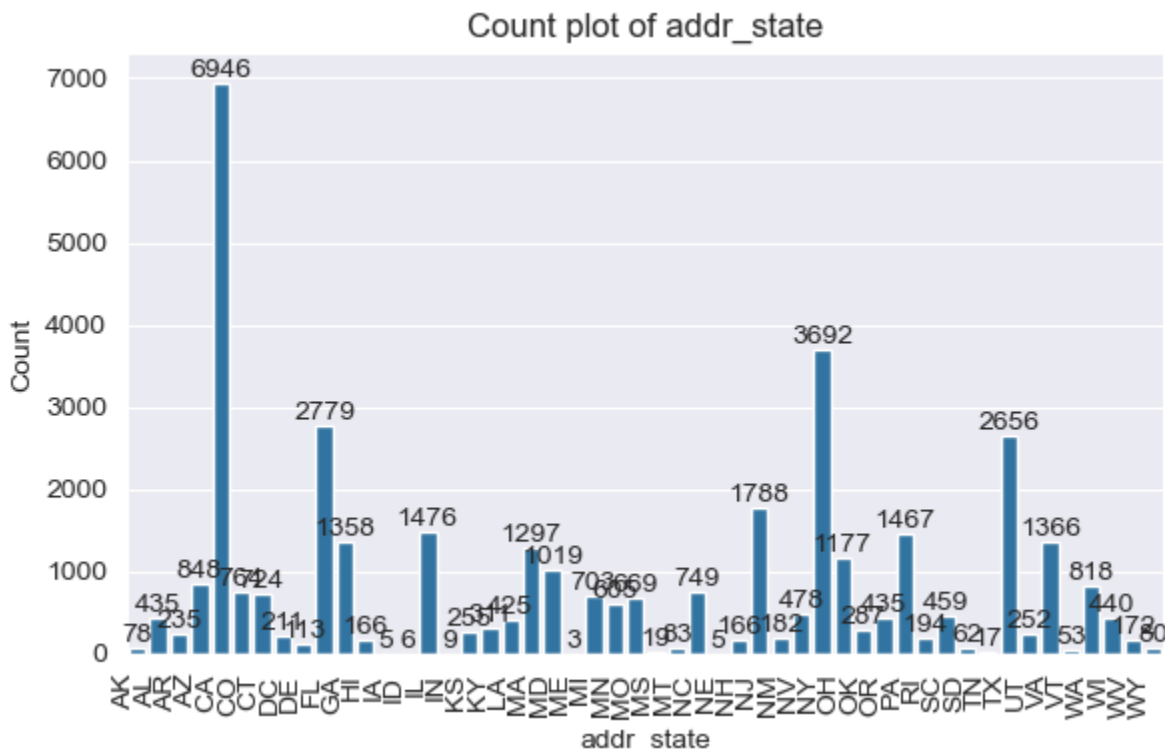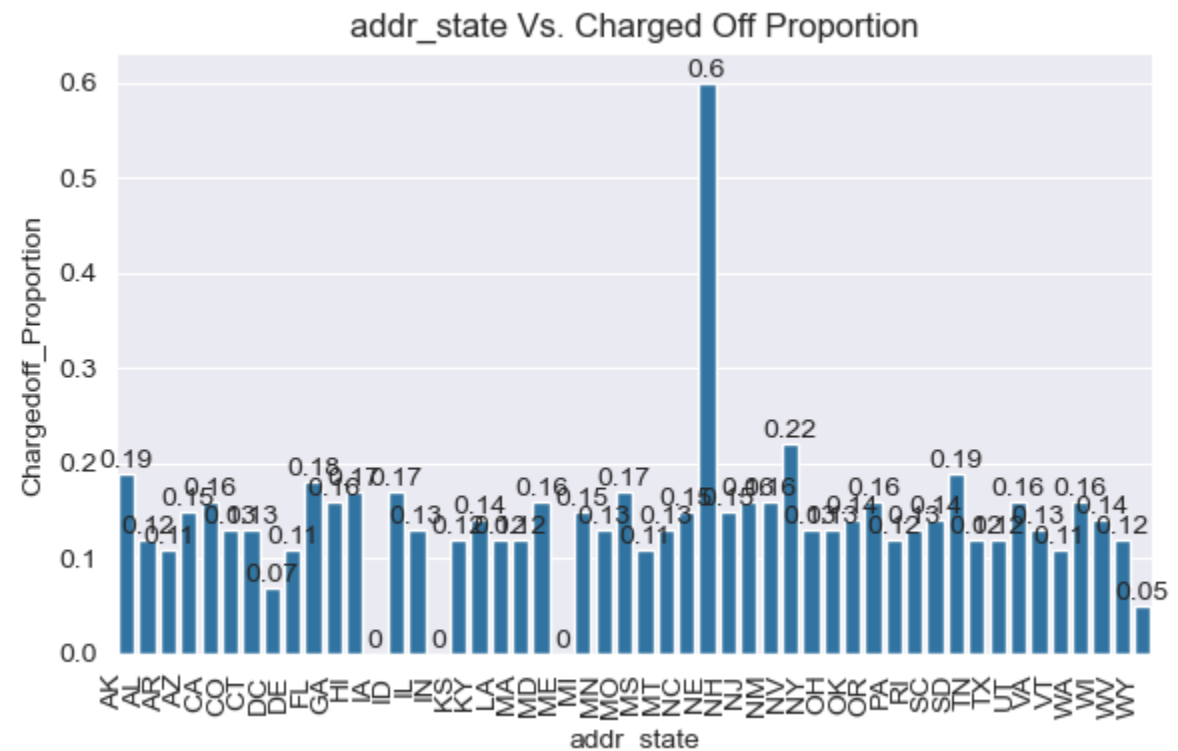# Bivariate – Charged Off Proportions - Loan status vs verification status
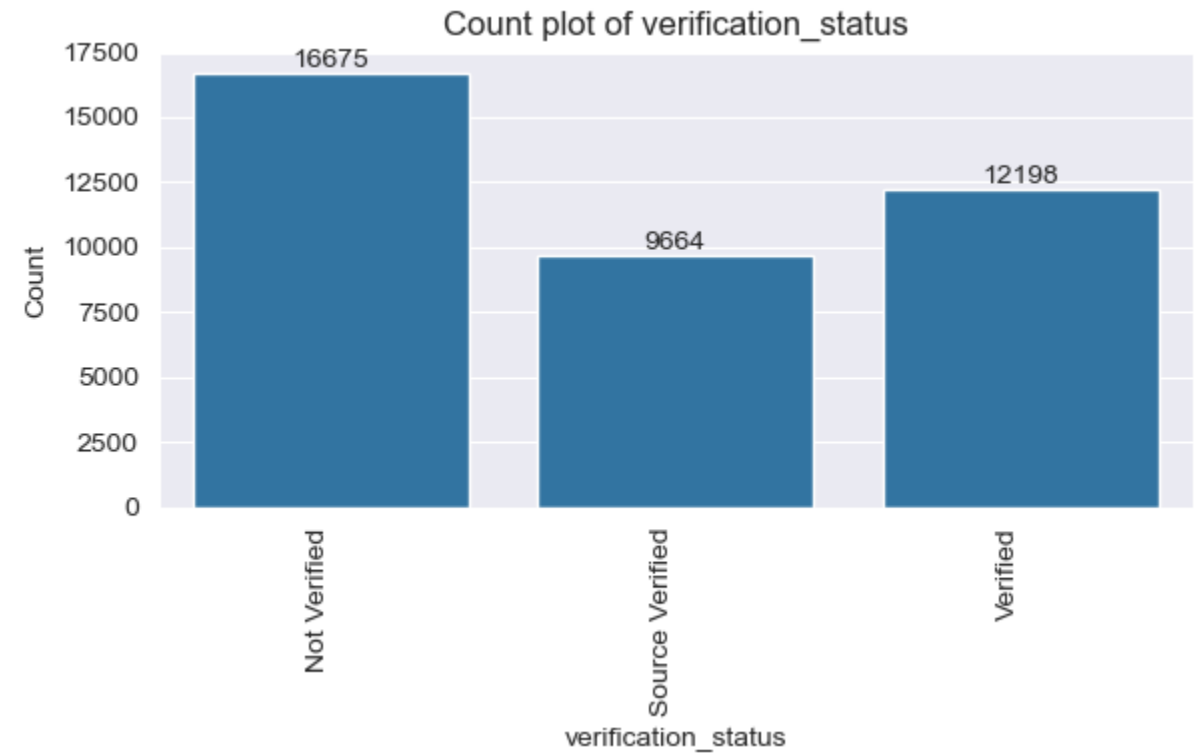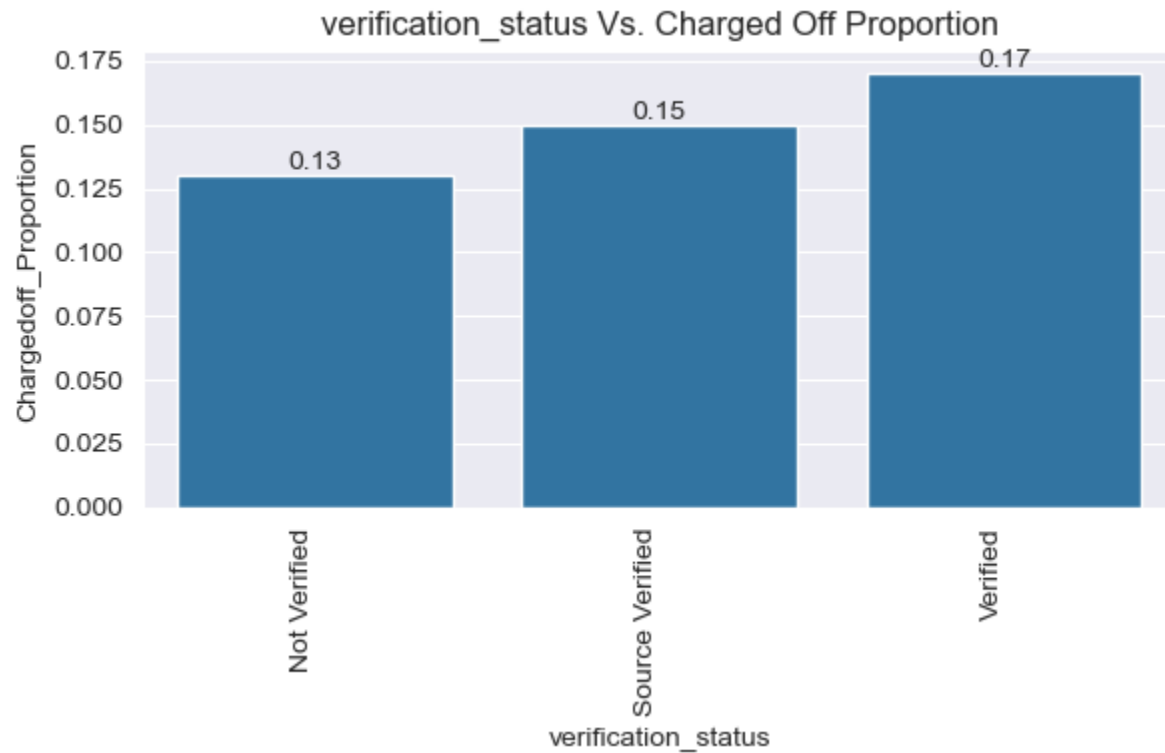
surprisingly verified items has higher default. Need to recheck on verification process.

