

# **LENDING CLUB CASE STUDY**

Ravi Khandelwal

Hari Satuluri

# AGENDA

Problem Statement

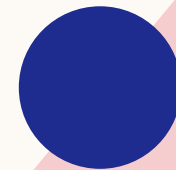
Driving variables

Univariate Analysis

Bivariate Analysis

Multivariate Analysis

Recommendations



# PROBLEM STATEMENT

Analysis required to take a decision for loan application received by lending club.  
The analysis should help making decision which would avoid both financial loss to investor and business loss to lending club and investor.



# **DRIVING VARIABLES**

# DRIVING VARIABLES

5

Driver variables which would help to analyze if the customer can be a defaulter

- **loan\_amnt/funded\_amnt/funded\_amnt\_inv :**

If the customer is requesting a higher loan amount we need to understand his purpose and other things as higher the loan amount the risk would be higher to be a defaulter

- **term :**

Term period customer trying to repay the loan could be driving factor, if customer taking huge loan and trying to close it in short term that could be a defaulter, need analysis in that variable

- **grade :**

We can derive some data as which graded customers are being highly defaulters, grades are given for a reason right.

- **home\_ownership :**

The place customer stays could lead him as defaulter, like if the customer is not having own house tracking them for repayment could be difficult, so they could become defaulter.

- **annual\_inc :**

customer with less income and high loans could become a defaulter.

# DRIVING VARIABLES

6

- **verification\_status :**

This is something same reason as home ownership, if the customer is not verified that would be difficult to track for repayments and could become a defaulter.

- **purpose :**

Purpose of the loan could give us information as if the customer is trying to manage his expenses or spending for luxury, with that we could analyze the behaviour of defaulter

- **zip\_code/addr\_state :**

with these we could analyze the behaviour based on living region

- **dti :**

Debt to Income Ratio is another highly informative parameter, which would tell us customers current financial status and the behaviour to be a defaulter.

- **pub\_rec :**

This would give us the general behaviour of the customer if the customer is having any derogatory issues in public

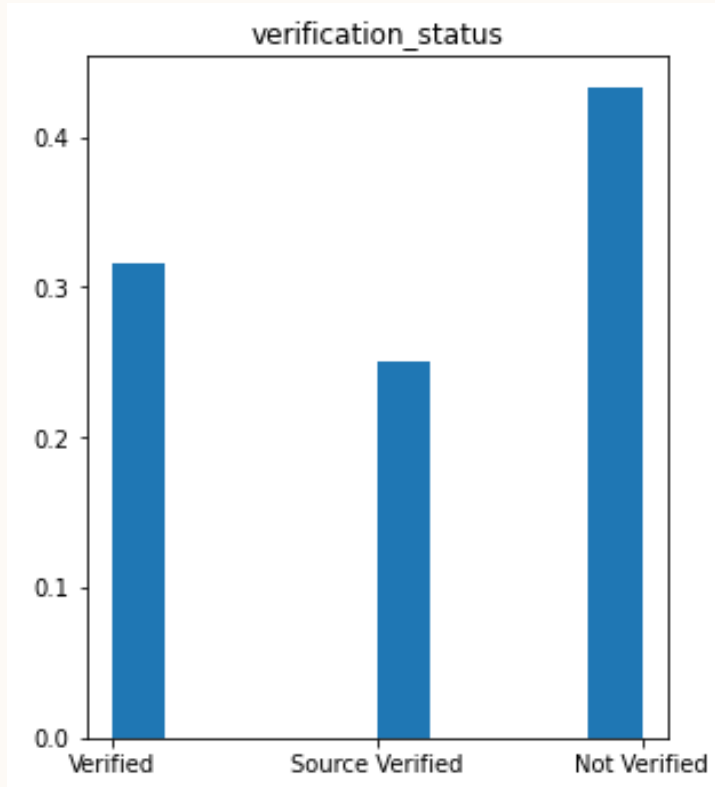
- **pub\_rec\_bankruptcies :**

This variable helps us to analyze the behaviour of the defaulter as he might be having bankruptcies before

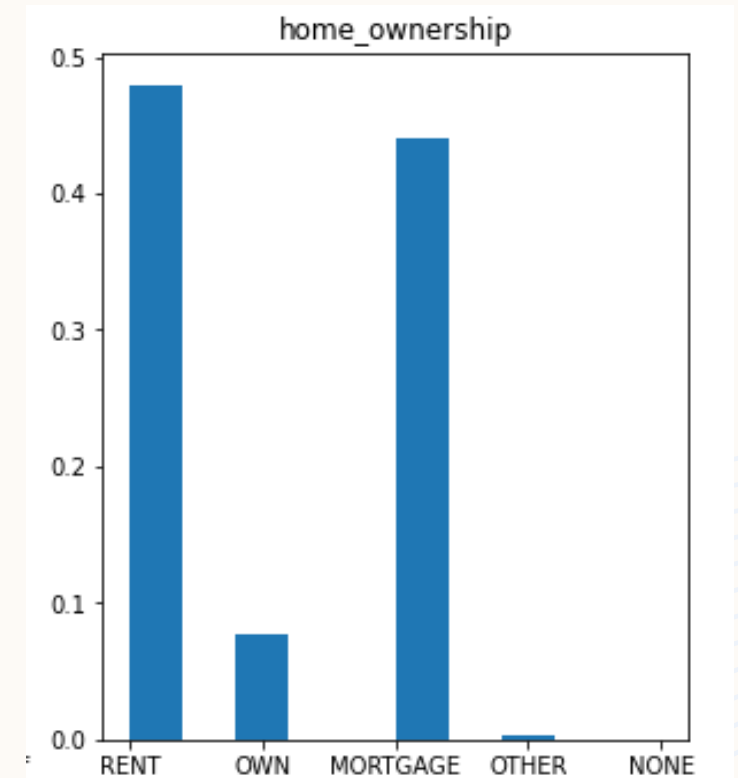
# UNIVARIATE ANALYSIS

# UNIVARIATE ANALYSIS - home\_ownership & verification status

8



- Most loan applications are from the people staying in rented house followed by mortgage properties.
- Less than 10% people who have own house applied for loan.
- By looking histogram of verification status, most people lies in 'not verified' status.
- Suggested to follow strict verification process, as not verified could lead to defaulter

