

# GRAMENER CASE STUDY

Presented By (Group Members):

1. Abhishek Singhal
2. Hemant Badgujar
3. Jetendra Mulinti
4. Prem Kumar R

# Purpose Process Payoff

## **Purpose:**

- Understand the risk analytics in banking and financial services and use the data provided to minimise the risk of financial loss while lending loan to customers.

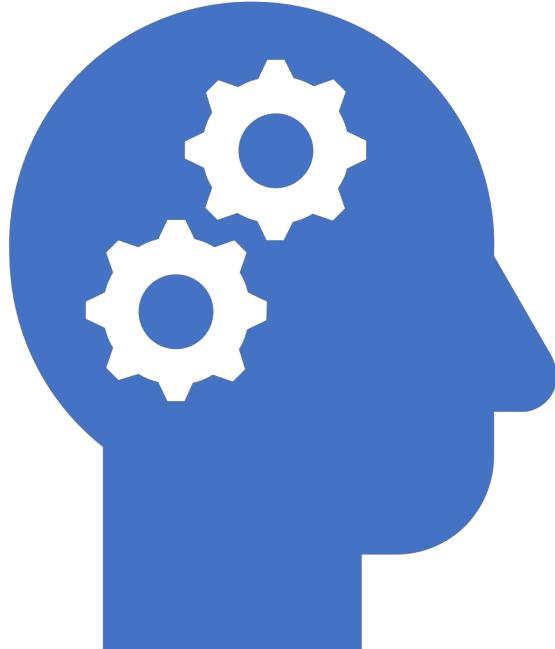
## **Process:**

- Identification of Applicants which can turn into defaulter at the latter stage.
- Identification of Parameters which contributes to knowing the defaulters in paying off the loan

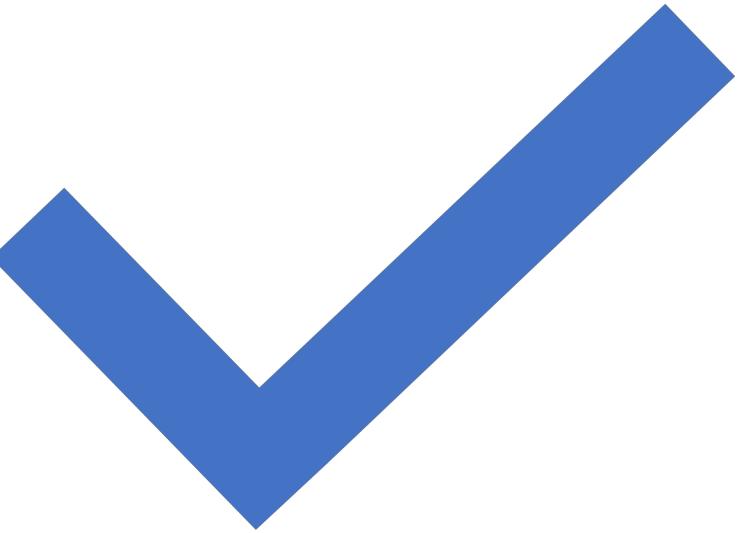
## **Payoff:**

- Analyzing the data and give the output or the key metrics which can be utilized to prevent the financial loss.

# Action Points



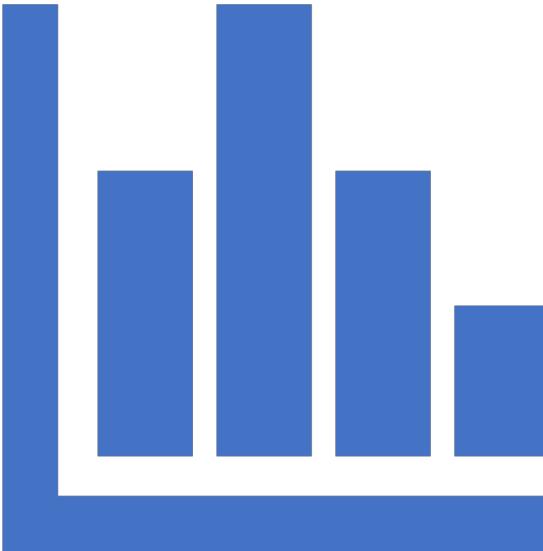
- Business Understanding- understanding the key points to be considered before providing any loan.
- Preparing Metadata- understanding each attributes from business perspective
- Data Cleaning- removing unwanted columns & rows and turning the data in the form in which it can be better analyzed.
- Analyzing the columns on which we can do the Exploratory Data Analysis.
- Univariate Analysis
- Bivariate Analysis



# Cleansing of Data

- Dropping the columns having high percentage of null values in it
- Dropping the loan status – “Current” since we cannot decide based on current status customers if they will pay off the loan or be defaulters.
- Dropping the redundant columns which are not going to affect our analysis, e.g.
  - URL as they are not of any use until we have the login id and password.
  - Interest rate column since, we got Grade and Subgrade giving us the good idea about the interest rates
- Column such as employee title needed manual inspection- as the column values had identical values but due to some characters due to which it was considered as unique, which we identified using pivot in excel and hence we changed them.

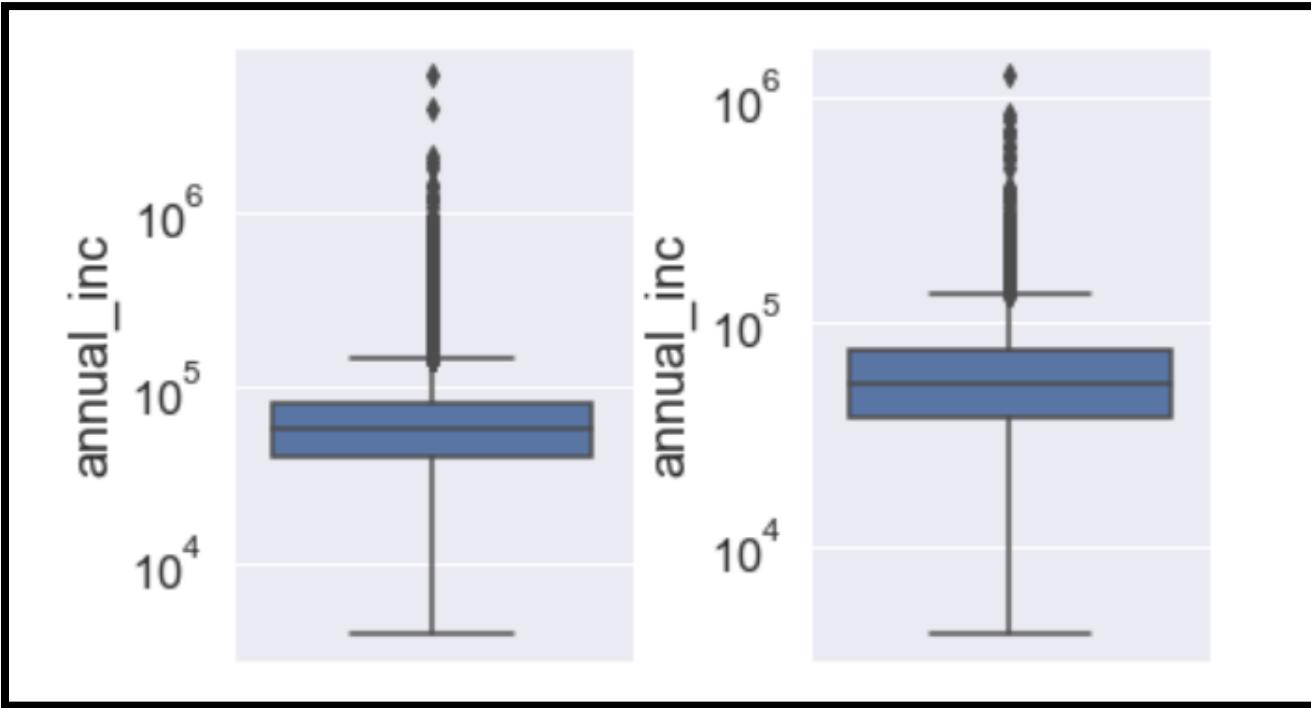
# Analysis



- Exploratory Data Analysis:
  - Univariate Analysis
    - Categorical Variable
    - Numeric Variable
    - Quantitative Variable
  - Segmented Univariate Analysis
  - Bivariate Analysis
    - Continuous variable
    - Categorical Variable

# Univariate Analysis:

## Analysis of Annual Income for Fully paid vs Charged off

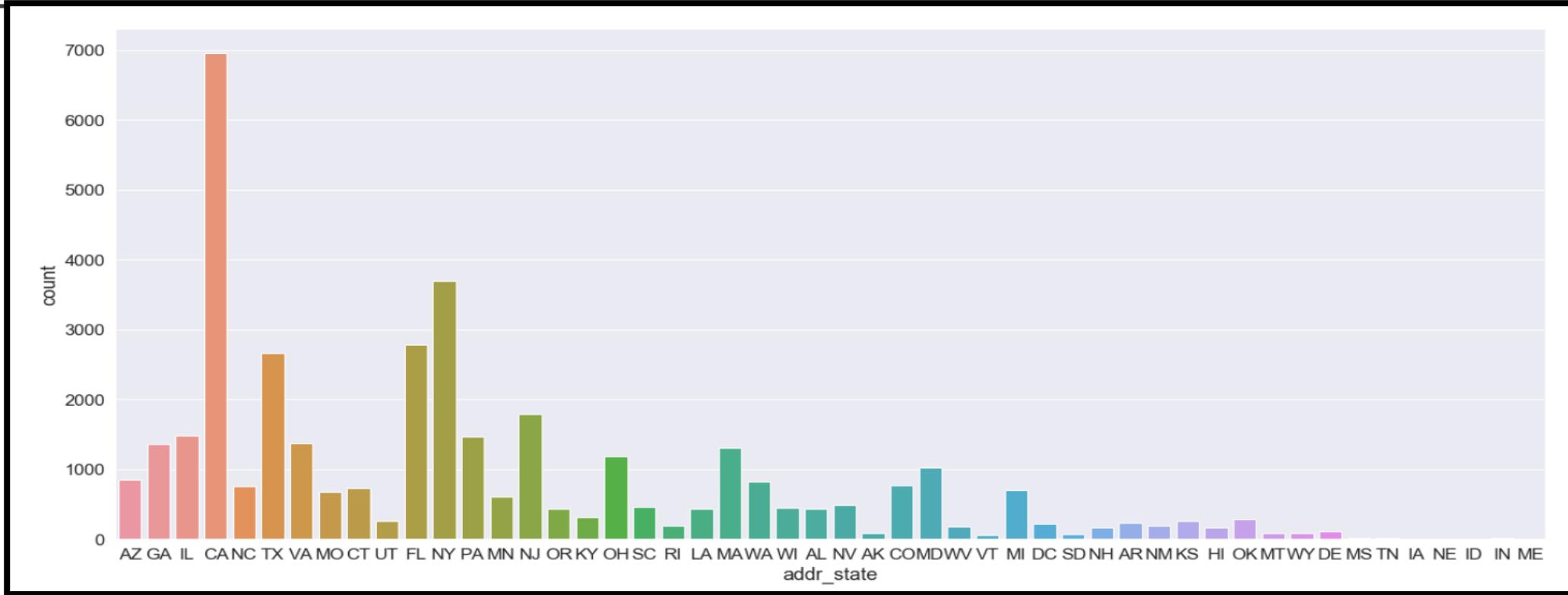


Understanding the spread of annual income data for loan statuses: “Fully Paid” & “Charged Off”. Since we’re going to do the analysis only on these two statuses as we’re not sure about the “Current” status customers as to whether they’ll completely pay off the loan or prove defaulter in later stage.

