Lending Club Case Study

The case study analyzing different factors affecting loan default through exploratory data analysis

Problem Statement

- **Context**:
  - 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.
  - Like most companies in this business, they run a risk of either loss of business or customer default which is the biggest risk and contributor to financial loss.
  - Hence the company would always be interested identifying these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- **Problem Statement / Business Objective:**
  - To analyze the given dataset to identify possible factors of loan default and create a generic profile of a risky customer
  - Scope: The current scope is identification of risky applicants using exploratory data analysis. No predictive model is expected.

Analysis Approach

- The analysis approach is broken down into the following steps:

  - ***Understanding overall data structure.***

  - ***Data handling, cleaning and standardization which will also involve:***
    - Data filtering and cleaning
    - Missing value and Outlier treatment
    - New derived column creation

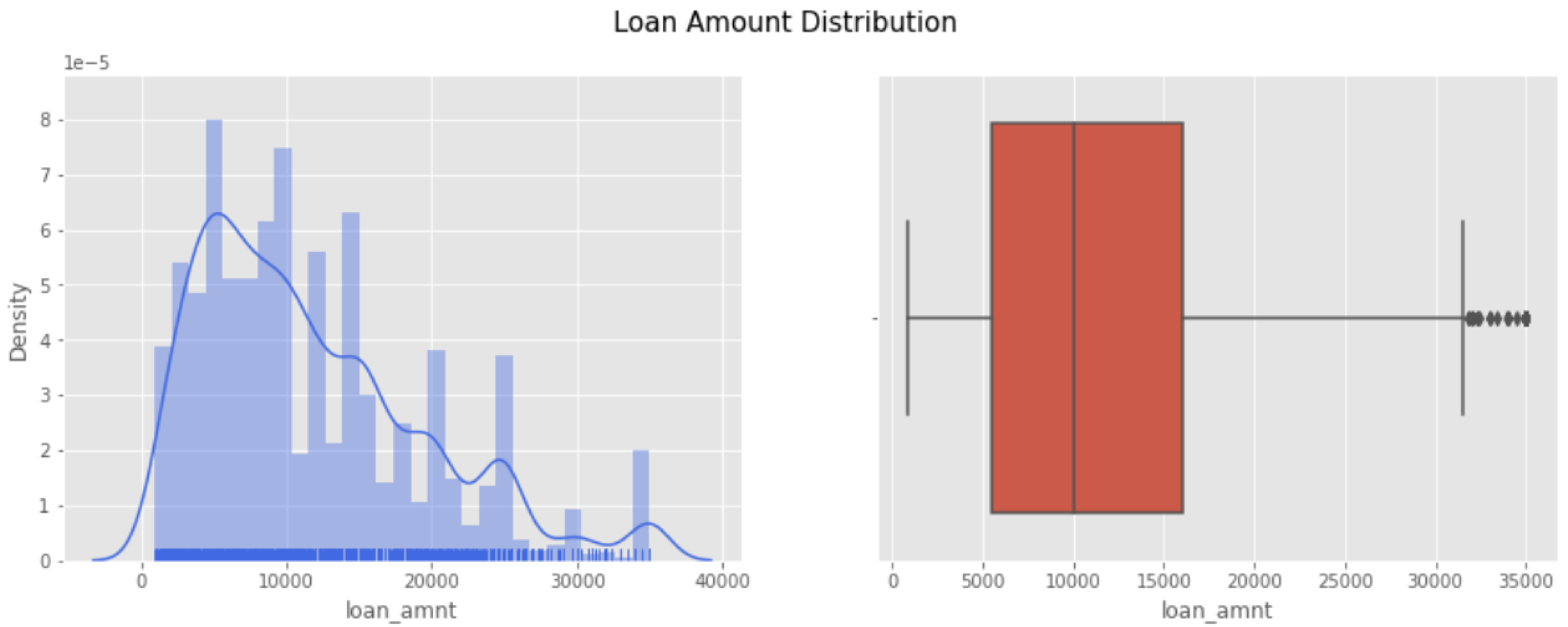
  - ***Exploratory Data Analysis which is broken into:***
    - Univariate Analysis – Defaulter data only
    - Segmented Univariate Analysis – Both paid off and defaulter data
    - Bivariate and Multivariate Analysis - Defaulter data only

  - ***Summary, findings and insights***

High Level Summary

- Overall data has got record of 38,577
- As per the data the current default rate is **~15%**.
- Annual Income : Whilst 75% of the borrowers have their annual income to be less than 80k there are certain high-income individuals. They can be treated as outliers and excluded from the analysis.
- Loan Amount : 75% of the borrowers are applying for less than 15k.
- Committed amount & and Funded Amount: comparing the values with the loan amount we can see that most of the applied amounts are getting approved and paid out to the borrowers.
- Interest rate : borrowers are paying approximately in the range 10% to 20%.
- Debt to Income Ratio : is less than 30 which is the acceptable value in the industry.
- Delinquency records for last 2yrs : most of the customers have not had any reported delinquencies in the last 2 years.
- Open Credit Lines: there multiple open credit lines which means for the borrowers could be cause of concerns
- Revolving Balance : Revolving balance seems to a bit high. This is in a way is good for LC as revolving balance accrues interest, but higher revolving balance may indicate that a borrower is relying too much on credit which can hurt his credit score.
- Revolving Utilization : Revolving utilization rate compares your credit debt to your total credit limit, and it is normally recommended to keep the utilization rate as low as possible with anything below 30% considered acceptable. The data suggests that utilization is higher which could again be a cause of concern.

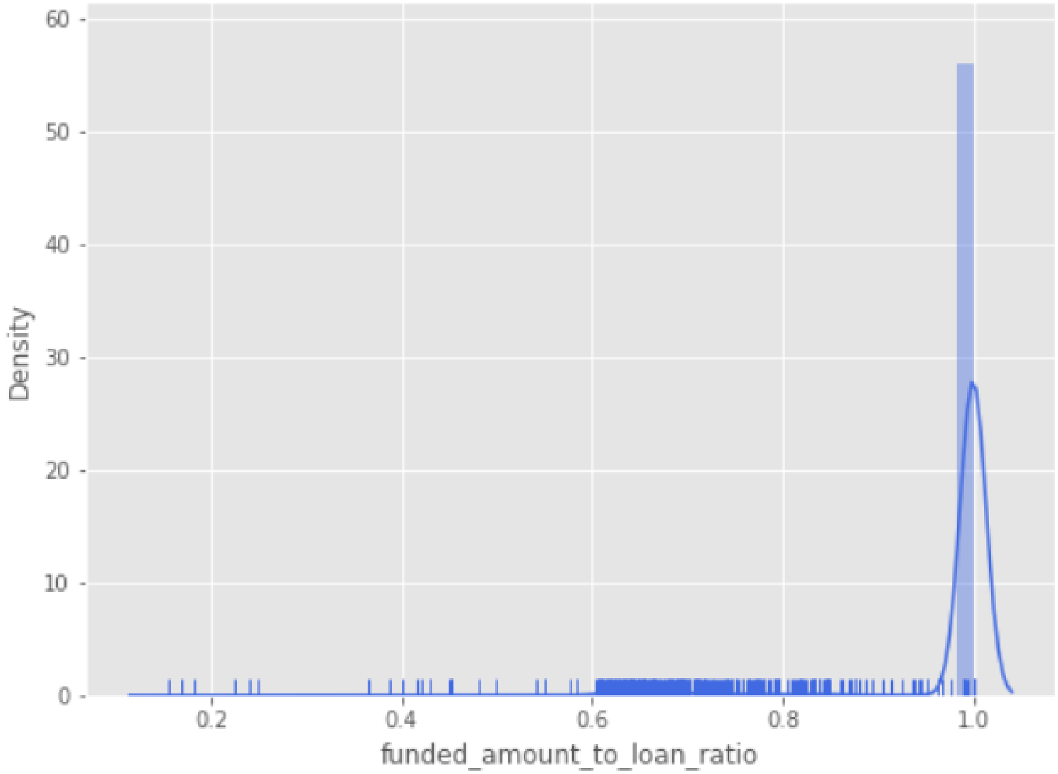
Univariate Analysis

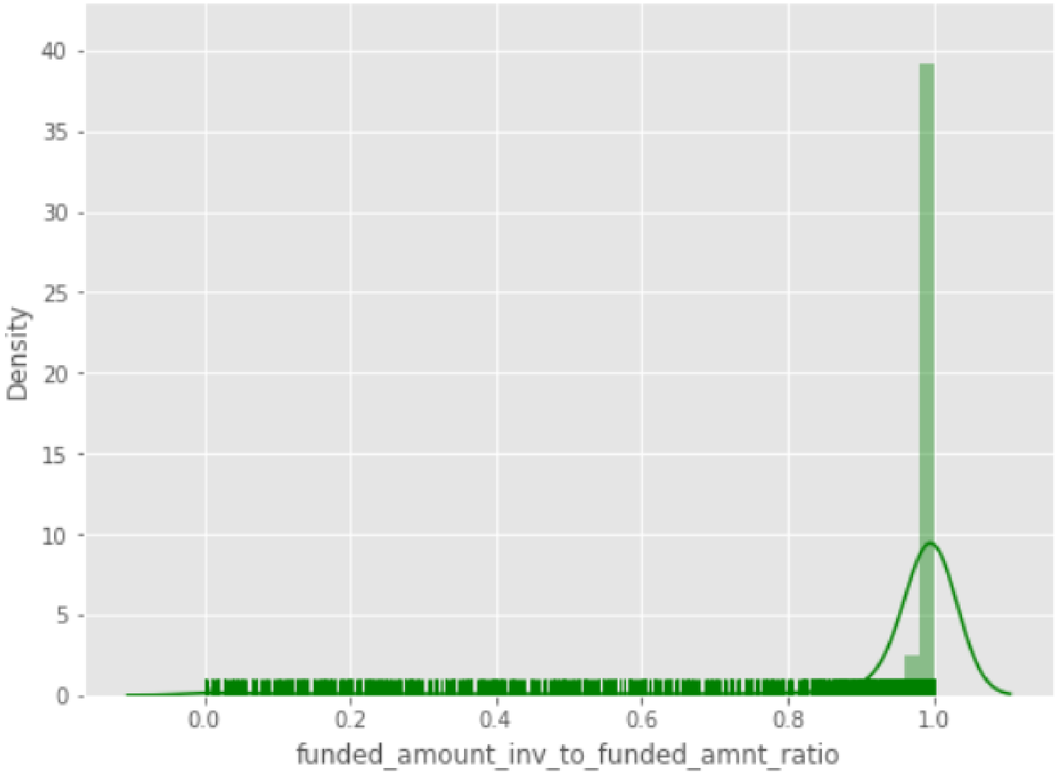


- 75% of the defaulters are applying for less than 15k loan amount.

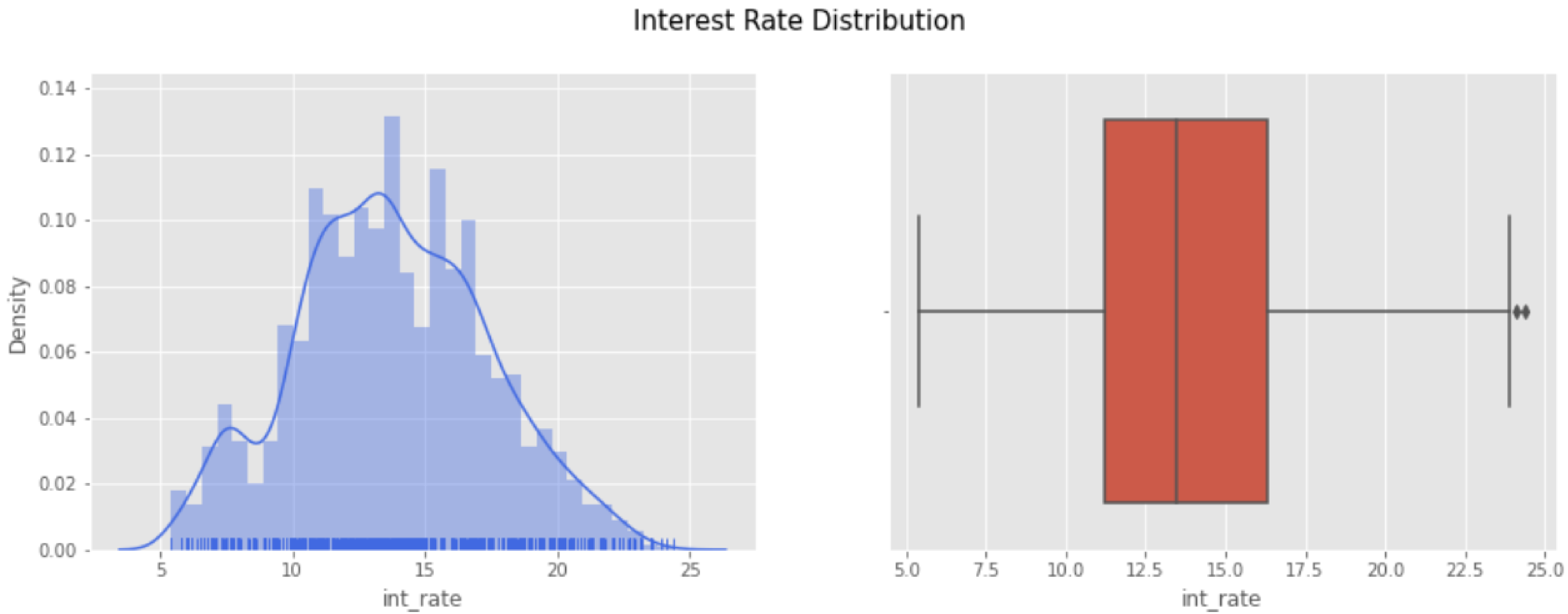
Comparison:

Committed Amount to Loan Amount
Committed Amount to Funded Amount

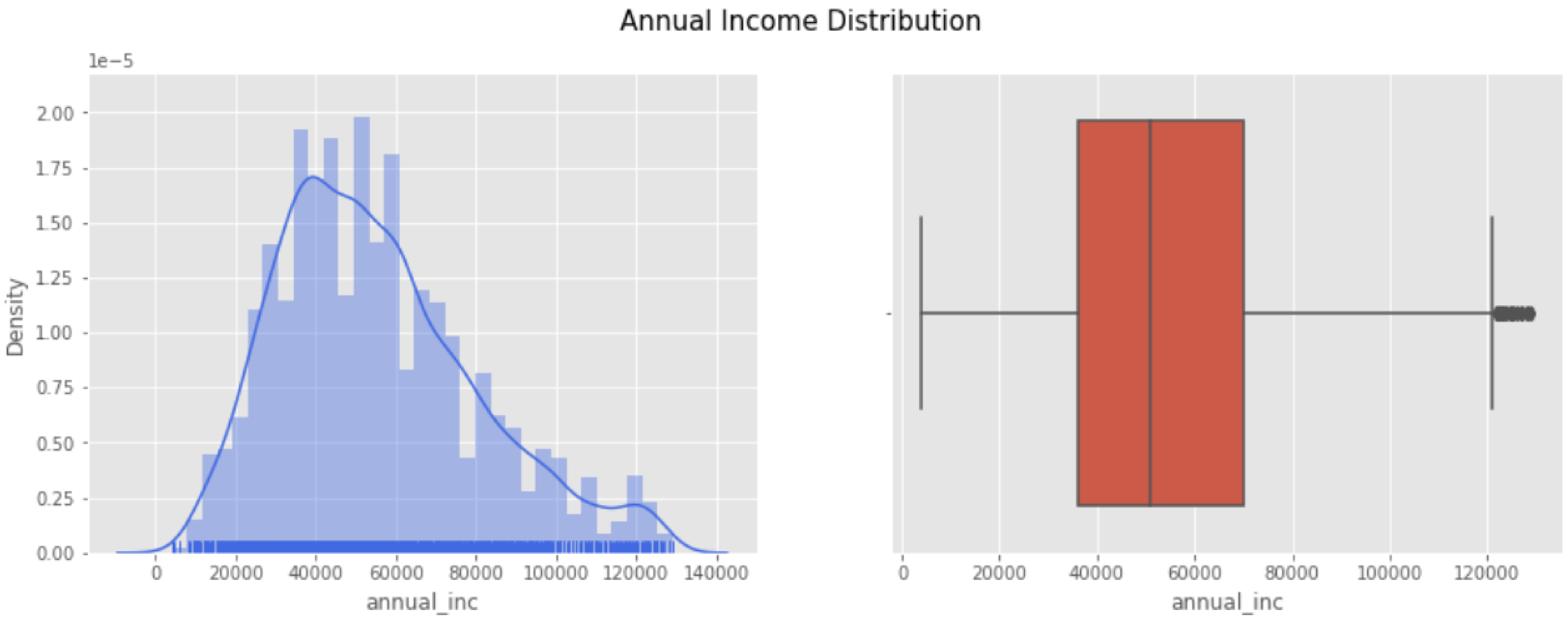




- Comparing the distribution of loan amount and funded amount we can say that most of the applied amounts have been approved and disbursed.

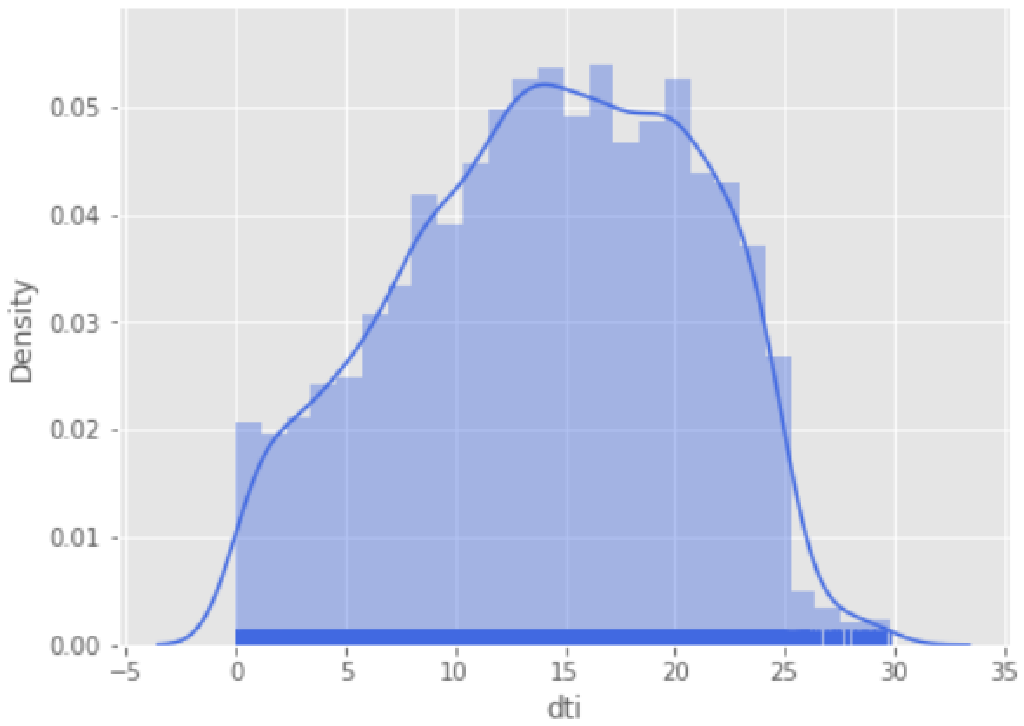


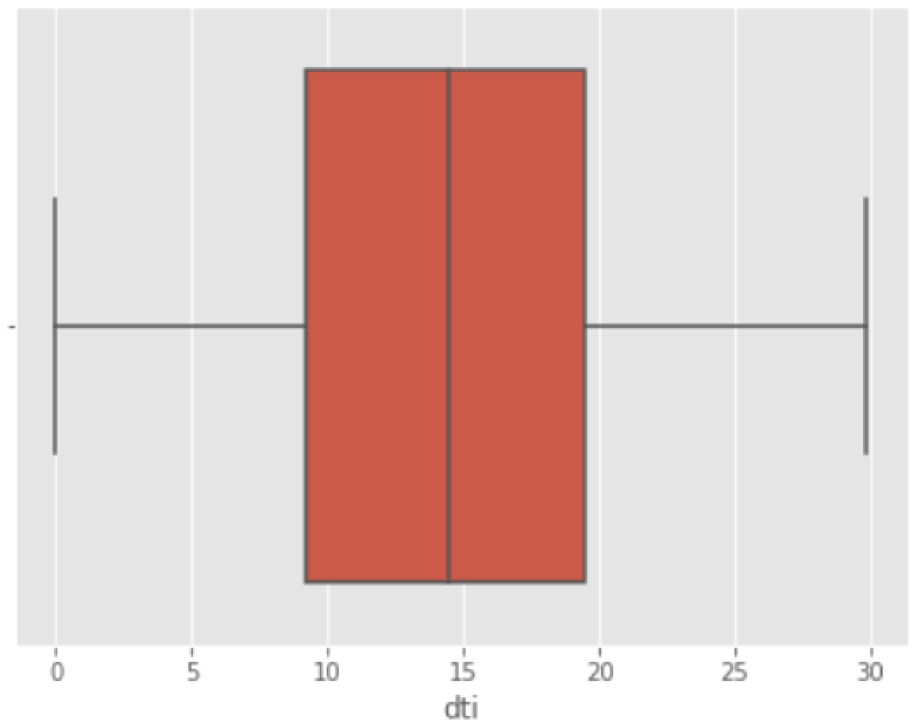
- Interest rates range between 9% to 16%. Some customers end up paying higher interest rates.



- Median income is below 50k. Majority of the defaulters are below 100k.
- As per the U.S. Census Bureau median household income in the United States in 2011 was 50,054 USD.
- Considering the above statistic and the fact that most the data from 2011 the defaulters seem to be mostly belonging to slightly higher income group.

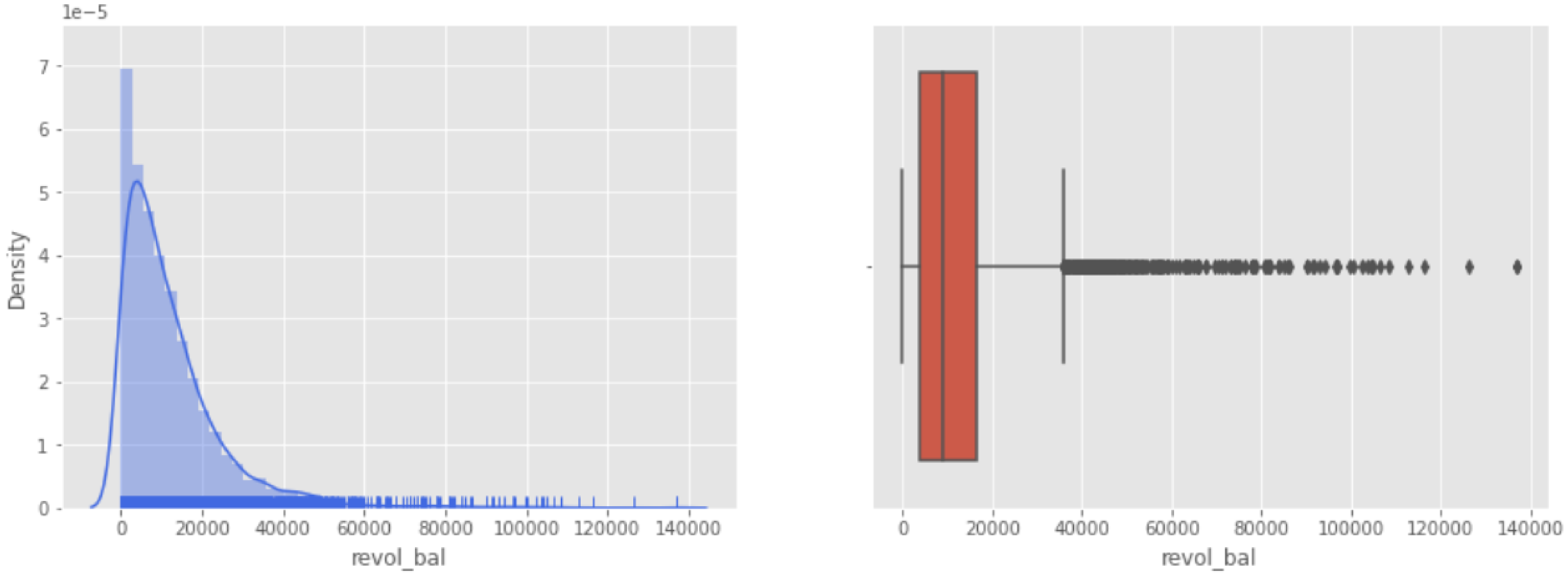
Debt to Income Ratio





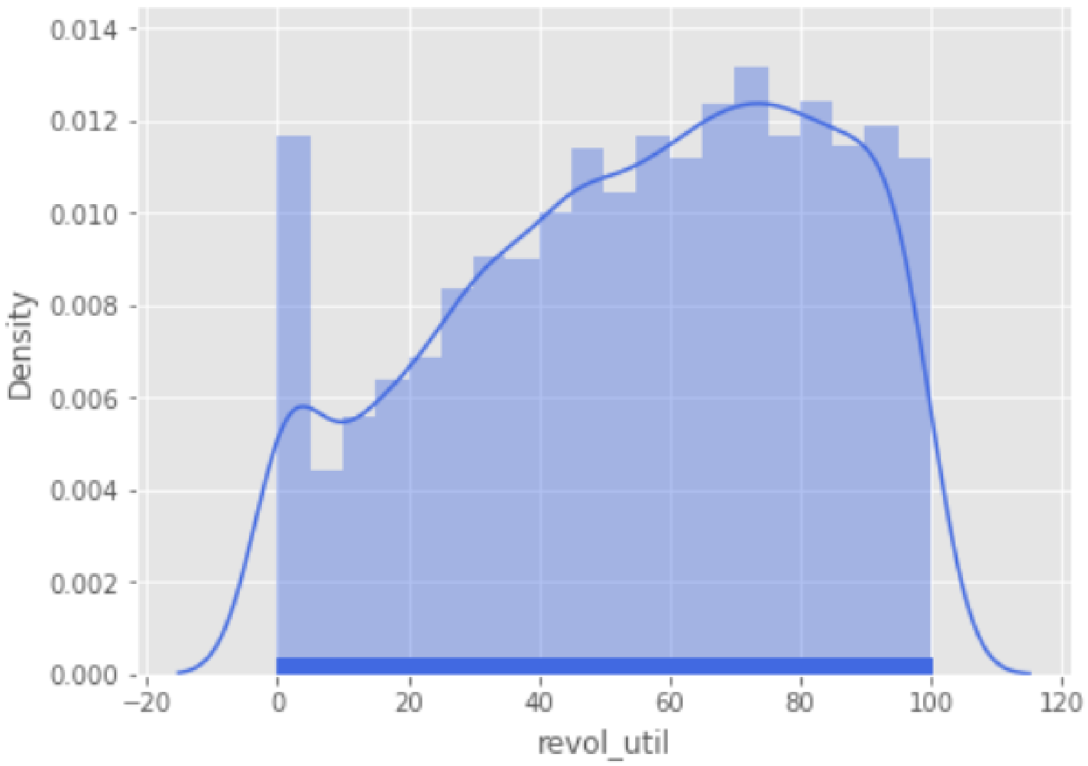
- Debt to income ratio is below 30 for all the defaulters which is acceptable rate as per the industry