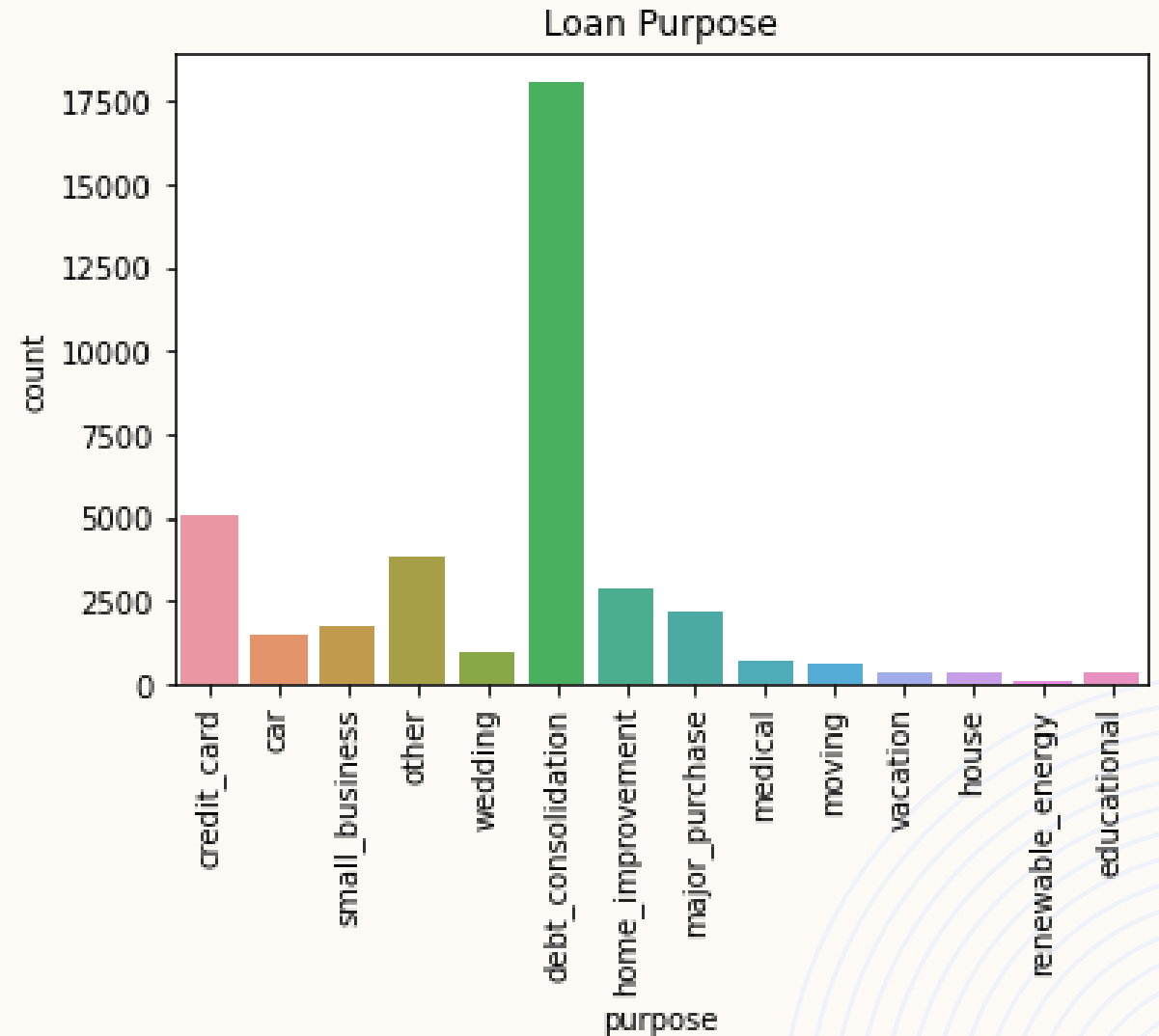




# UNIVARIATE ANALYSIS – loan purpose

9

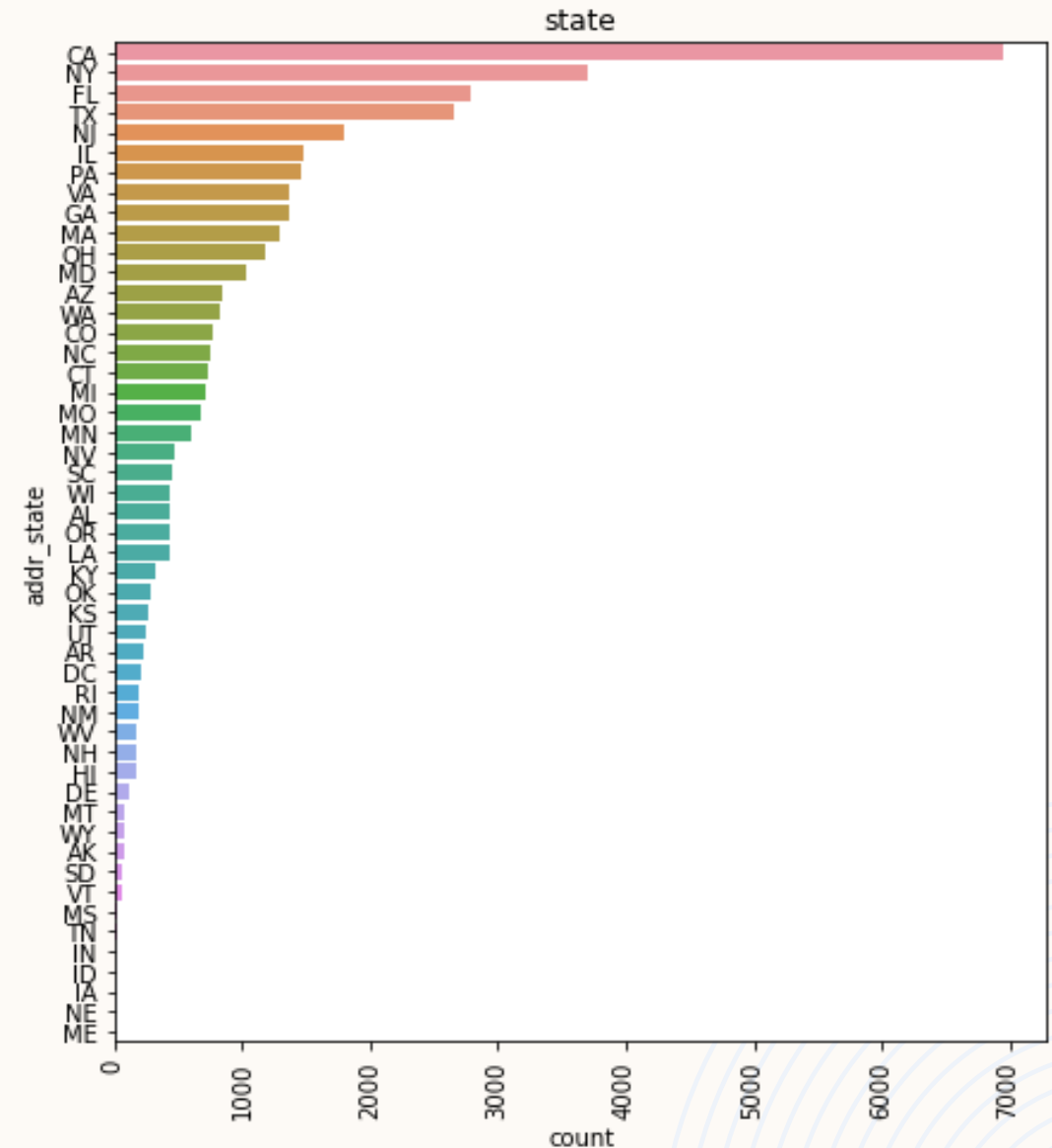
- By looking at this graph we can say most of the loan applications to lending club are for debt consolidation.
- When the customer is looking for debt consolidation, we might need to analyze how much income he gets and his DTI



