

Lending Club Case Study

SubhajitPal And Ramesh

Problem Statement

For company like Lending Club there are two ways in which the company can make a loss.

- 1. Not lending money to a borrower who would repay the loan.*
- 2. Lending money to a borrower who will not repay the loan.*

The main cause of loss for the company Lending Club turns out to be issuing loans to borrowers who will not repay the loan(ie. they will default). In this exploratory data analysis, we will try to find out the characteristics which will help to identify the risky borrowers who are most likely to default. Thus through this analysis we will try to mitigate the risk of losing money for Lending Club.

Approach to analyse the data set

- *Data Understanding And Identifying the Key Variable*
- *Cleaning The Data*
- *Univariate And Segmented Data Analysis*
- *Bivariate Analysis*
- *Recommendations*

Cleaning the data

- Deleted all the columns which have all null values
- We are observing that the columnns namely `acc_now_delinq`, `delinq_amnt`, `chargeoff_within_12_mths`, `initial_list_status`, `application_type`, `policy_code`, `collections_12_mths_ex_med`, `pymnt_plan`, `tax_liens` are categorical variables with only one unique value which doesn't provide much to the analysis of the loan status both in the case of univariate and bivariate analysis so we can remove these columns
- Also `member_id` and `url` variable are total unique values and we can remove them as well
- Deleting the columns which are related to Applicants demographics

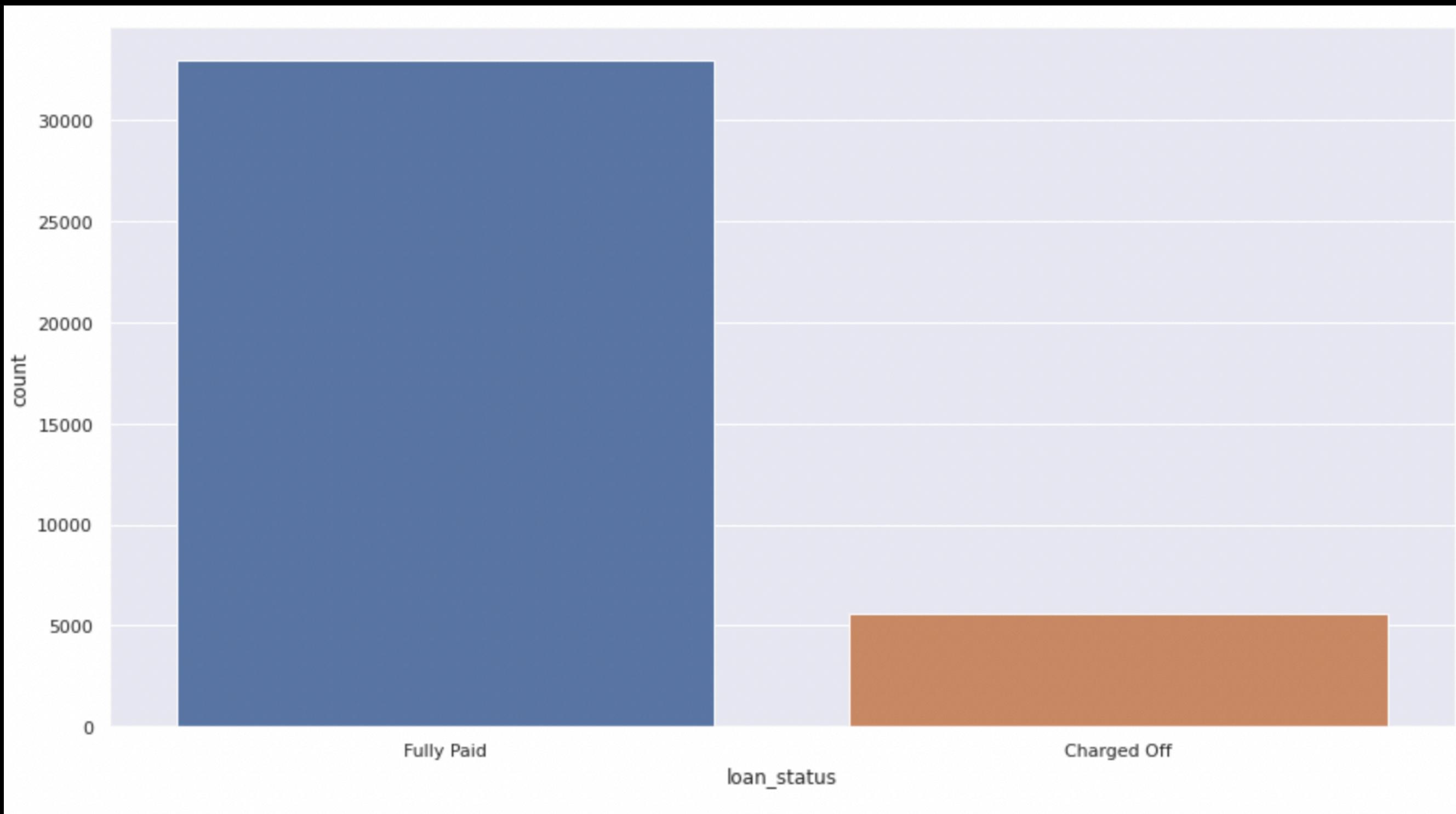
Univariate Analysis

We will going to analyse the data using univariate analysis with all driving variables which are mentioned below.

- *loan_amnt*
- *dti_cat*
- *int_rate_cat*
- *annual_inc_rate*
- *purpose*
- *issue_month*
- *term*
- *grade*
- *verification_status*
- *emp_length*
- *pub_rec_bankrupties*

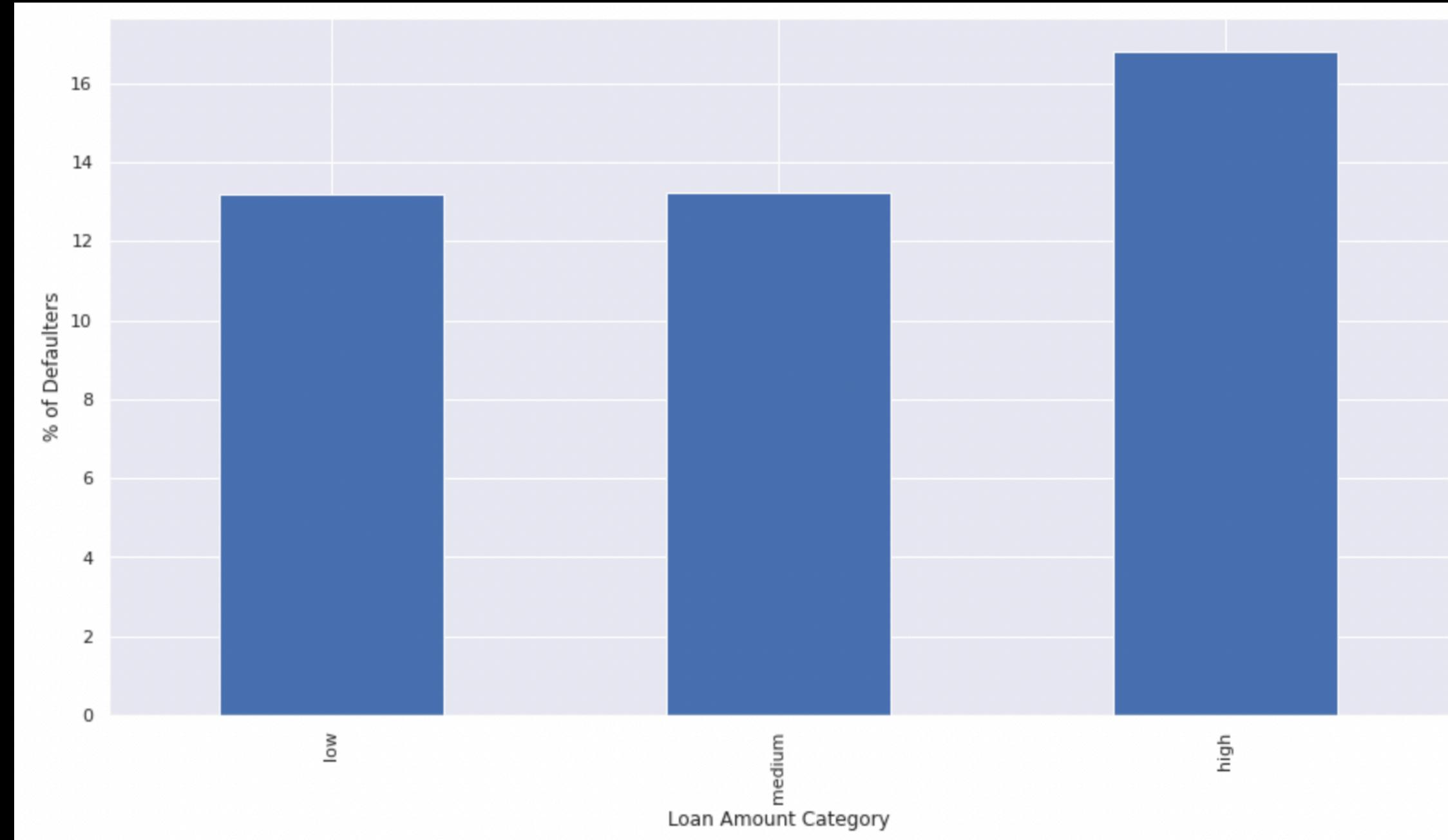
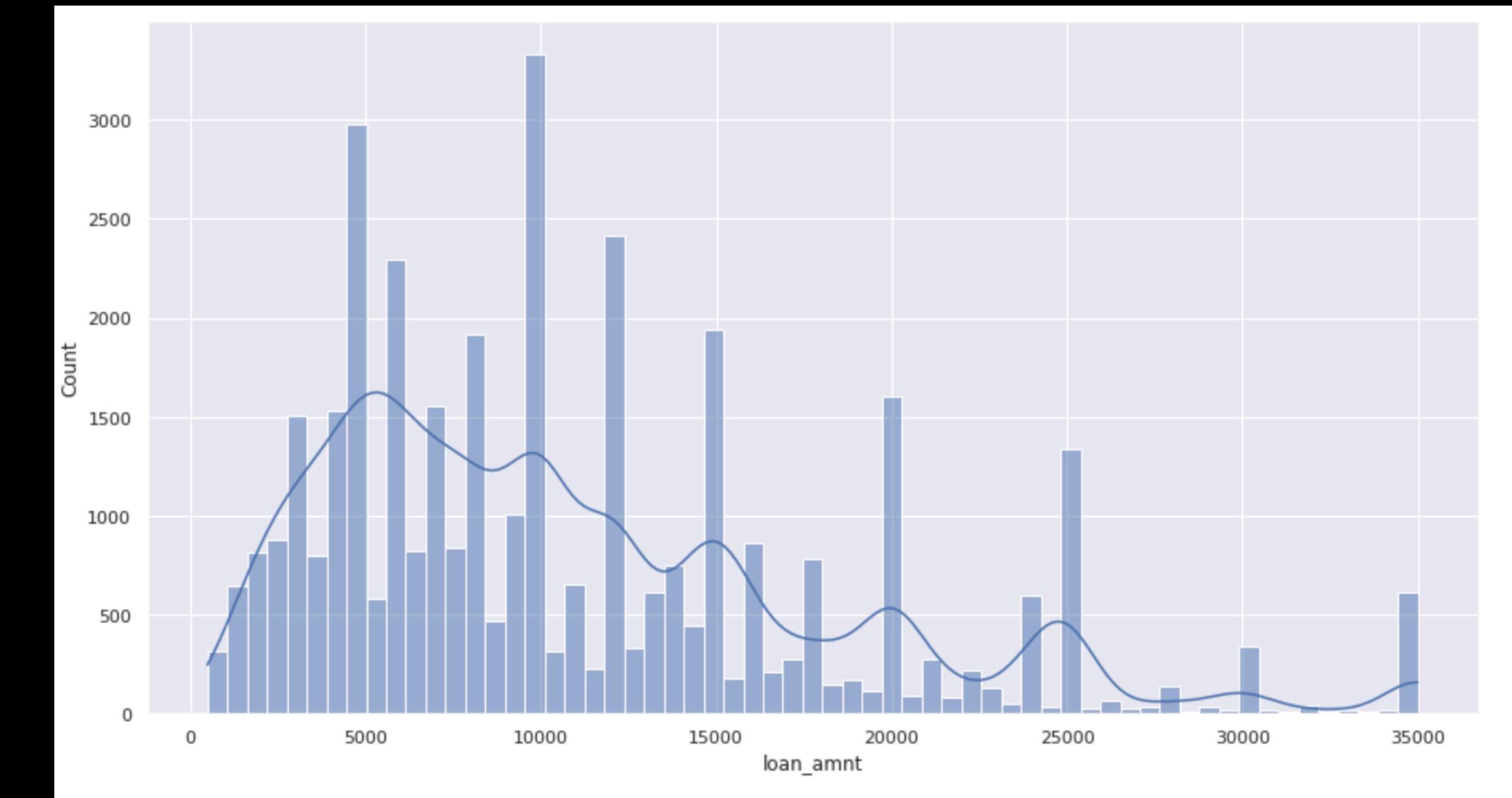
loan_status