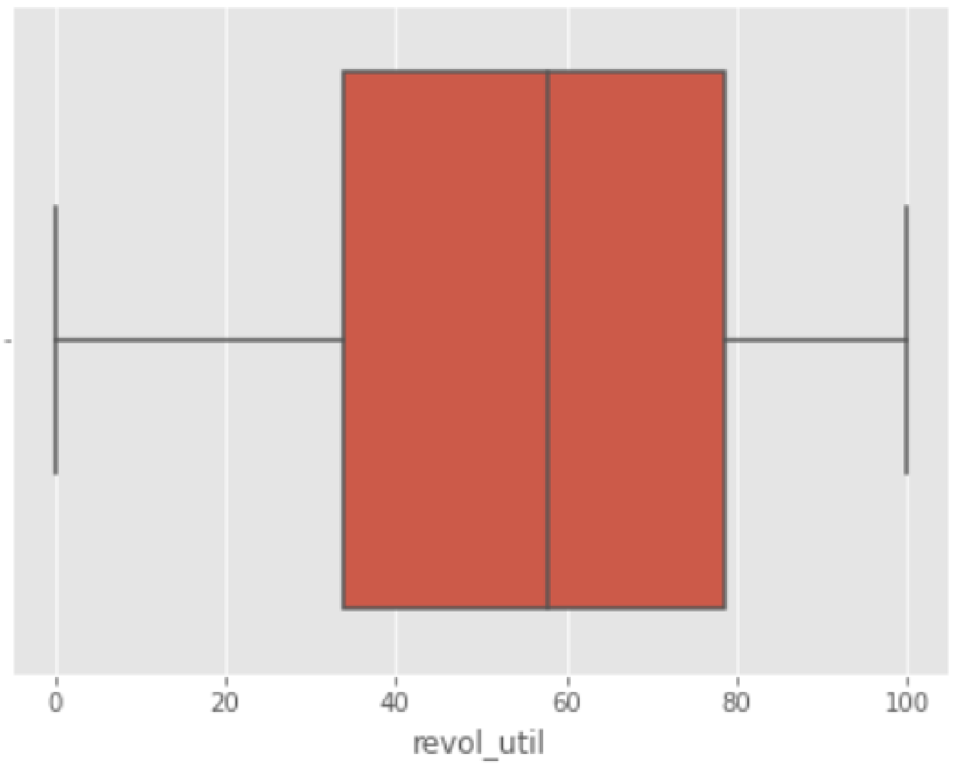
Revolving Balance Distribution



- Revolving balance seems to be below 20k for majority of the defaulters.

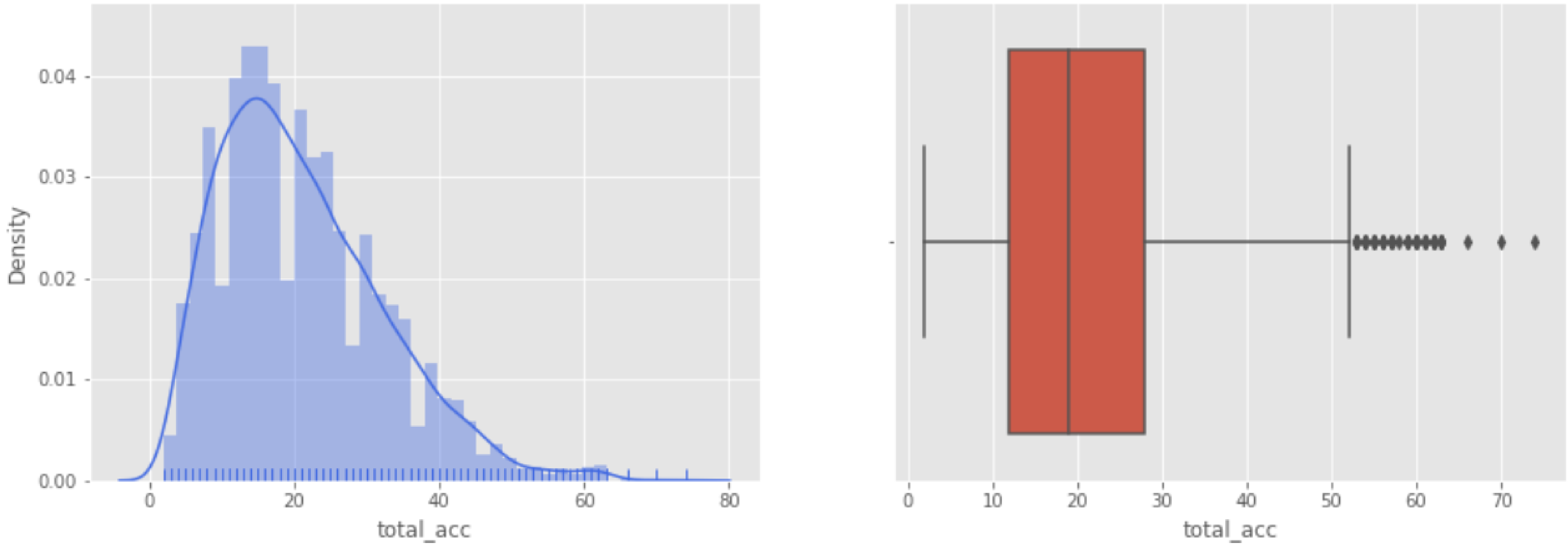
Revolving Utilization Distribution



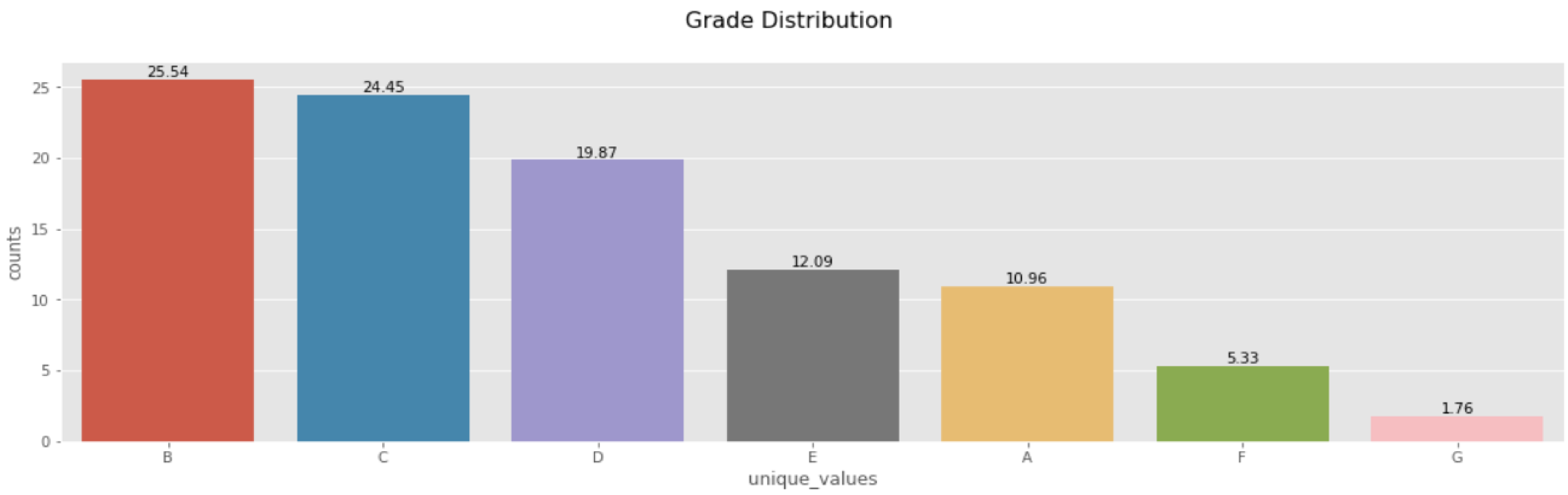


- We can see that the frequency is higher towards the higher utilization rates. majority the defaulters have the utilization rate between 30% to 80% which could be a cause of concern

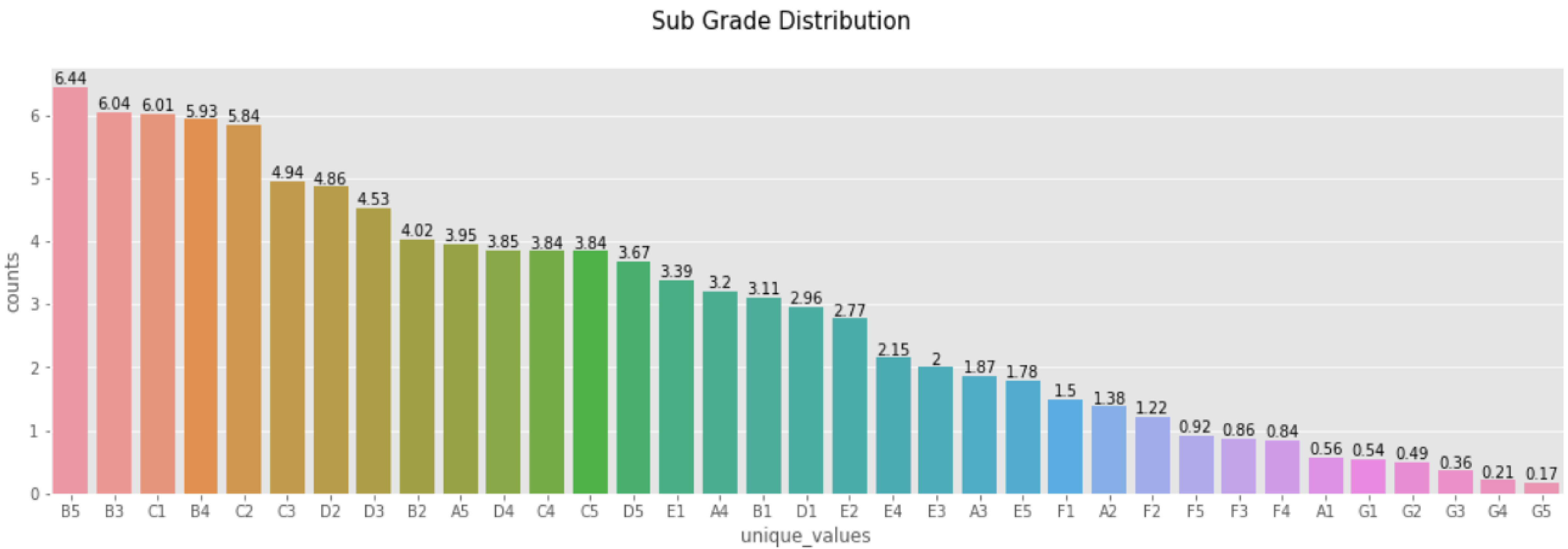
Total Accounts Distribution



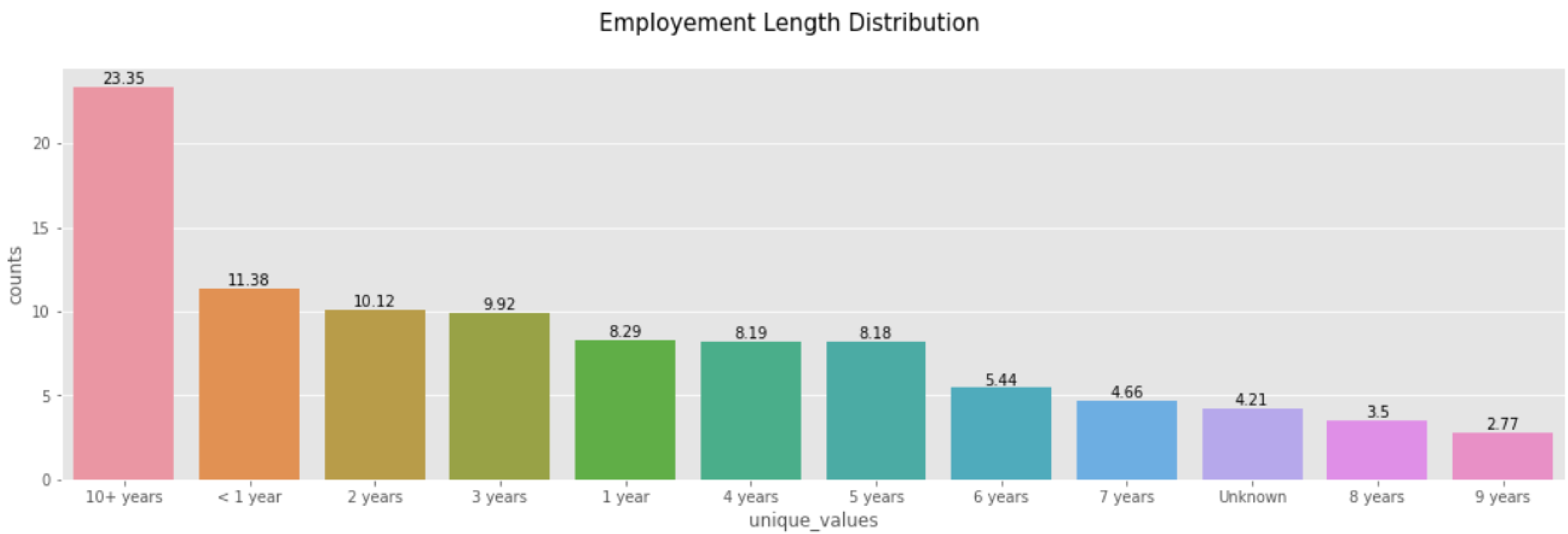
- As per understanding the total accounts would include open and closed credit lines. most of the defaulters seem to be having between 14 to 30 credit lines at any given point in time