# Bivariate – Charged Off Proportions - Loan status vs Loan Purpose

small business and debt consolidation has higher default rates.
some other types shows higher default, but their count is low to take any decision.

Bivariate – Charged Off Proportions - Loan status vs Applicant State

New York has higher default probability with decent count.
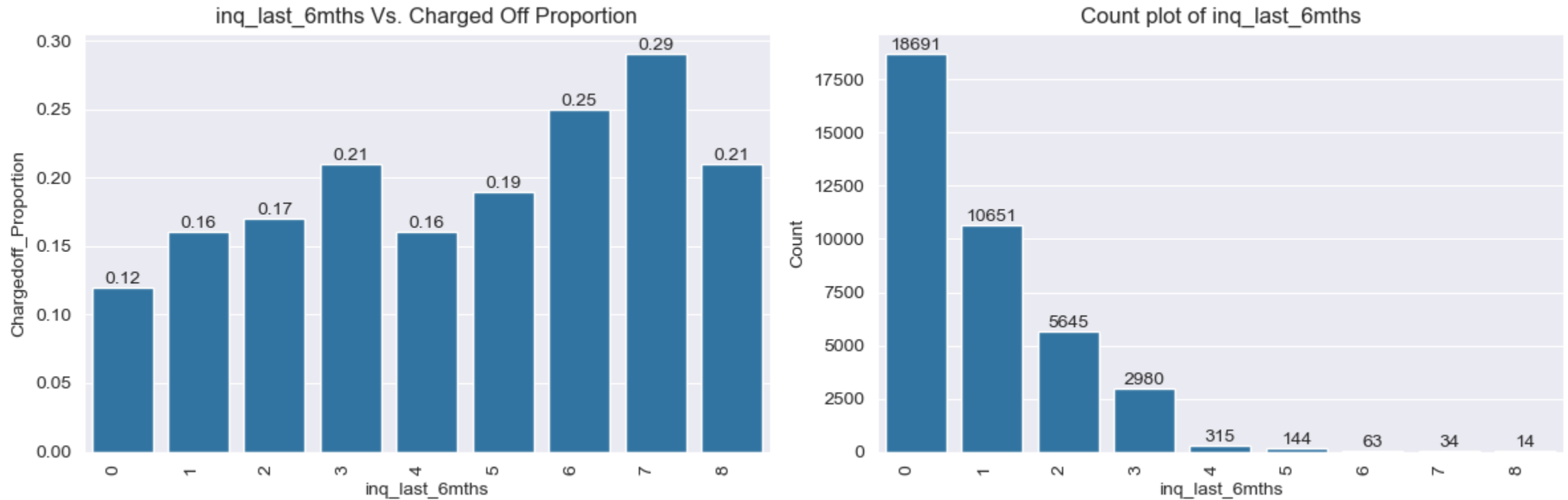NH items are low to consider that.

Bivariate – Charged Off Proportions - Loan status vs delinquency in 2yrs
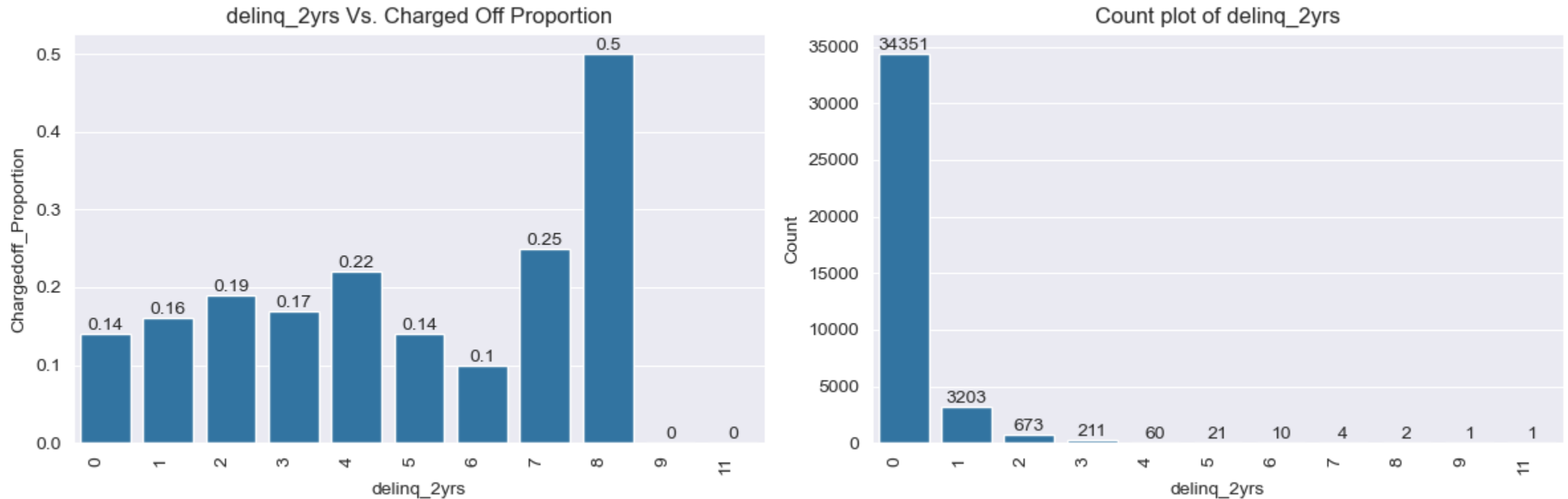
As delinquency cases increases, default rate increases.
**If applicant has more than 1 case, it is risky considering number of applicants.**