# Univariate Analysis:

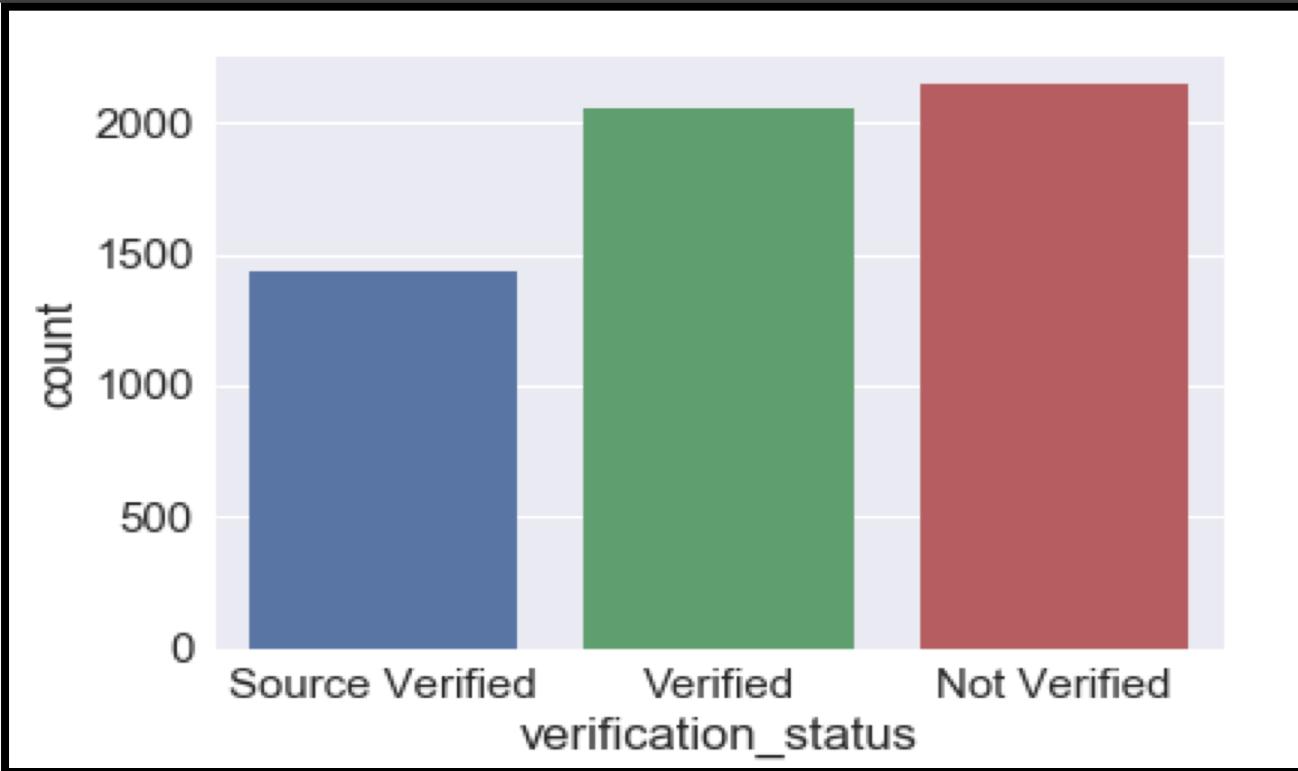
---

## Frequency of loans data – State wise



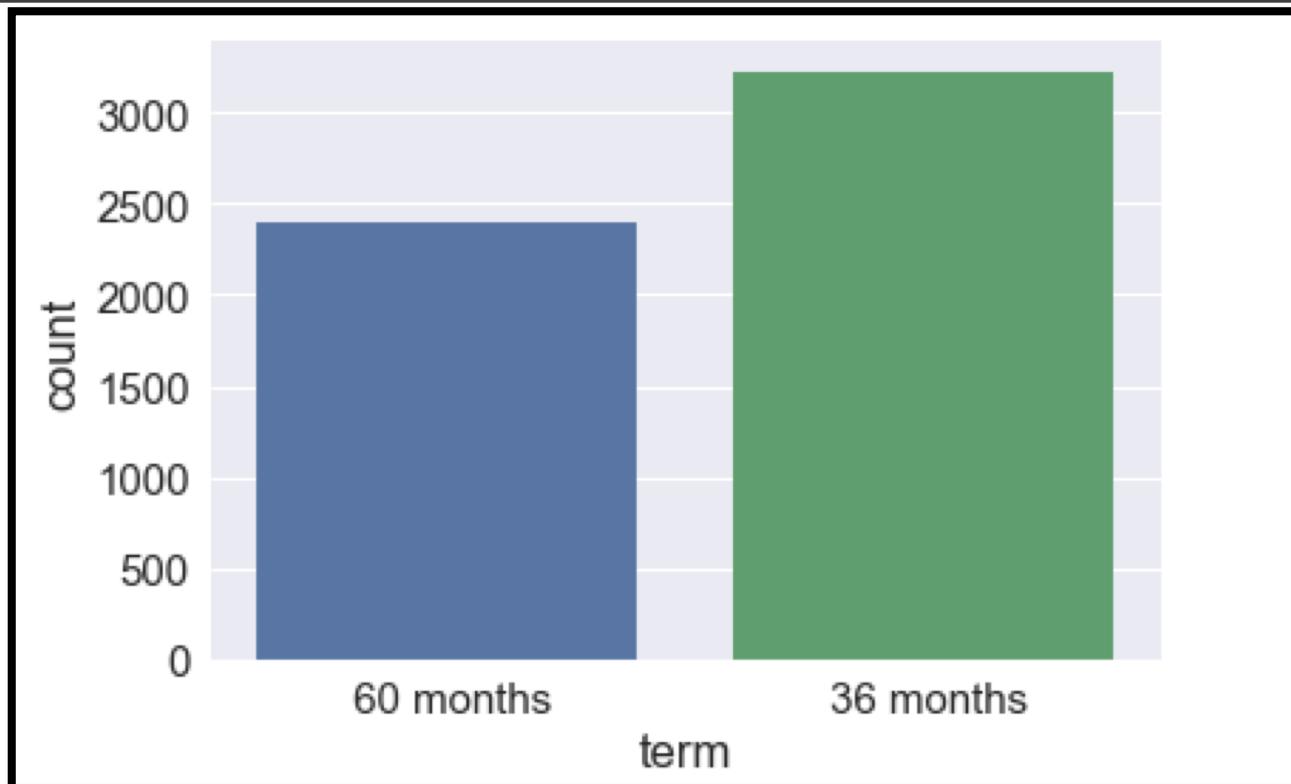
Observation- Most loan takers are based out of California, New York, Florida, New Jersey and Texas

# Univariate Analysis: Frequency of loans data – Verification Status wise



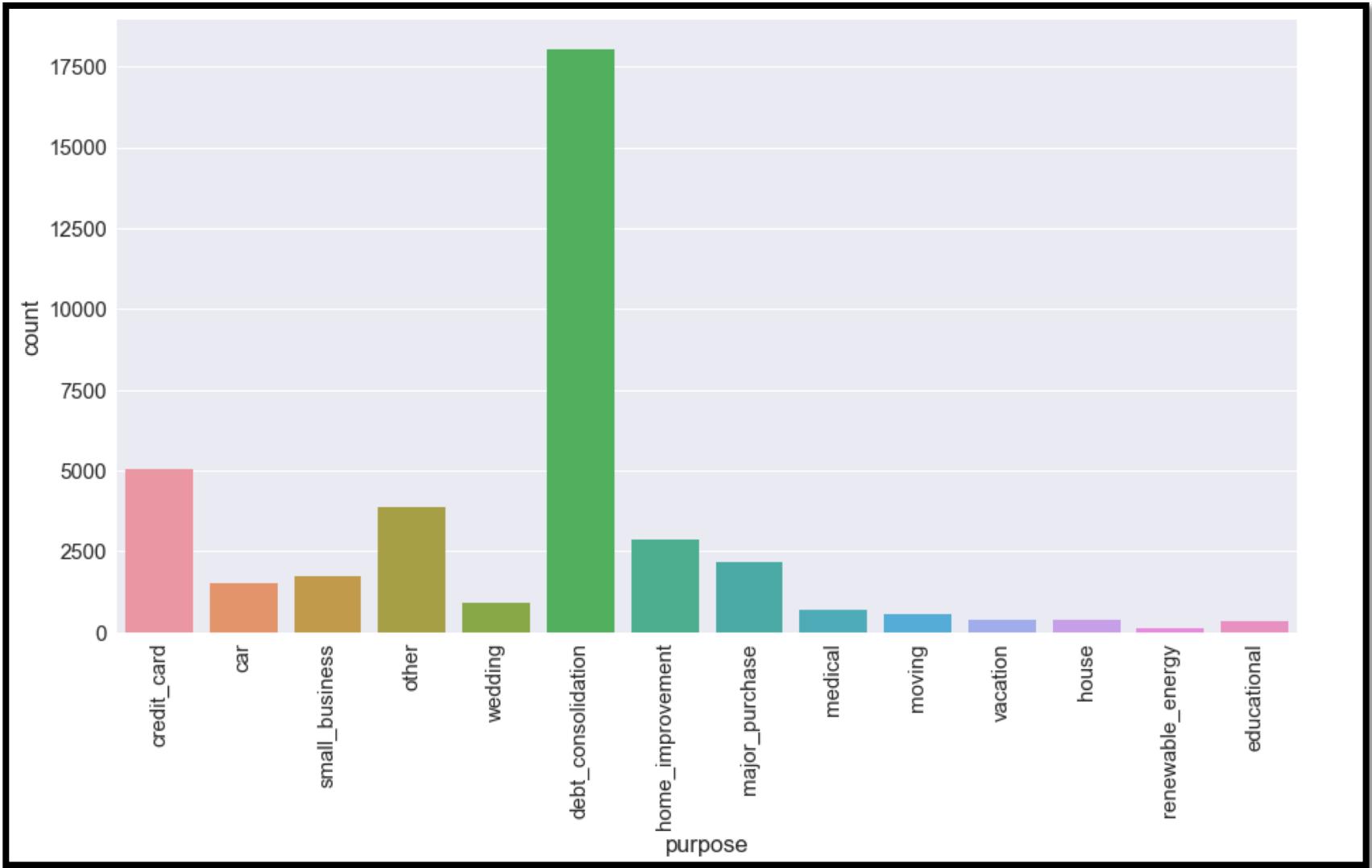
Observation: Most defaulters source was “Not Verified”

## Univariate Analysis: Frequency of loans data – loan term wise



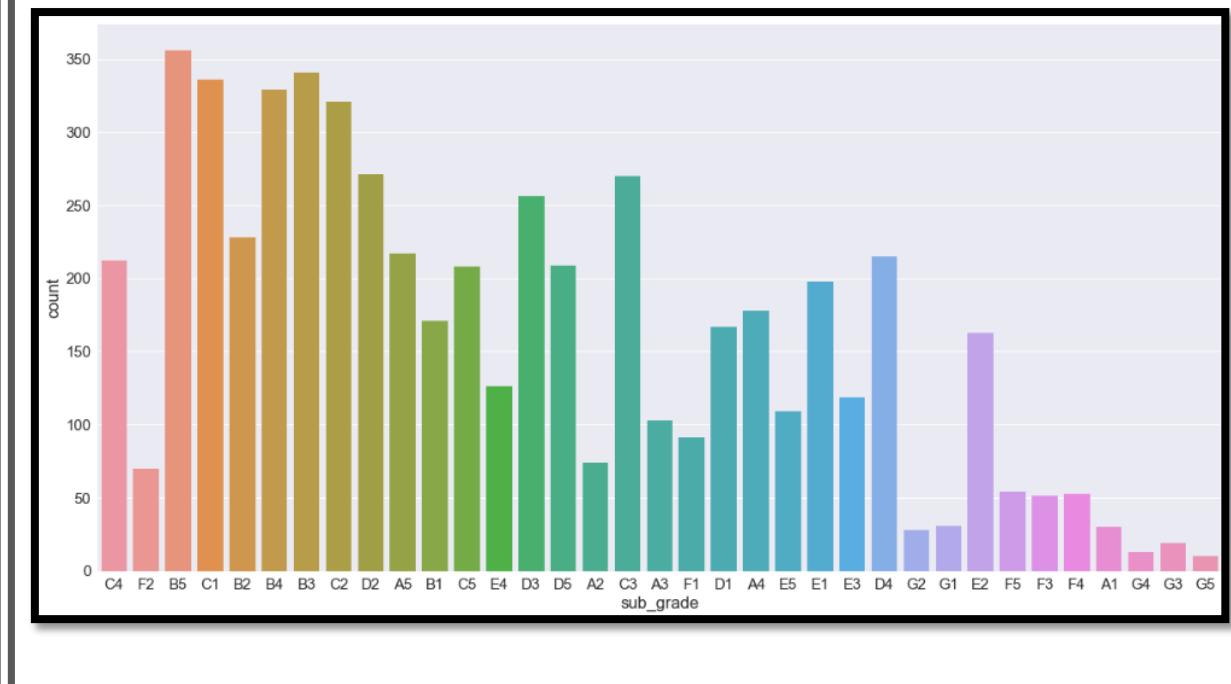
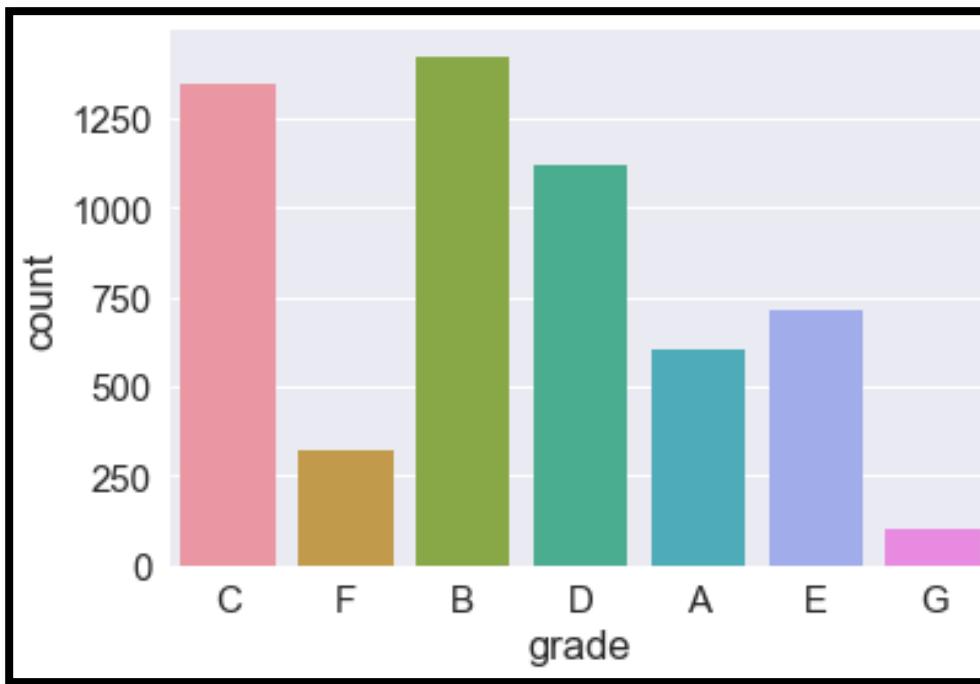
Observation: Most defaulters tenure of loan is 36 months