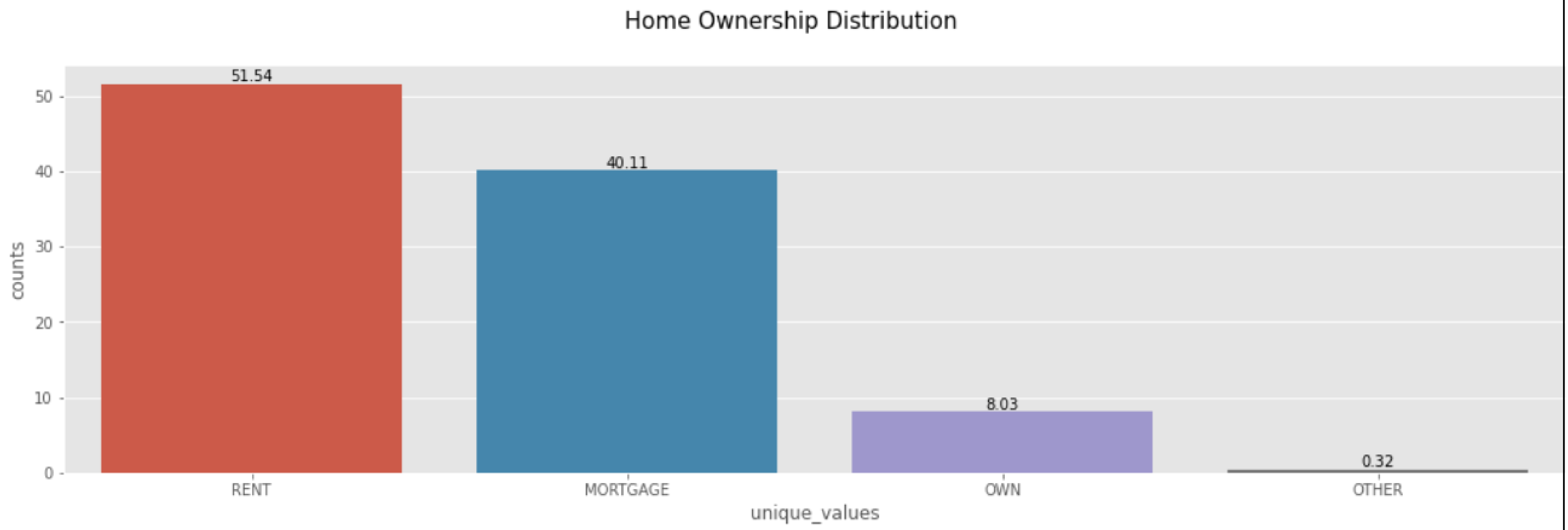
- 70% of the defaulters have grade B, C & D loans.
- Possibility to re-look at the loan grading and making it more stringent.
- This will also require a evaluation of trade off between potential business loss if people are evaluated strigently getting lower grades and there by denied loans



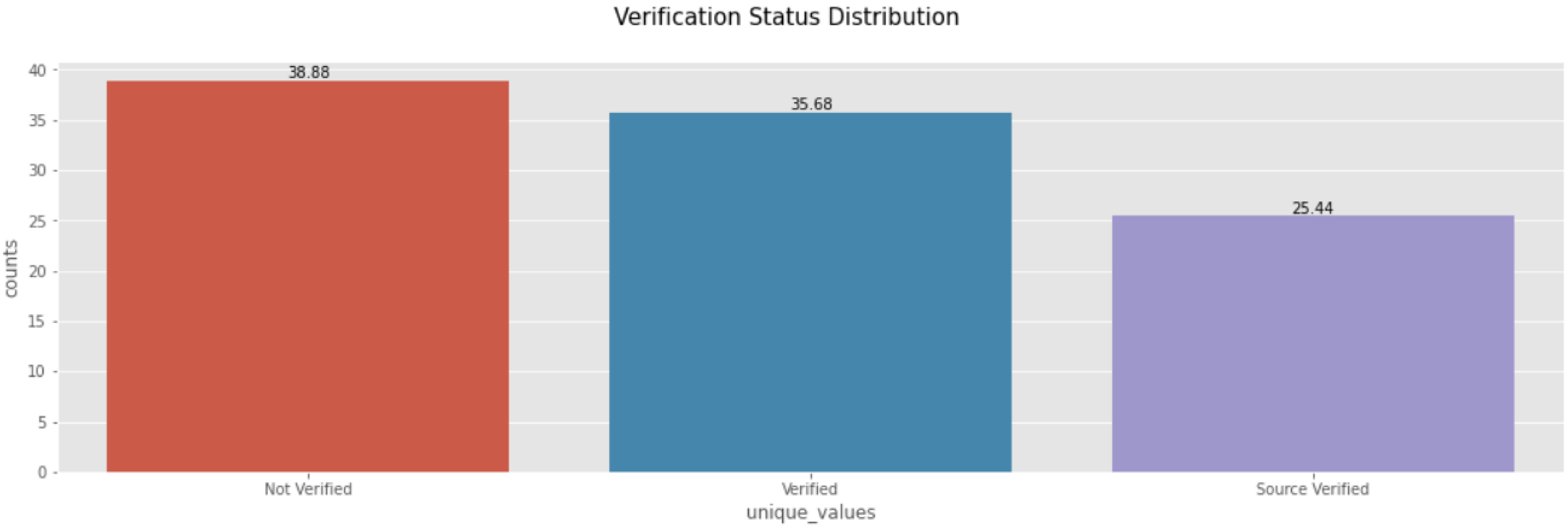
- we can see that mostly lower sub-grades (for.eg. B5, B3, C1 etc.) within the grades have a higher default



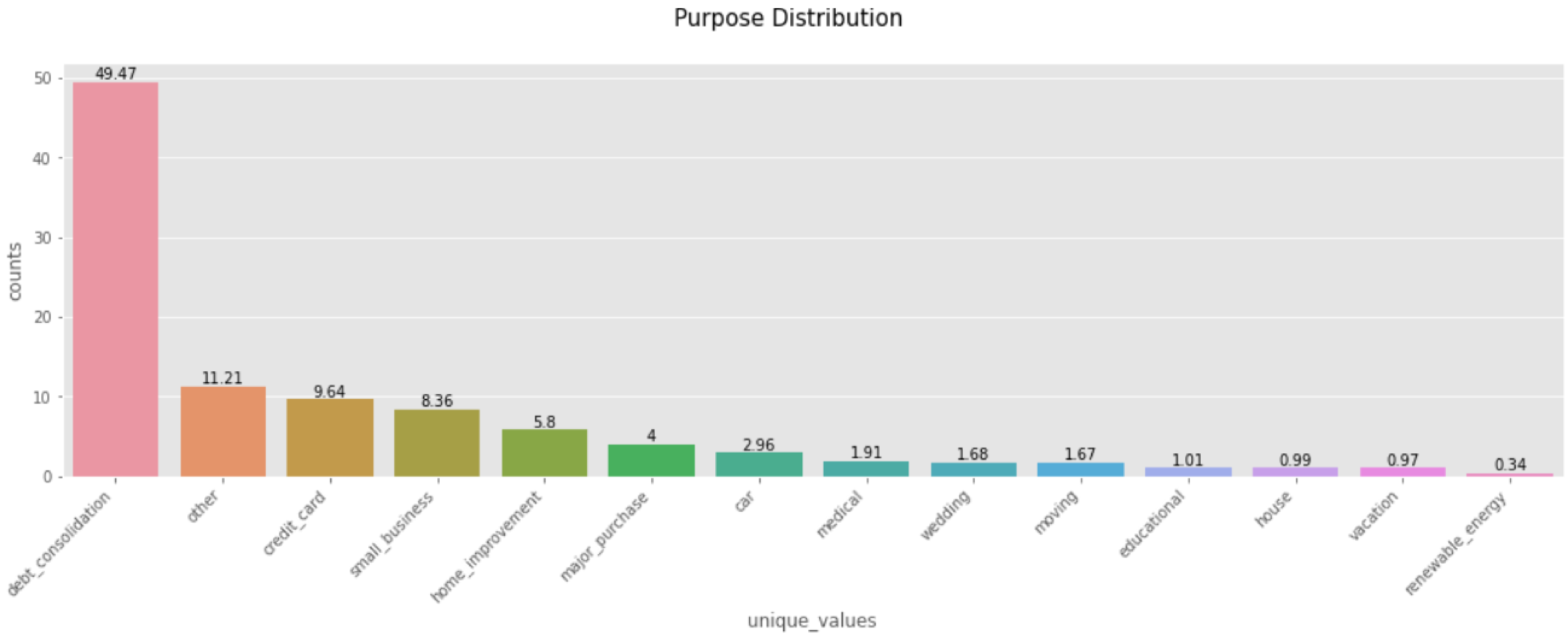
- People with more than 10+ years employment experience seem to be defaulting the most. This could also be because:
  - This could also be because these people are given a preferance and most their loans are sanctioned
  - These are people nearing retirement or retired and have lesser sources of incomes
- Apart from that bucket we can see that as the employement length increases defaulters are decreasing



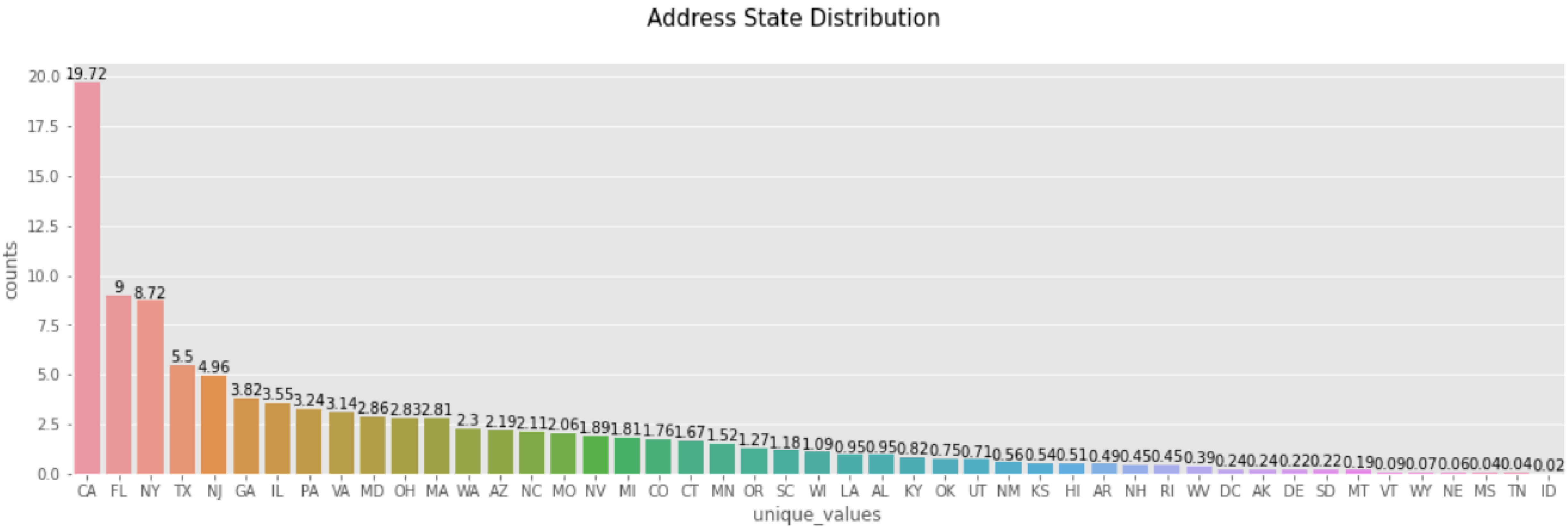
- An overwhelming 91% of the defaulters are either living in a rented apartment or a mortgaged house.