The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years

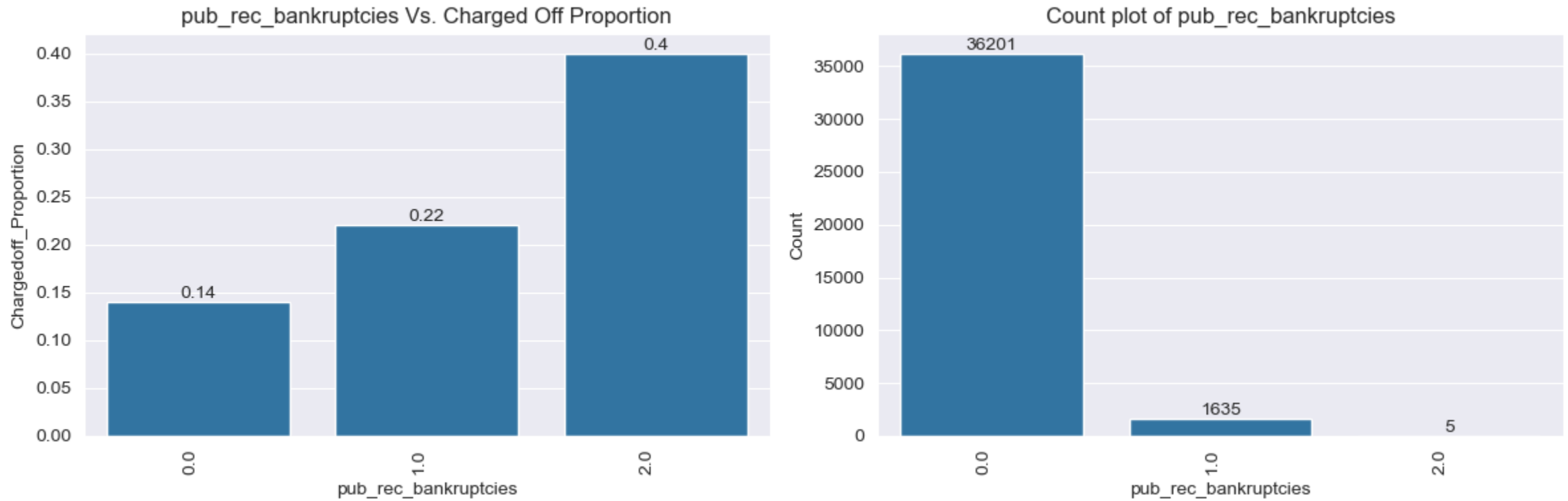# Bivariate – Charged Off Proportions - Loan status vs Inquiries in last 6mnths

As inquiries increases, default rate increases.
If applicant has more than 2 case, it is risky considering number of applicants.

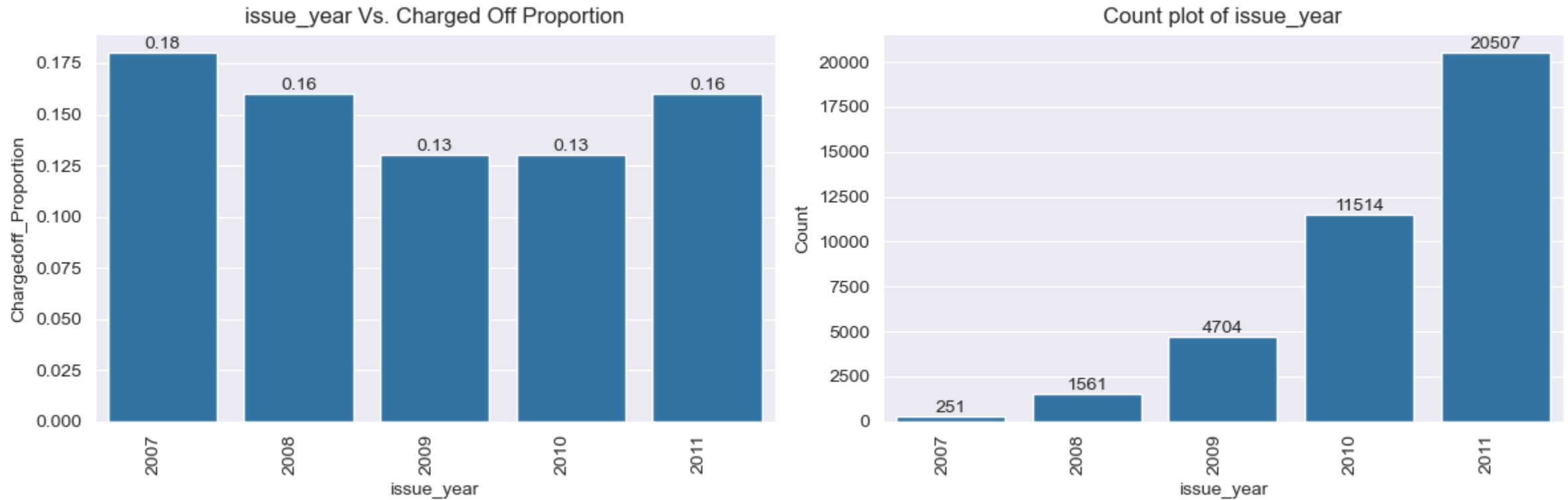# Bivariate – Charged Off Proportions - Loan status vs Number of derogatory public records
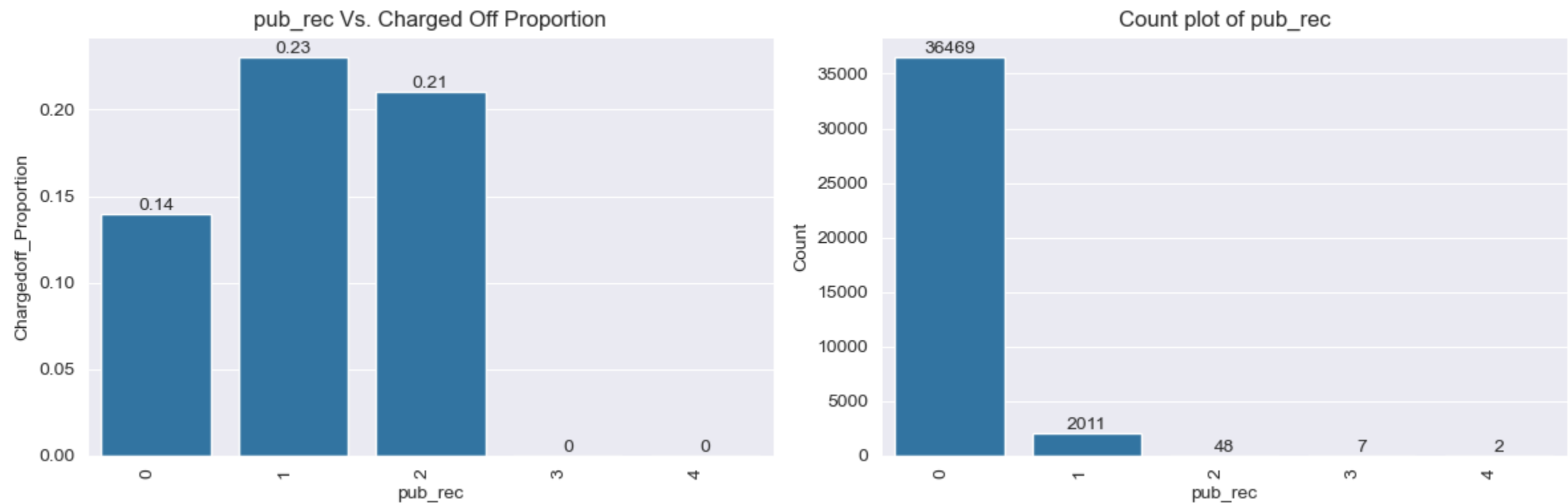
more than 0 public records can be good cutoff. As records increases, default probability increases.

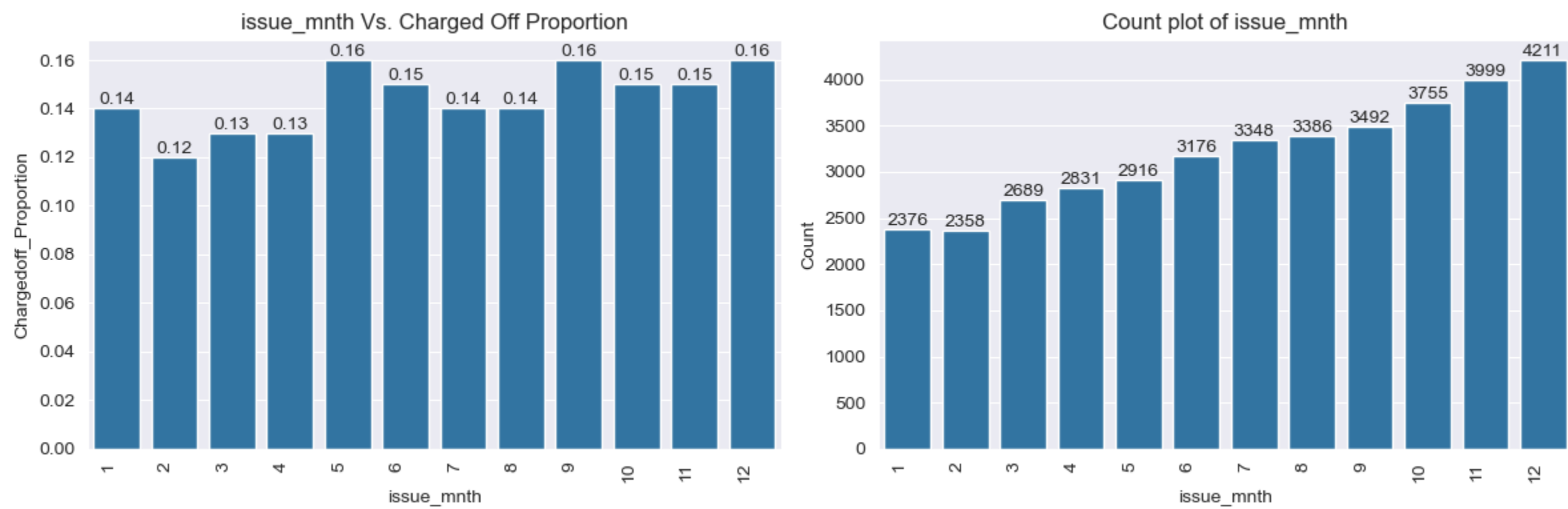# Bivariate – Charged Off Proportions - Loan status vs public recorded bankruptcies