# UNIVARIATE ANALYSIS – state

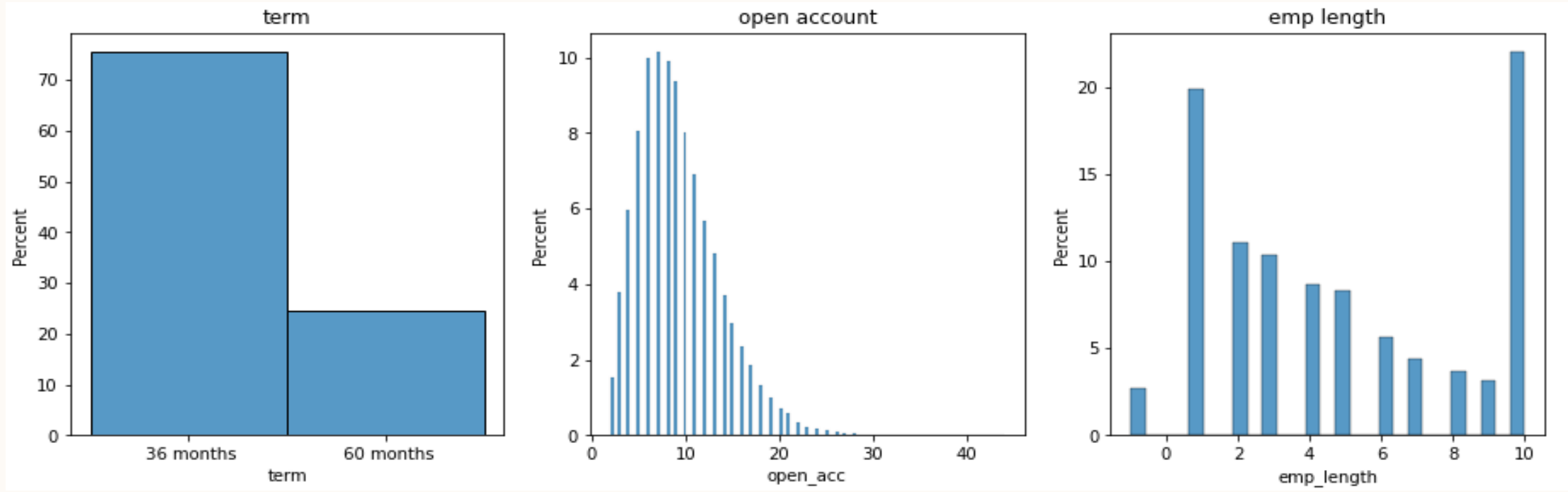
10

- By this histogram we see that most of the loan applications are from CA state.
- We found out that 50% loan application belong to top 6 state out of 50 states.
- Six state - CA,NY,FL,TX,NJ,IL



# UNIVARIATE ANALYSIS – term, open\_account, emp\_length

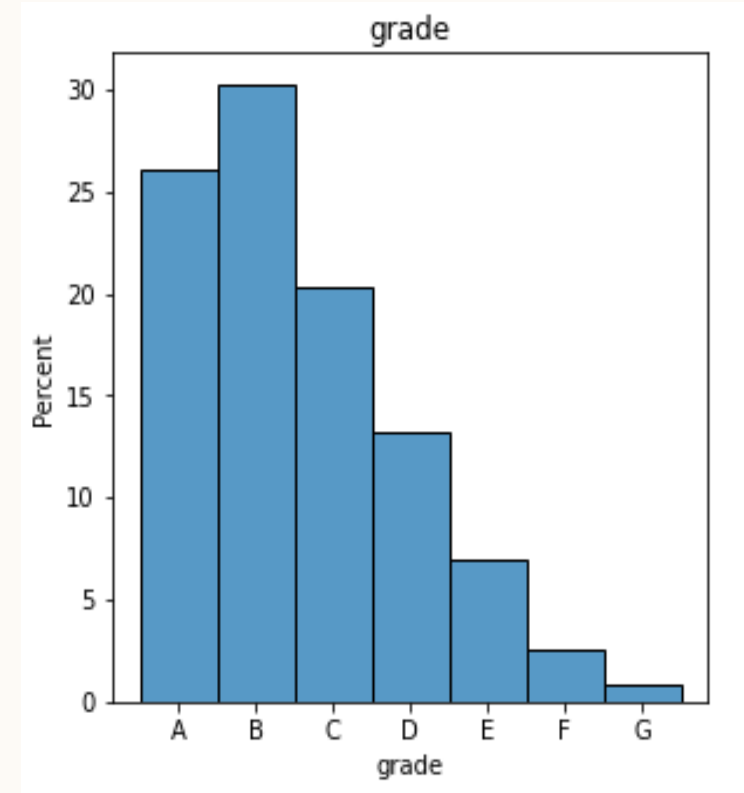
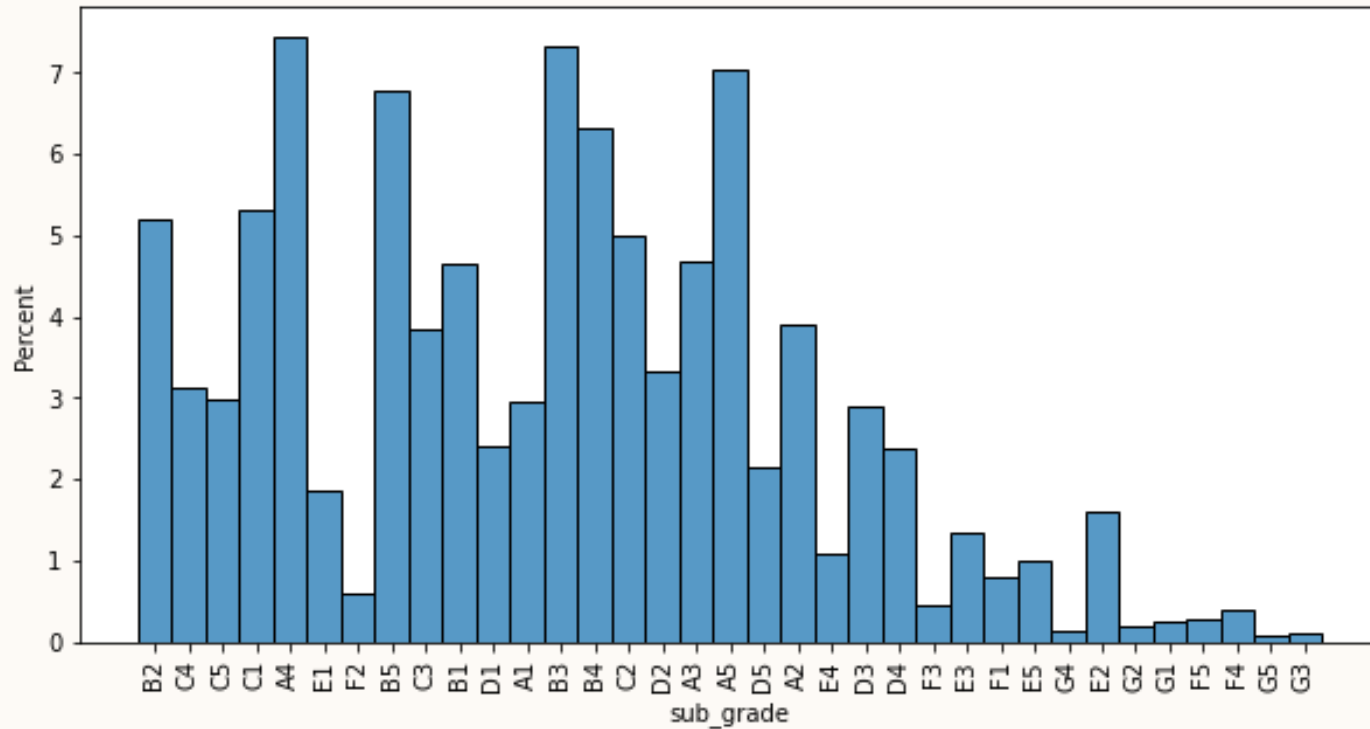
11



- With emp\_length hist, we observe as employment term increases, they are less likely to apply loan application.
- As we can view on Open Account Histogram, Initially people open account increases, loan application increase, But after reaching at maximum point around 9, frequency of loan applications start decreases on higher size.

# UNIVARIATE ANALYSIS – grade & sub\_grade

12



- From Grade categorical histogram, loan applications follow negatives trends as grade increase from A to G.
- May be loan disbursement for lower grades is at lower scale when compared to higher grades.
- Analysis should be done while approving loan to lower grades.