- We can observed that the defaulters for the given data set is approximately about 14% of the total customers.



loan_amnt

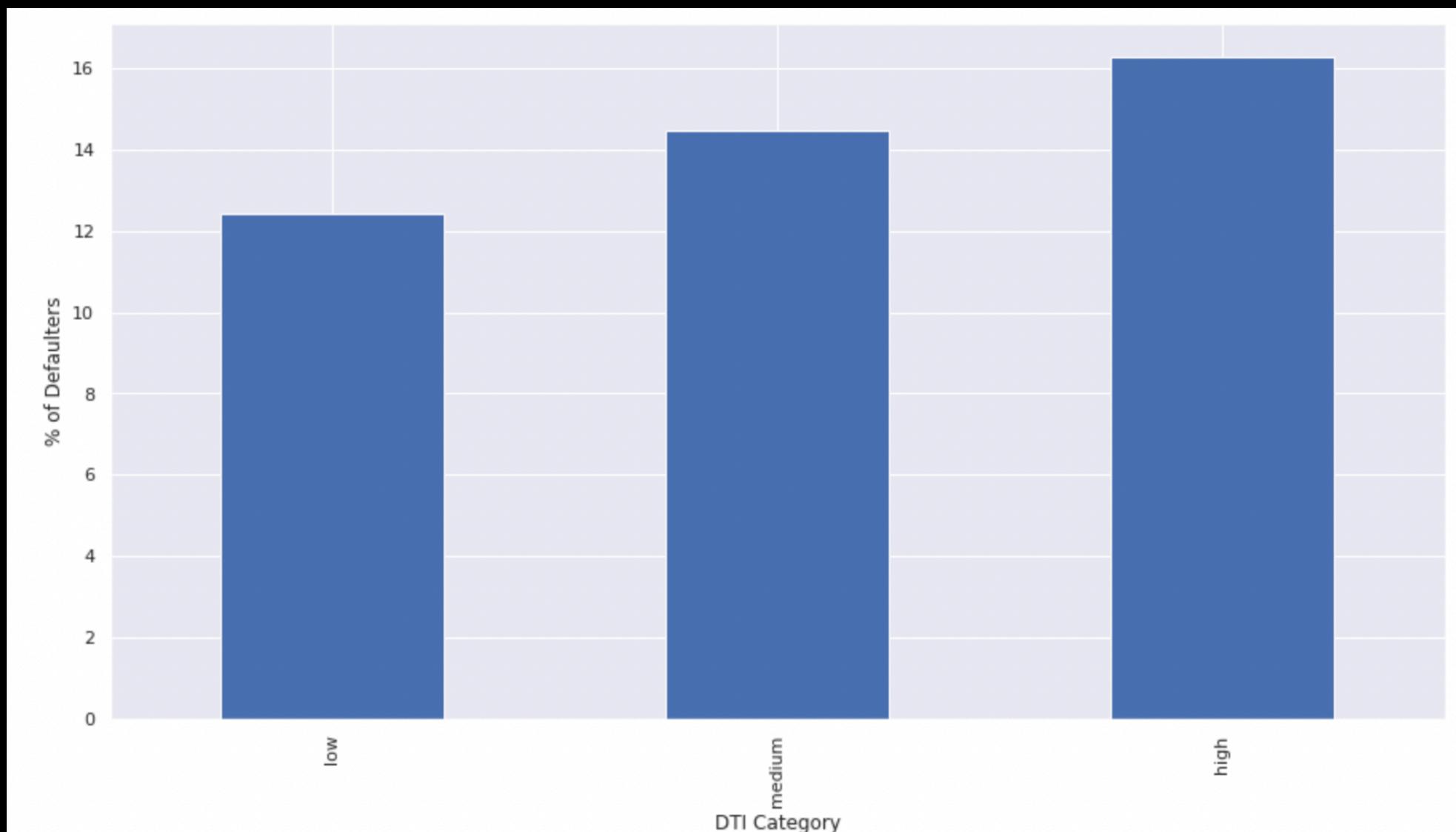
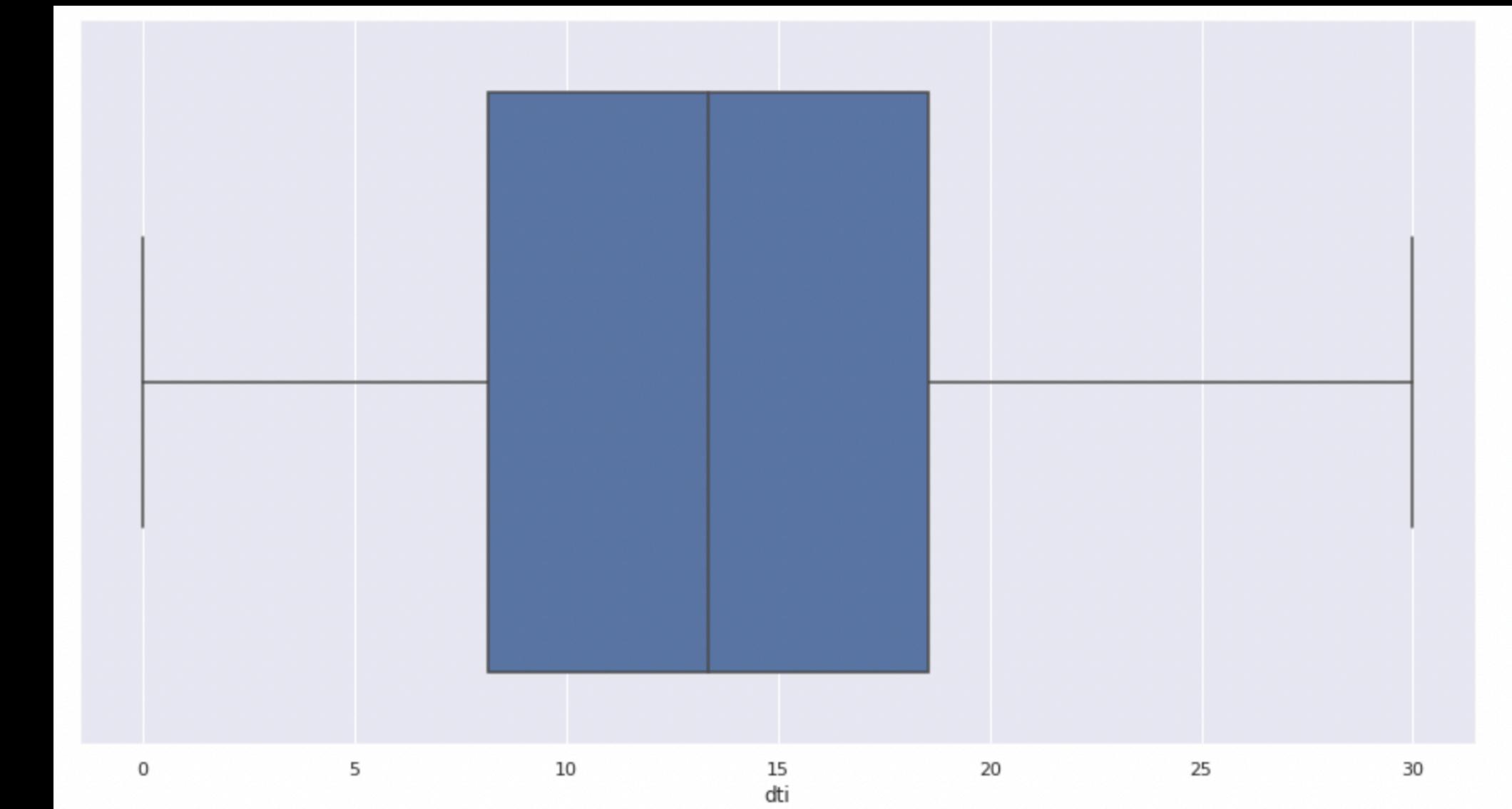
- The applied loan amount distribution is not standard normal distribution
- **We can observe that funding amounts are higher around each 5000 boundary.**
- We will use 5000 as bucket size for later analysis.