As cases increases, default rate increases. good cutoff is more than 0.

# Bivariate – Charged Off Proportions - Loan status vs Issue date Year

As year progress from 2007 to 2011, loans are increased. From 2007 to 2009 charged off proportion decreased. After that, it increased. So, trend may continue from 2011. need to be careful.

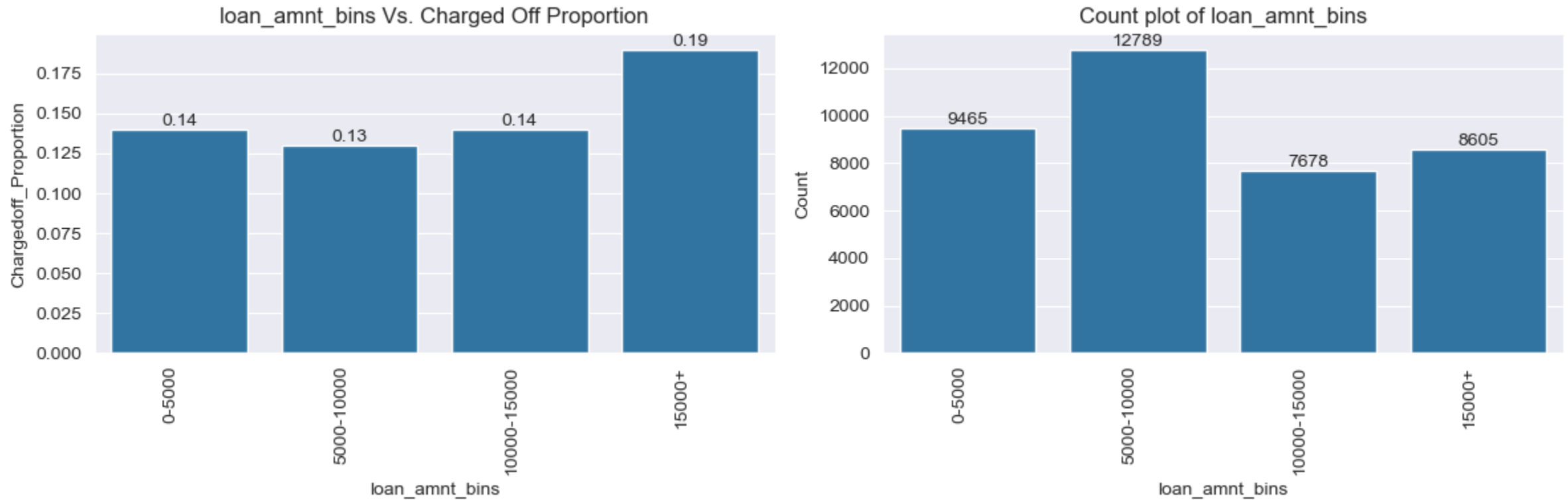# Bivariate – Charged Off Proportions - Loan status vs Issue date Month

Towards the end of the year, loans are increased. But no trend in default probability over months

Bivariate – Charged Off Proportions - Loan status vs Earliest Credit account opening Year
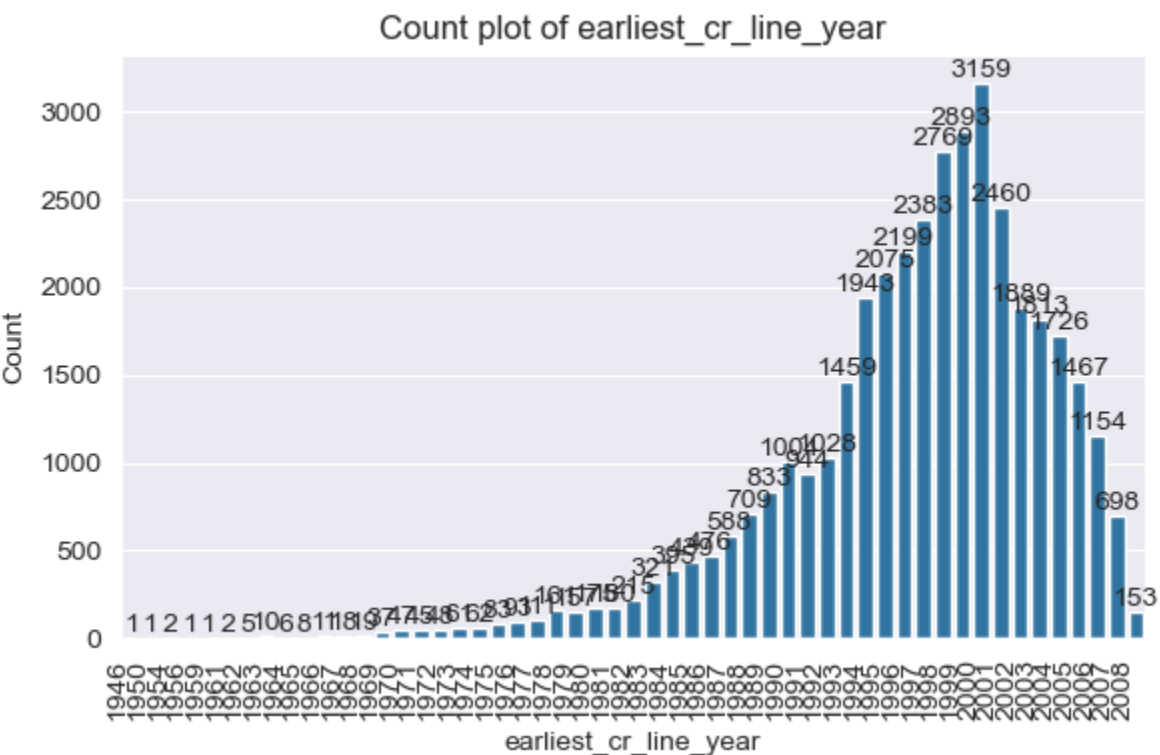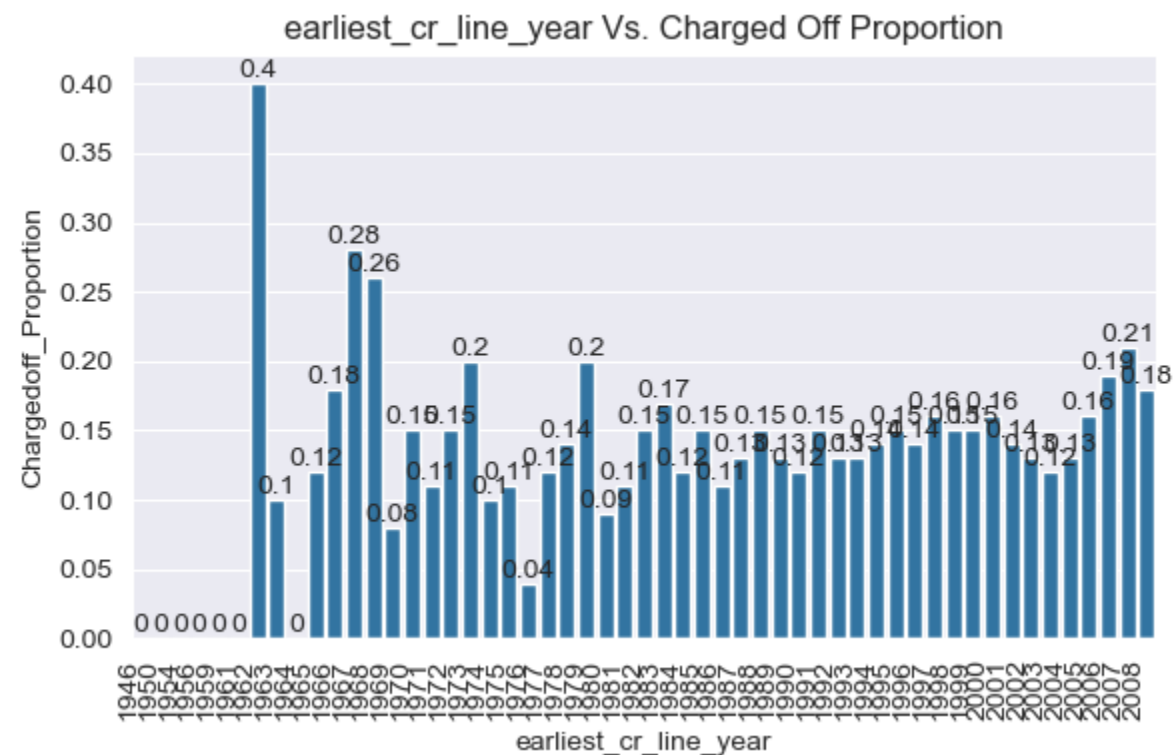