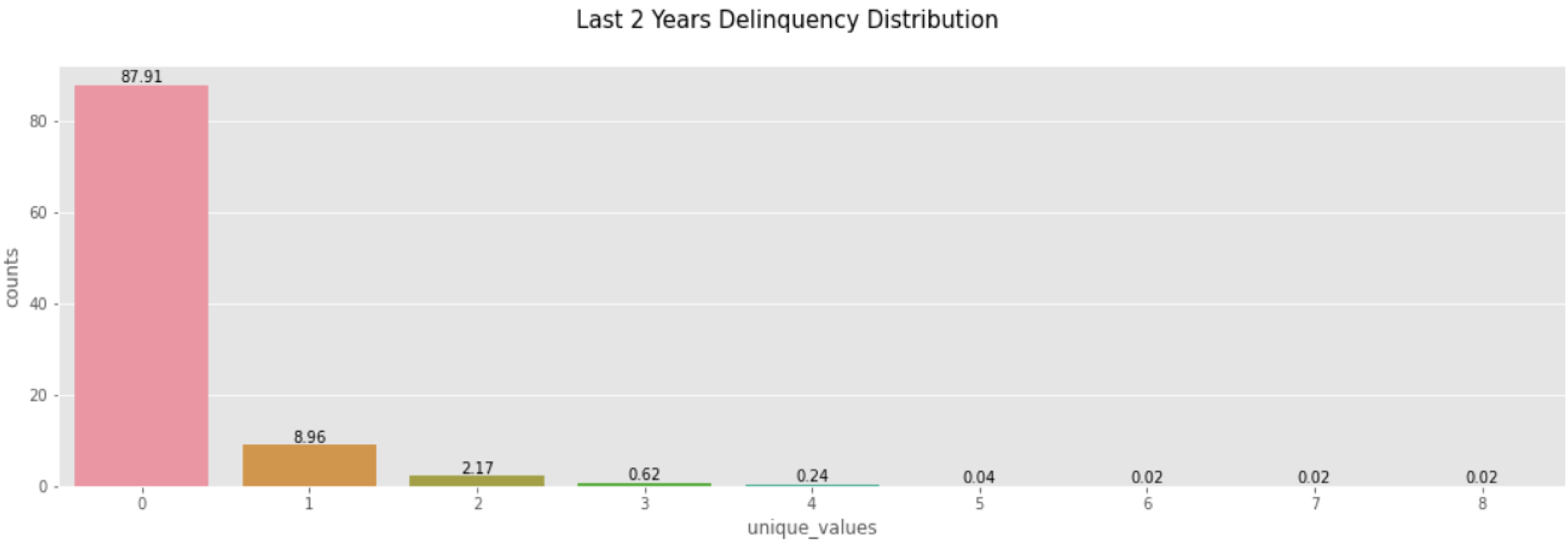
- Only 25% of declared income has it's sourced verified which leads to higher chances of fraudulent declaration



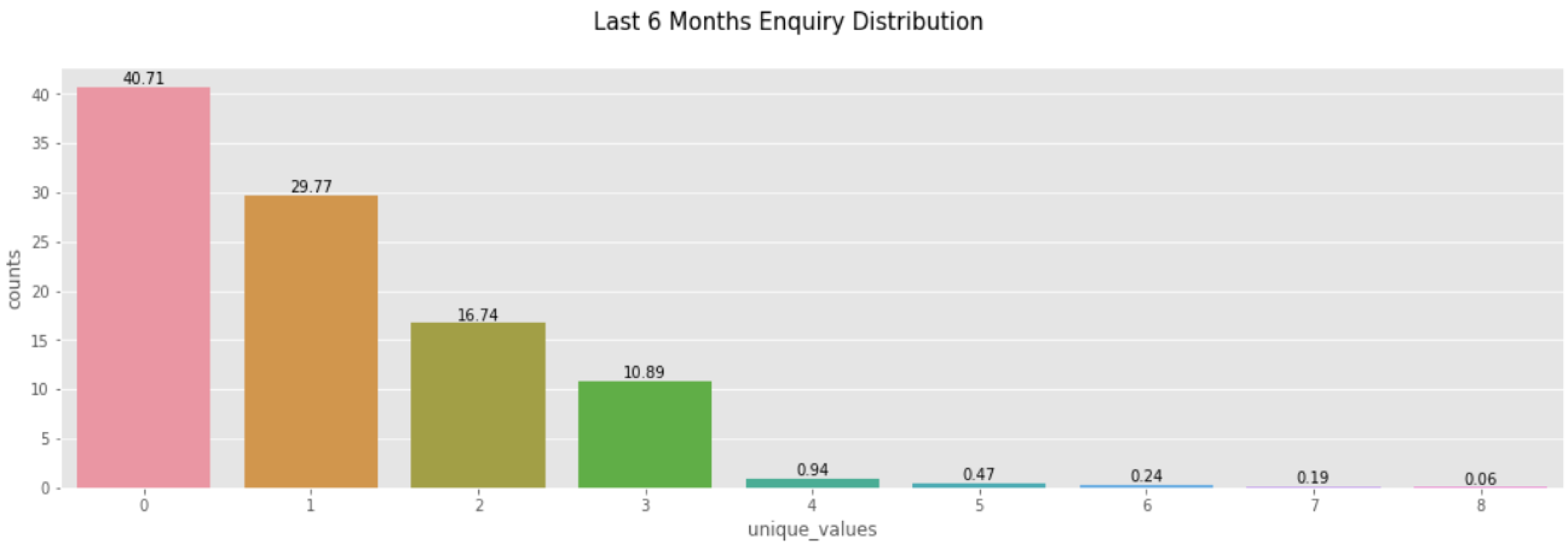
- Almost 50% of the defaulters are taking loan for debt consolidation purpose which could essentially mean taking a loan to pay of other loan or liabilities.
- ~11% coming from others category which could also be a possible incorrect data categorization issue



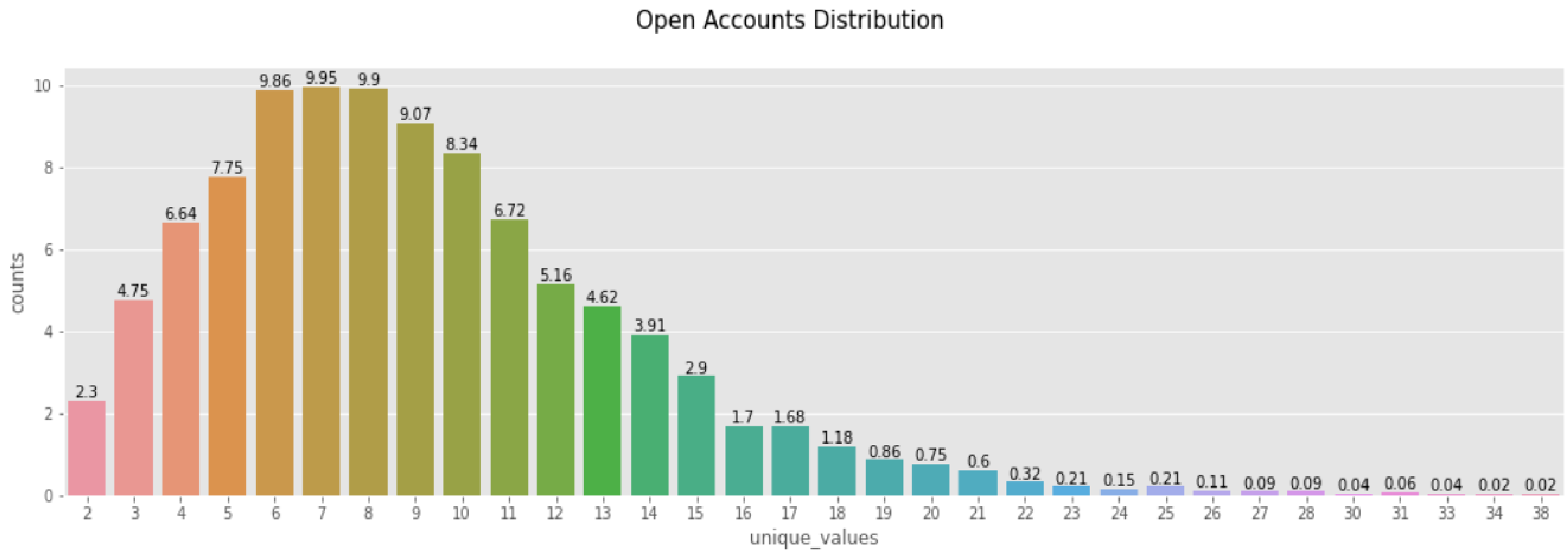
- Top 5 states namely California, Florida, New York, Texas and New Jersey contributing ~50% of the total default



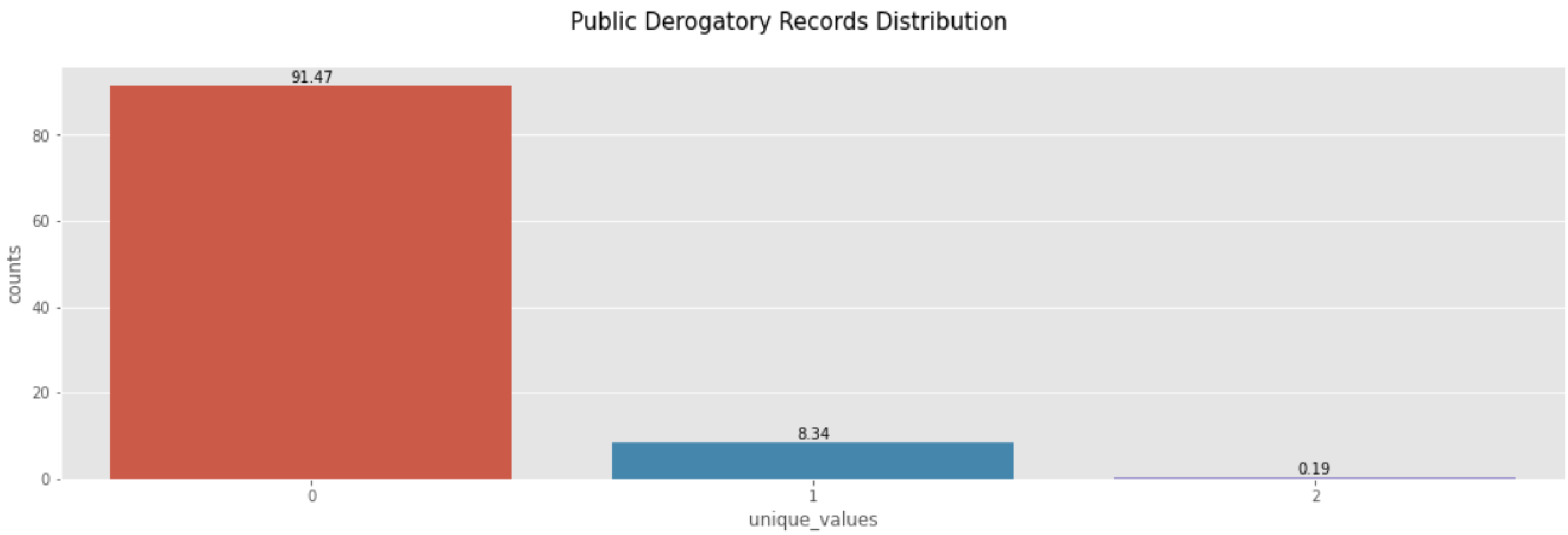
- No instances of 30 days past due in the last 2 years for majority of the defaulters. So most of seem to paying their dues on time within 30 days of due date



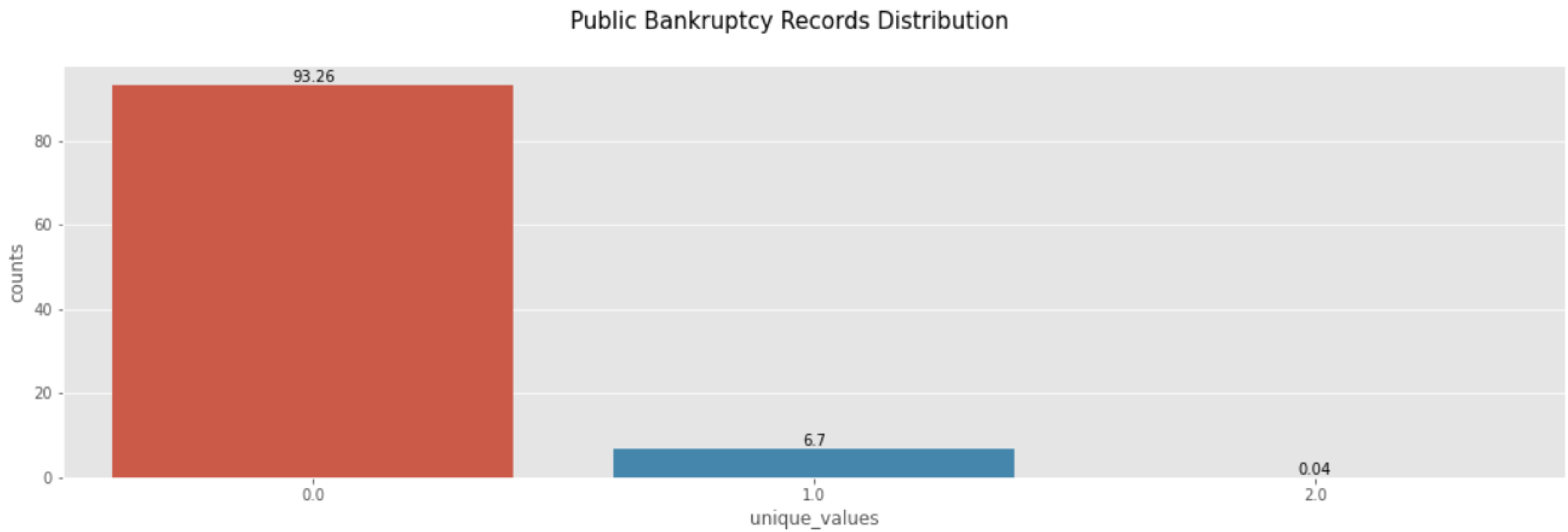
- 60% of the defaulters have had a loan enquiry in the last 6 months. In ~30% instances more than 2 enquiries