Observation - We can observe that higher the loan amount the *ratio of defaulters are high*

Dti_cat

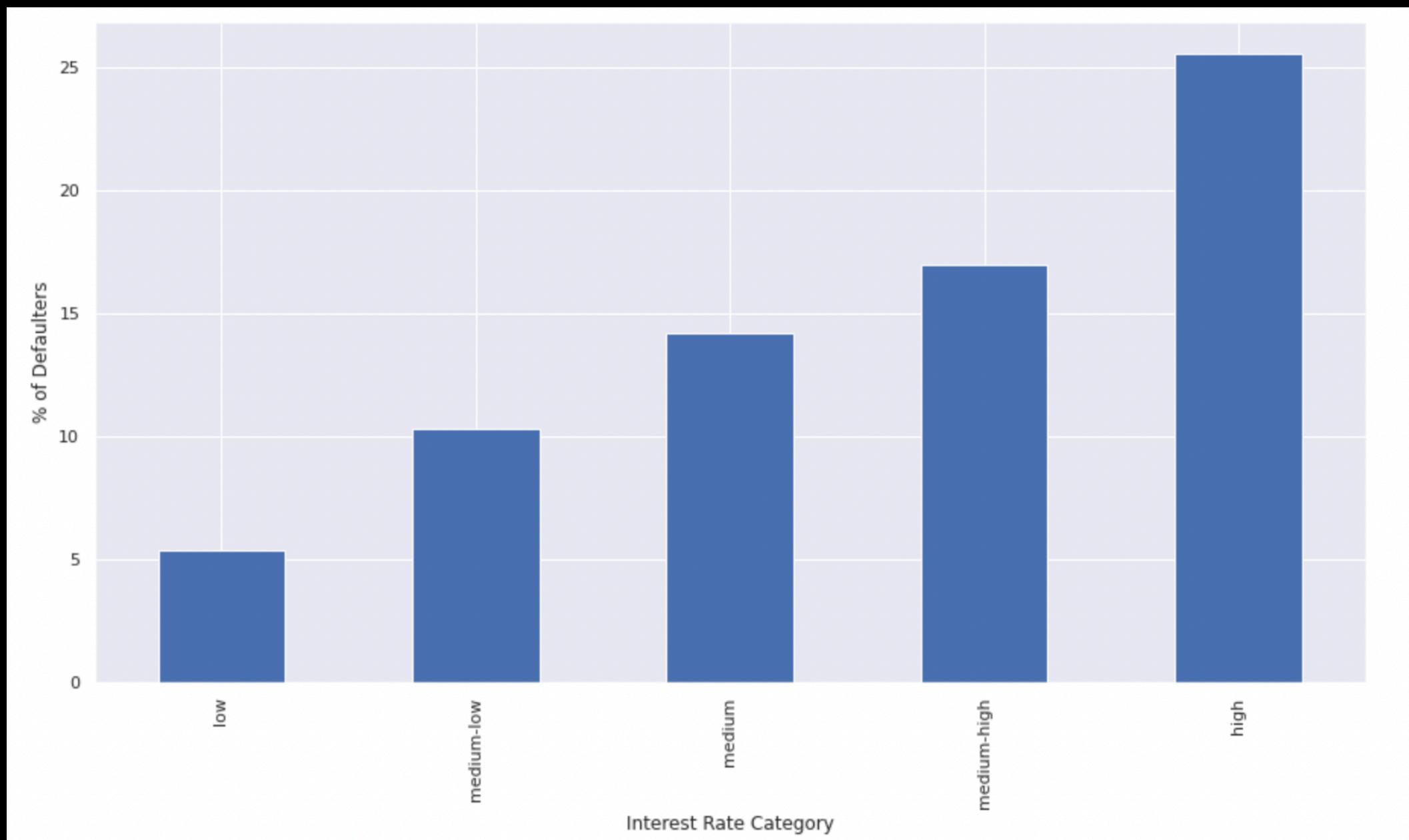
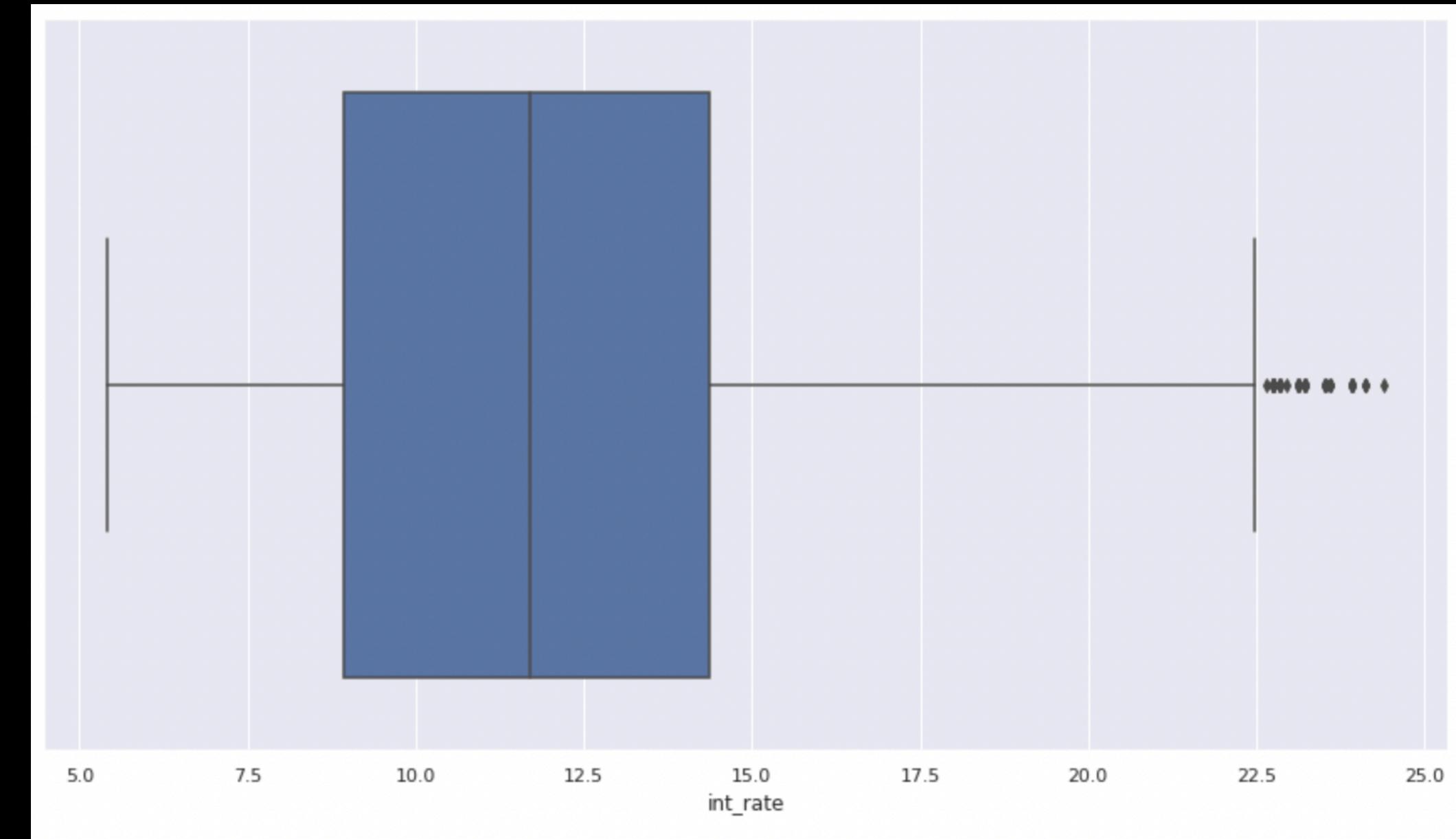
- We made *Dti_cat* as categorical variables for our analysis. We made the bins with respect to quartile ranges
- We binned the column to low, medium, and high for analysis according to quartile ranges



Observation - We can observe that If higher the debt to income ratio higher the defaulting rate