Loan applicants opened their first credit account starts from 1946.
Many people opened their first credit account in 2001. the trend was raising till 2001, after that it dropped. But no trend is observed in charged off proportion, it was high in 1963, but account were low, no decision can be taken.

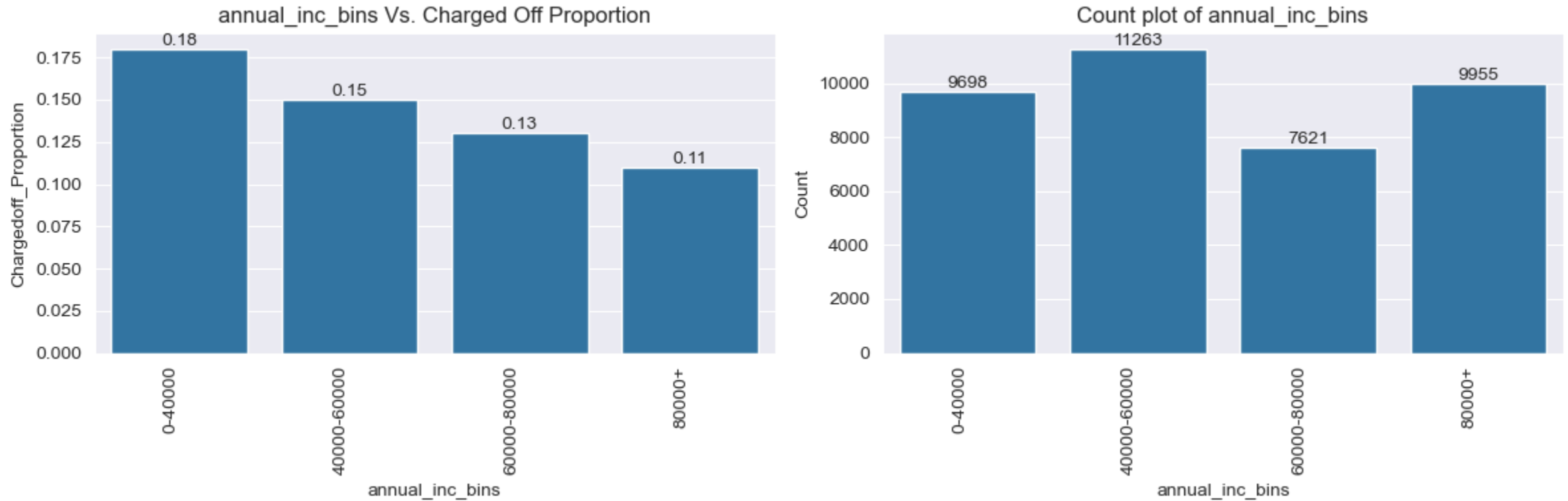# Bivariate – Charged Off Proportions - Loan status vs Loan Amount Bins

Rs. 5000 -10000 is the common loan amount. But 15000+ loans has higher probability of default.

Bivariate – Charged Off Proportions - Loan status vs Loan interest rate Bins
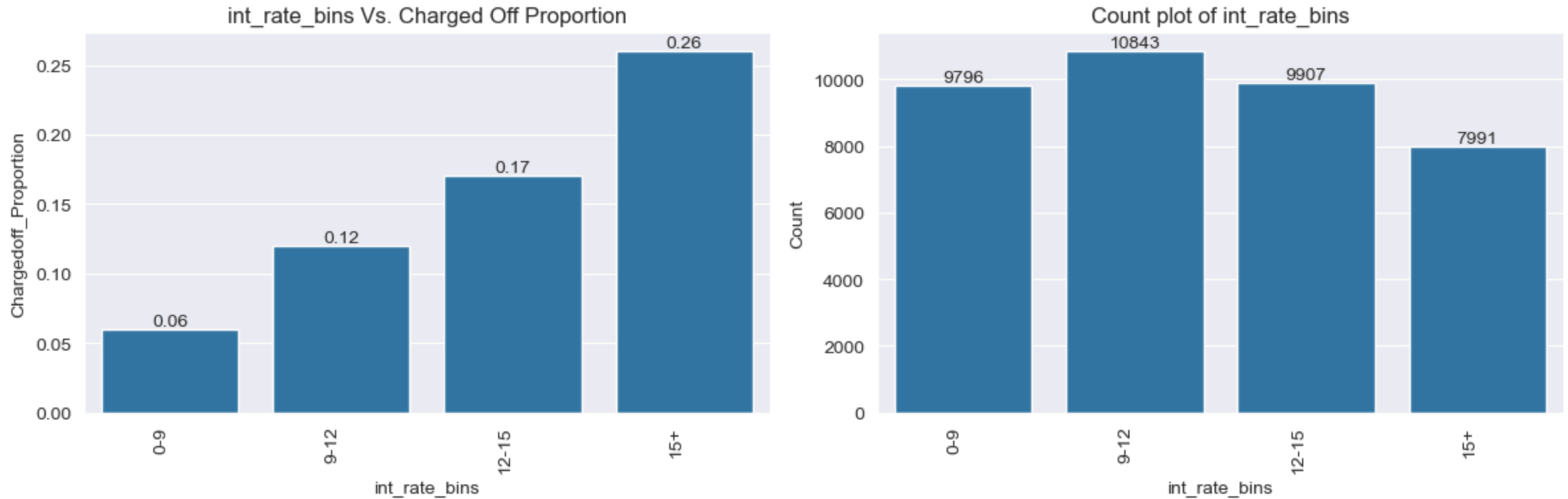

9-12% is the common interest rate. Default probability increases as interest rate increases. 15+ interest rates are risky.

Bivariate – Charged Off Proportions - Loan status vs Annual Income Bins

**Rs. 40,000 - 60,000 is the common annual income. Default probability reduces as annual income reduces. Rs. 0 – 40000 has higher default probability.**

Bivariate – Charged Off Proportions - Loan status vs Debt to Income Ratio Bins
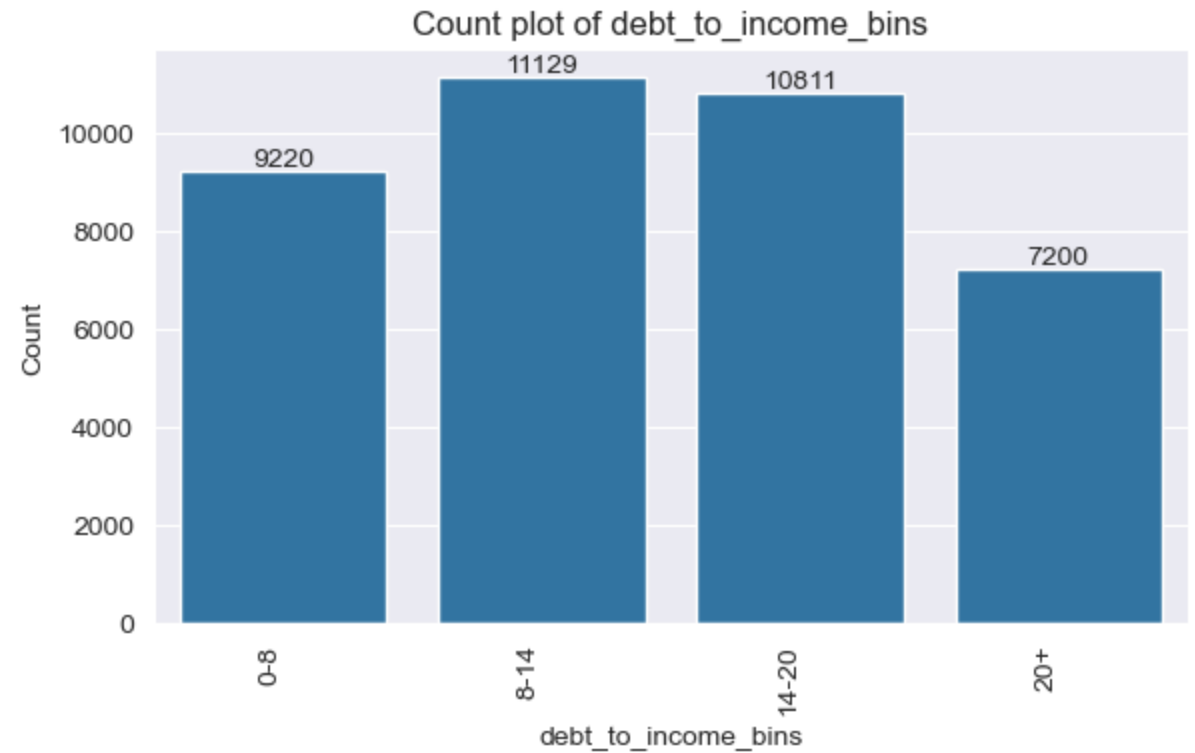
8-14 is the most common debt to income ratio. As debt-to-income ratio increases, default probability increases. 20+ debt to income ratio is risky.