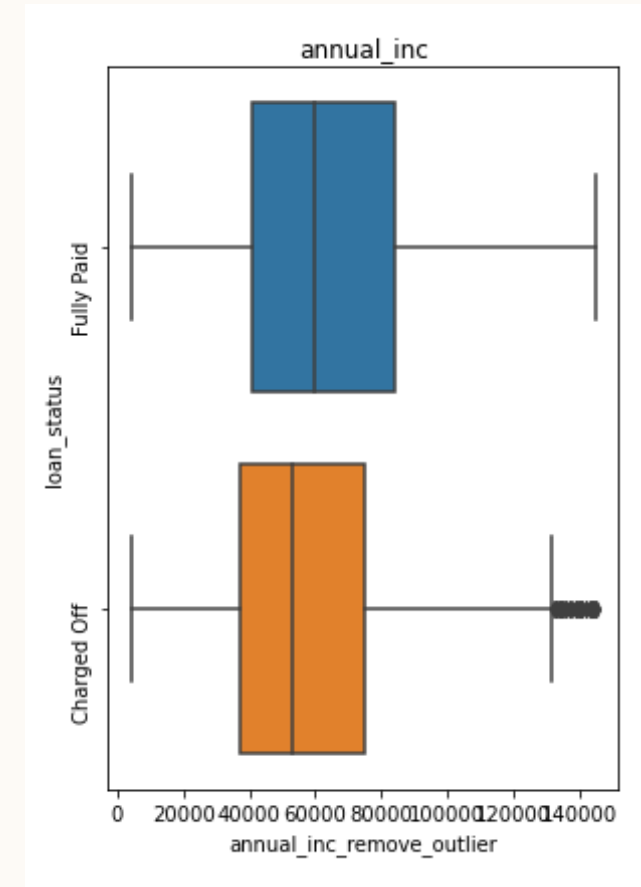
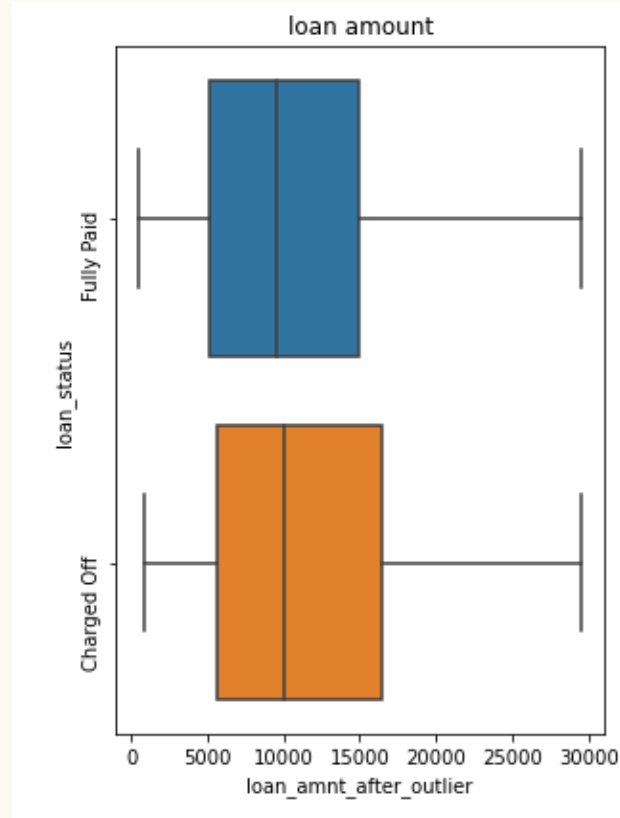


# **BIVARIATE ANALYSIS**

# BIVARIATE ANALYSIS – loan\_amount, annual\_inc

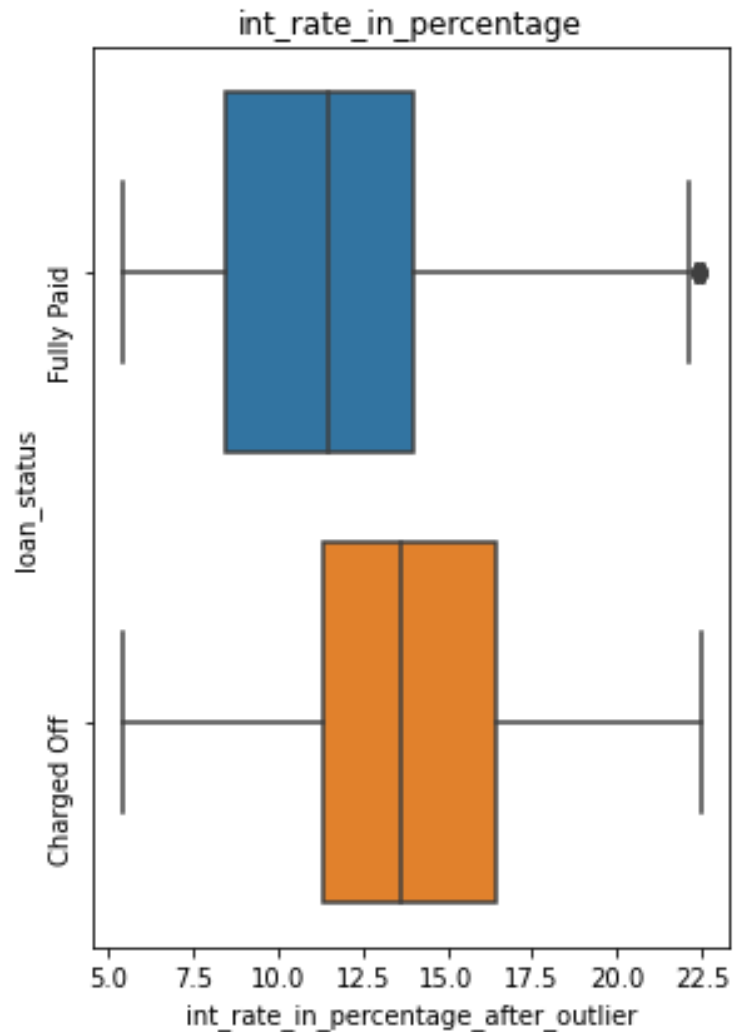
14

- It is visible from loan amount box plot, iqr is bigger for charged off loans, which means higher the loan amount risk is higher for being defaulter
- From Box plot of annual income, Q3 value of charged off loan is lower than fully paid loan. It means good income persons are more likely to close loan as fully paid.