int_rate_cat

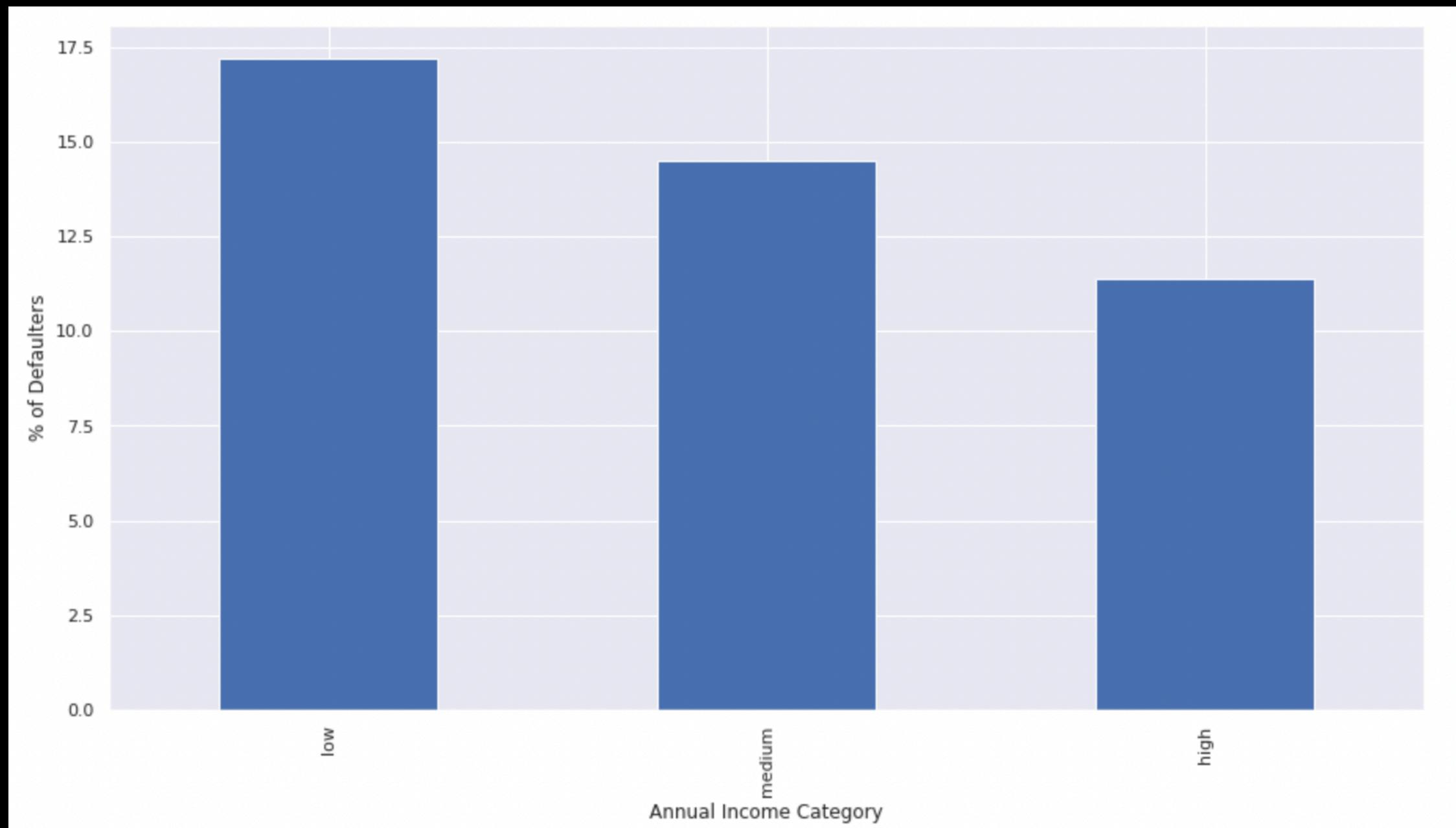
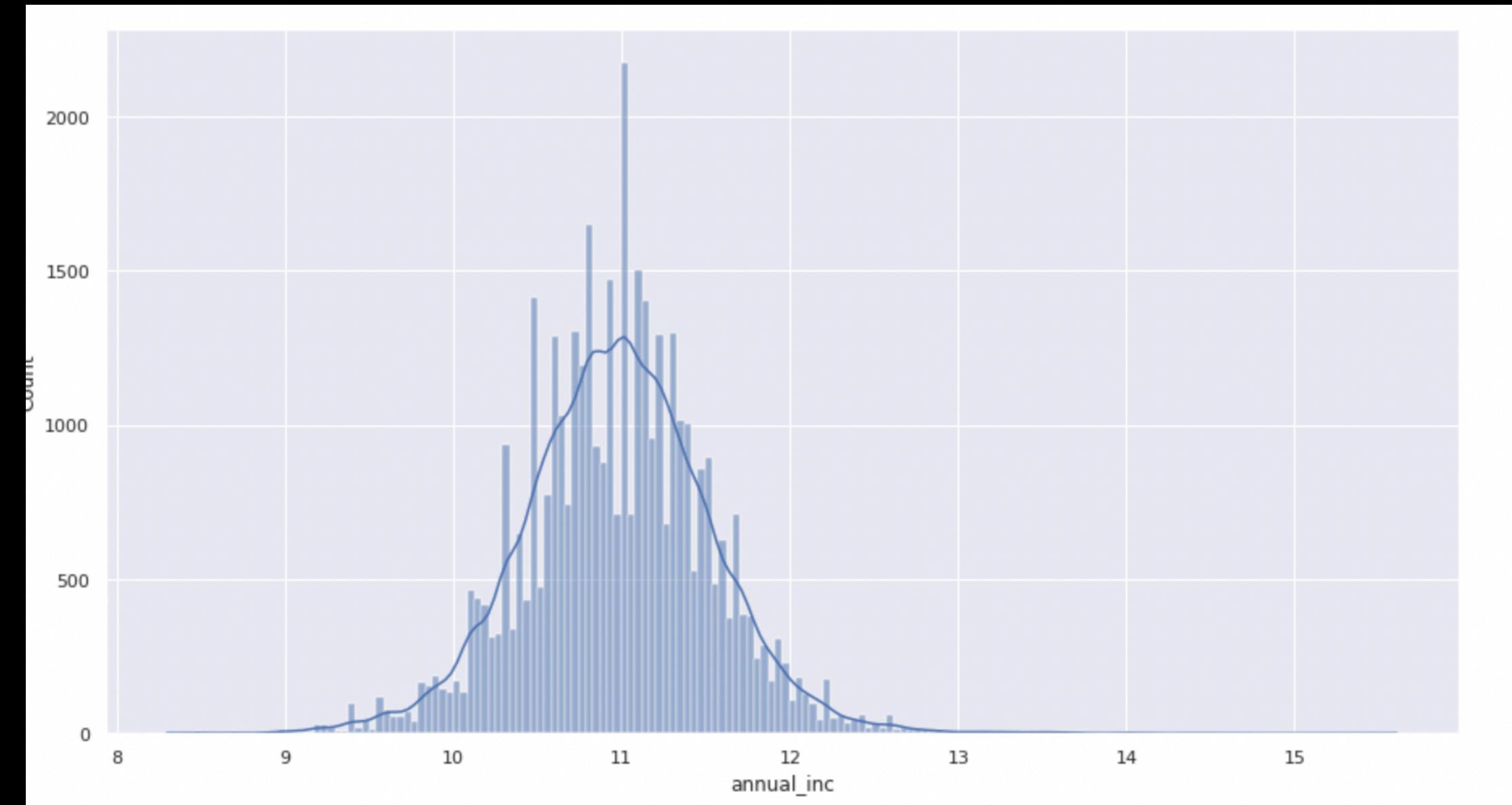
- We made *int_rate_cat* as categorical variables for our analysis. We made the bins with respect to quartile ranges
- We binned the column to low, medium-low, medium, medium-high and high for analysis according to quartile ranges



Observation - higher the interest rate higher the percentage of Defaulters

annual_inc_rate

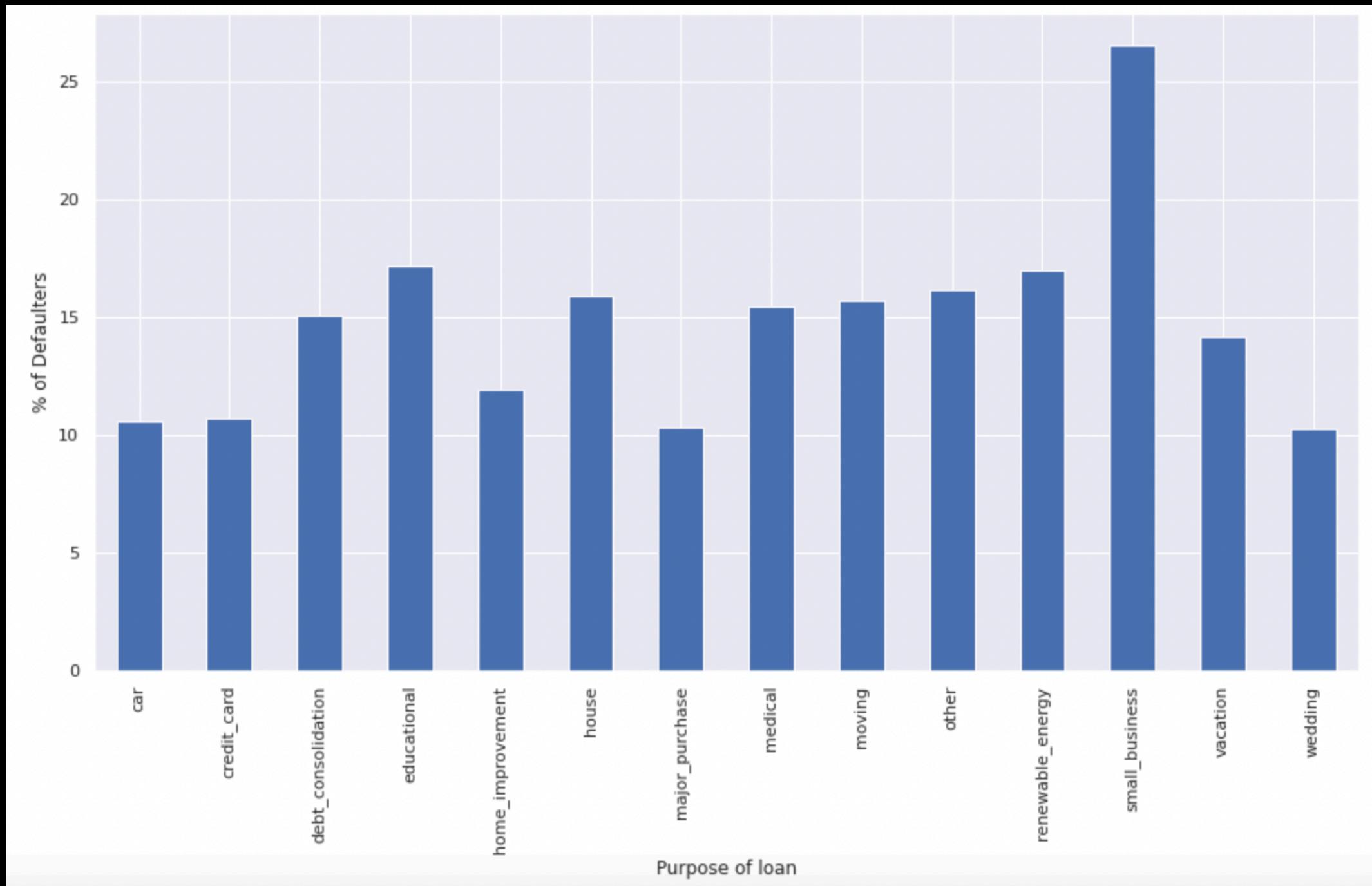
- Created a log normal distribution from the annual_inc variable for creating better categorical variable
- We made annual_inc_rate as categorical variables for our analysis. We made the bins with respect to quartile ranges
- We binned the column to low, medium, and high for analysis according to quartile ranges