Segmented Analysis – Segmented Loan Purpose
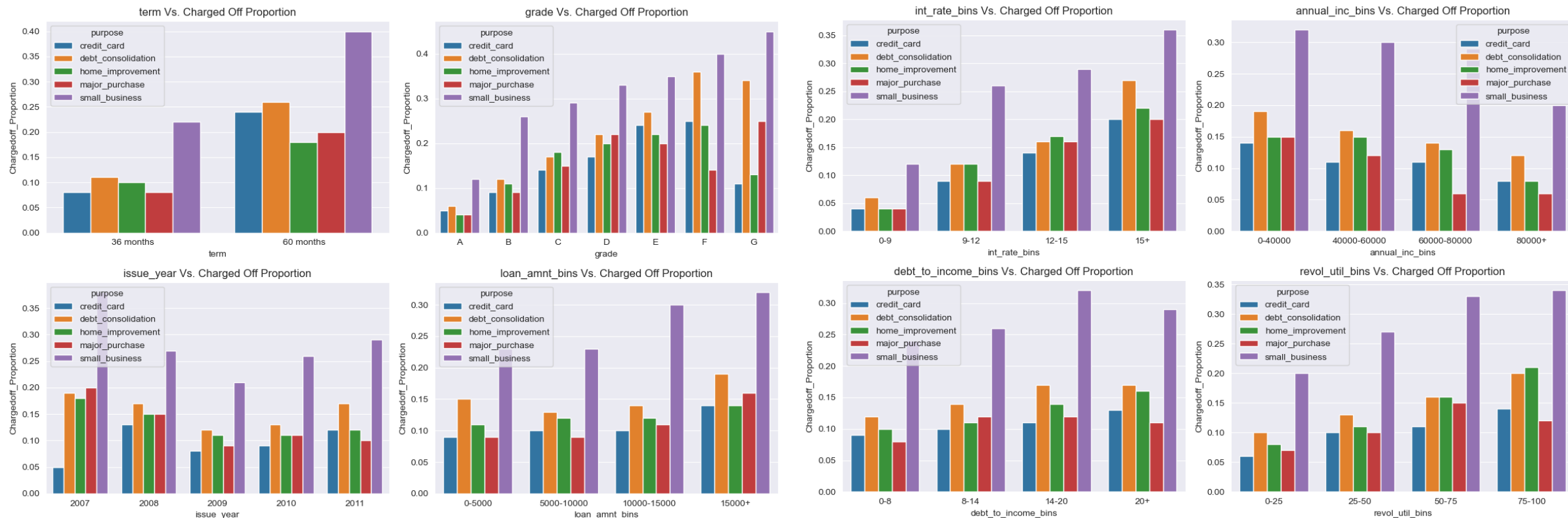
Among all the plots, observations are consistent across different segments.
Small business loans are having higher default ratios. Debt consolidation is next to it.
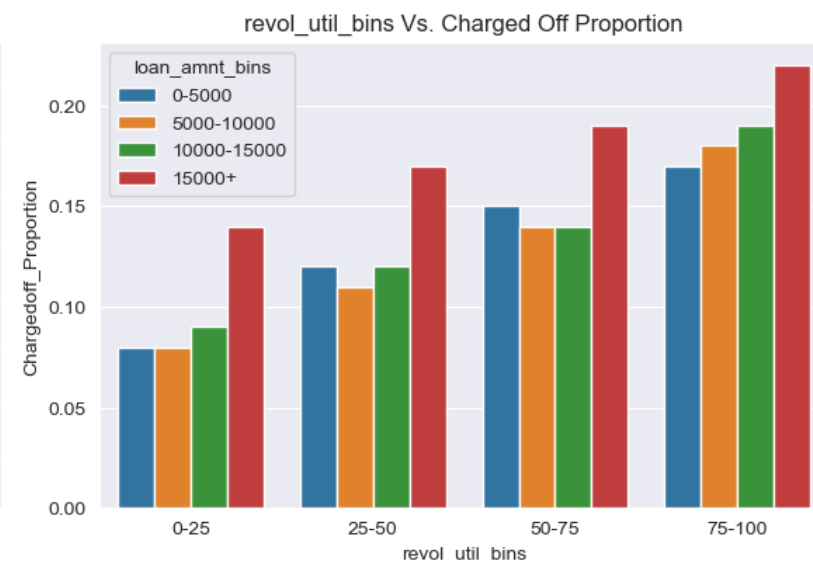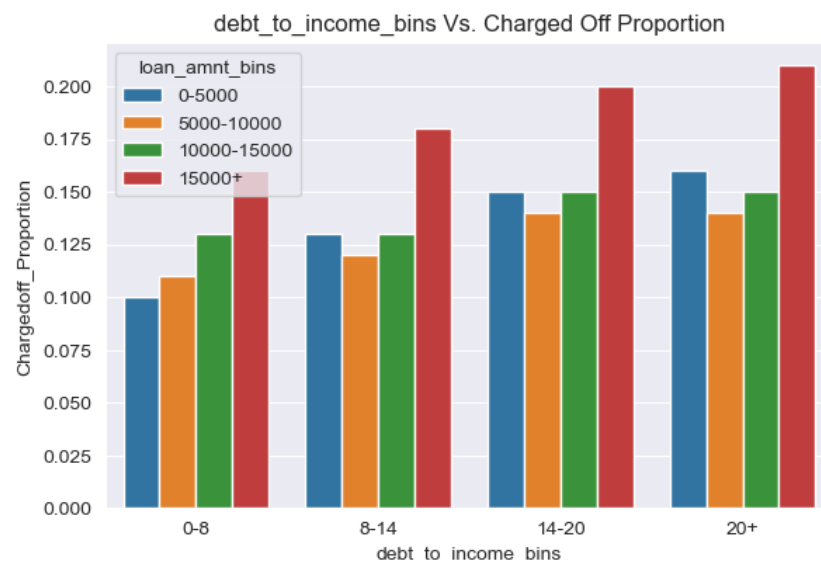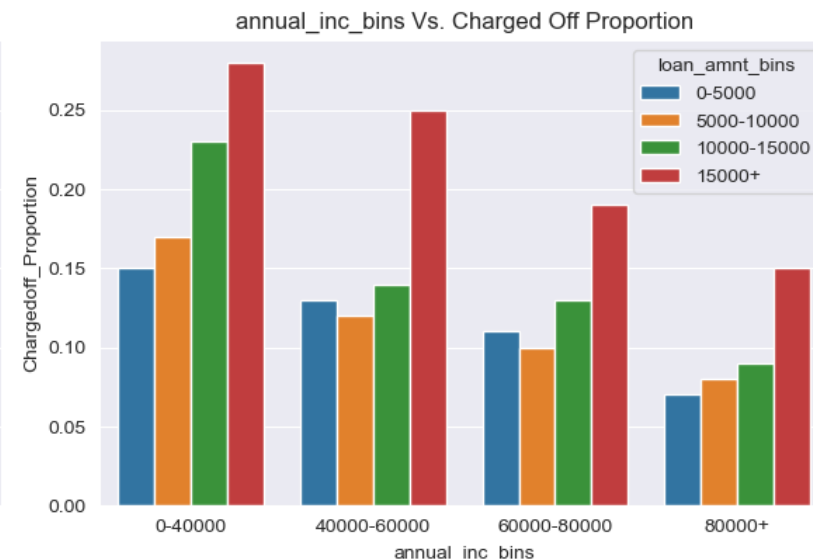
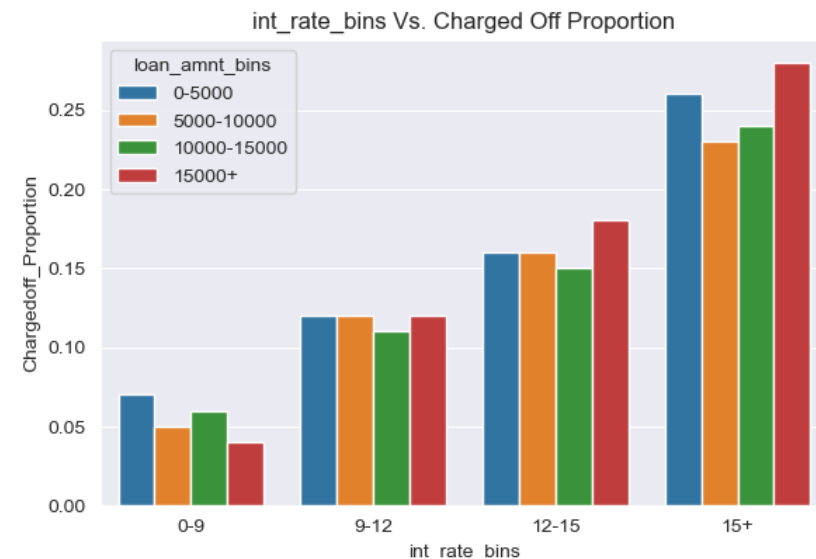# Segmented Analysis – Segmented Loan Amount Bins