## Univariate Analysis: Frequency of loans data – Loan purpose wise



Observation- Most common reason of loan was debt consolidation

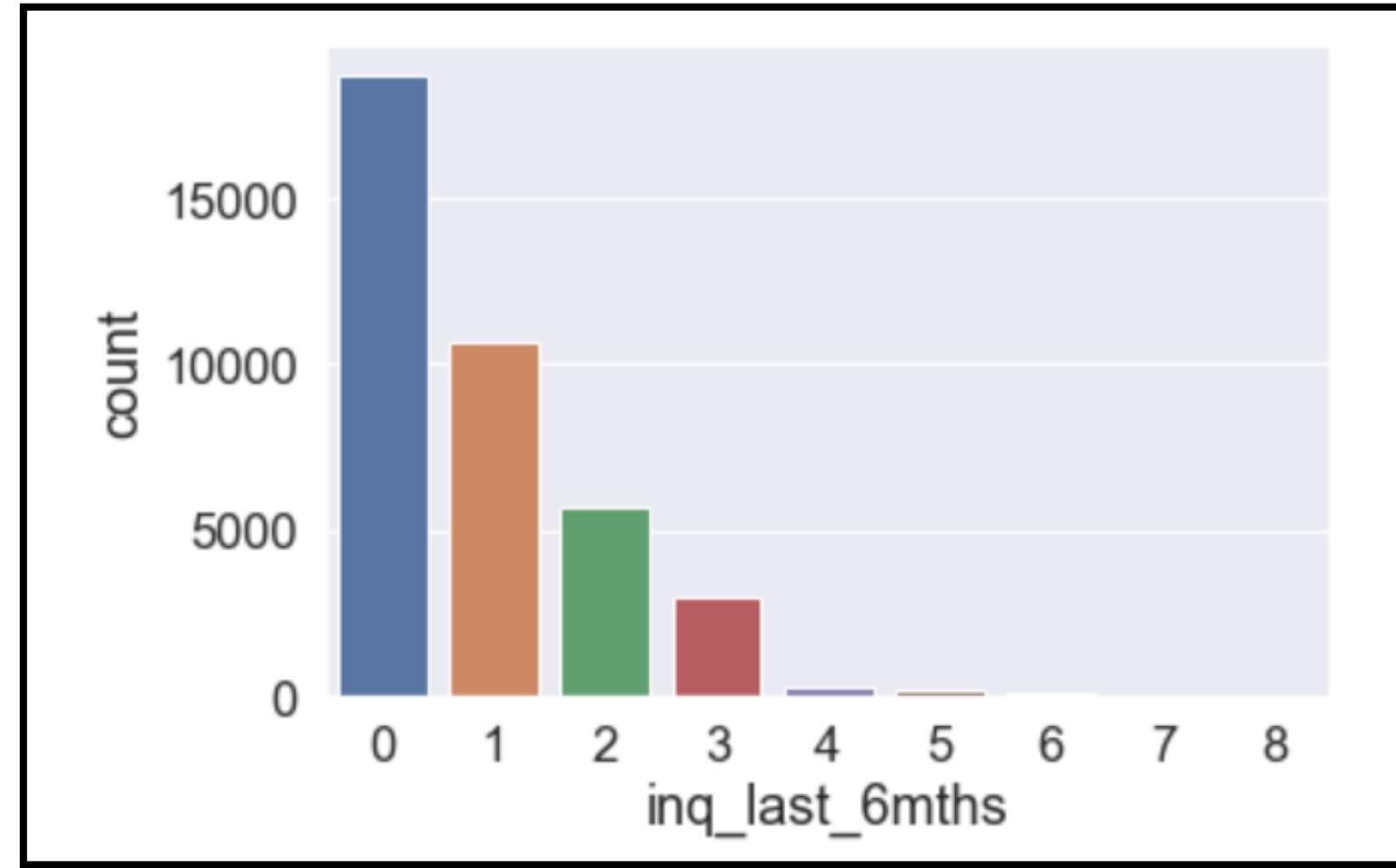
# Univariate Analysis: Frequency of defaulters data – Grade and Sub-Grade wise



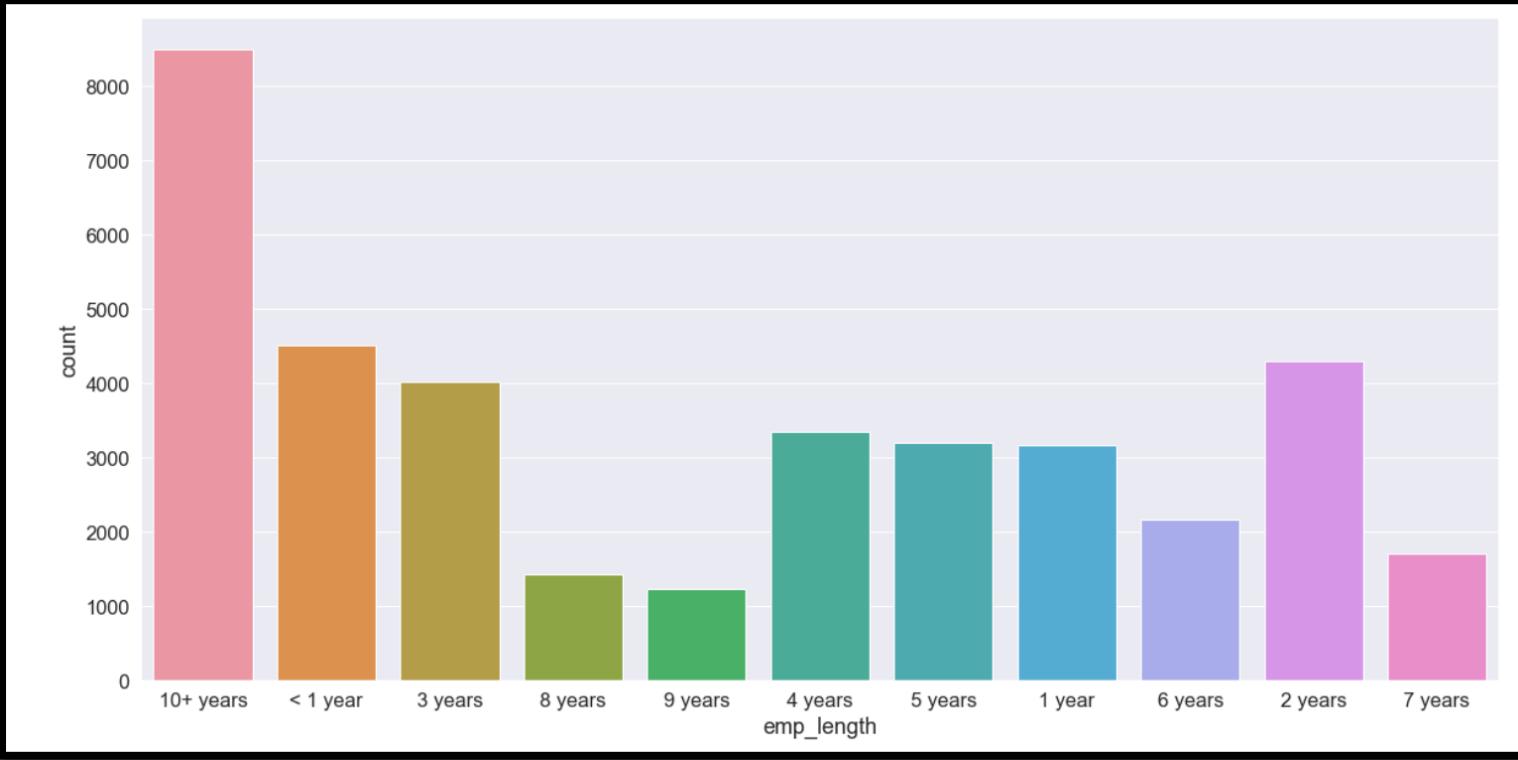
Observation - It appears that the most charged off cases are in B5, B3, C1, B4 & C2

Ref: <https://www.lendingclub.com/public/rates-and-fees.action>

Univariate  
Analysis:  
No of Inquires  
in last 6  
months

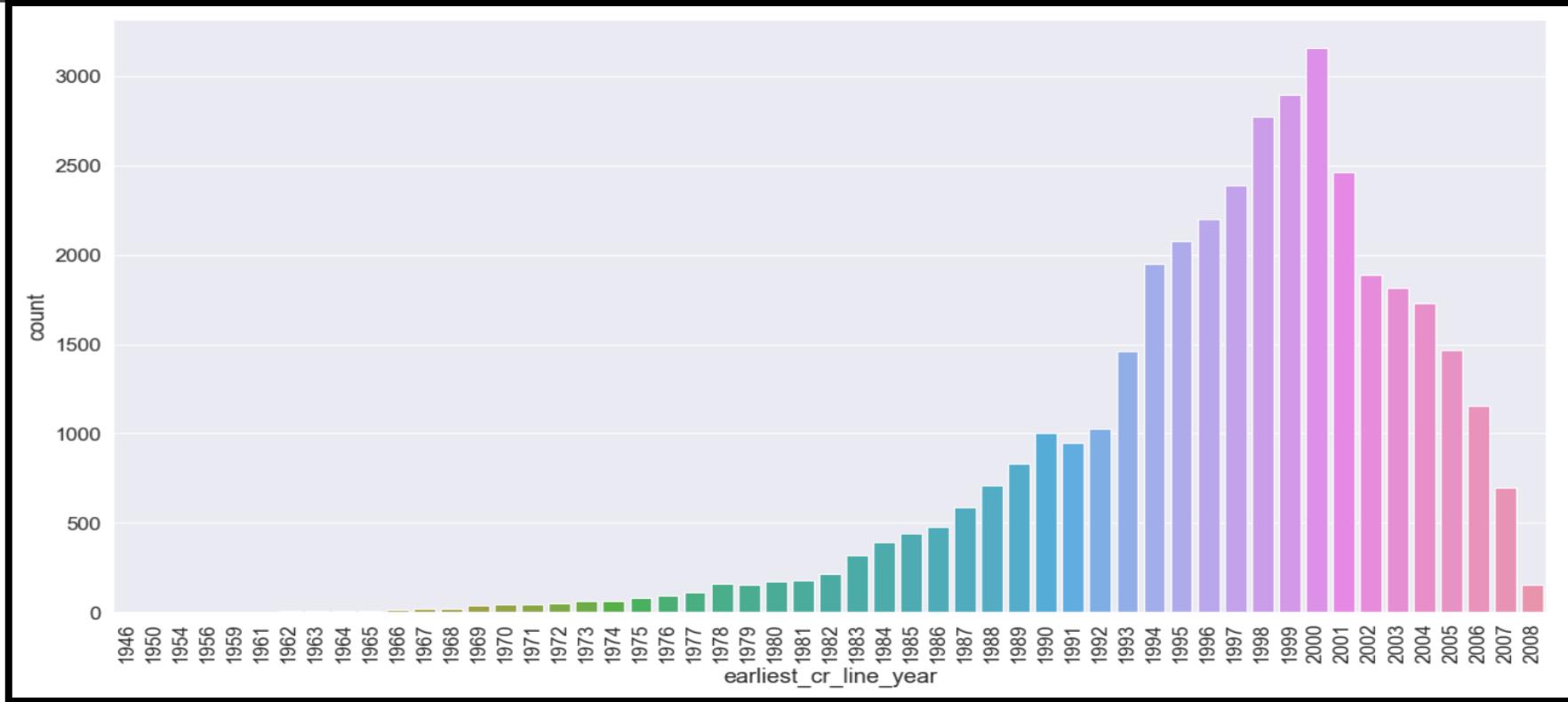


# Univariate Analysis: Frequency of loans data – employee length at Job wise



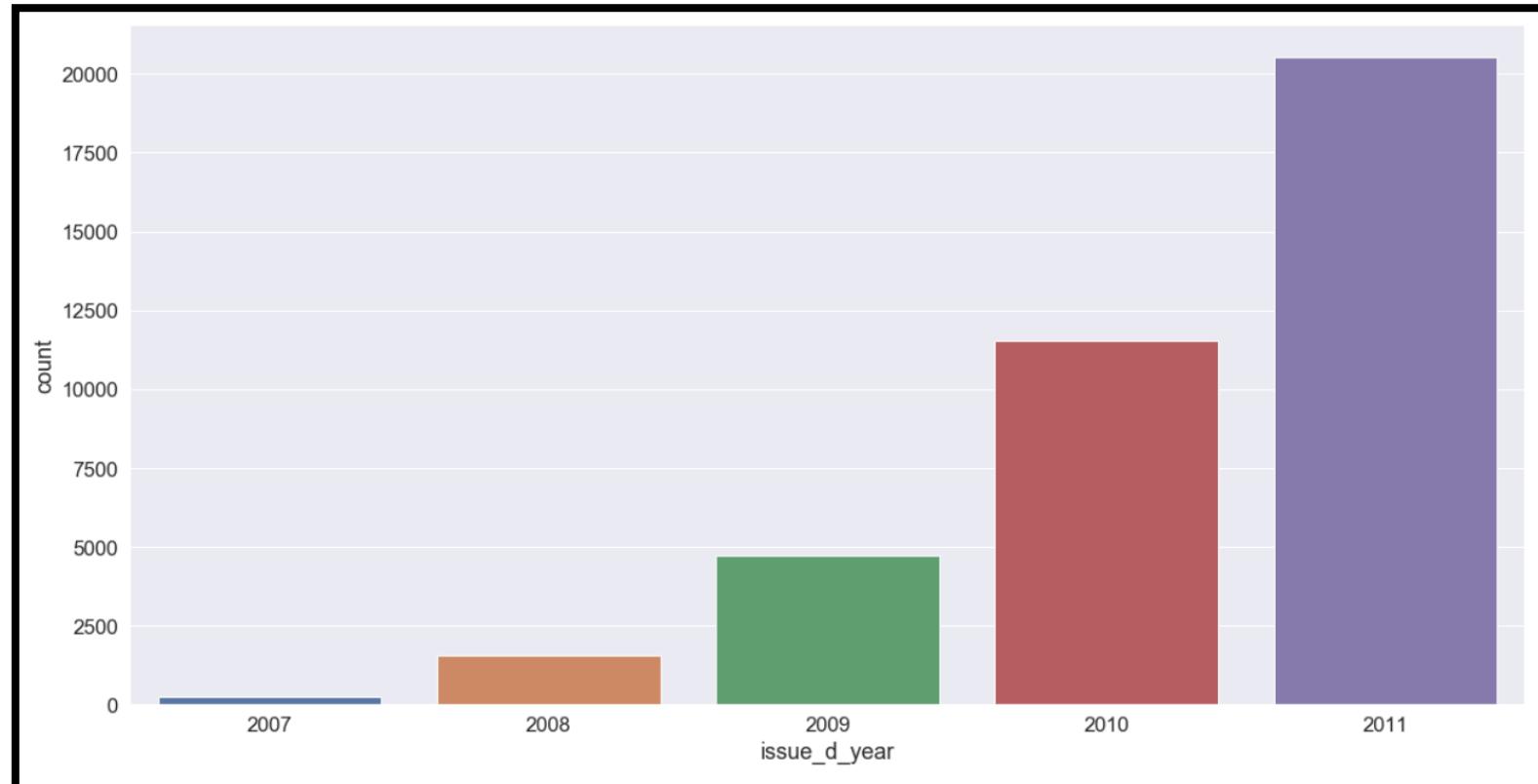
Observation - Most loan takers are employed for more than 10 years