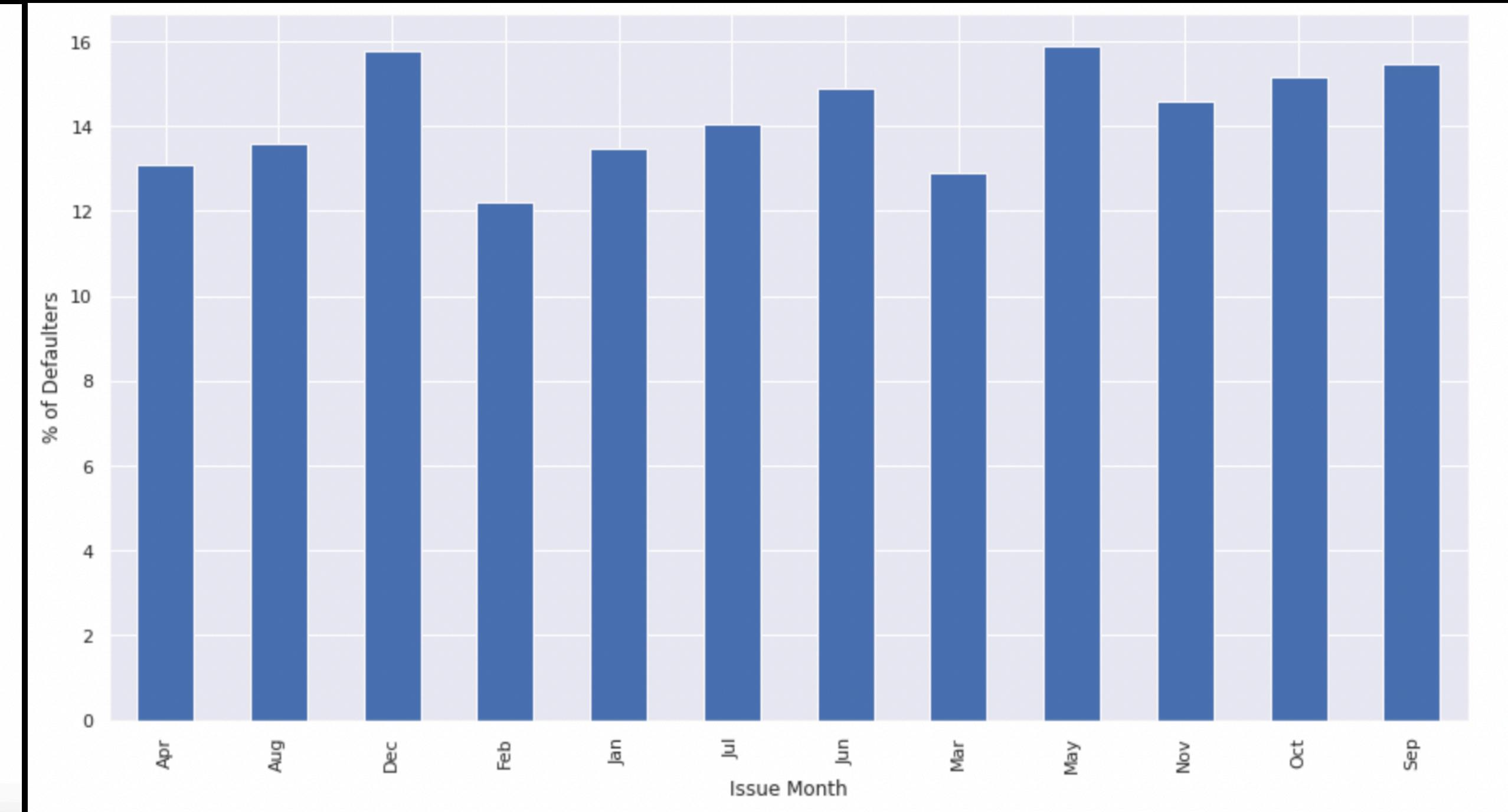
Observation

- We are having an interesting observation here that with the annual income being low the % of defaulters are high.

purpose



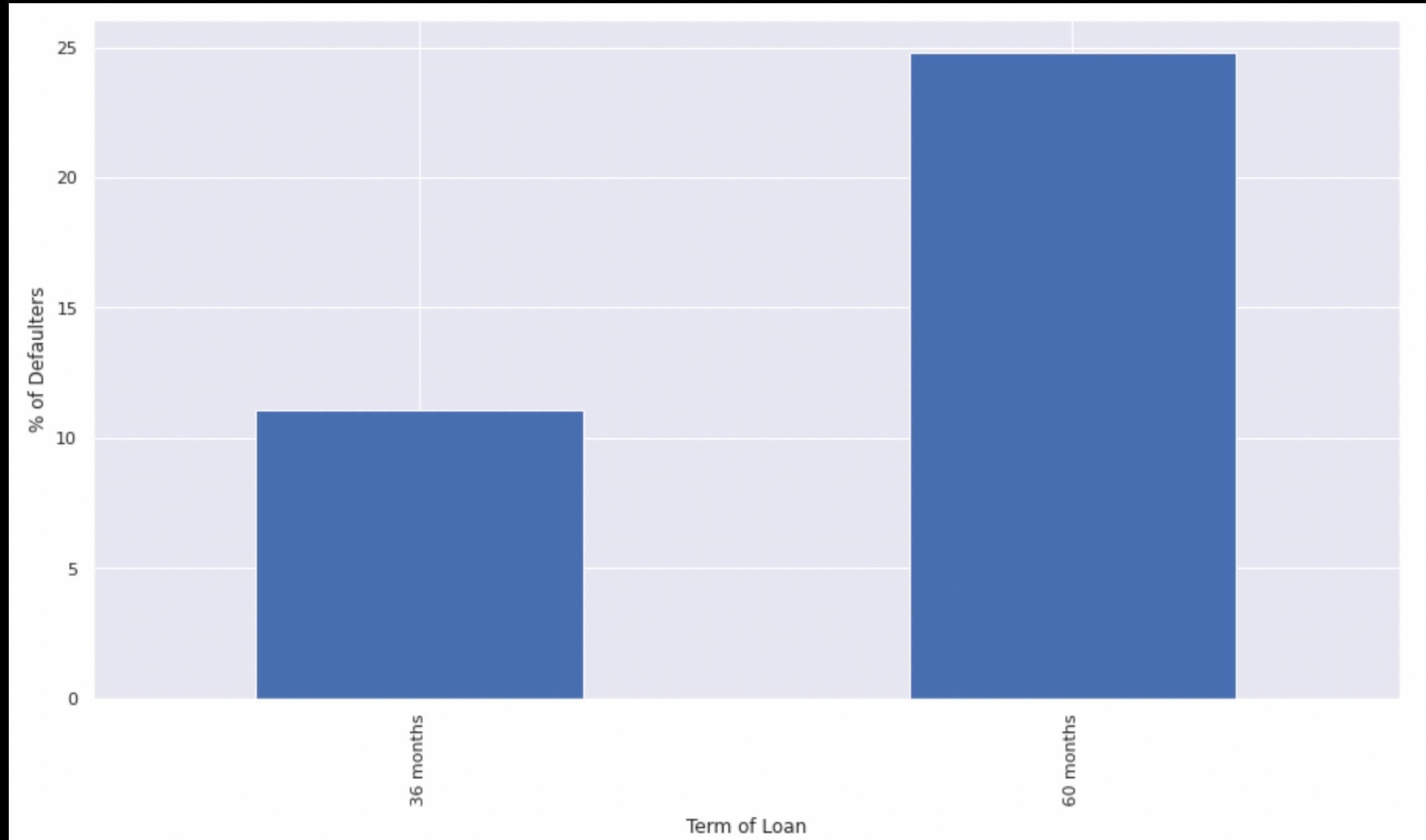
issue_month



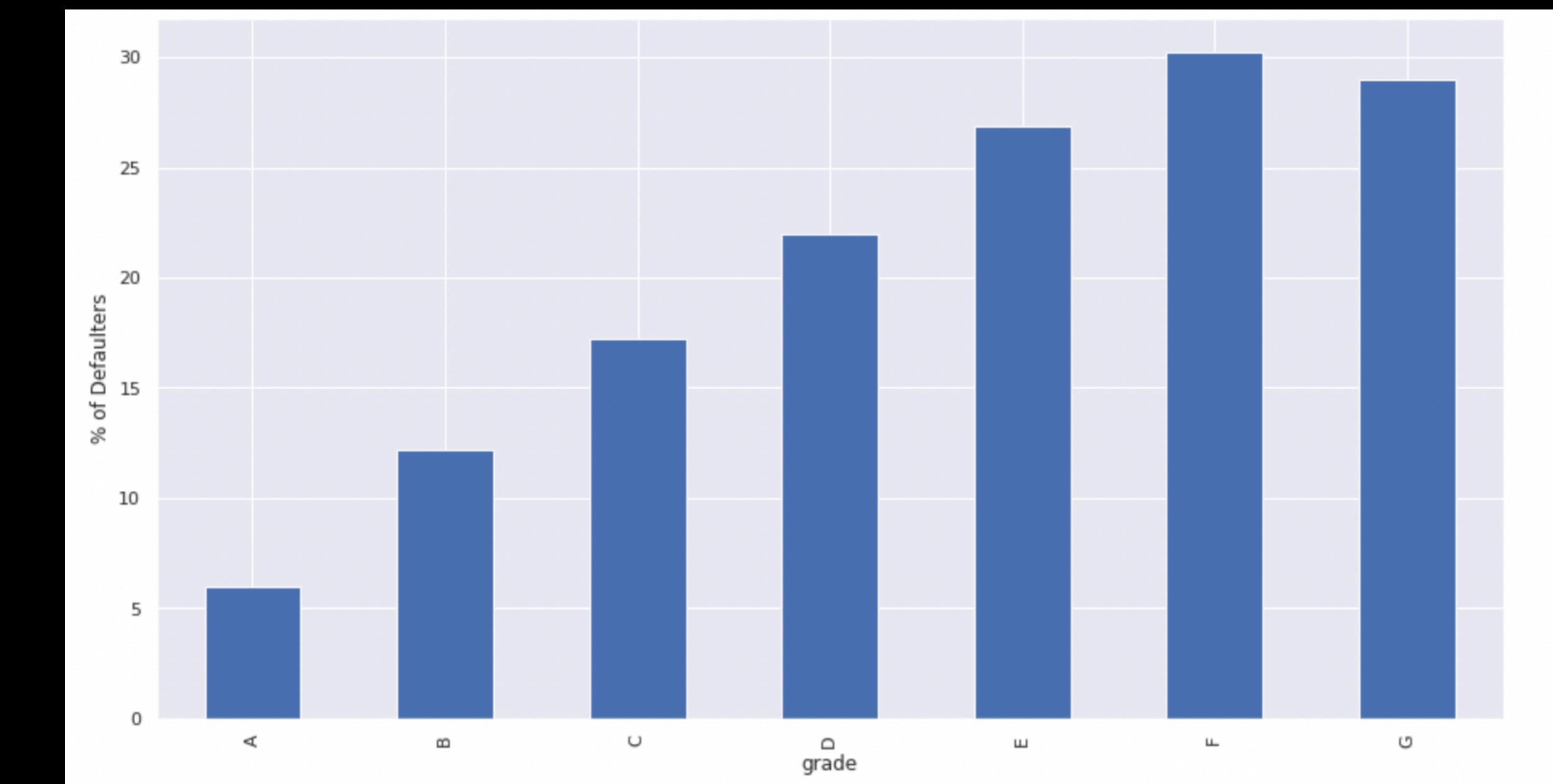
Observation - It is clearly evident that the % of default of small business category is much higher than the rest of the categories

Observation - based on the issue month of the loan we are not getting any kind of strong inferences.