Higher Default Probability Ranges:

Rs. 15000+ loan amount and

- ❑ **15%+ interest rate**
- ❑ **0-40000 Annual income**
- ❑ **20+ debt to income ratio**
- ❑ **75-100% revolving credit utilization**
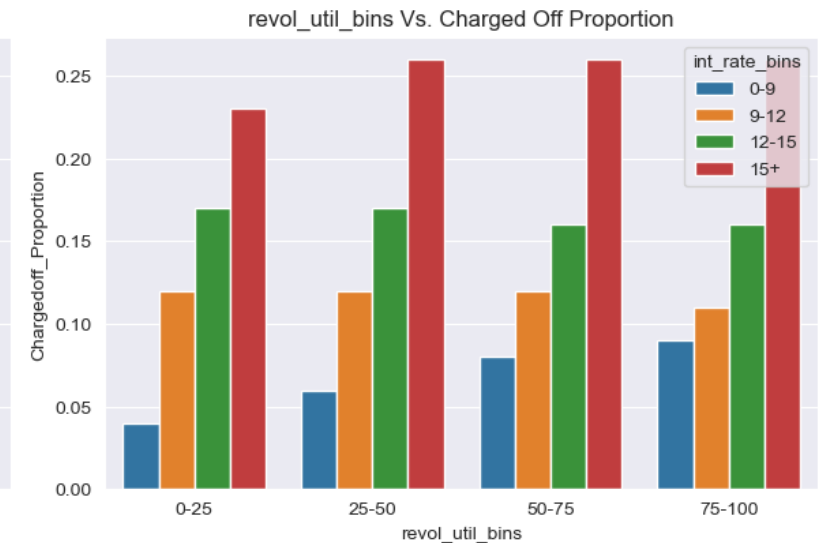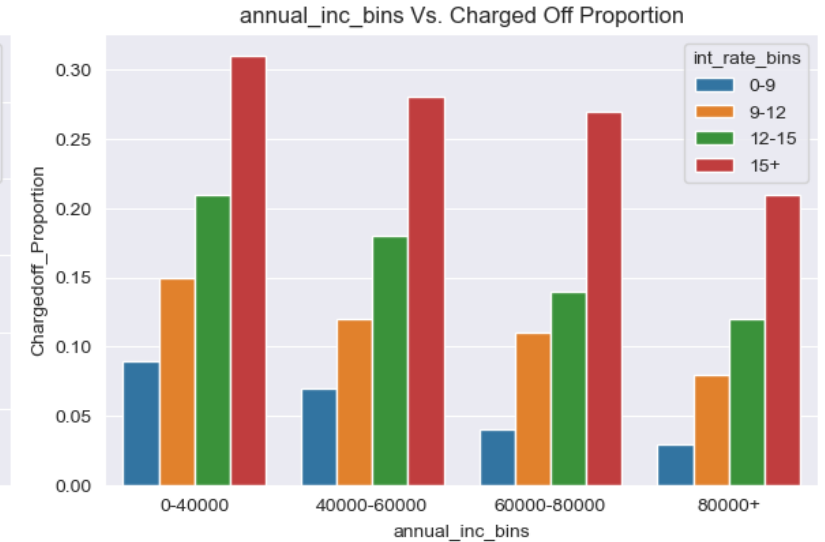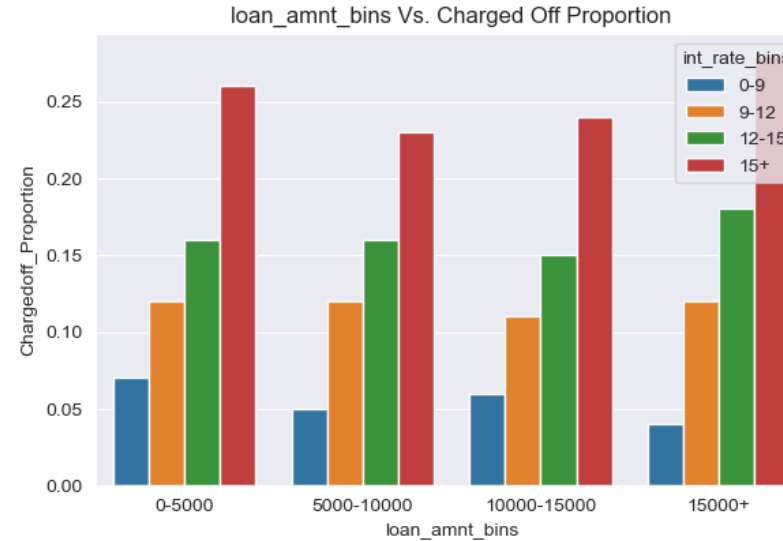
observations are consistent across different segments

Segmented Analysis – Segmented Interest Rate Bins

Higher Default Probability Ranges:

Rs. 15%+ interest rate and

- ❑ **0-40000 Annual income**
- ❑ **20+ debt to income ratio**
- ❑ **75-100% revolving credit utilization**