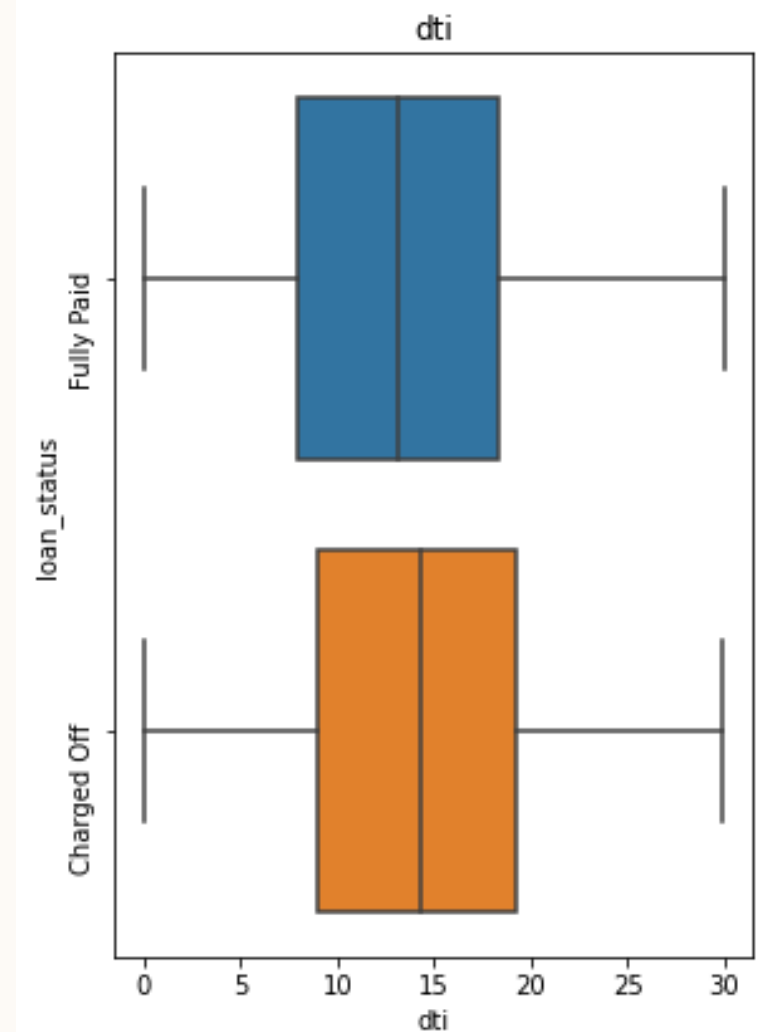



## BIVARIATE ANALYSIS – int\_rate, dti

15



- From int\_rate graph it's clear information that higher interest rate loan is more likely to end up with charged off.
- From Box plot of Debt-to-income ratio, q1, median, and q3 having larger value in charged off loan. higher debt to income ratio is not good.



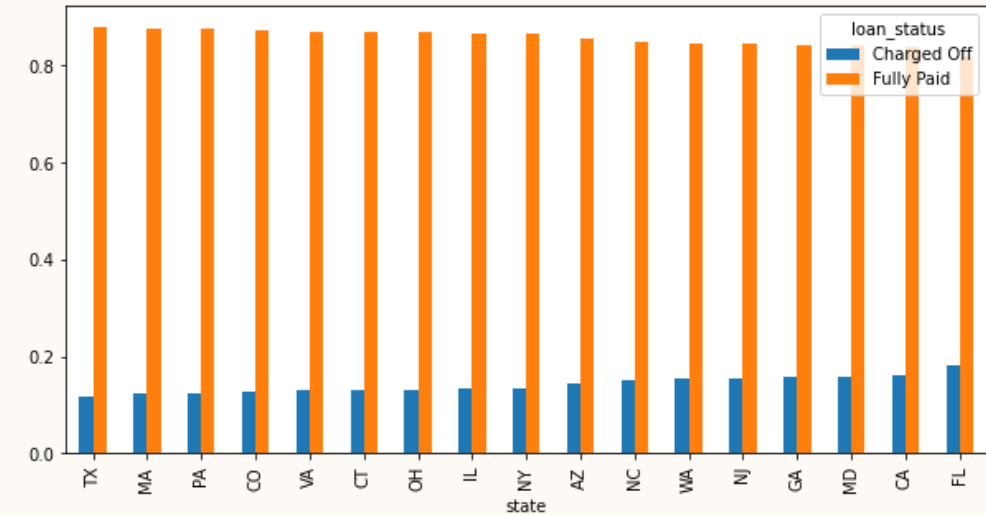
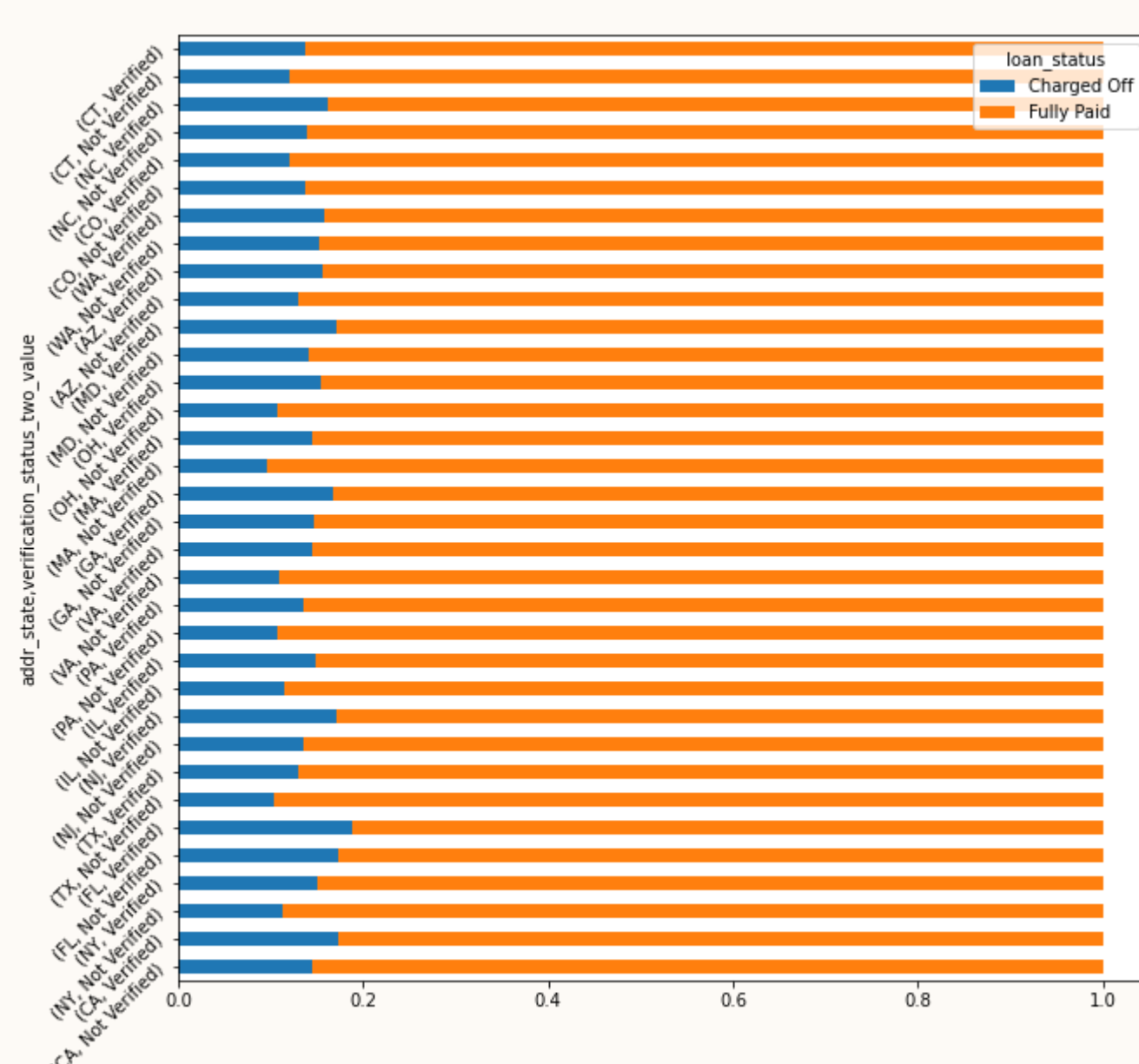
The background features a vertical line that divides the space. To the left of this line, there are concentric white circles on a light green background in the upper left, and a light pink triangle in the lower left. To the right of the line, there is a light blue semi-circle in the upper right and a light red triangle in the lower right. The title text is centered on the white background to the right of the vertical line.