Term



Grade



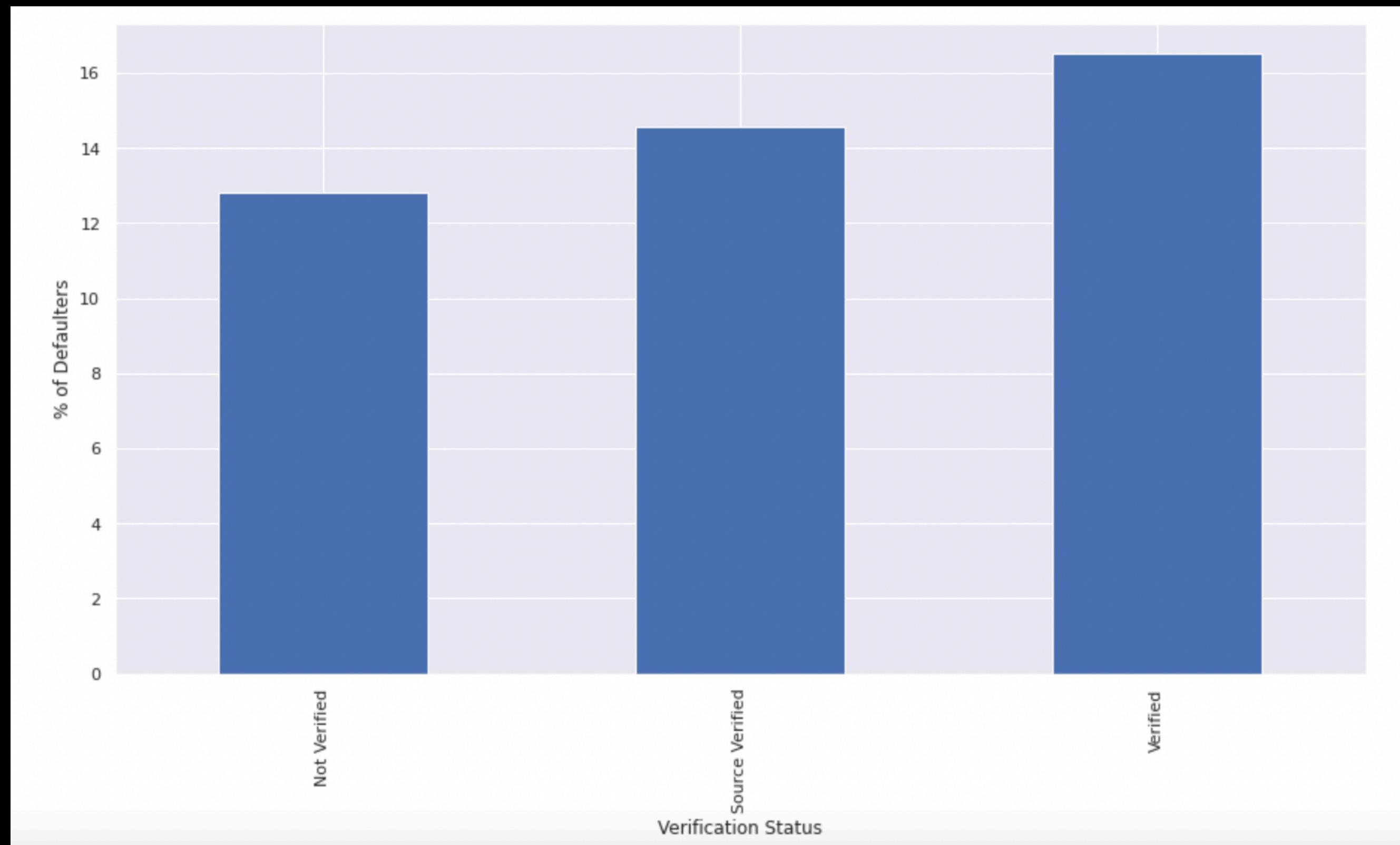
- **Observation** - The % of defaulters for 60 months is high. So the higher term loans have a higher chance of default. Lending club should approve short term loans

- **Observation** - Grade E onwards are risky, and less numerous. Lending Club should either refuse loans or charge high interest rates.

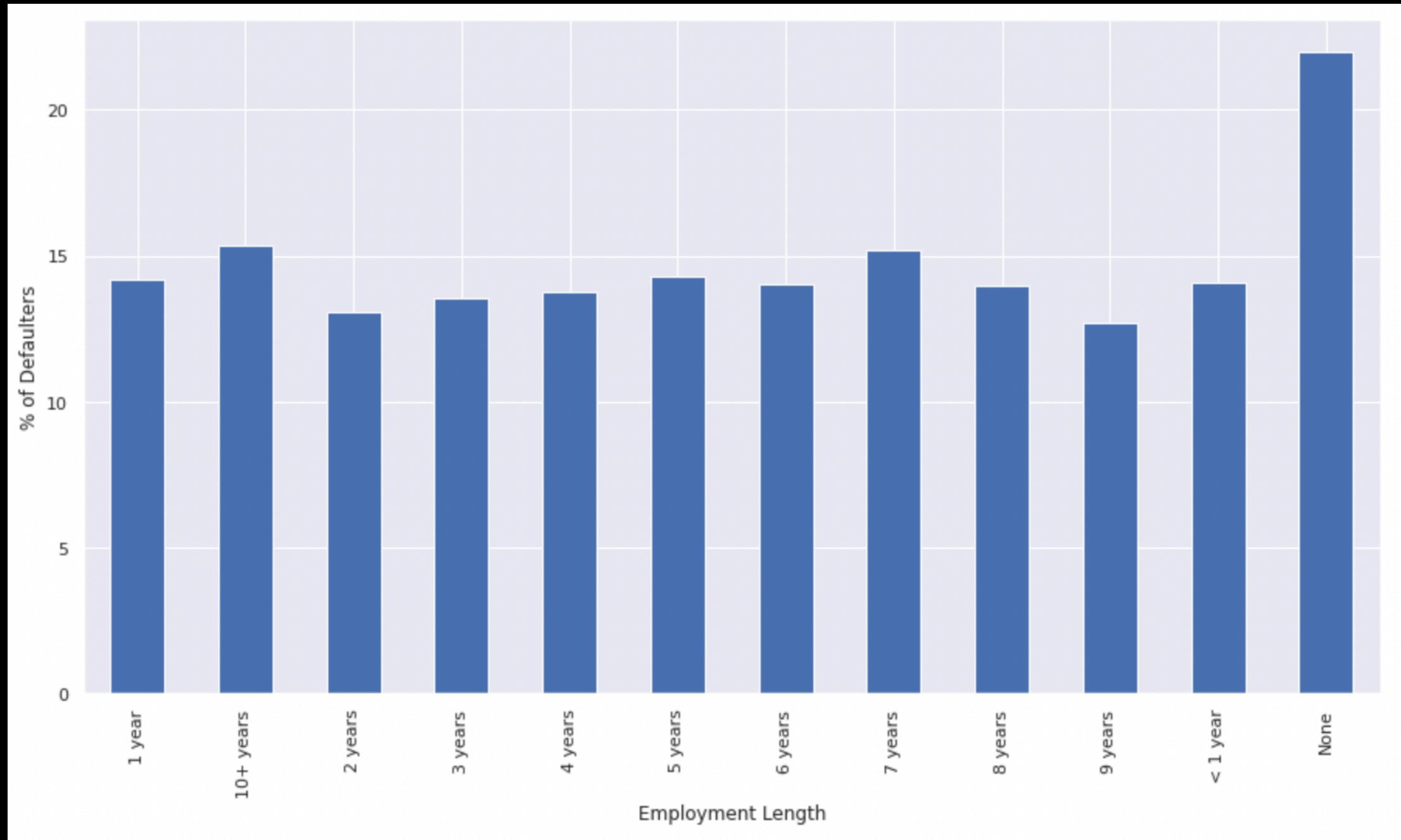
verification_status

Observations -

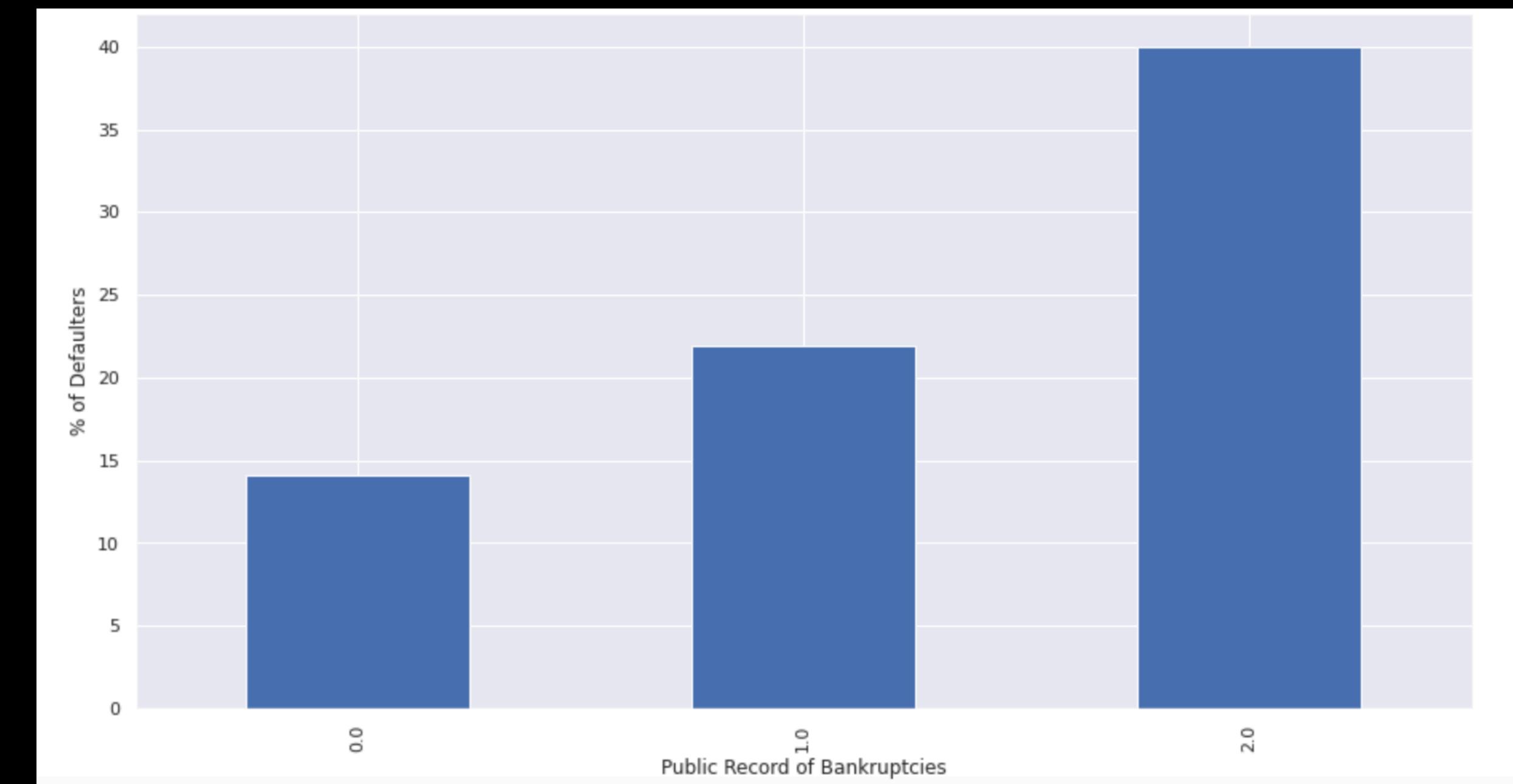
- Verification status gives us an interesting insight. It shows that there is a higher percentage of defaulters amongst the borrowers whose source of income was verified by a 3rd party source or by LC.
- This can mean any one of the two things:
 1. The verification process has some flaws
 2. A verified source of income does not help us to concur that a borrower will less likely default and thus has no impact in risky applicant determination
- The later point may be considered to be more likely, as a third party source and LC both making error in validation is highly unlikely.