observations are consistent
across different segments

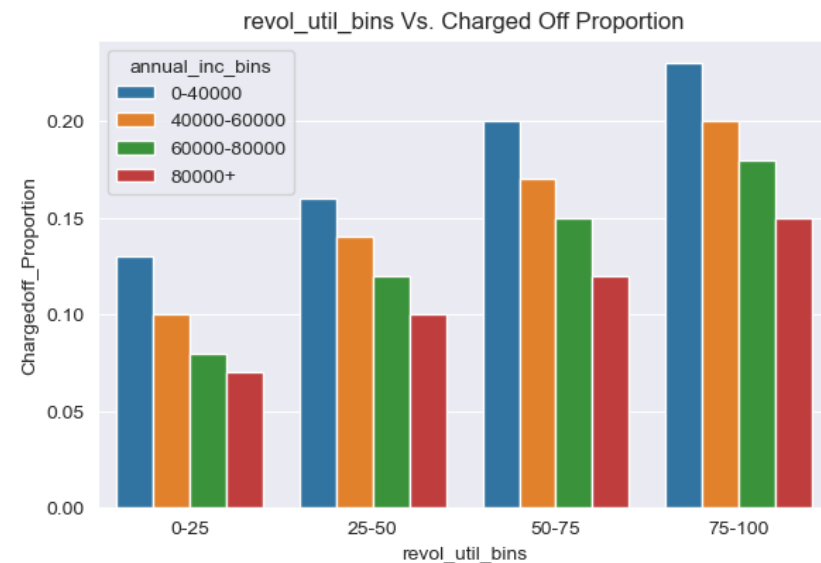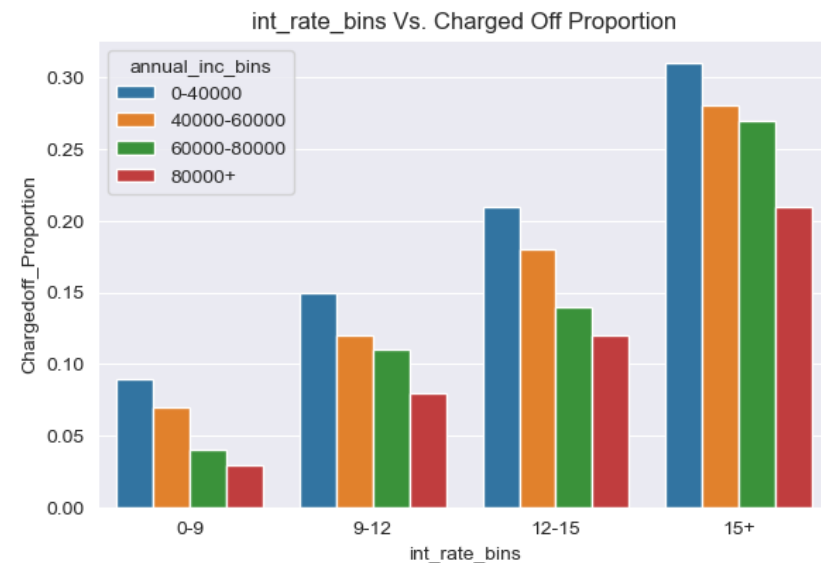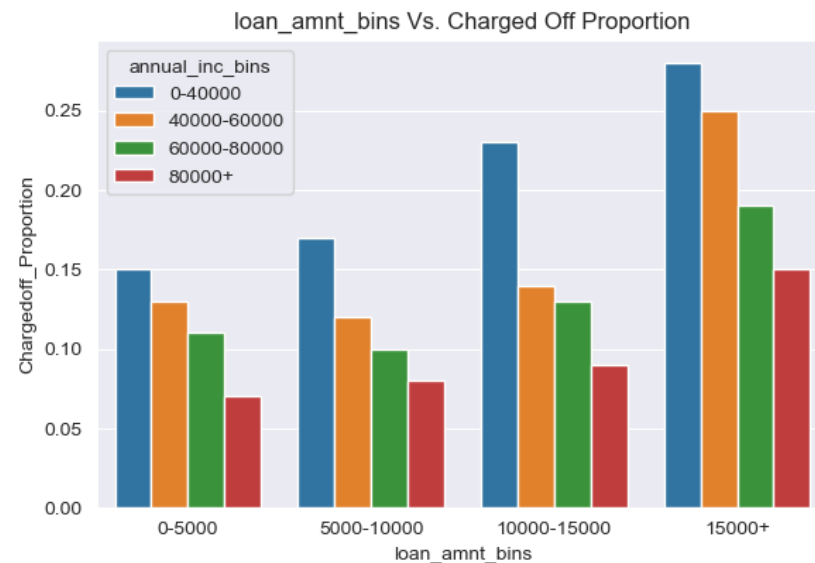# Segmented Analysis – Segmented Annual Income Bins

Higher Default Probability Ranges:

Rs. 0-40000 Annual income and

❑ **20+ debt to income ratio**

❑ **75-100% revolving credit utilization**

observations are consistent across different segments

Multi variate Analysis – Correlation Plot

Observations

- ❑ Loan amount, funded amount, funded amount invested are strongly correlated.
- ❑ Loan amount is positively correlated to Annual income.
- ❑ loan amount is not correlated to debt-to-income ratio and revolving utilization.
- ❑ Interest rate is higher for people with high revolving credit utilization.
- ❑ Interest rate is not correlated to annual income, revolving credit balance, total and open credit accounts.
- ❑ Annual income is correlated positively to revolving balance. And it is not correlated to revolving utilization.
- ❑ Revolving utilization is not correlated to total and open accounts.

>0.4 and <-0.4 is used as reasonable positive or negative correlation.

-0.1 to 0.1 is used for no correlation.
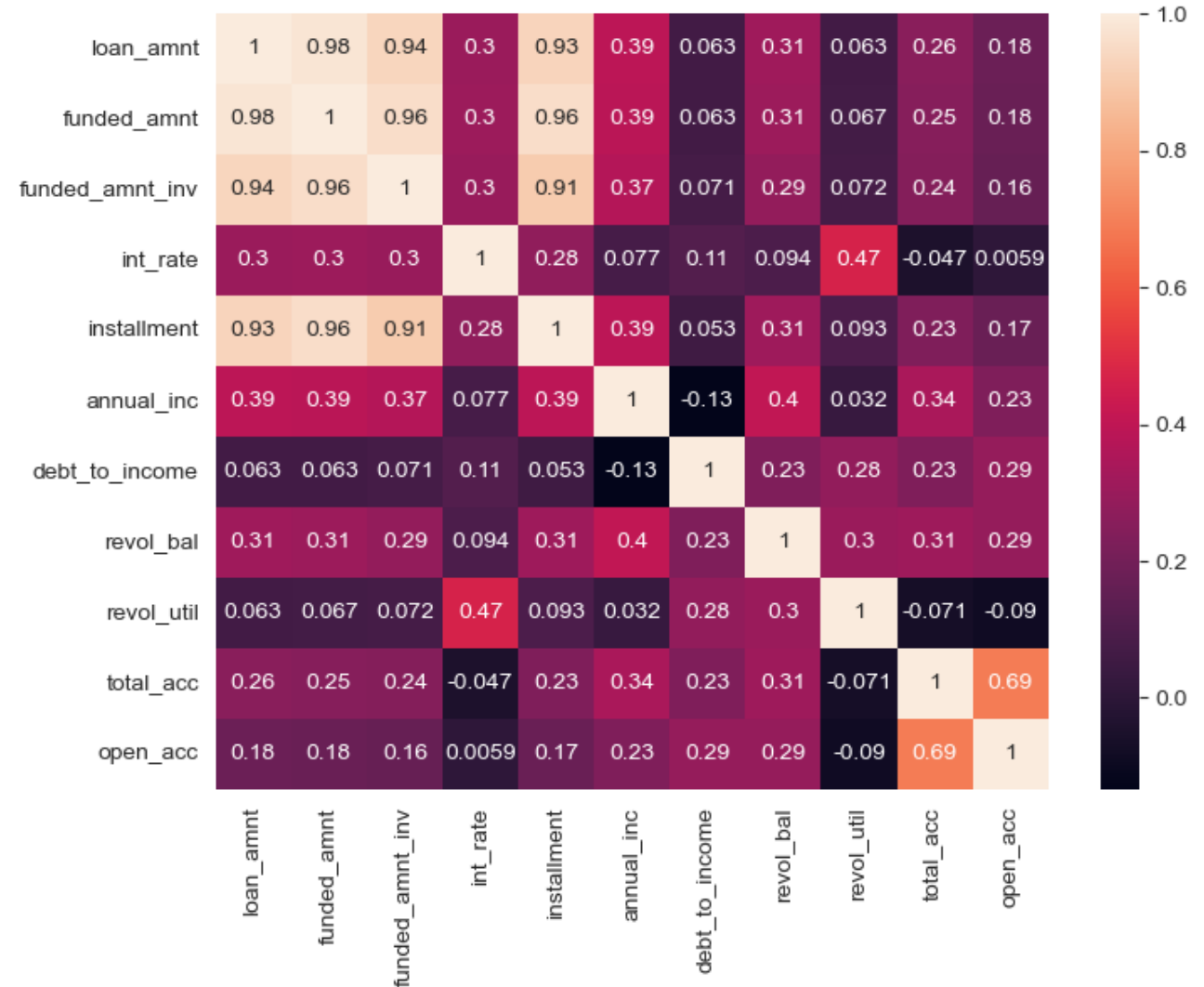
Summary & Recommendations

Summary & Recommendations

- ❑ **Loan default rate** is **14.5%**.

- ❑ Loan Purpose: Top items are debt consolidation, credit card, home improvement, Major Purchase, Small Business.

- ❑ Loan Term: **60 months** has higher probability of default.

- ❑ Loan Grade: Probability of default increases as grade **increases** from **A to G**. LC has done good job in deciding grades.

- ❑ Loan Purpose: **small business and debt consolidation** has higher default probability.

- ❑ Issue date: As year progress from 2007 to 2011, loans issued are increased. From **2007 to 2009** charged off proportion decreased. After that, it increased again.

- ❑ Loan Amount: Rs. 5000 -10000 is the common loan amount. But **15000+ loans** has higher probability of default.

- ❑ Interest rate: 9-12% is the common interest rate. Default probability increases as interest rate increases. **15%+ interest rate** has higher default rate.

- ❑ Annual Income: Rs. 40,000 - 60,000 is the common annual income. Default probability increases as annual income reduces. Rs. **0-40,000 annual income** has higher default rate.

- ❑ Debt to income ratio: 8-14 is the most common debt to income ratio. As debt-to-income ratio increases, default probability increases. **20+ debt to income ratio** has higher default probability.

- ❑ Revolving credit utilization: 25-75% is the common revolving credit utilization rate. As credit utilization increases, default probability increases. **75% - 100% revolving credit utilization** has higher default probability.