# **MULTIVARIATE ANALYSIS**



## MULTIVARIATE ANALYSIS – addr\_state vs varification status with loan\_status

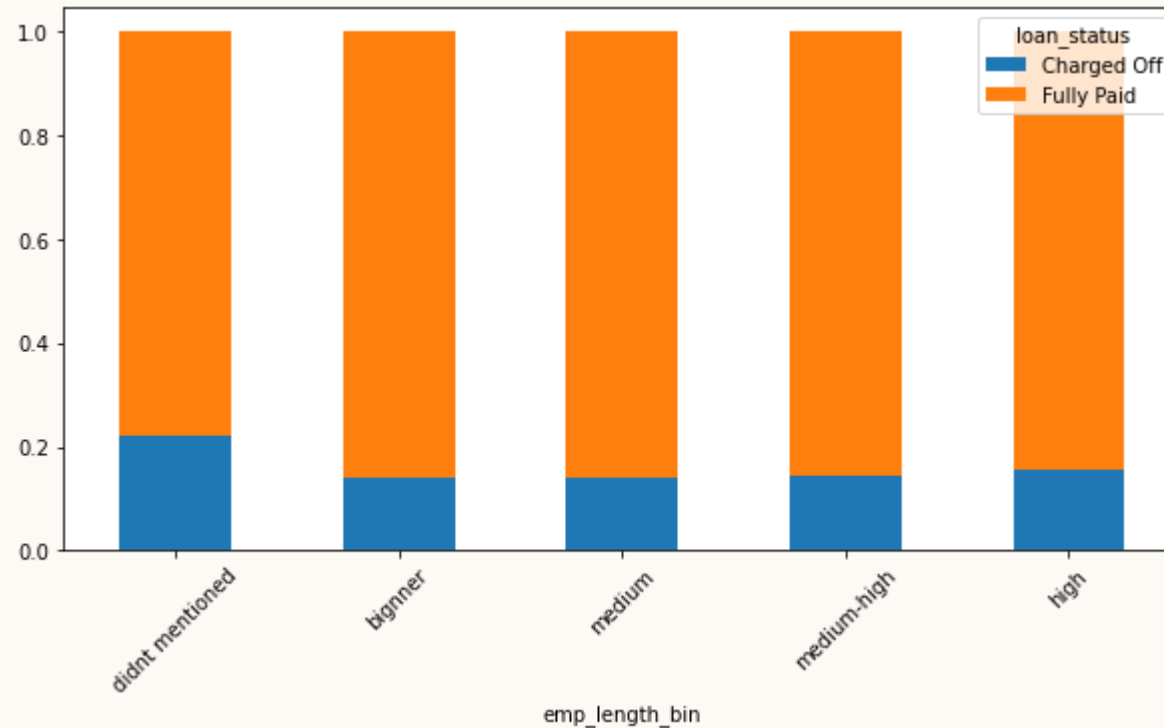
17



- State FL is having most defaulter loans.
- State TX is having most fully paid loans.
- Verification process is not good because blue line (charged off) for verification status are bigger than not verified.

## MULTIVARIATE ANALYSIS – emp length vs loan\_status

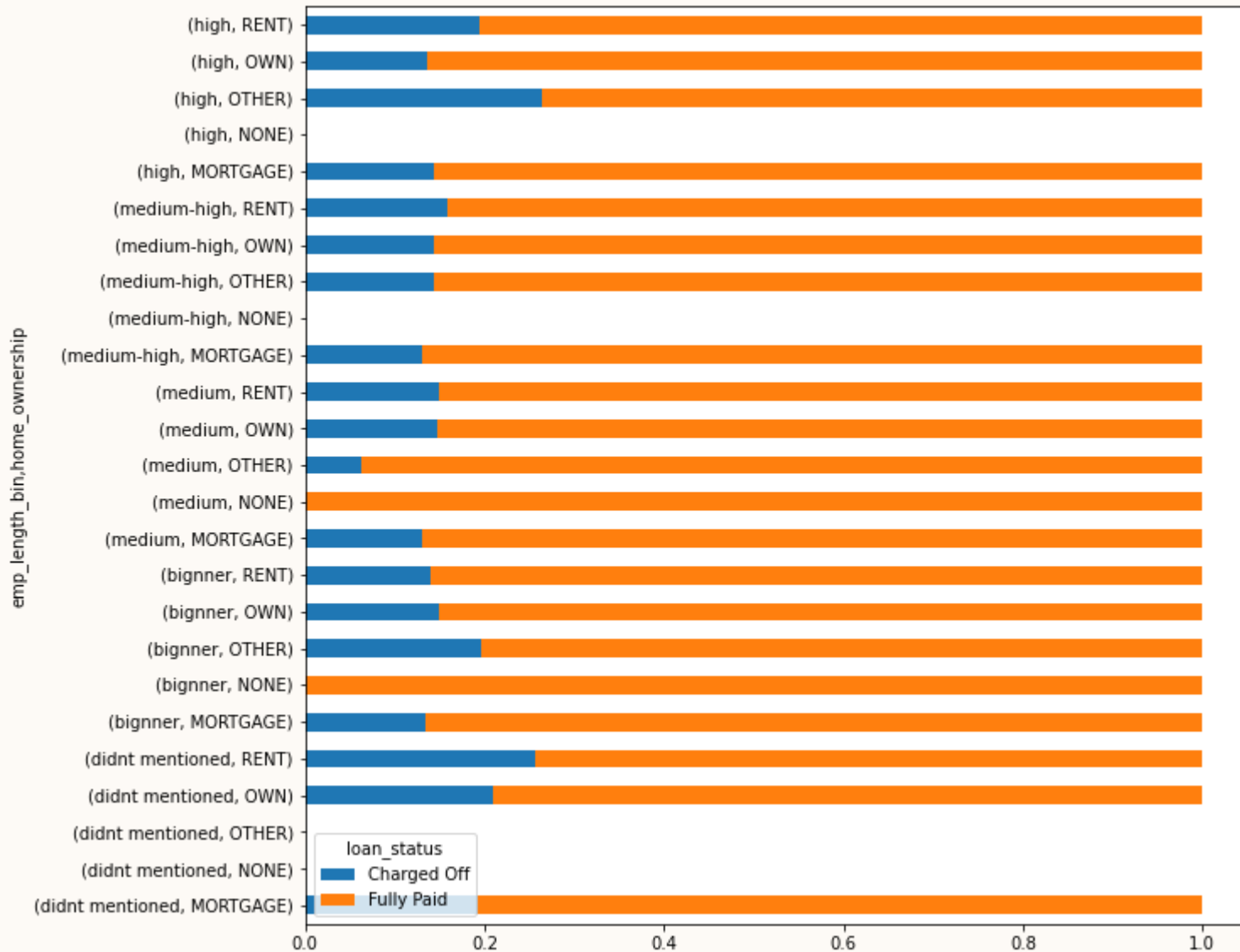
18



- Applicants who doesn't mention Employee length are having high defaulters compared with others