# Univariate Analysis: Frequency of loans data – earliest credit line wise



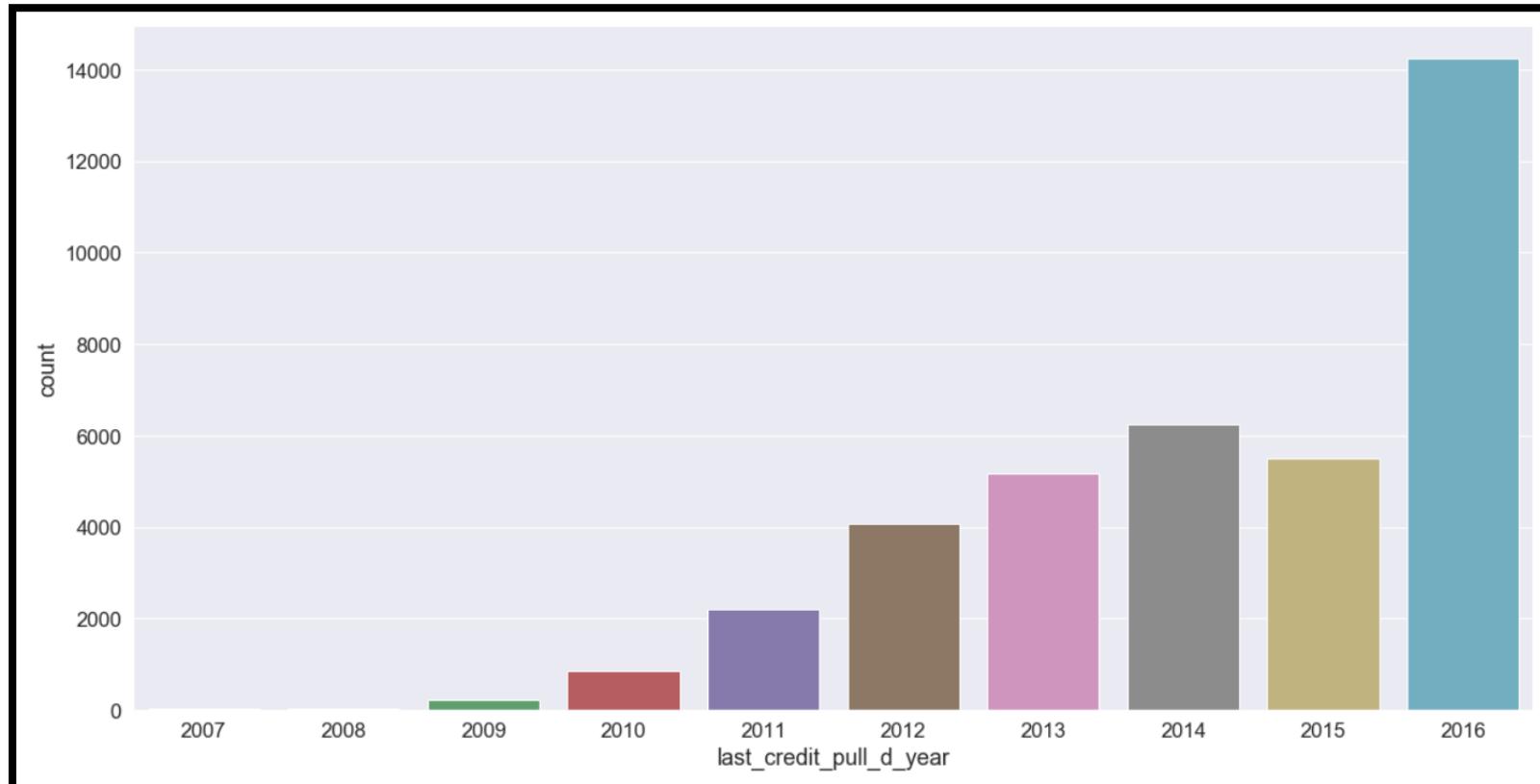
Observation: Early Credit Lines seems to be increasing until 2000; later it appears that fewer Credit Lines were opened.

# Univariate Analysis: Frequency of loans data – loan issued date wise



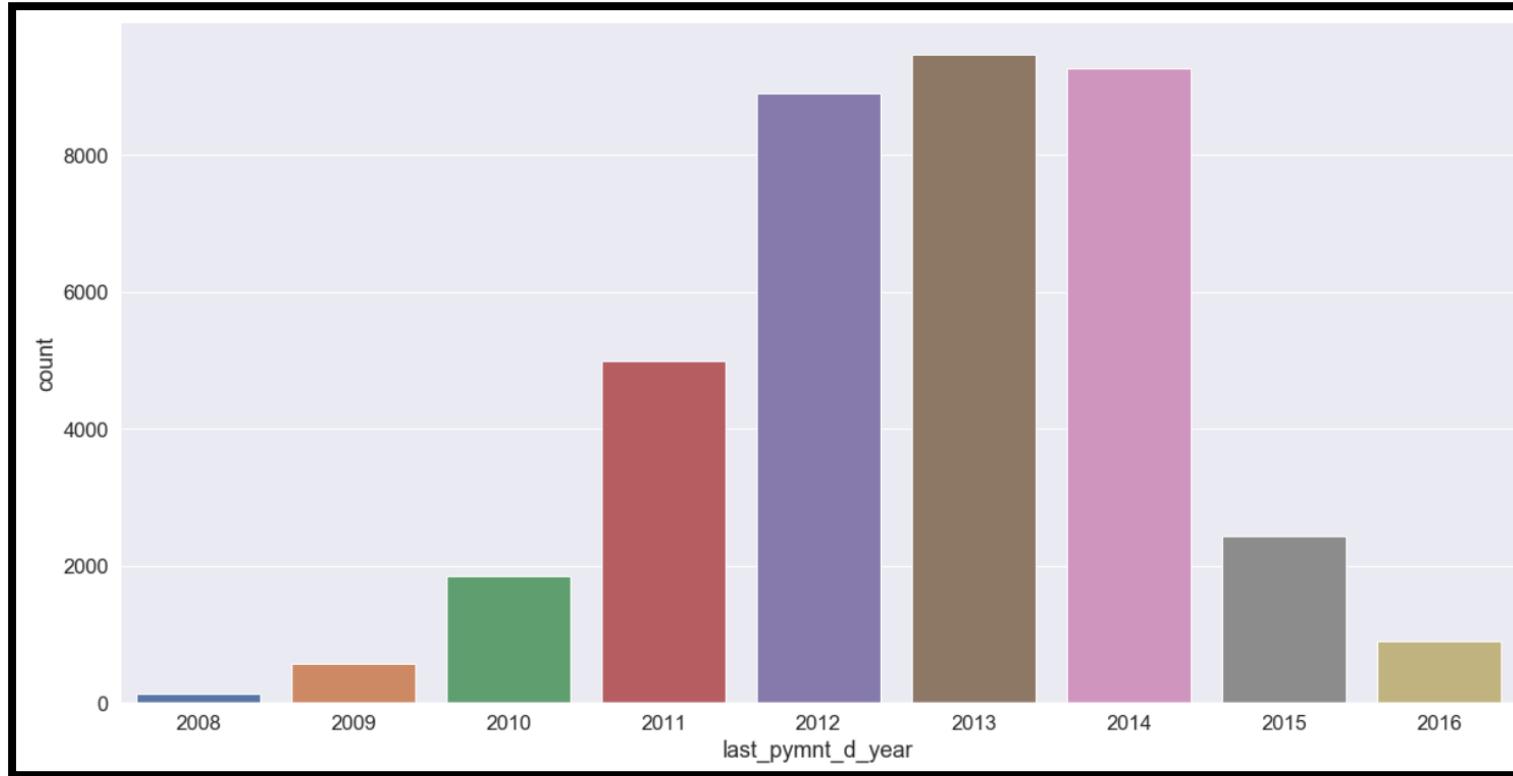
Observation – increasing trend for loan issued yearly

# Univariate Analysis: Frequency of loans data – last credit pull date wise



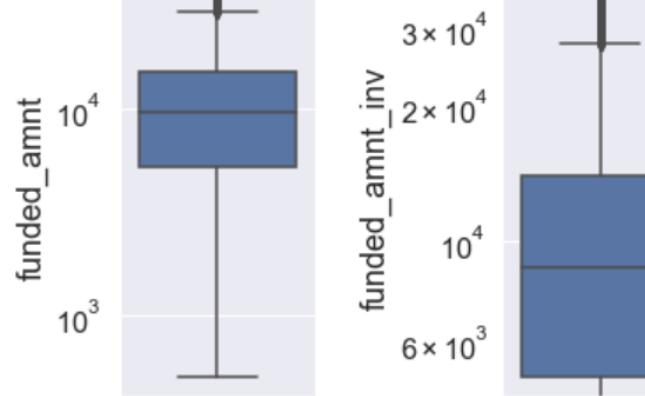
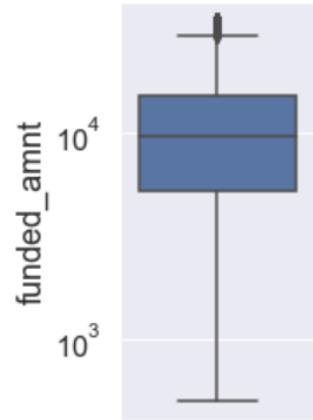
Observation: increasing trend in the no. of credit pulls by creditors every year

# Univariate Analysis: Frequency of loans data – last payment date wise

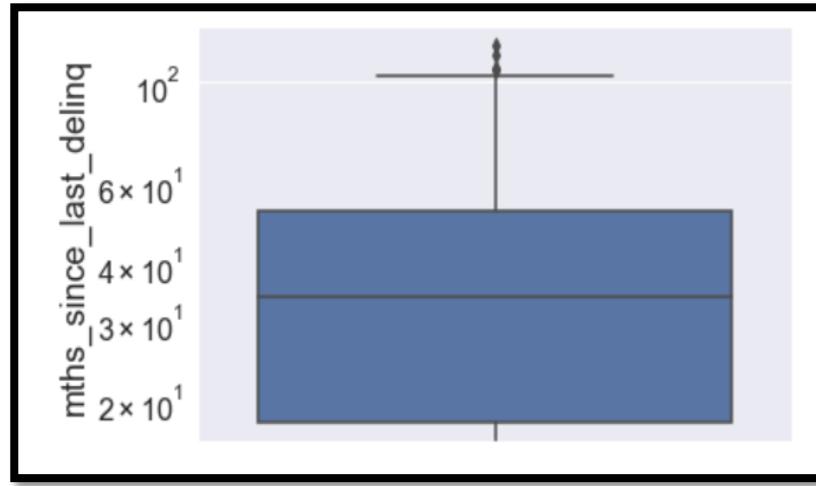


Observation: last payment date seems to be at peak in 2012 to 2014

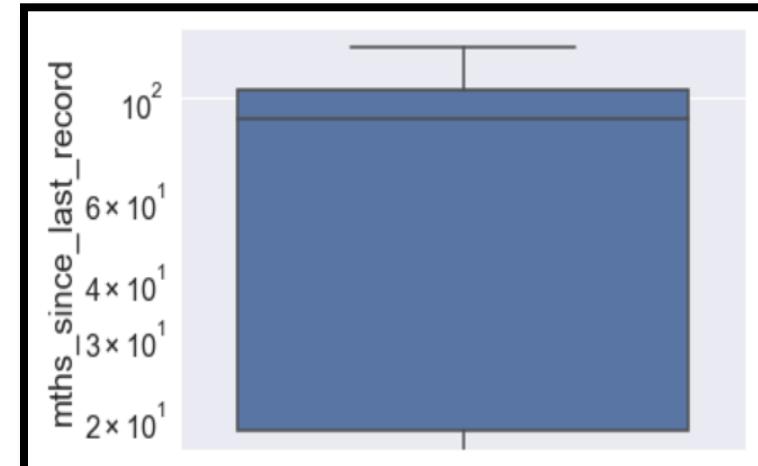
## Box plots to understand the spread of data