emp_length



pub_rec_bankruptcies



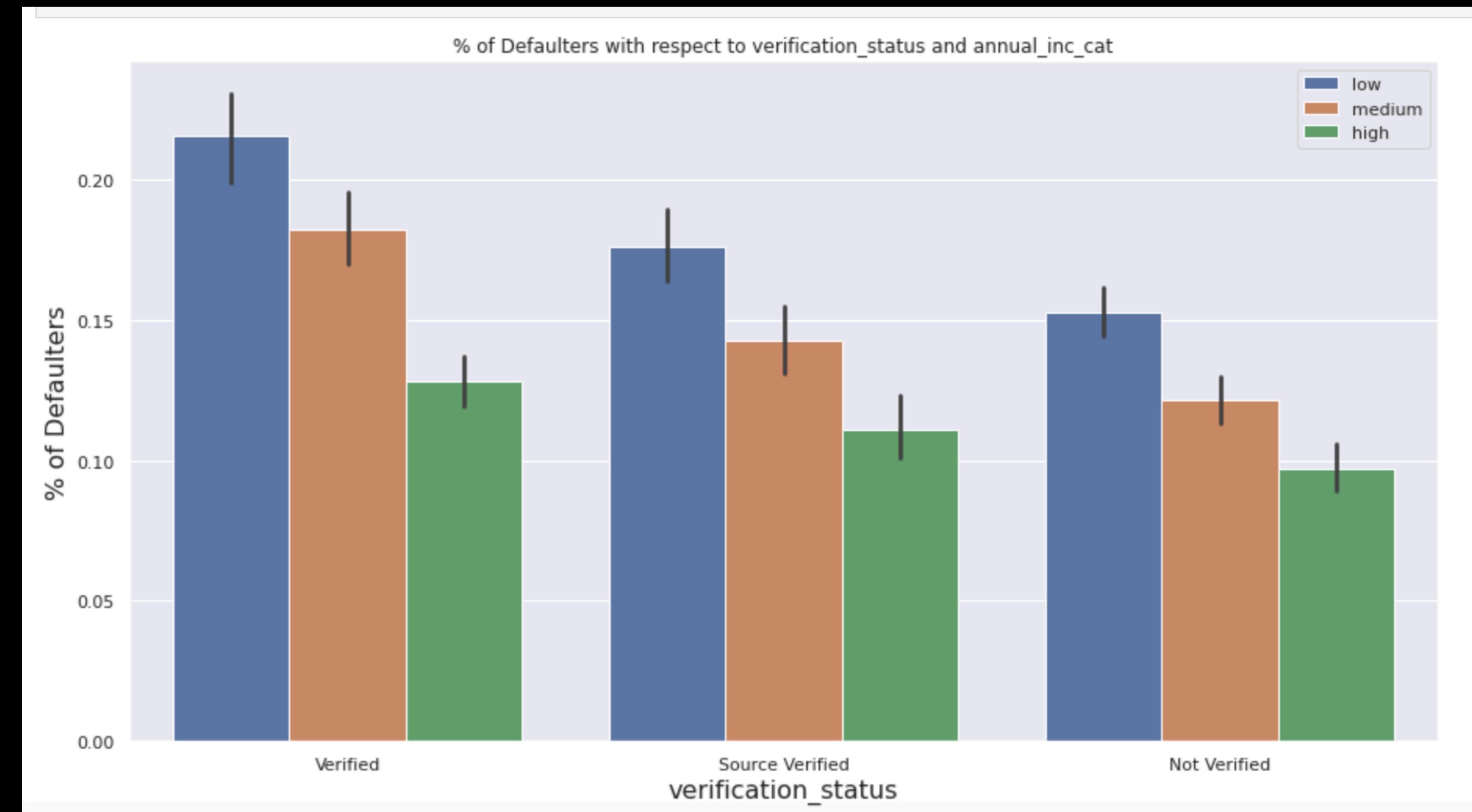
- **Observation** - The Default rate for persons with no employment Length is higher.

Observation - The percentage of Charged Off loans is markedly higher when the borrower has a prior record of bankruptcy.

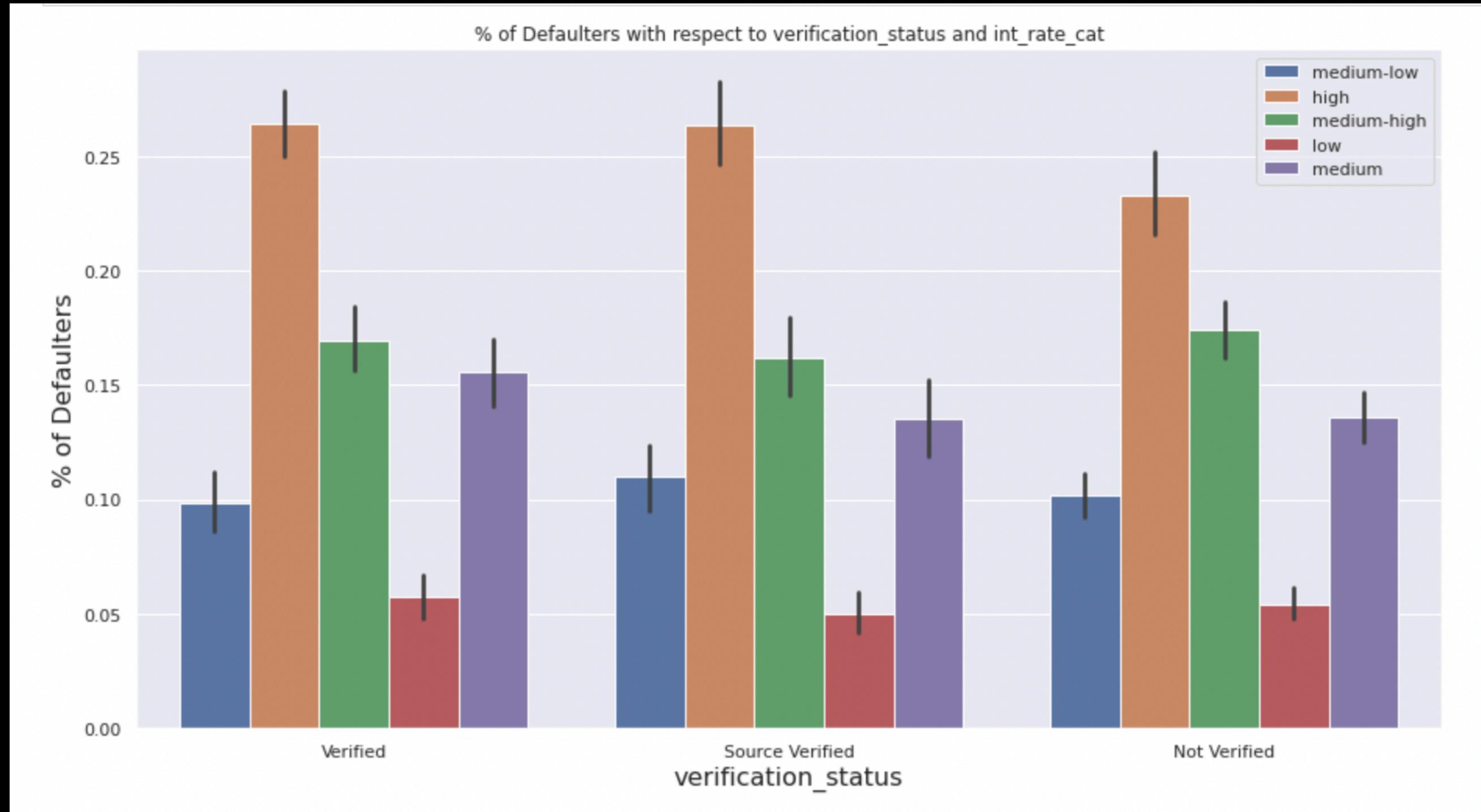
Bivariate Analysis

Verification_status and annual_inc_cat

- **Observation** - We can conclude from the above verification status and annual income plot that low annual income with non verified status customers are more likely to default



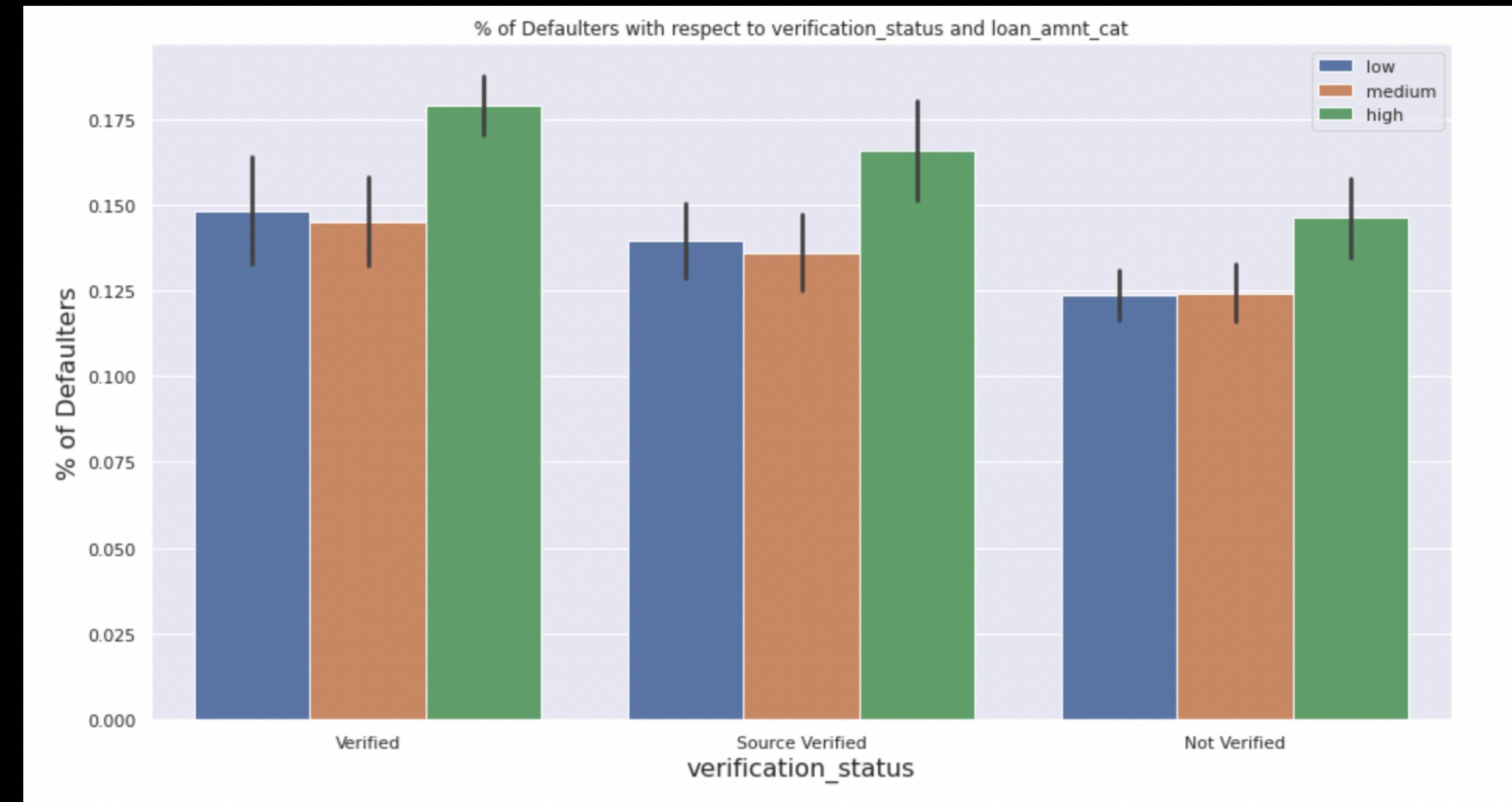
Verification_status and int_rate



- **Observation** - We can conclude from the above verification status and int rate category plot that high interest rate with non verified status customers are more likely to default