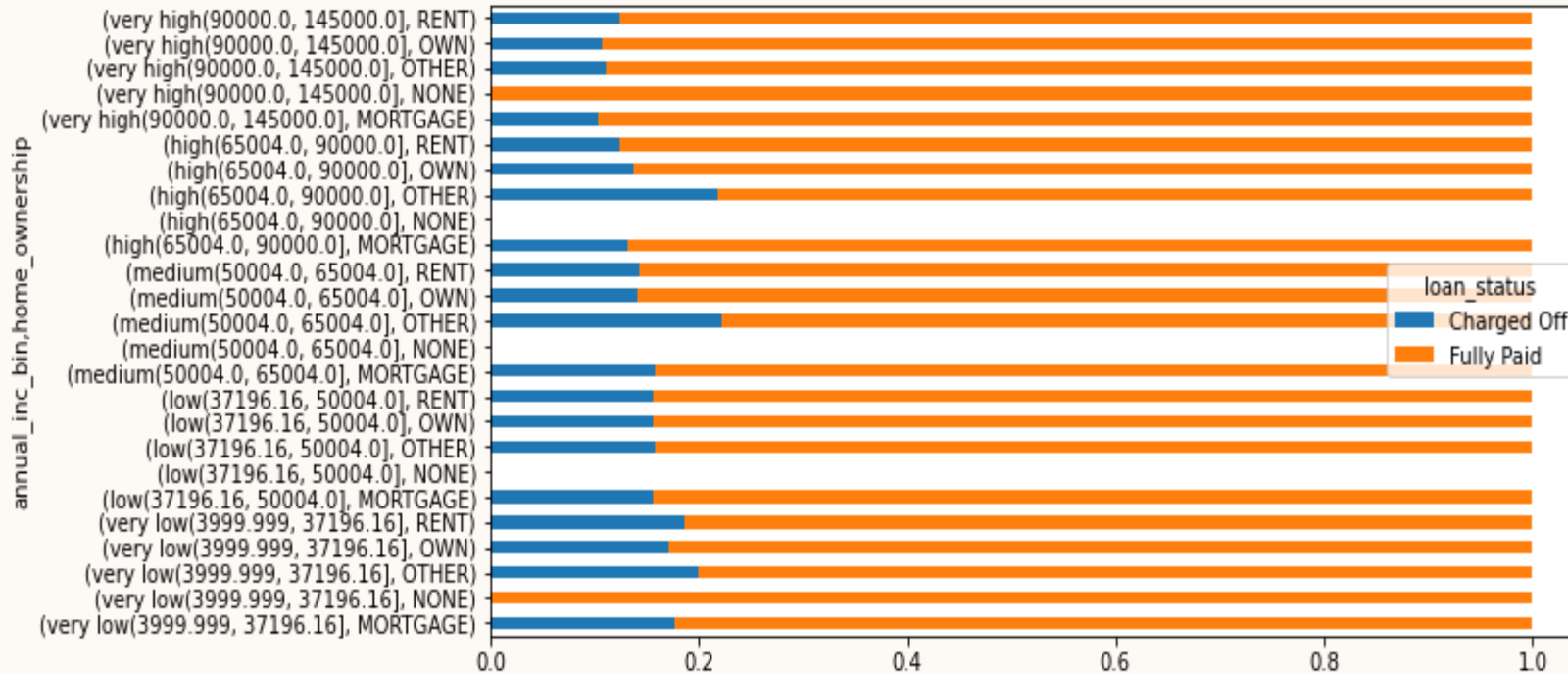
# MULTIVARIATE ANALYSIS – employee length,home\_ownership vs loan\_status

19



- loan applications of people who did not mentioned employee length with home ownership rent are more risky application. (as visible at blue bar of didn't mentioned vs rent)
- Beginner employee length with other home ownership and high employment length with other are with more risk.

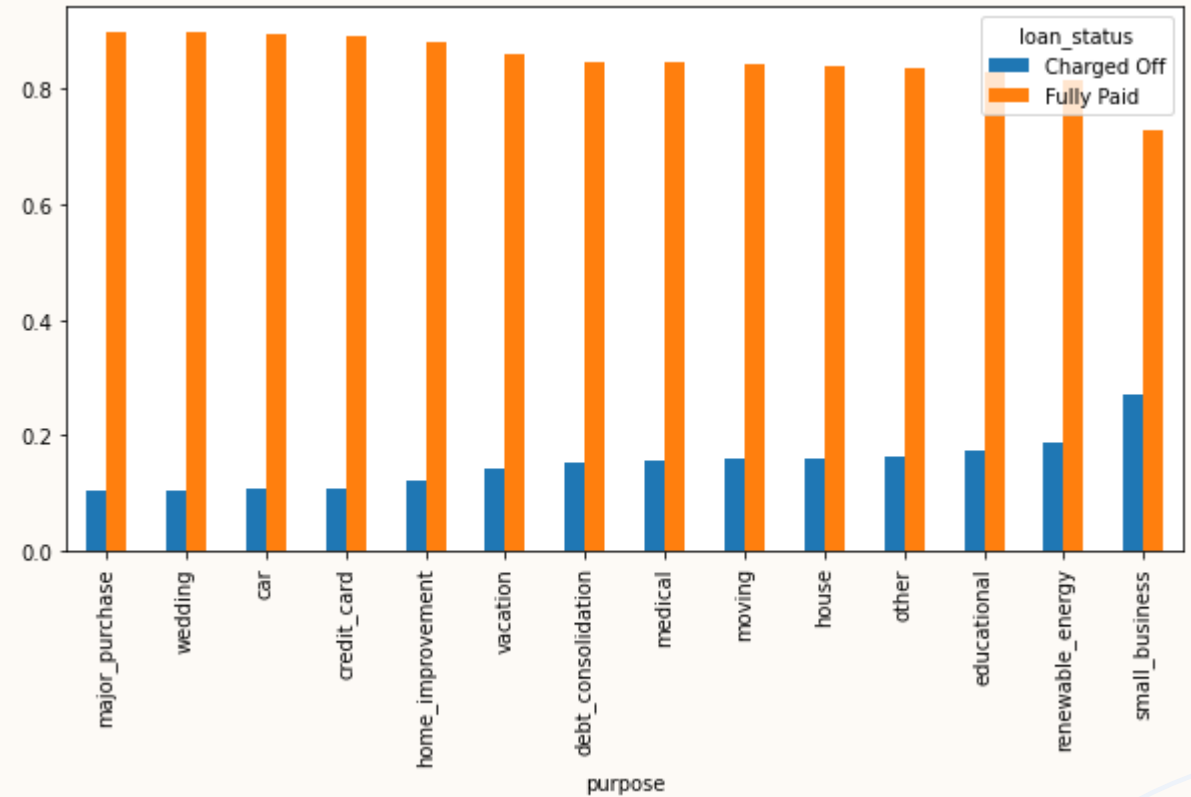
## MULTIVARIATE ANALYSIS – employee income,home\_ownership vs loan\_status 20



- It is evident that People with 'other' Home ownership are more risky no matter what they belong to any income category.
- High chance of loan application from very low income going to be a defaulter.
- People with mortgage home ownership are likely to complete loan successfully.