Funded Amount & Funded  
Amount invested

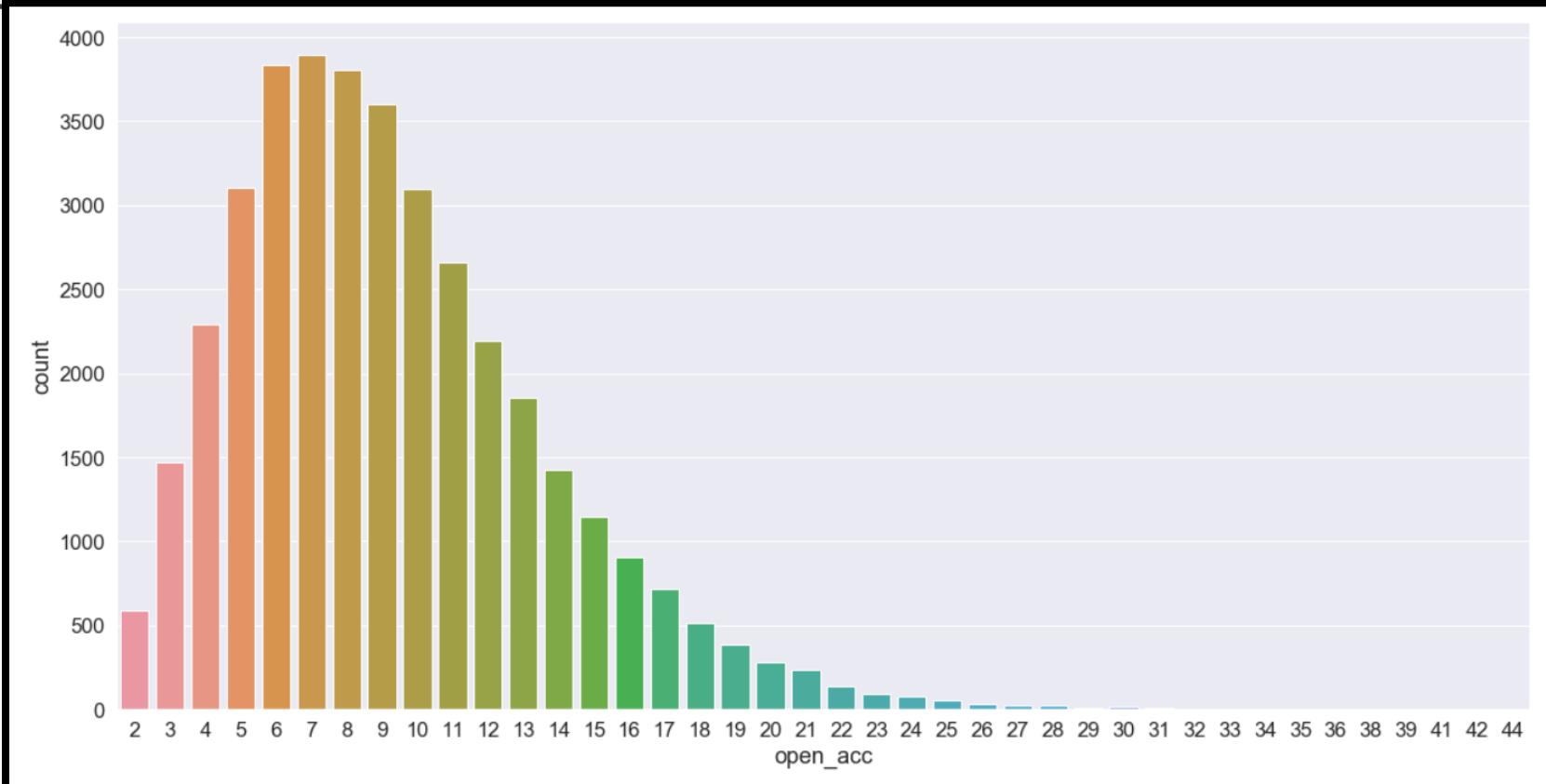


Months since  
last delinquency



Month since last  
record

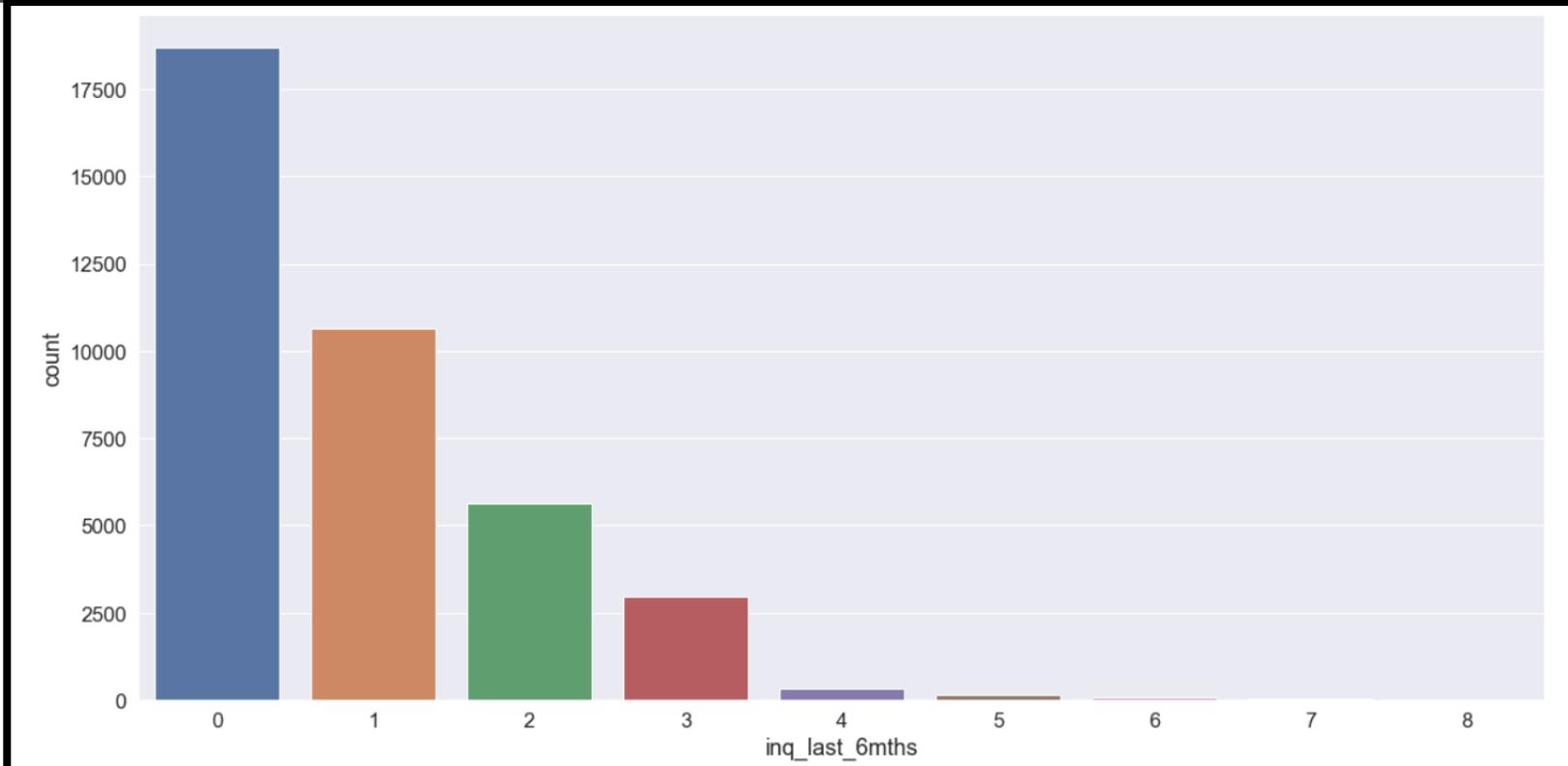
# Univariate Analysis: Frequency of loans data – number of open credit lines wise



Observation: number of opened credit lines for borrowers seems high at the count of 7 & 8

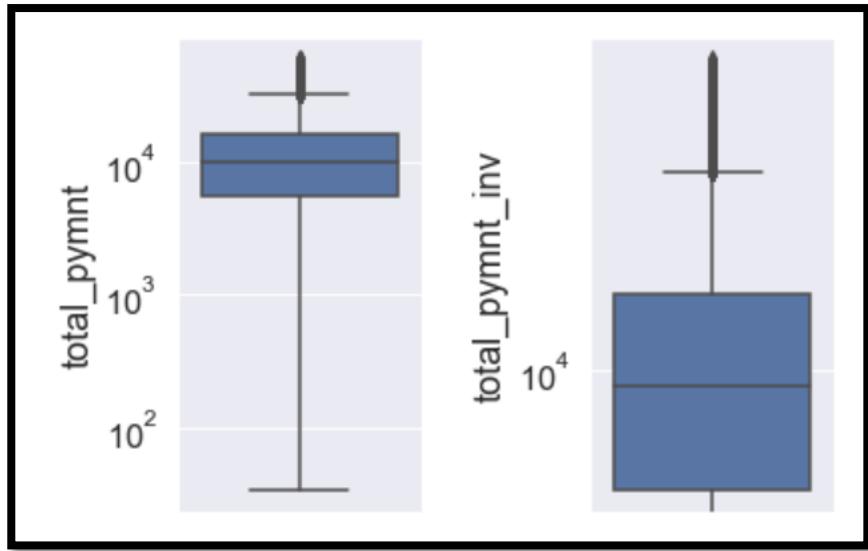
## Univariate Analysis:

Frequency of loans data – number of inquiries in last 6 months wise

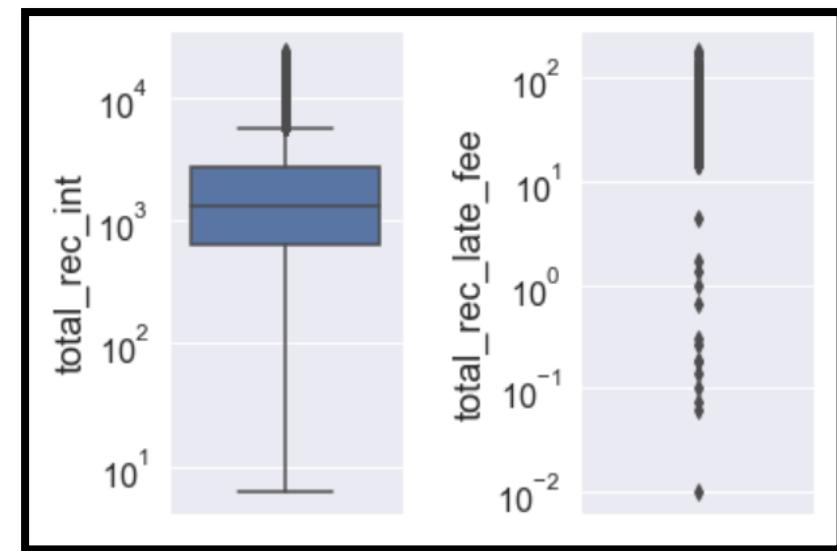


Observation: there are high number of 0 values signifying no inquiries in last 6 months

## Box plots to understand the spread of data

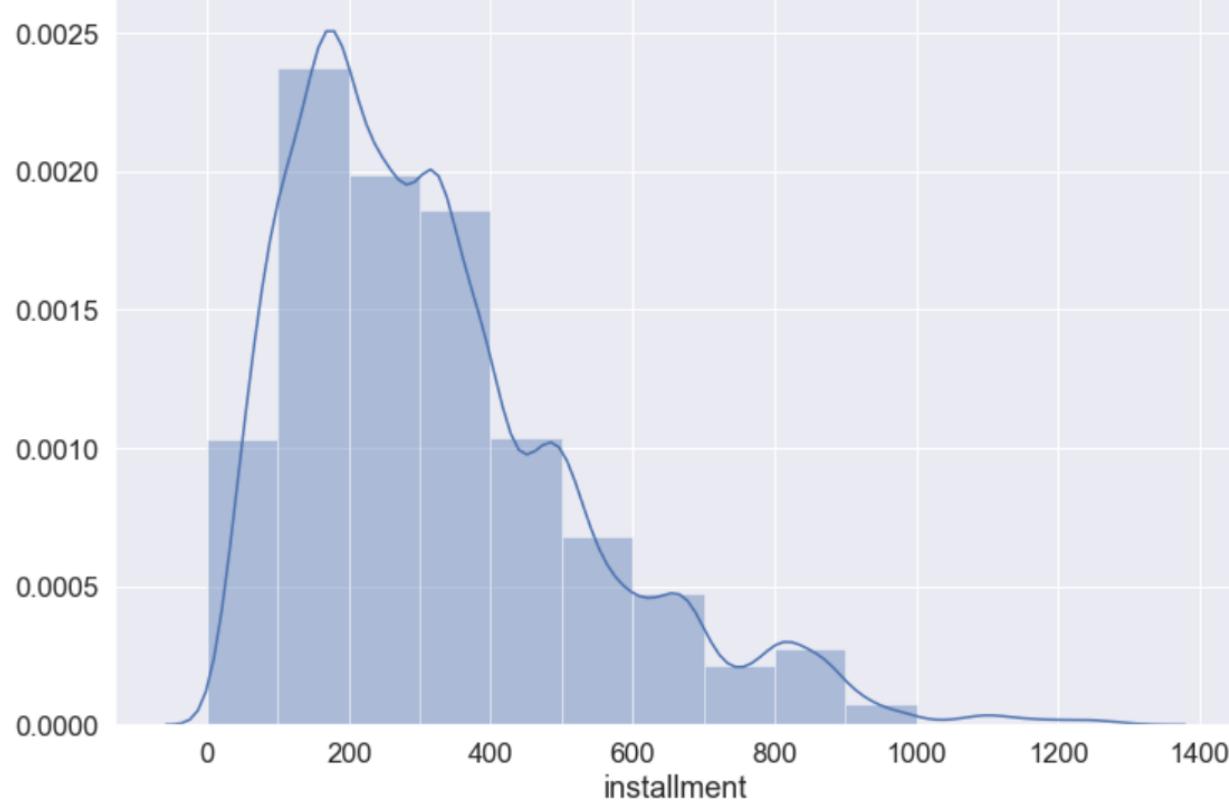


Total Payment received & Total payment received from the investors part

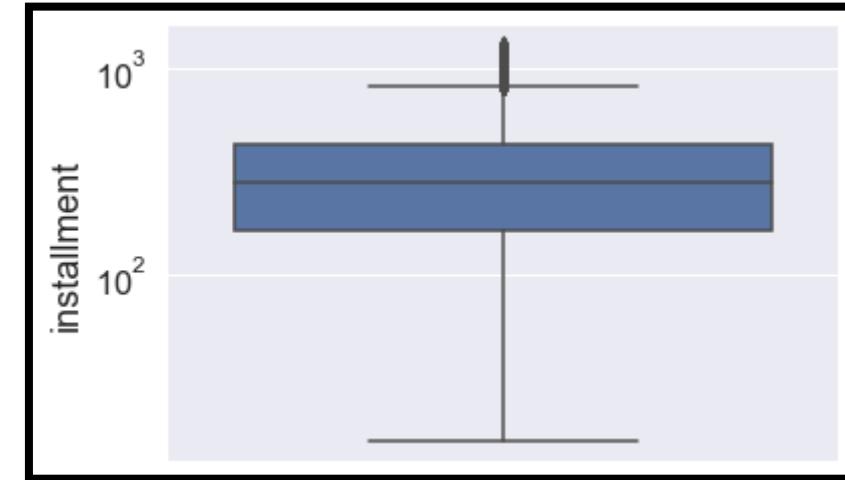


Interest received & Late fee received

# Univariate Analysis: Frequency of loans data – installment wise

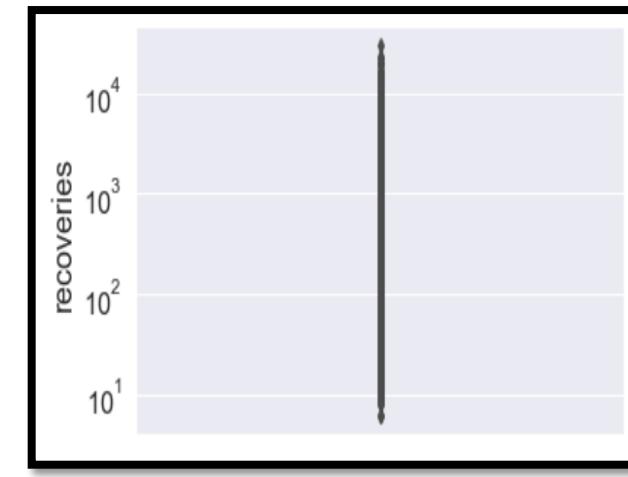
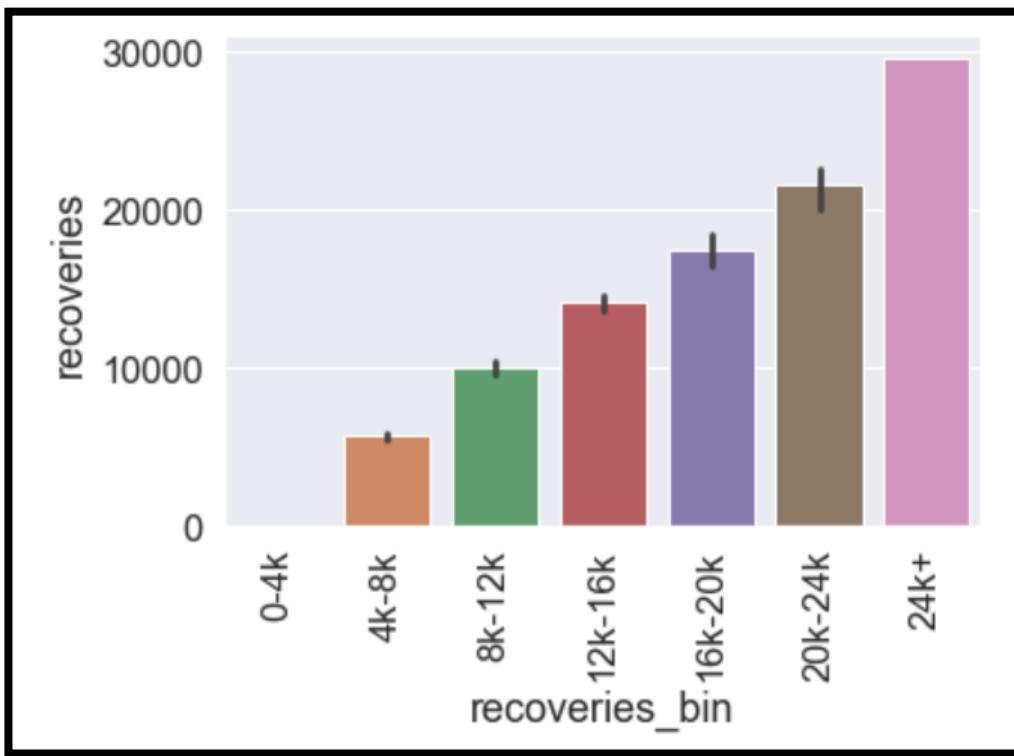