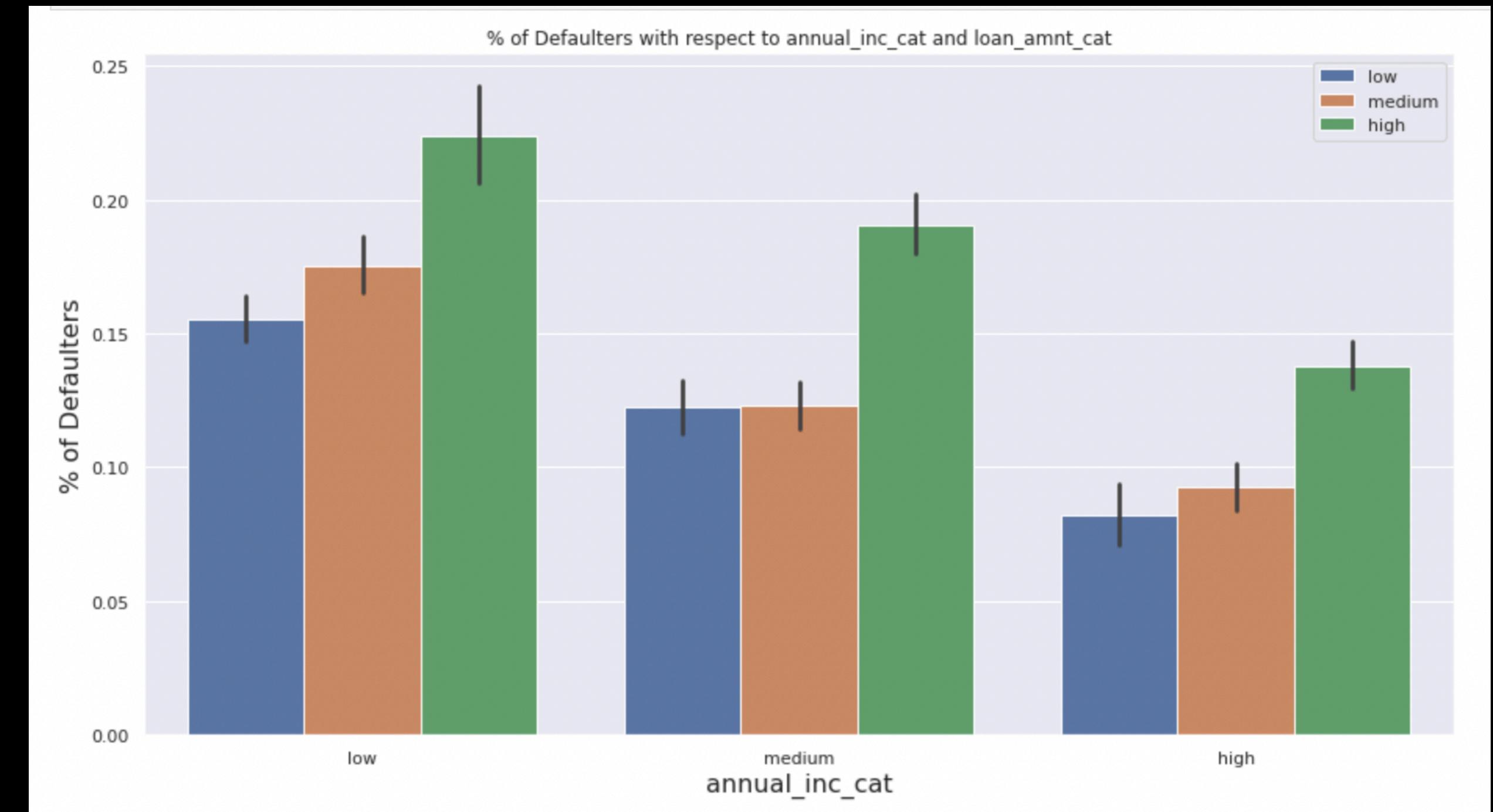
Verification_status and loan_amnt_cat

- **Observation** – We can conclude from the beside verification status and loan amount plot that high loan amount with non verified status customers are more likely to default



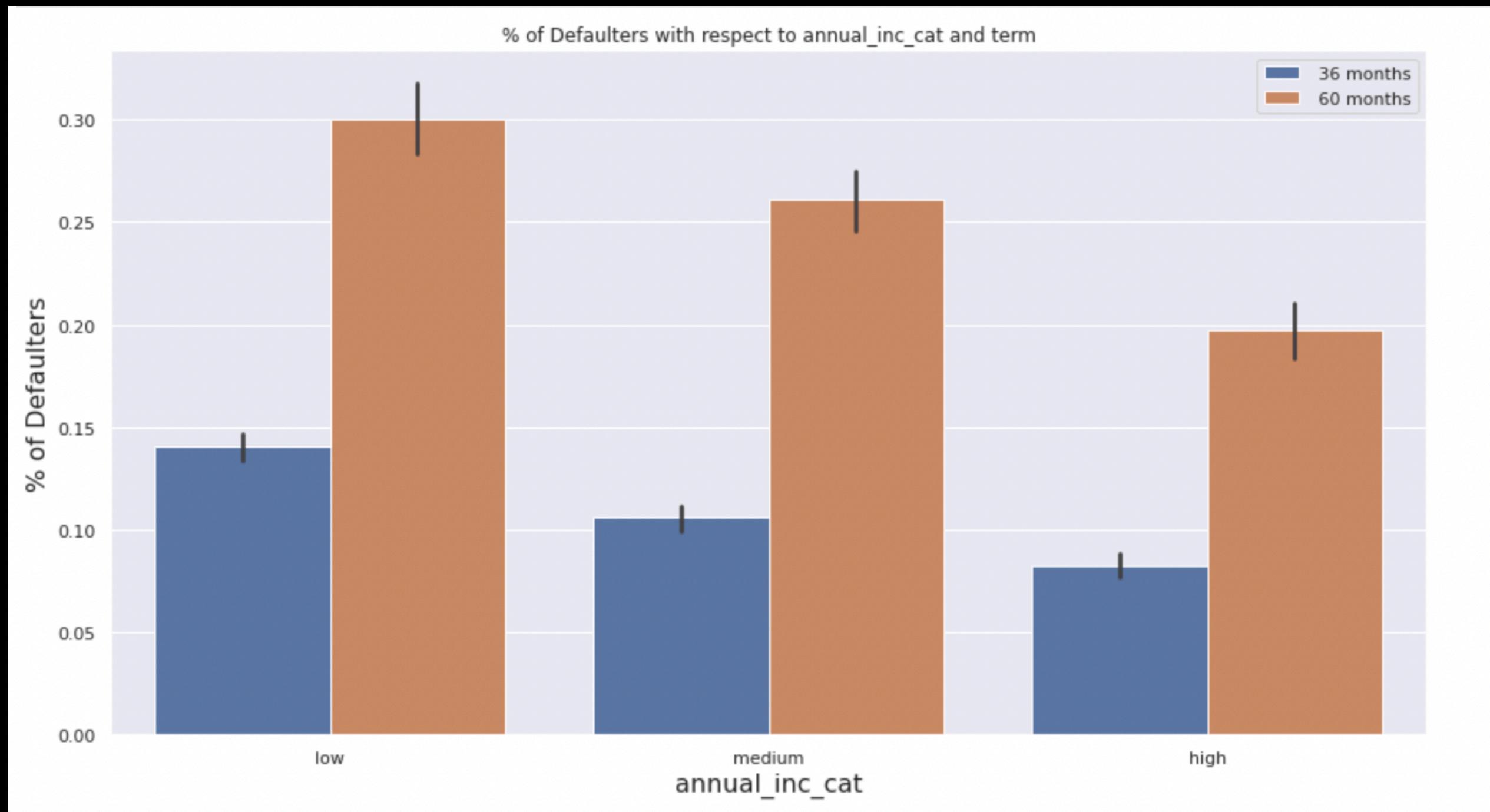
annual_inc_cat and loan_amnt_cat

- **Observation -** We can conclude from the beside annual income and loan amount plot that high loan amount with low annual income customers are more likely become defaulters



annual_inc_cat and term

- **Observation** - We can conclude from the beside annual income and term plot that higher the term with low annual income customers are more likely become defaulters



Recommendations :

- The Driving metrics for our analysis are, **annual_inc,dti,verification_status,int_rate,term,grade,purpose,loan_amnt**.
- Based on the Analysis when the DTI of a borrower is high, the person is more likely to default. So LC must consider a borrower with DTI less than **9.23** which is our start of low dti value.
- Lending Club should consider accepting more loans of grade A and B. It should be cautious of loans falling in grades E,F and G
- If the loan is for purpose of small business there is a high chance of default.
- LC need to consider the number of public bank-ruptcies for a specific borrower higher the number higher the chance of default.
- Lending Club should accept more loans for the term of **36 months** as the % of charged off loans is less and the no. of loan applicants are more. Also the avg fully paid loan amount is increasing over the years and avg charged off lon amount is decreasing over the years.
- Lending Club should be wary of the loans where the purpose is Small Business as the percentage of a loan being charged off is maximum **27.1%**. Accepting loans for the purpose of Weddings,major purchase, car and credit card is highly recommended.
- Lending Club should consider accepting more loans from applicants whose annual income is greater than **116000** as their probablity of charge off is minimum.