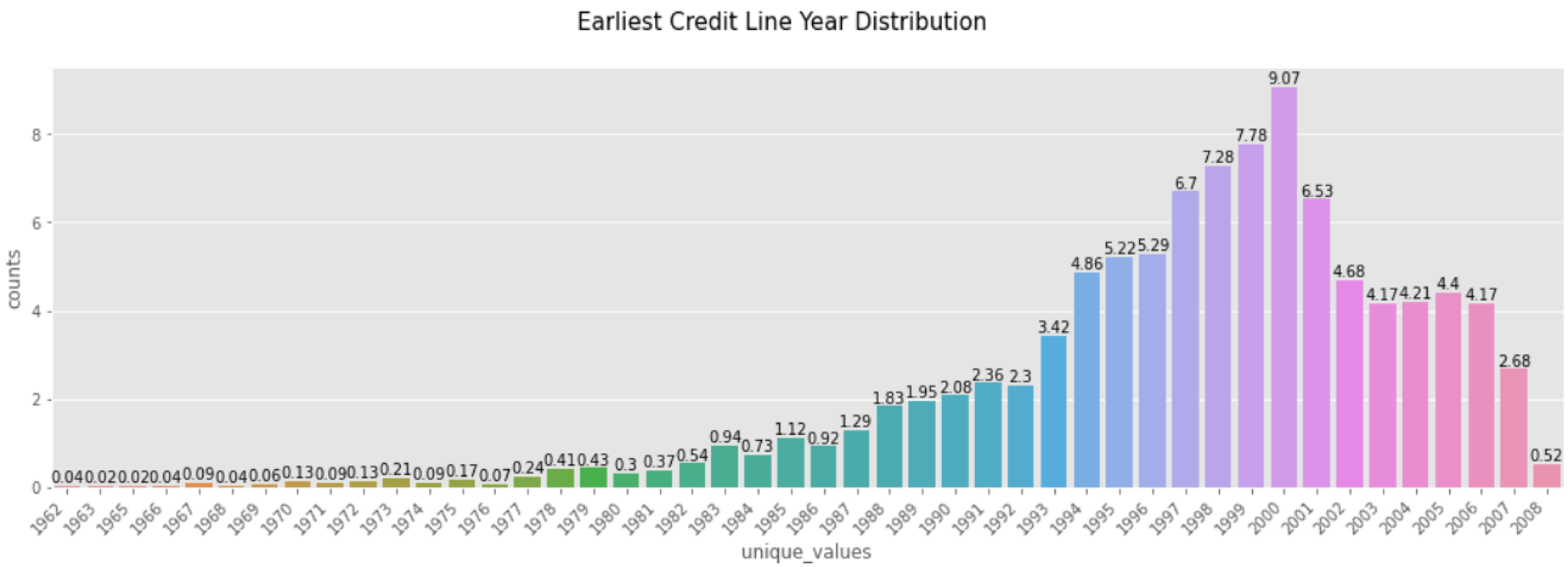
- Majority of the defaulters have between 2 to 20 open credit lines



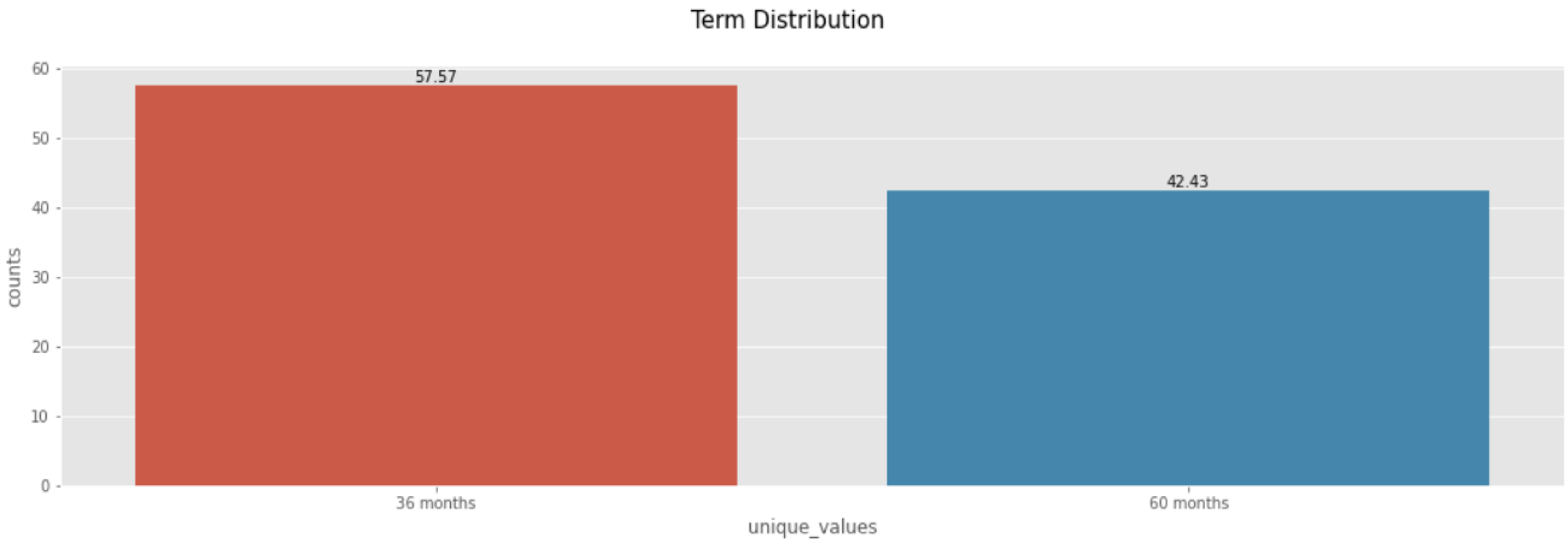
- Most of the defaulters don't have a derogatory public record in the past.
- If we go by the data most of the defaulters are **first time defaulters**.



- This chart validates the finding of the previous chart



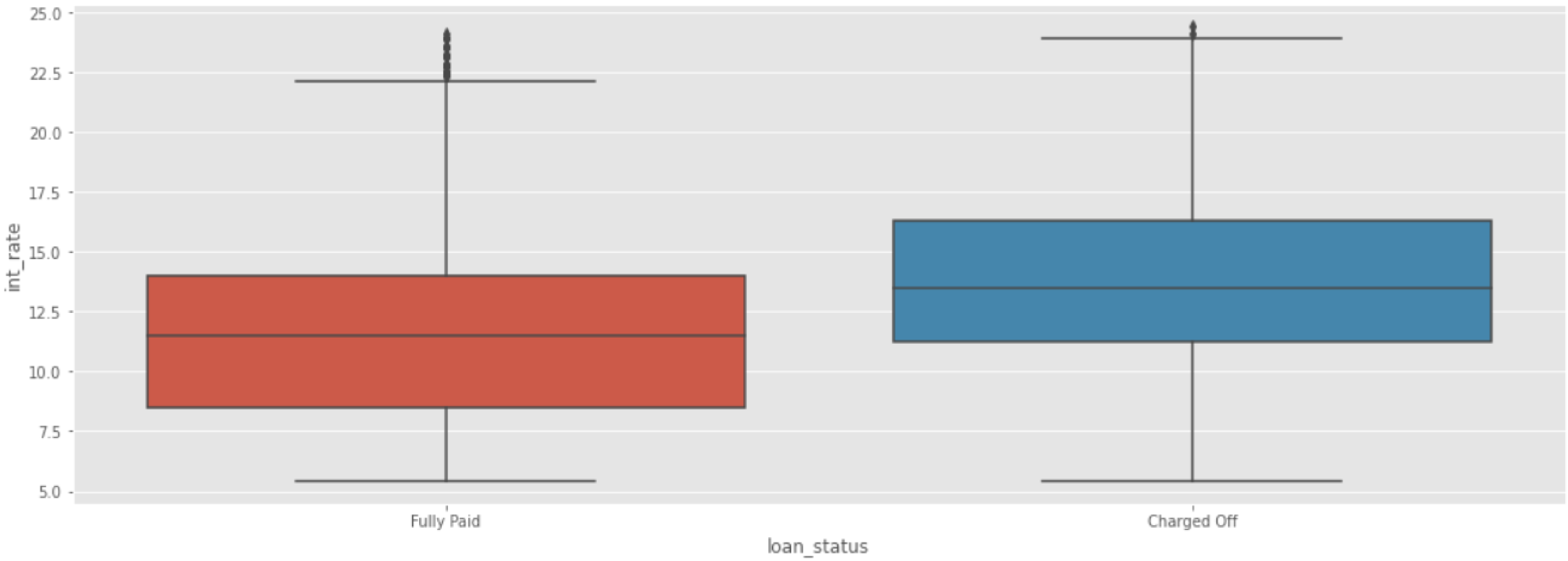
- Majority of the defaulters have their credit lines starting from the late 80s with the number peaking in 2000. Assuming that the person needs to 18 years and above to take a loan can we say that people started to take more loans as the credit industry became more structured.