Box plot for installment shows highly spread of data



Box plot for installment shows  
highly spread of data

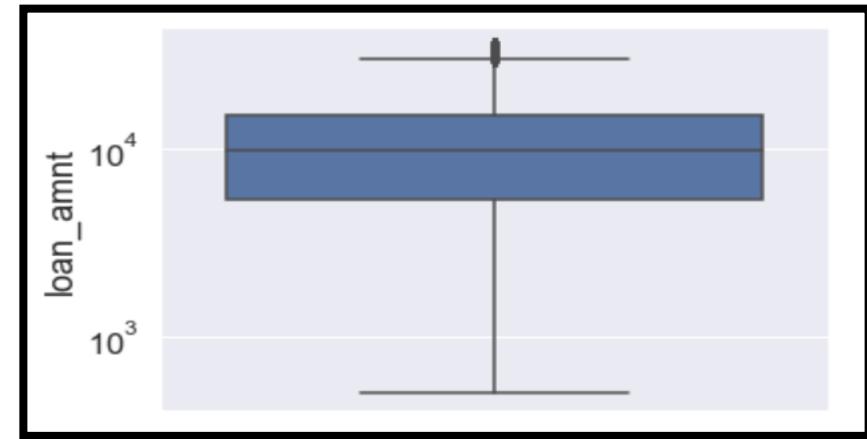
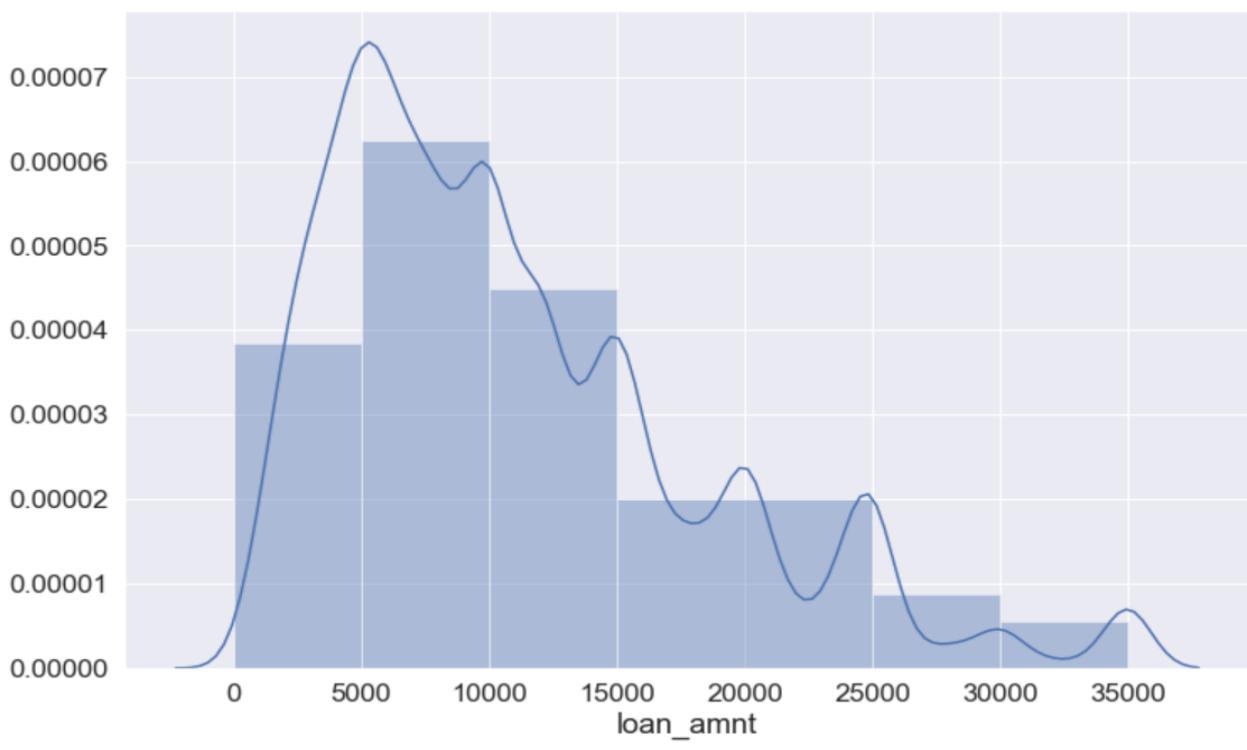
## Univariate Analysis: Frequency of loans data – recovery wise



Highly spread in data for Post charge off gross recoveries

Observation: number of recoveries seems to be increasing with the increase in amount

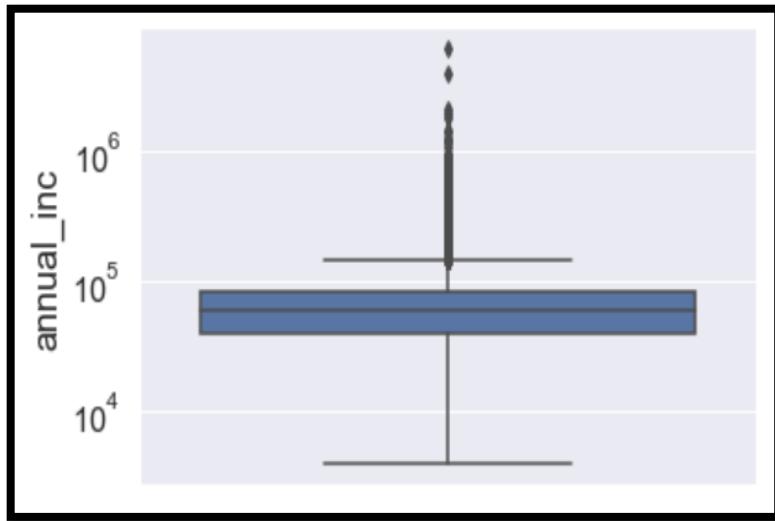
# Univariate Analysis: Frequency of loans data – loan amount wise