## MULTIVARIATE ANALYSIS – Purpose vs loan status

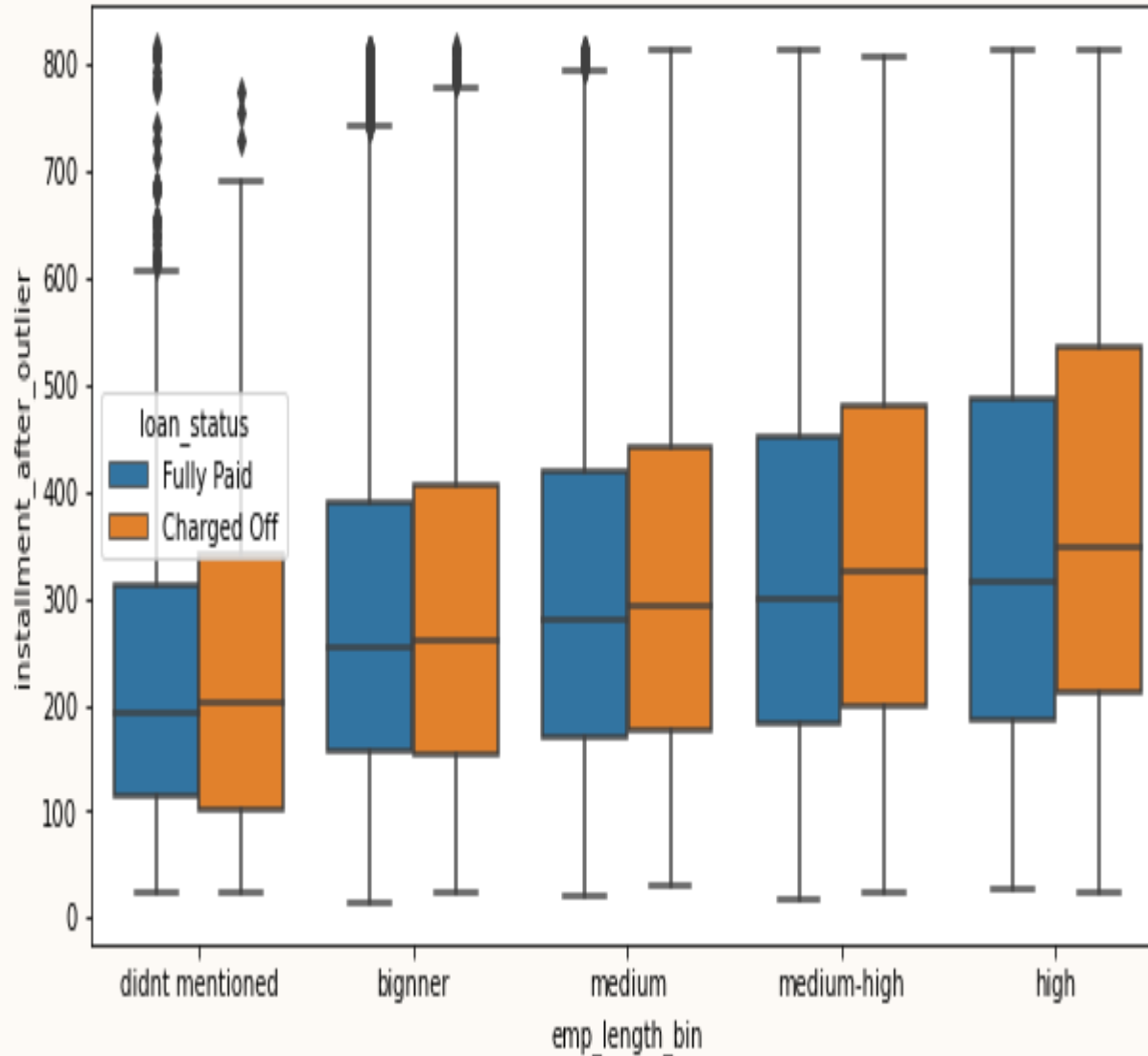
21



- Loan application with small\_business purpose are high on charged off side, it's clear that business can be successful or might become loss.
- Thorough analysis of business is also very much required for those kind of loans.

## MULTIVARIATE ANALYSIS – Installment vs home\_ownership against loan\_status

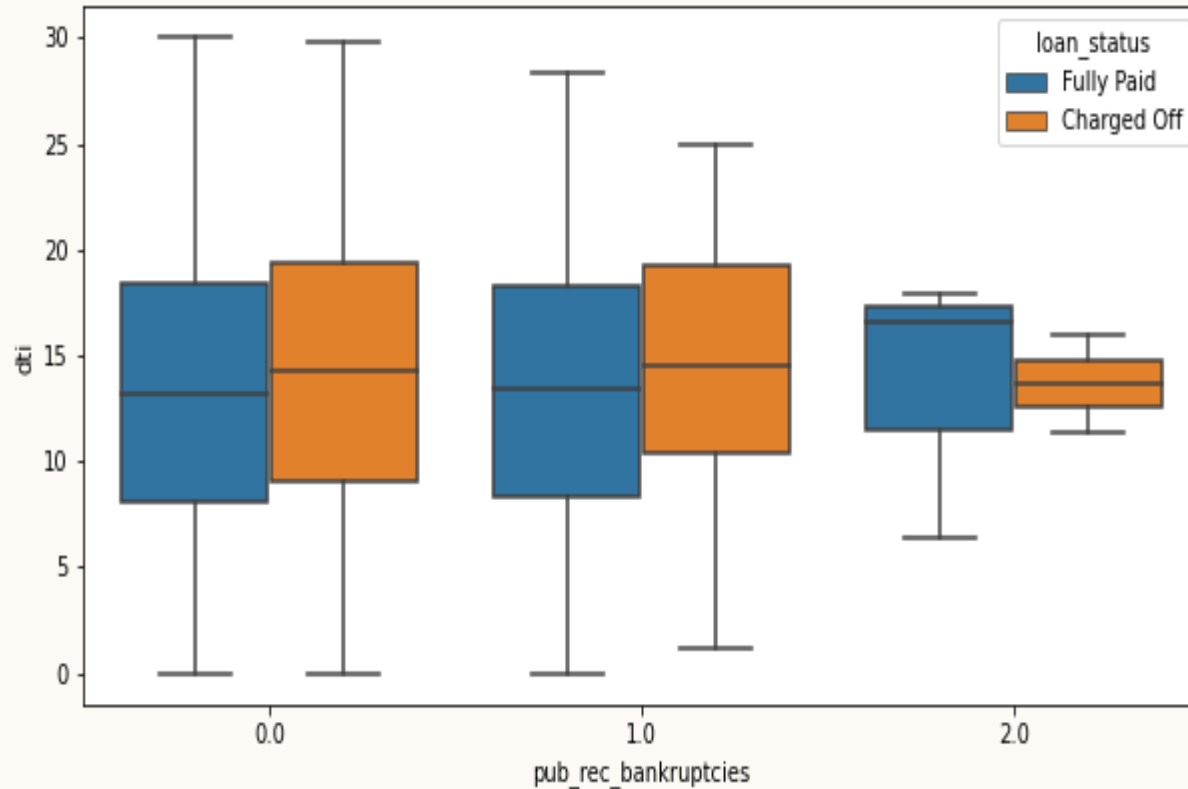
22



- Median of installment of each category of not paying people is higher than paid off people, Same happening with third quartile. Which means people of any category with lower installments are fully payable customers
- But after looking at Q1 of each category, there are some who are also not able to make pay lower emi as well.

## MULTIVARIATE ANALYSIS – dti vs pub\_rec\_bankruptcies variable against loan\_status

23



- Bigger debt to income ratio and higher number of public bankrupt means more chance to become loan defaulter.
- But little deviation toward total 2 bankrupts. It may be because number of loan given to such people is very few.



# RECOMMENDATIONS



# RECOMMENDATIONS

25

- Applicants with High DTI and low income or high bankruptcies in past or requesting high loan amount with higher EMI or lower tenure could be a defaulter
- When consolidated applications as per states, customers from FL are more likely to be defaulters and on positive note customers from TX state are more likely to close their loans.
- Customers who has own houses or mortgage homes are more likely to finish their loans on time, but to be vigilant on company financials customers who specify home ownership as other are most likely to charge off their loans no matter of their employment length or annual income.
- Applications which haven't specified their employment experience are more riskier as most of those applications are charged off.
- When it comes with verification of applications in lending club, seems that verification process is not so perfect, because we could see most of the charged off loans verification is completed successfully.





# **THANK YOU**

Ravi Khandelwal

Hari Satuluri