- More than half of the defaulters have 36 months term

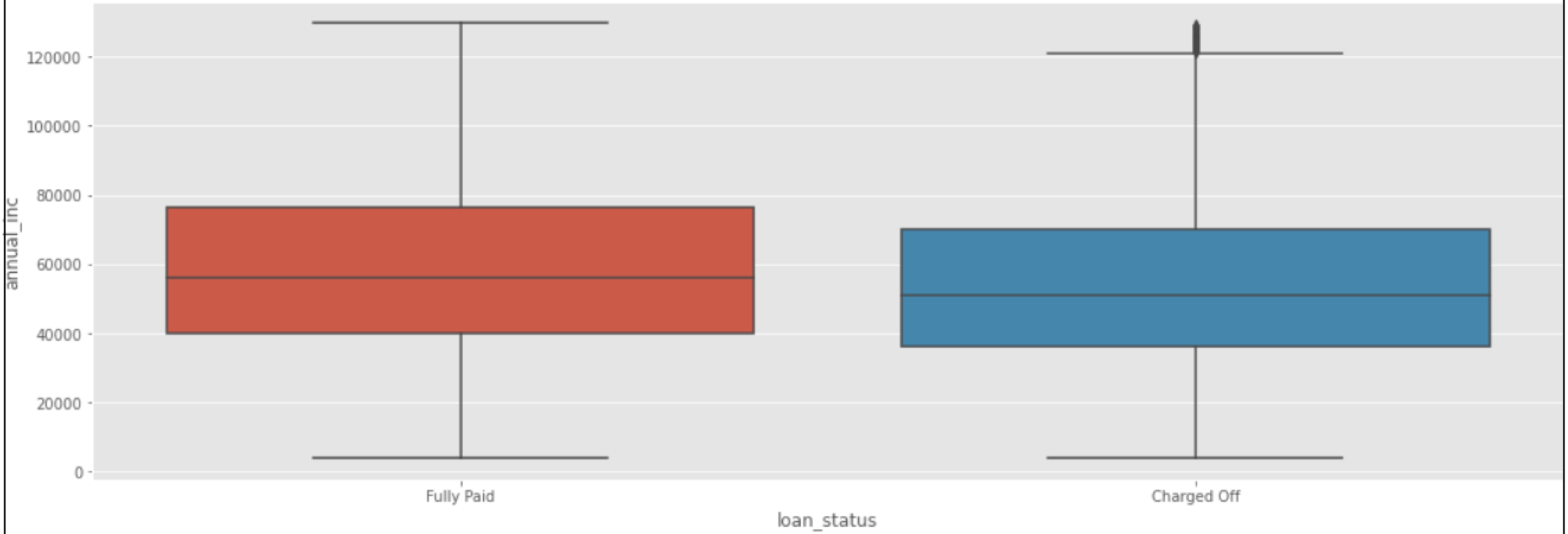
Segmented Univariate Analysis

Loan status and Interest Rate comparison

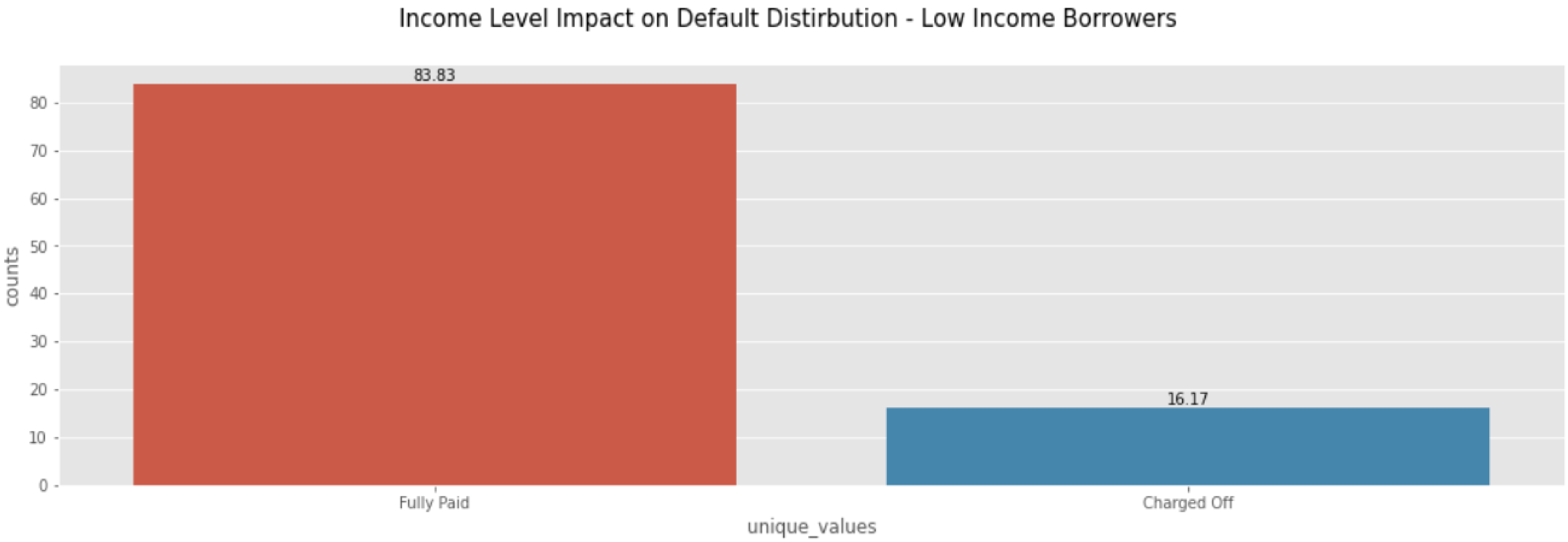


- Charged off customers seemed to be paying higher interest than the fully paid customers.
- Would higher interest rate be one of the reasons of default ?

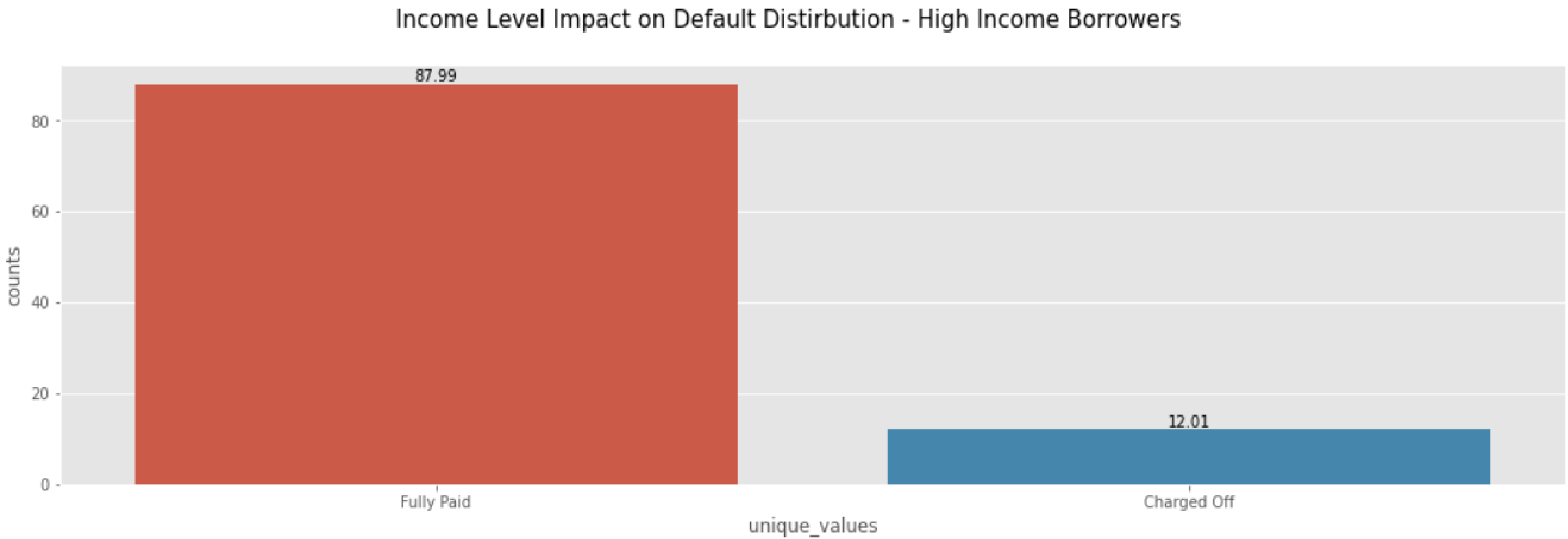
Loan status and Annual Income comparison



- Charged off borrowers having lower income than fully paid borrowers



- Default rate increases slightly for lower income levels

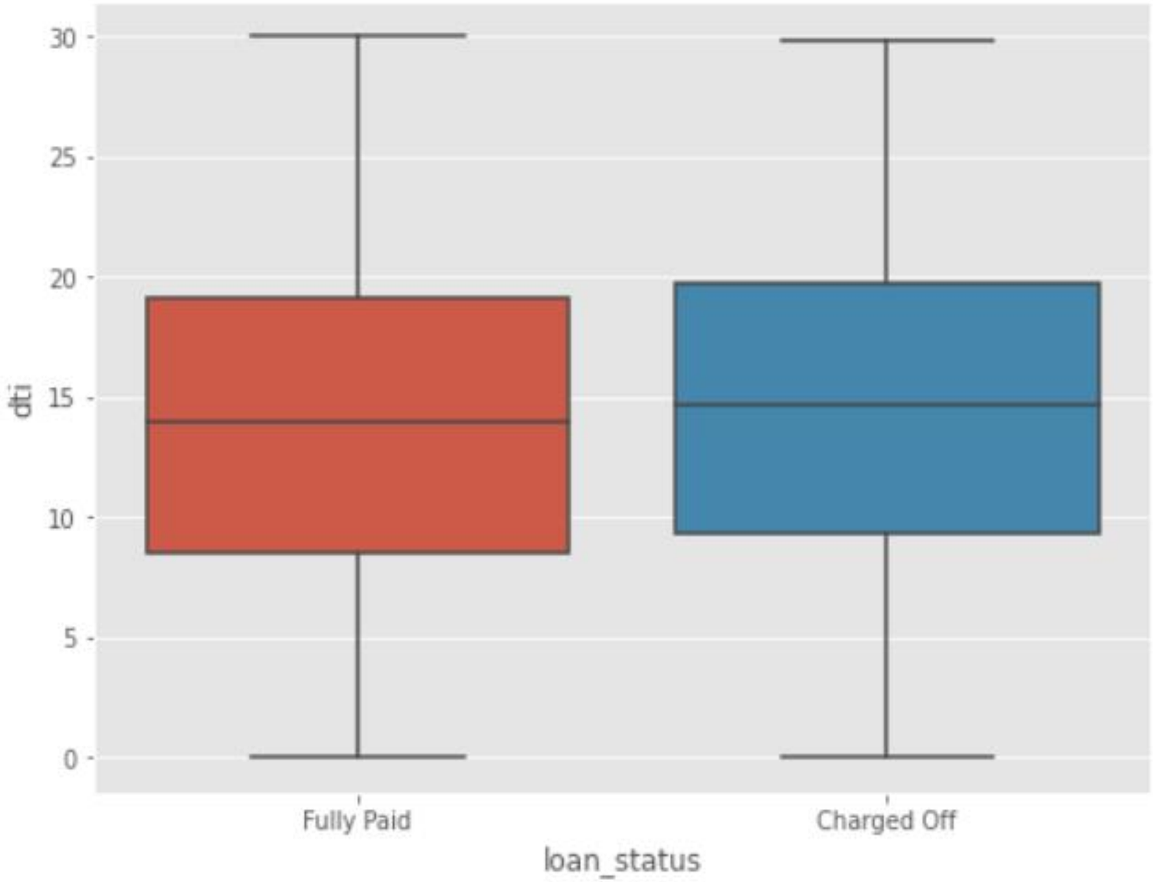


- Default rate goes down considerably (by ~400 bps) for higher income levels. This is quite intuitive.

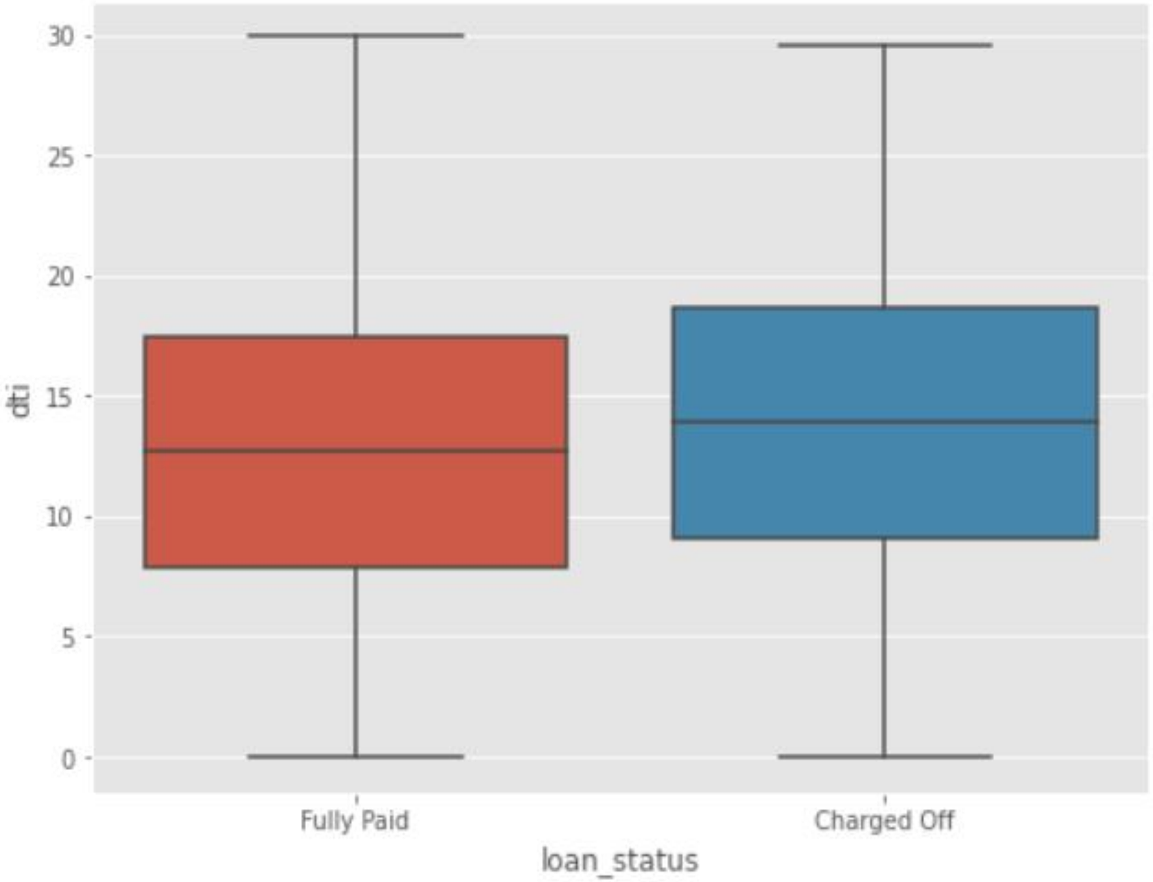
Comparing debt to income ratio across salary levels