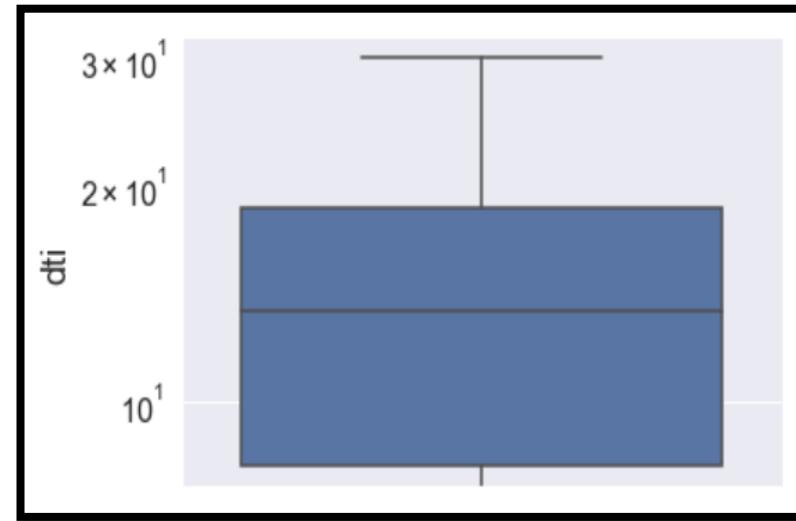
Highly spread in data for loan amount

Observation: high frequency of loan amounts between 5K to 15K

## Box plots to understand the spread of data

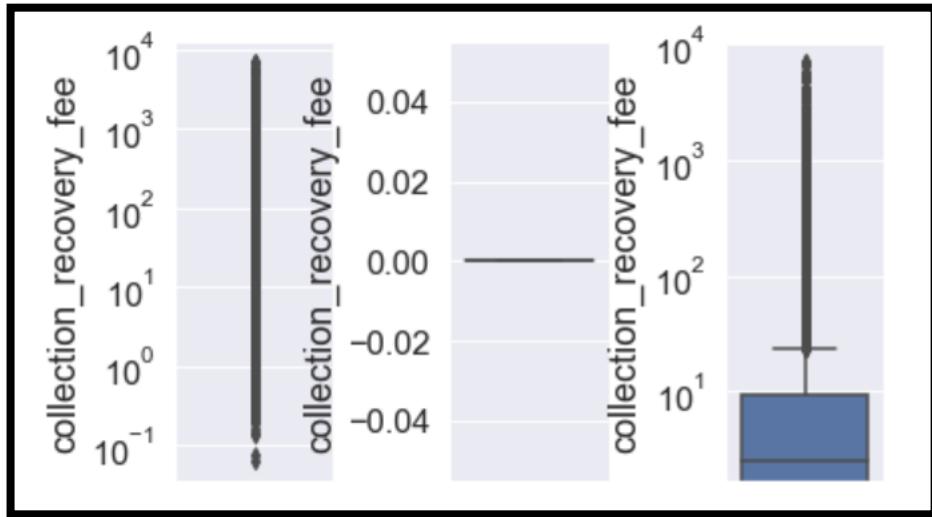


High spread of data in annual income of borrower

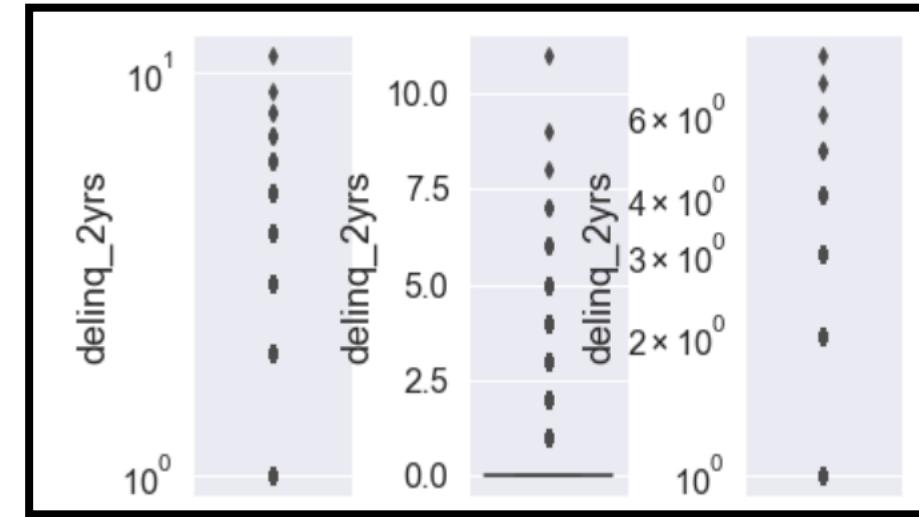


High amount of data for dti lie in between 50 to 200

## Box plots to understand the spread of data

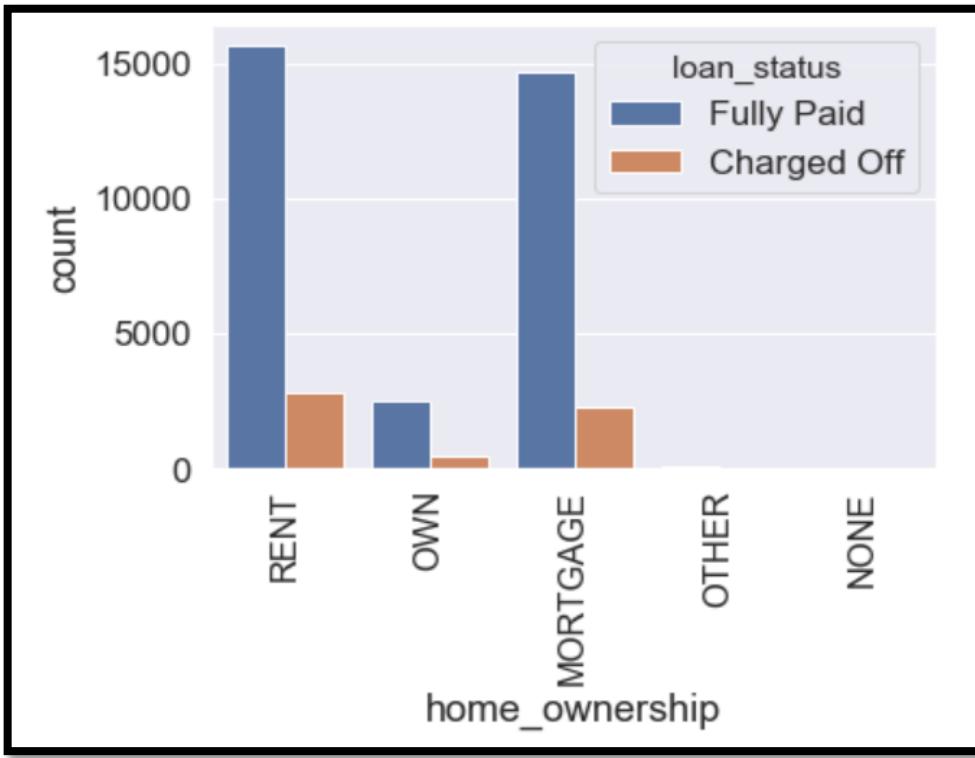


Data is spread very widely and can be removed while analysis

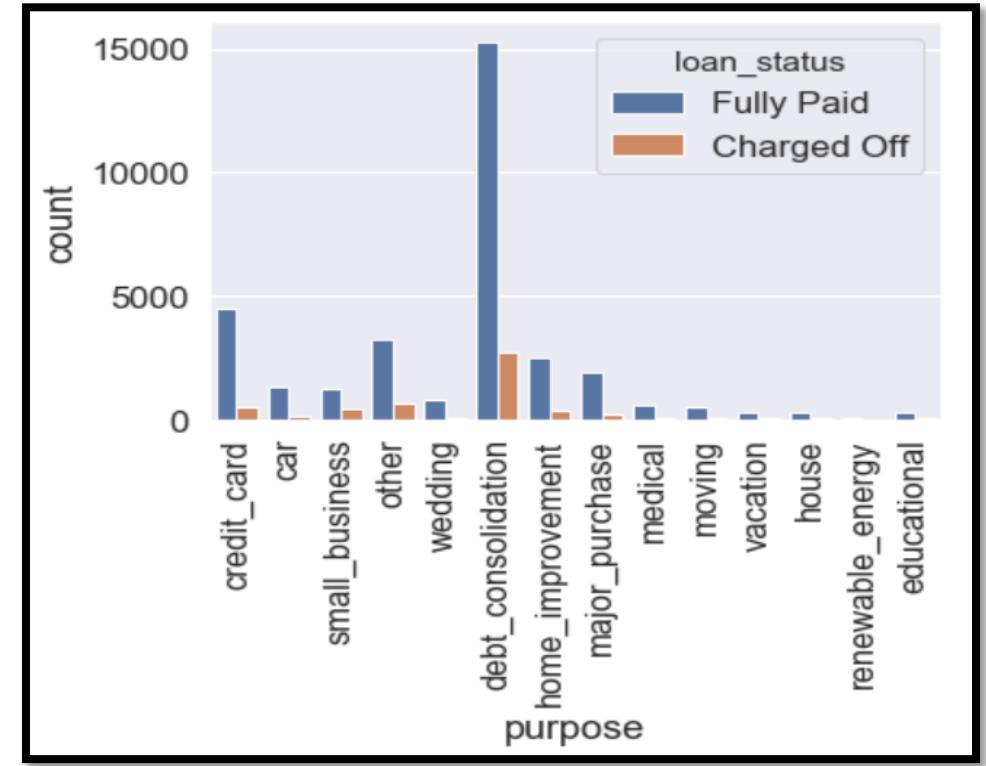


Data is spread very widely and can be removed while analysis

## Bi-variate Analysis

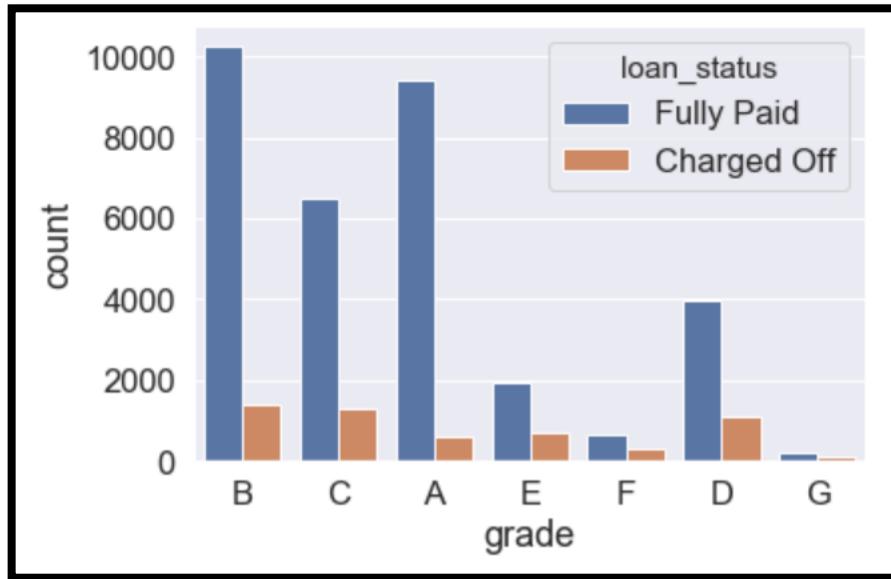


Ownership status – “Rent” and “Mortgage” is highly susceptible to be defaulters

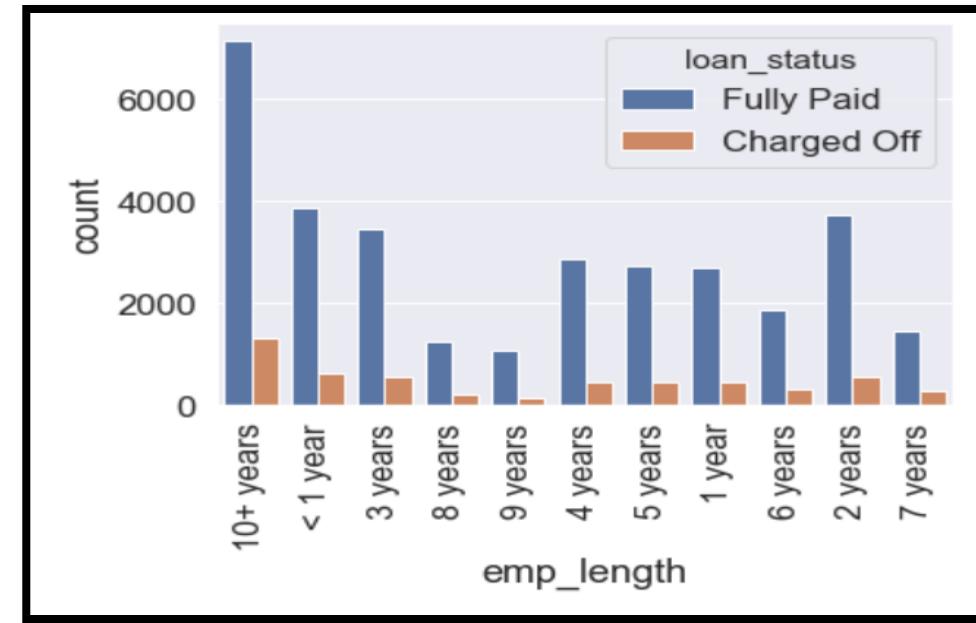


Borrowers with “Debt\_consolidation” is most likely to default

## Bi-variate Analysis

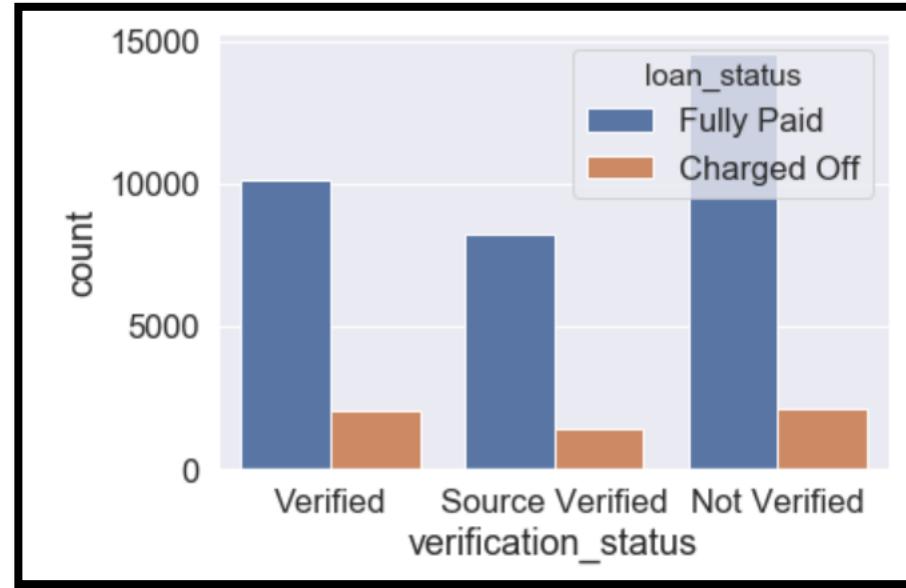


Grade B & C is most likely to default



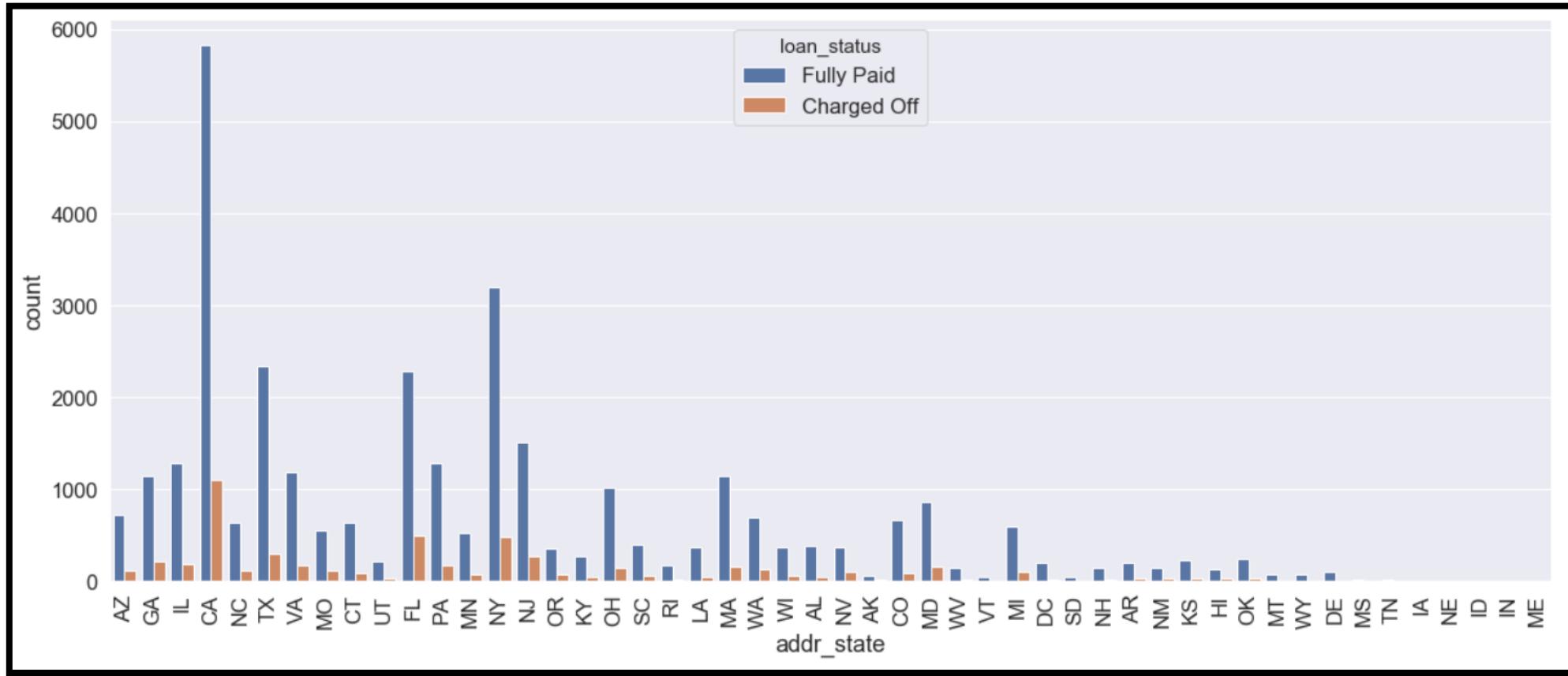
Borrowers employed for 10+ years are most likely to default

## Bi-variate Analysis



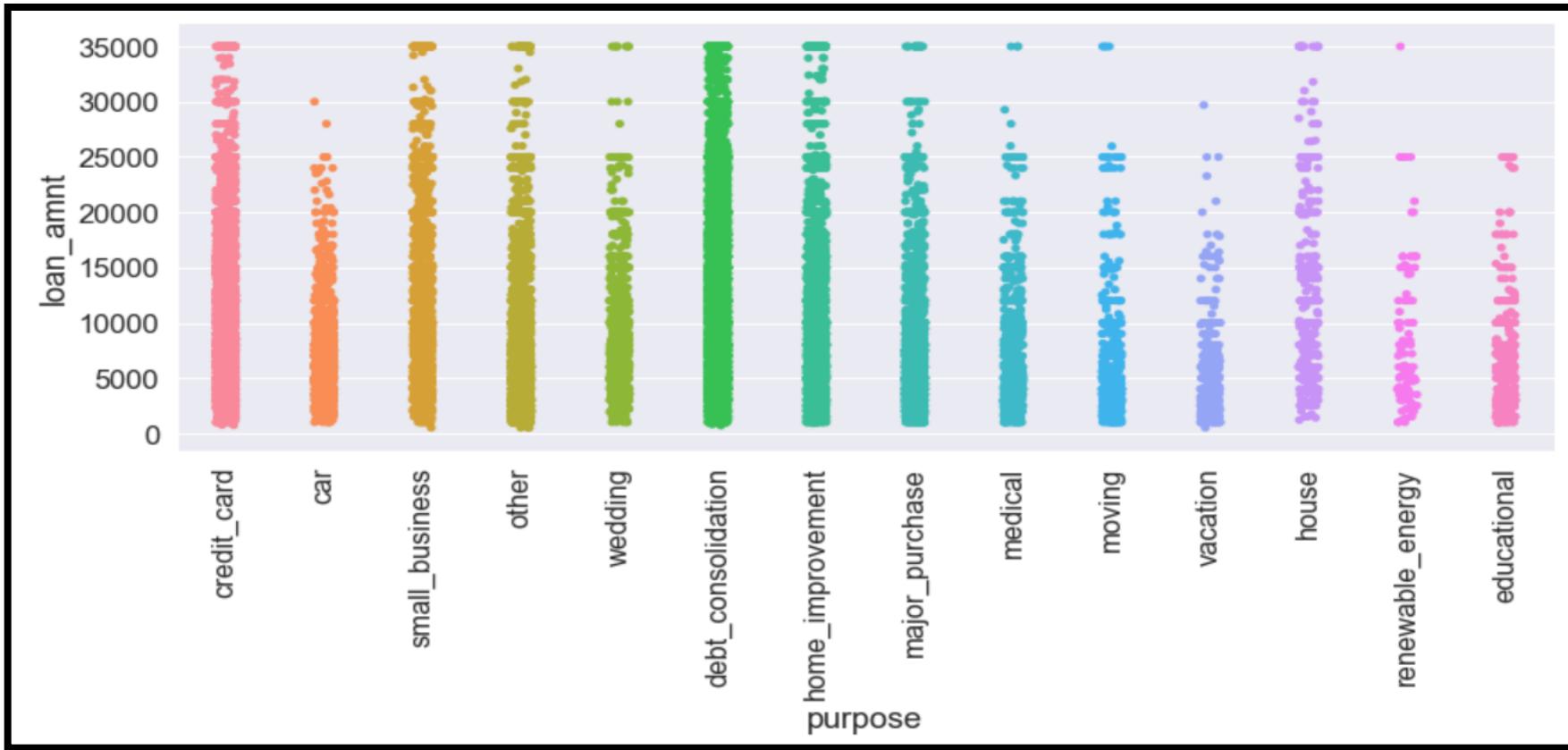
In spite of verification status, there are high number of defaulters