Debt To Income Ratio Across Income Levels

Low Income Borrowers – less than 70k



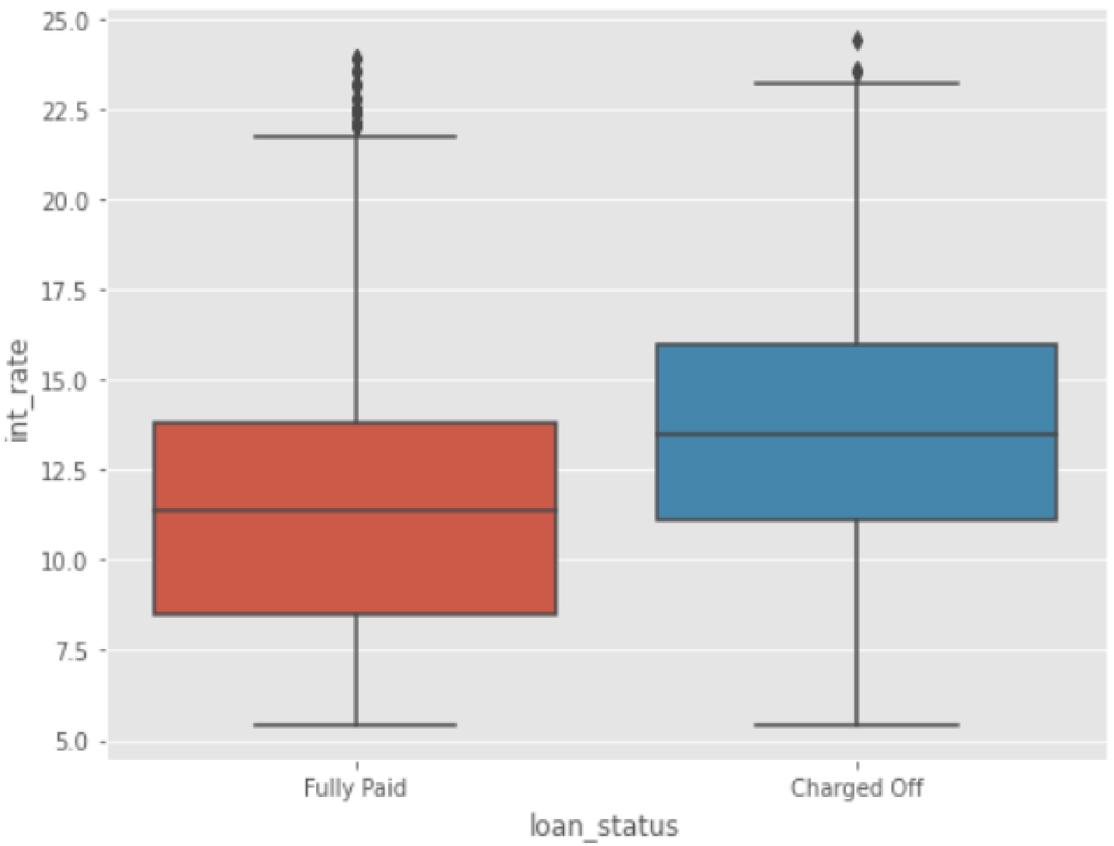
High Income Borrowers – more than 70k



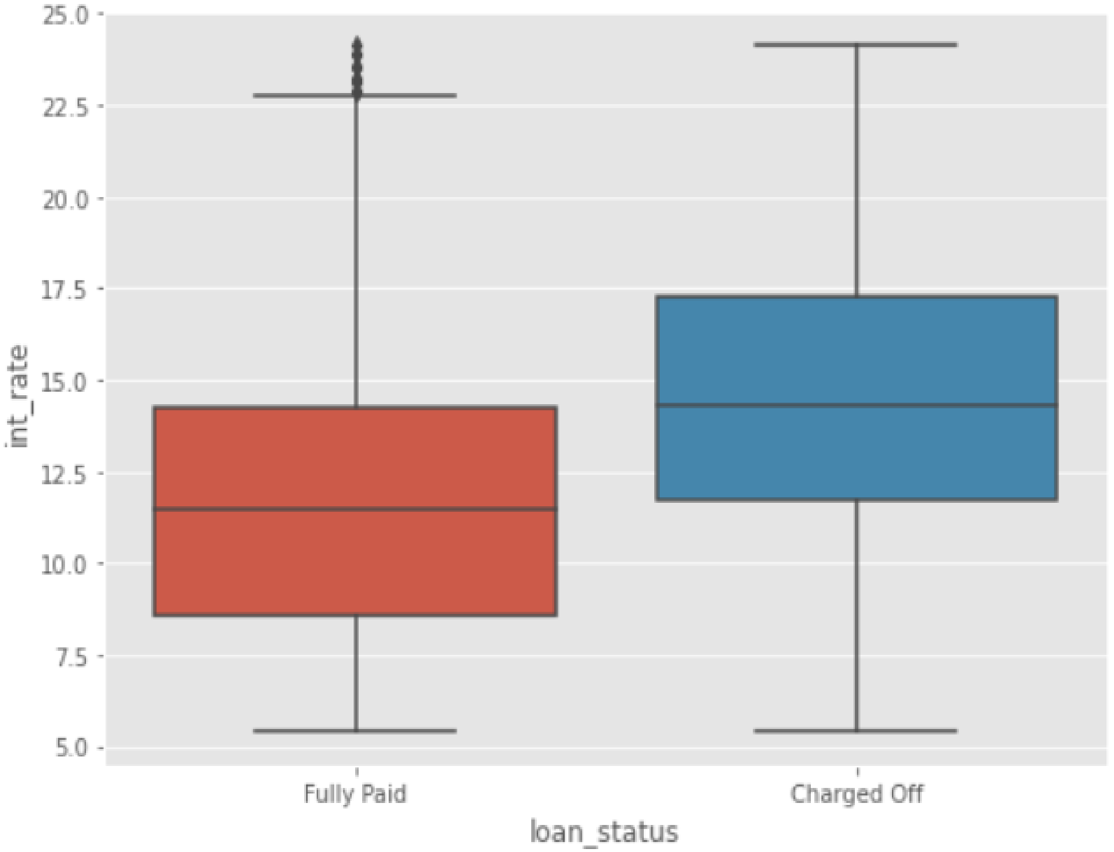
- Defaulters with lower salary levels have a high debt to income ratio

Debt To Income Ratio Across Income Levels

Low Income Borrowers – less than 70k



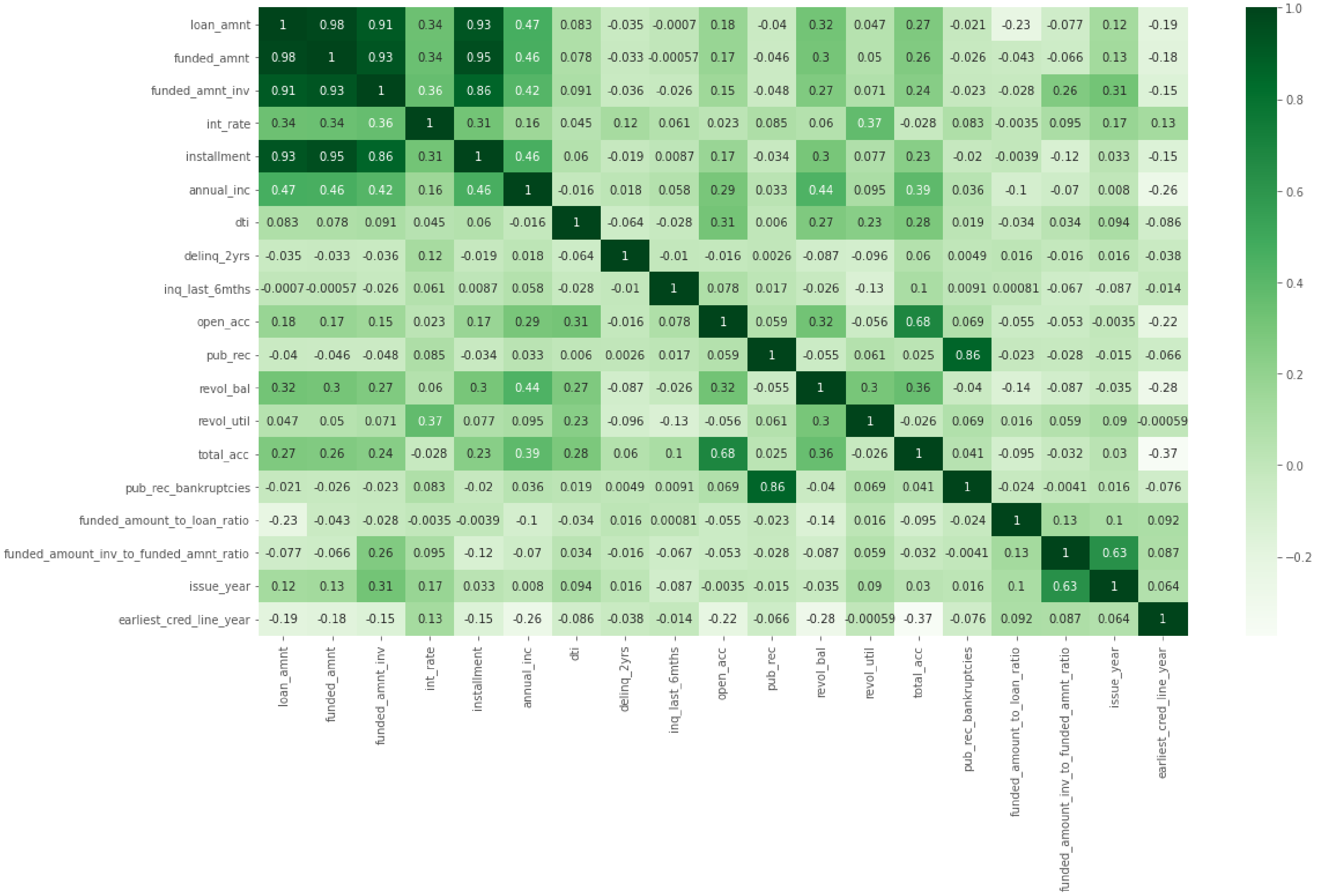
High Income Borrowers – more than 70k



- Defaulters having higher income levels are paying higher interest rate

Bi-Variate & Multivariate Analysis

Correlation Grid



- Correlation does not provide any additional insight. Most of the high correlations are intuitive.

grade	A	B	C	D	E	F	G
verification_status							
Not Verified	309	602	575	393	140	42	17
Source Verified	138	344	303	299	170	81	25
Verified	139	419	429	370	336	162	52

- Defaulters are high for Not Verified and Verified status for B, C and D grades

grade	A	B	C	D	E	F	G
home_ownership							
MORTGAGE	255.00	553.00	516.00	400.00	256.00	128.00	36.00
OTHER	2.00	5.00	2.00	5.00	2.00	1.00	NaN
OWN	52.00	112.00	106.00	77.00	49.00	22.00	11.00
RENT	277.00	695.00	683.00	580.00	339.00	134.00	47.00

- Defaulters are high for mortgaged and rented properties for B, C and D grades

verification_status	Not Verified	Source Verified	Verified
home_ownership			
MORTGAGE	800	446	898
OTHER	10	1	6
OWN	182	113	134
RENT	1086	800	869

- Deafulters are almost double for verified income borrowers between mortgaged and rented properties

verification_status	Not Verified	Source Verified	Verified
purpose			
car	68	51	39
credit_card	197	131	187
debt_consolidation	916	652	1076
educational	37	4	13
home_improvement	136	77	97
house	23	16	14
major_purchase	96	62	56
medical	50	29	23
moving	31	44	14
other	309	144	146
renewable_energy	9	4	5
small_business	152	105	190
vacation	24	15	13
wedding	30	26	34

- Debt Consolidation by far the riskiest purpose across income verification categories

home_ownership	MORTGAGE	OTHER	OWN	RENT
purpose				
car	68.00	1.00	15.00	74.00
credit_card	214.00	2.00	37.00	262.00
debt_consolidation	1070.00	6.00	190.00	1378.00
educational	11.00	NaN	5.00	38.00
home_improvement	233.00	1.00	39.00	37.00
house	15.00	NaN	3.00	35.00
major_purchase	70.00	NaN	26.00	118.00
medical	40.00	NaN	10.00	52.00
moving	8.00	1.00	6.00	74.00
other	168.00	2.00	58.00	371.00
renewable_energy	9.00	NaN	1.00	8.00
small_business	195.00	4.00	32.00	216.00
vacation	12.00	NaN	4.00	36.00
wedding	31.00	NaN	3.00	56.00

- People living in reneted properties with debt consolidation highest contribution to default

- **Major Observations & Conclusions:**

- **Factors Impacting Default**

  - Most of the applied loan amounts are committed and disbursed. Possible review required for the sanction process.
  - Higher interest could be one of the major factors leading to default
  - Debt to income ratio is less than 30 which is the acceptable value in the industry.
  - most of the borrowers have not had any reported delinquencies in the last 2 years.
  - There are multiple open credit lines which means for the borrowers which could be cause of concerns
  - Revolving balance seems to a bit high. This is in a way is good for LC as revolving balance accrues interest, but higher revolving balance may indicate that a borrower is relying too much on credit which can hurt his credit score.
  - Revolving utilization rate compares your credit debt to your total credit limit, and it is normally recommended to keep the utilization rate as low as possible with anything below 30% considered acceptable. The data suggests that utilization is in the range of 30% to 80% which could again be a cause of concern.
  - Possibility to re-look at the loan grading and making it more stringent. This will also require an evaluation of trade off between potential business loss if people are evaluated stringently getting lower grades and there by denied loans
  - Only 25% of declared income has it's sourced verified which leads to higher chances of fraudulent declaration. This the verification process requires a closer look.

- **Generic Profile of a Defaulter**

  - Having income below 70k.
  - People with more than 10+ years employment experience seem to be defaulting the most. Apart from this bucket we can see that as the employment length increases defaulters are decreasing.
  - Having a rented apartment or a mortgaged house.
  - Applicants from either of these 5 states namely California, Florida, New York, Texas and New Jersey. **Although this is not a clinching factor.**
  - Have had a loan enquiry in the last 6 months.
  - Have between 2 to 20 open credit lines.
  - Don't have any derogatory public record in the past making them first time defaulters (as per the data).
  - Don't have any public delinquency record in the past.
  - Having known earlier credit lines starting from the late 80s till about 2000.
  - Opting for short term loans.