## Bi-variate Analysis



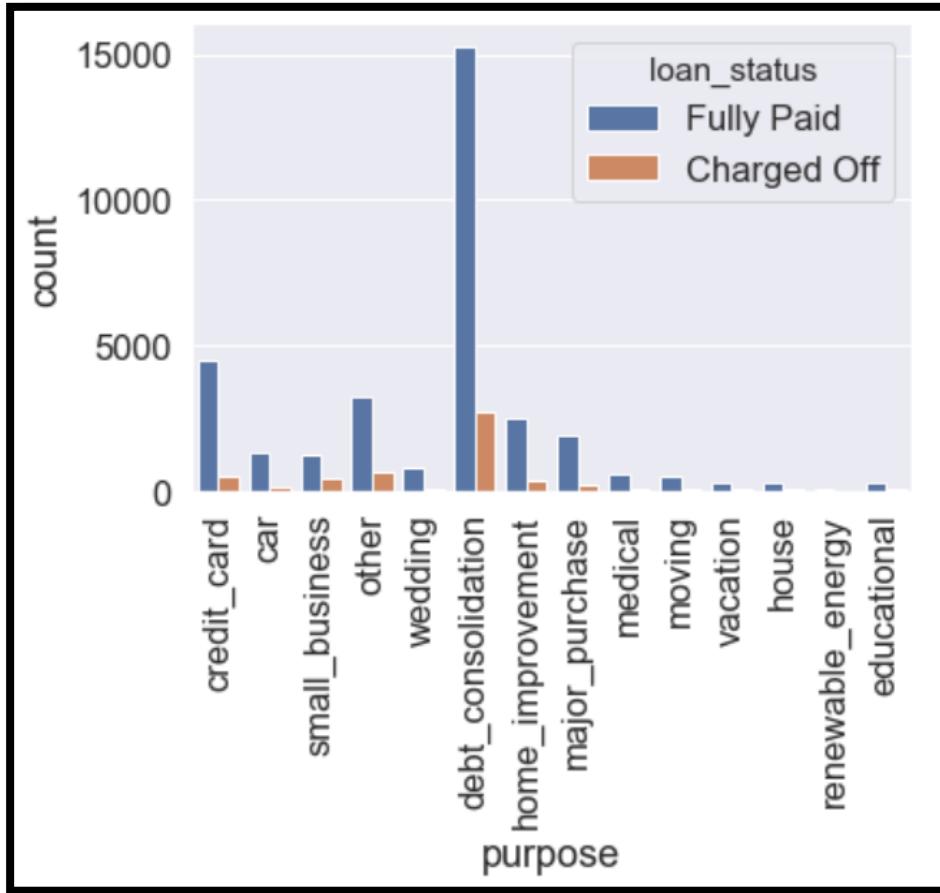
California is the most risky state to lend loans

## Bi-variate Analysis



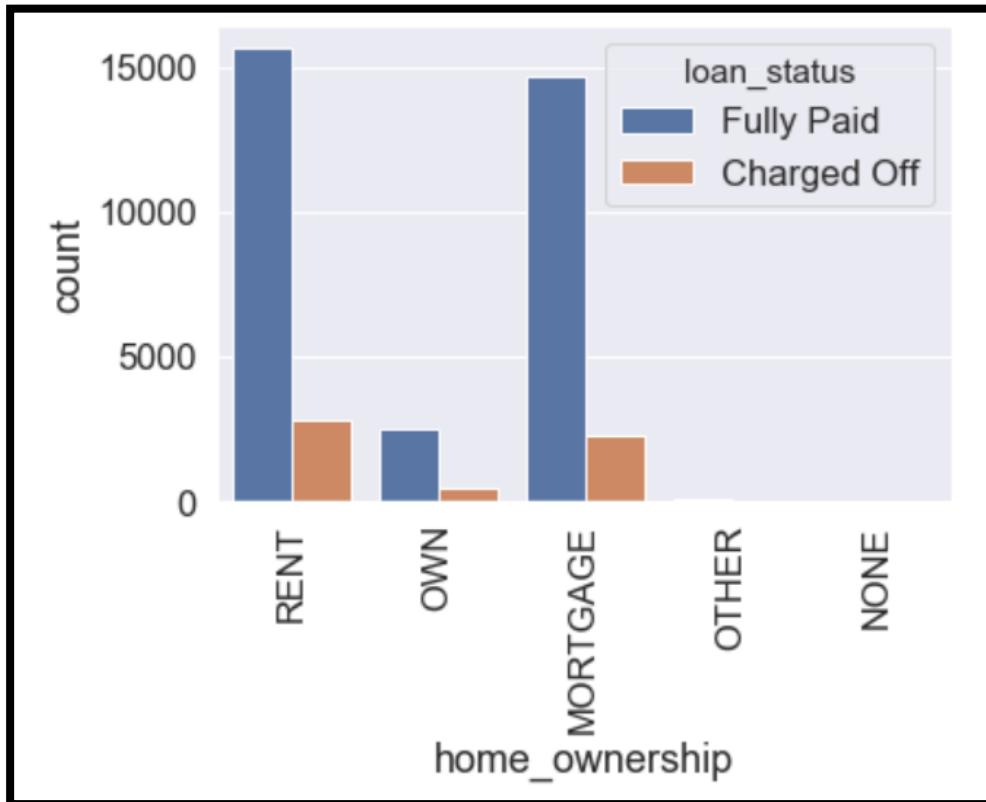
Most defaulting borrowers mention purpose as debt consolidation

# Conclusions drawn based on the Observations



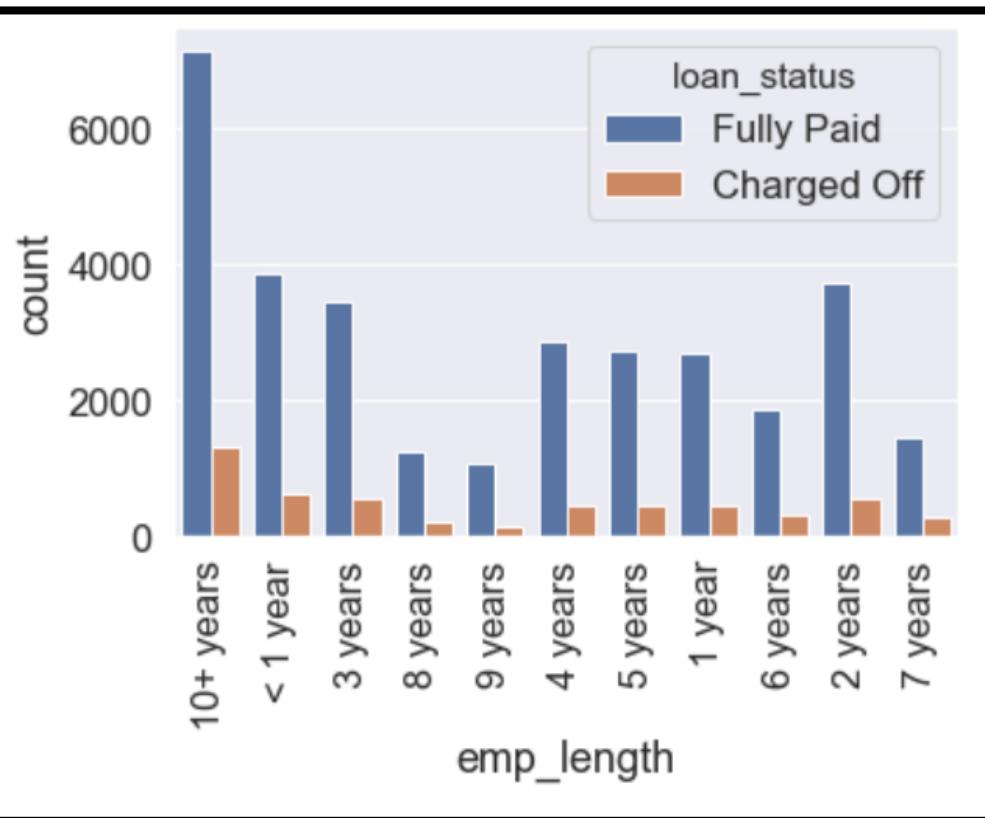
- “Debt\_Consolidation” is the most commonly mentioned purpose for both the loan statuses (Fully Paid and Charged Off)

# Conclusions drawn based on the Observations



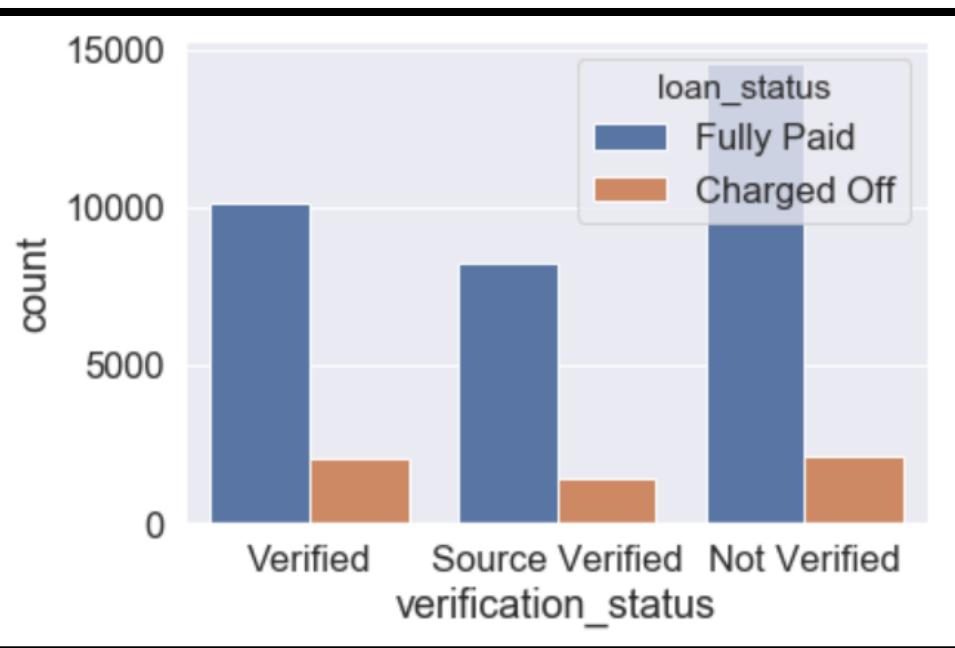
- Most defaulting borrowers are with ownership - “Rent”

# Conclusions drawn based on the Observations



- The common tenure of defaulters is having 10+ years of employee experience

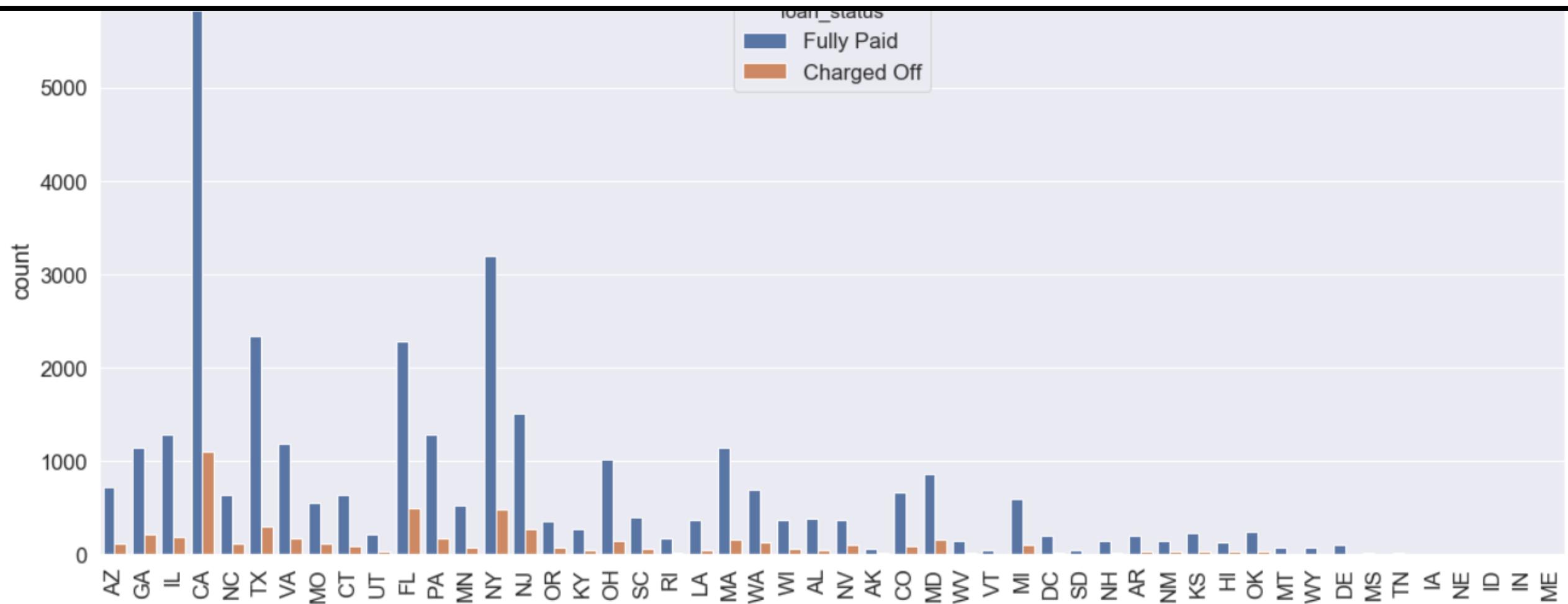
# Conclusions drawn based on the Observations



- The Verification Status of the defaulters is “Not Verified” in most of the cases.

# Conclusions drawn based on the Observations

- “California” – State is showing the most defaulters





Thank You!